UNIVERSIDAD DE COSTA RICA

CIMPA Centro de Investigación en Matemática Pura y Aplicada

XXI Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias

III LACSC: Congreso Internacional en Estadística Computacional

SIMMAC

III Latin American Conference on Statistical Computing

Programa y Resúmenes
27 de febrero al 2 de marzo, 2018,
San José, Costa Rica

20 aniversario CIMPA

XXI Simposio **I**nternacional de **M**étodos **M**atemáticos **A**plicados a las **C**iencias

San José, 27 Febrero al 02 de Marzo, 2018/ February-March, 2018

Comité Organizador / Organizing Committee

Javier Trejos Zelaya (Coordinador), Mario Villalobos Arias (UCR), Alex Murillo Fernández(UCR), Álvaro Guevara(UCR), Eduardo Piza Volio(UCR), Esteban Segura Ugalde(UCR), Fabio Sánchez Peña(UCR), Juan Gabriel Calvo Alpízar(UCR), Luis Barboza Chinchilla(UCR), Maikol Solís Chacón (UCR) y Oldemar Rodríguez Rojas(UCR).

Comité Científico / Scientific Committee

Adilson Xavier (Rio de Janeiro, Brasil), Alex Murillo (CIMPA, UCR), Alexey L Sadovski (Texas A&M, U.S.A.), Álvaro Guevara (CIMPA, UCR), Andreas Griewank (Berlin, Germany), Álvaro Corvalan (Buenos Aires, Argentina), Claudio Cuevas (Santiago, Chile), Eduardo Piza (CIMPA, UCR), Edwin Diday (Paris IX, France), Ellina Grigorieva (Texas WU, U.S.A.), Graciela González (CIMAT, Mexico), leana Castillo Arias (UCR), José A. Ramírez (CIMPA, UCR), Juan Manuel Otero (Havana, Cuba), José Hernández (Caracas, Venezuela), Luis Barboza (CIMPA, UCR), Luka Neralic (Zagreb, Croatia), Manuel Núñez (Connecticut, U.S.A.), Mario Villalobos (CIMPA, UCR), Marlos Viana (Illinois, U.S.A.), Michel Théra (Limoges, France), Oldemar Rodríguez (CIMPA, UCR), Onésimo Hernández (CINVESTAV, Mexico), Patrick Groenen (Rotterdam, Netherlands), Pedro Méndez (CIMPA, UCR), Romina Cardo (Buenos Aires, Argentina), Rosanna Verde (Napoli, Italy), Santiago Cambronero (UCR), Sergio de los Cobos (UAM, Mexico) y Theodore Hill (Georgia Tech, U.S.A.).

Encargada de Organización

María Luisa González Campos, CIMPA (UCR)

Asistentes de Organización

Jessica Pérez Aguilar, Felipe Escalente Guido, Laura Soto Sanabria, Adler Moisés Rodríguez Fallas, Carlos Darío Badilla Cerdas, César Andrés. Rojas Monge, Daniel Granados Campos, Francisco Arce Leitón, Isaac Fernando López Rodríguez, Jazmín Alexandra Cruz Sojo, Kimberly Karina Víctor Castro, Laura Marisol Mora González, Luis Angelo Bermúdez Ayales, Luis Fernando Mejías Molina, María José Martínez Montero, Merilyn Castillo, Oscar Luis Espinoza Bonilla, Pablo Luis Vivas Corrales, Stefanny Chacón Vargas, Stefanny Chacón Vargas, Stephanny Espinoza Herrera, Valeria Ayala Alonso y Valery Zúñiga Kondrashov.

Editores de LaTeX: Mario Villalobos Arias y Keythlyn Soto Madriz.

III Congreso Internacional Latinoamericano de Estadística Computacional

San José, 27 Febrero al 02 de Marzo, 2018/ February-March, 2018

Comité Organizador / Organizing Committee

Paulo Canas Rodrigues (Chair), Federal University of Bahia, Brazil, and University of Tampere, Finland, Javier Trejos Zelaya (Coordinador, UCR), Mario Villalobos Arias (UCR).

Comité Científico / Scientific Committee

Paulo Canas Rodrigues (Chair), Federal University of Bahia, Brazil, and University of Tampere, Finland; Alba Martinez Ruiz, Universidad Católica de la Santísima Concepción, Chile; Ana Ahedo, University Buenos Aires, Argentina; David F. Muñoz, Instituto Tecnológico Autónomo de México, Mexico; Eufrásio de Andra de Lima Neto, Federal University of Paraíba, Brazil; Emilio Porcu, Universidad Técnica Federico Santa María, Valparaíso, Chile; Flavio Augusto Ziegelmann, Federal University of Rio Grande do Sul, Brazil; Francisco de A. T. de Carvalho, Federal University of Pernanbuco, Brazil; Francisco Louzada Neto, Universidade de São Paulo, São Carlos, Brazil; Javier Trejos, CIMPA & School of Mathematics, Universidad de Costa Rica, Costa Rica; Ronny Vallejos, Universidad Técnica Federico Santa María, Valparaíso, Chile; Vanda Lourenço, Nova University of Lisbon, Portugal

Encargada de Organización

María Luisa González Campos, CIMPA (UCR)

Asistentes de Organización

Jessica Pérez Aguilar, Felipe Escalente Guido, Laura Soto Sanabria, Adler Moisés Rodríguez Fallas, Carlos Darío Badilla Cerdas, César Andrés. Rojas Monge, Daniel Granados Campos, Francisco Arce Leitón, Isaac Fernando López Rodríguez, Jazmín Alexandra Cruz Sojo, Kimberly Karina Víctor Castro, Laura Marisol Mora González, Luis Angelo Bermúdez Ayales, Luis Fernando Mejías Molina, María José Martínez Montero, Merilyn Castillo, Norma Segura Corella, Oscar Luis Espinoza Bonilla, Pablo Luis Vivas Corrales, Sebastián Flores Alvarado, Stefanny Chacón Vargas, Stefanny Chacón Vargas, Stephanny Espinoza Herrera, Valeria Ayala Alonso y Valery Zúñiga Kondrashov.

Editores de LaTeX: Mario Villalobos Arias y Keythlyn Soto Madriz..

AGRADECIMIENTOS

El Centro de Investigación en Matemática Pura y Aplicada (CIMPA) de la Universidad de Costa Rica, agradece a las siguientes instituciones y entidades que ayudaron e hicieron posible la realización del XXI Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias y del III Congreso Latinoamericano de Estadística Computacional:

- Rectoría de la Universidad de Costa Rica.
- Vicerrectoría de Investigación de la Universidad de Costa Rica.
- Vicerrectoría de Administración de la Universidad de Costa Rica.
- Vicerrectoría de Acción Social de la Universidad de Costa Rica.
- Vicerrectoría de Vida Estudiantil de la Universidad de Costa Rica.
- Escuela de Matemática de la Universidad de Costa Rica.
- Facultad de Educación de la Universidad de Costa Rica.
- Facultad de Ciencias de la Universidad de Costa Rica.
- Sede del Atlántico de la Universidad de Costa Rica.
- Oficina de Divulgación e Información de la Universidad de Costa Rica.
- Oficina de Asuntos Internacionales y Cooperación Externa.
- Editorial de la Universidad de Costa Rica.
- Sección de Transportes de la Oficina de Servicios Generales de la Universidad de Costa Rica.
- Sección de Seguridad y Tránsito de la Oficina de Servicios Generales de la Universidad de Costa Rica.
- Centro de Informática de la Universidad de Costa.
- El Instituto francés para América Central (L'Institut français d'Amérique Centrale – IFAC)
- Asociación Internacional de Computación Estadística (LARS-IASC)
- Sociedad Latinoamericana de Clasificación y Análisis de Datos (SoLCAD).
- Sociedad Costarricense de Optimización e Investigación de Operaciones (SoCOIO).

Dr. Javier Trejos
Coordinador, SIMMAC

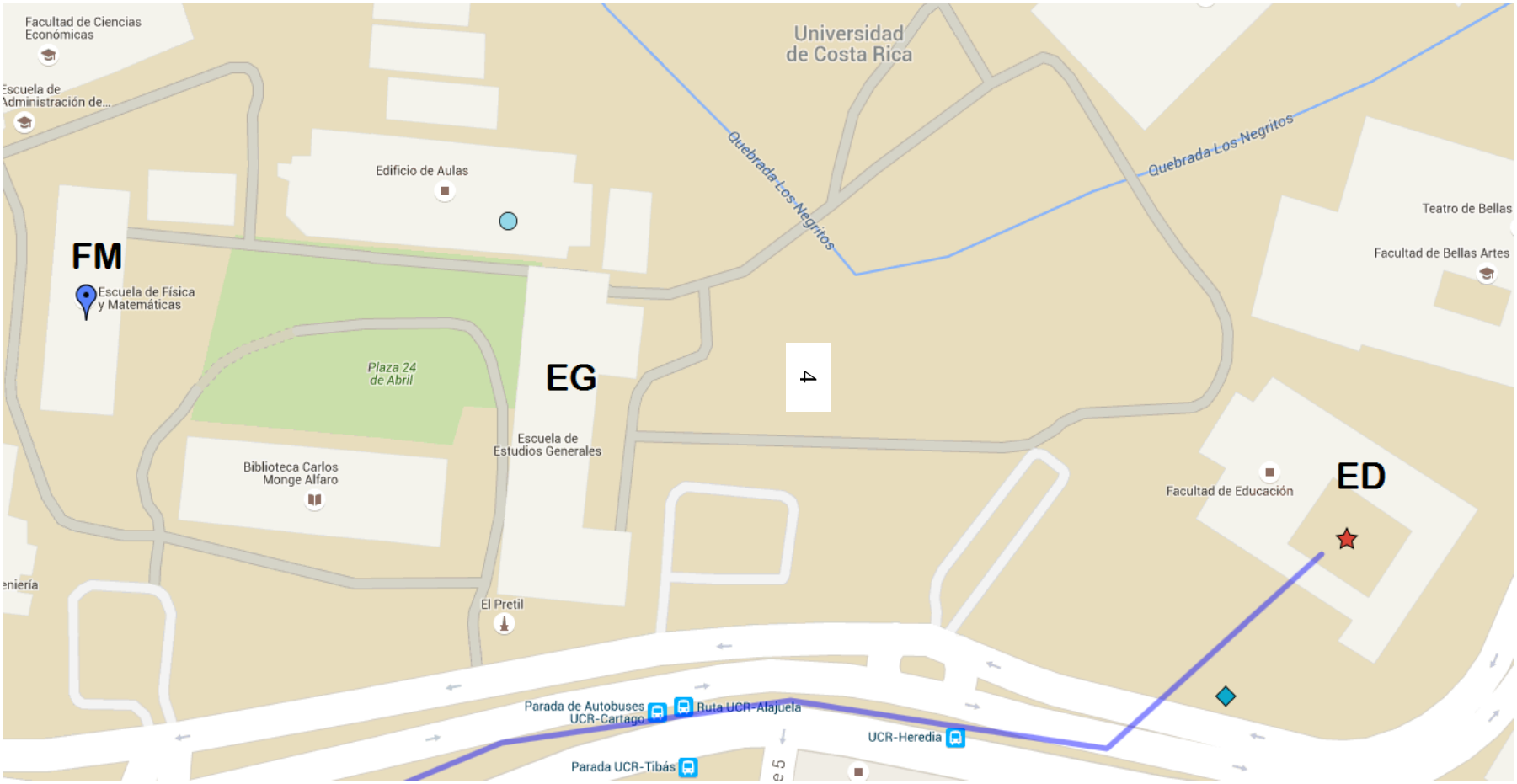
Sesiones / *Sessions*

Apl	Applications / Aplicaciones
Aprox	Approximation / Aproximación
Bio	Biomathematics / Biomatemáticas
Clas	Classification / Clasificación
DA	Data Analysis / Análisis de datos
DJ	Simulación Multiagente: Herramientas y Aplicaciones
DMin	Data Mining / Minera de Datos
DS	Dynamical Systems / Sistemas Dinámicos
FinMat	Financial Mathematics / Matemática Financiera
L3-1-1	LACSC-New developments in Genomic Selection and Prediction studies
L3-2-1	LACSC-Recent advances in Statistical Computing 1
L3-4-1	LACSC-Statistics and Applications
L4-1-1	LACSC-Statistical Computing for Data Science
L4-2-1	LACSC-Forecasting using Simulation
L4-3-1	LACSC-Recent advances in robust statistics and distribution theory
LACSC	LACSC-Statistical Computing
LACSC-3-1	LACSC-Copulas
Mod	Modeling / Modelación
Num	Numerical Analysis / Análisis Numérico
Opt	Optimization / Optimización
OR	Operations Research / Investigación de Operaciones
Prob	Probabilty / Probabilidad
Cart	Poster/Carteles
CONF	Conference
Tut	Tutorial
C-LACSC	Conference LACSC
SC-LACSC	Tutorial LACSC
R	Clustering and Visualization of Complex Data

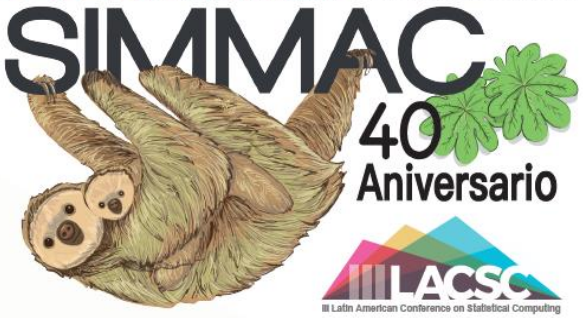
Aulas / *Rooms*

ED: Facultad de Educación / Education Faculty

Auditorio (ED) / *Auditorium*	Primer piso ED / *first floor ED*
Aula 112 (ED) / *Room* 1	Primer piso ED / *first floor ED*
Aula 113 (ED) / *Room* 2	Primer piso ED / *first floor ED*
Aula 111 (ED) / *Room* 2	Primer piso ED / *first floor ED*
Laboratorio 119 / *Laboratory 119*	Primer piso ED / *first floor ED*
Oficina del SIMMAC / *SIMMAC desk*	Primer piso ED / *first floor ED*
Oficina del CIMPA / *CIMPA office*	Edificio CIMPA-Matemática, Ciudad de la Investigación


Facultad de Ciencias Económicas
Escuela de Administración de...
Universidad de Costa Rica
Quebrada Los Negritos
Quebrada Los Negritos
Edificio de Aulas
FM
Escuela de Física y Matemáticas
Plaza 24 de Abril
EG
4
Teatro de Bellas
Facultad de Bellas Artes
Escuela de Estudios Generales
Biblioteca Carlos Monge Alfaro
Facultad de Educación
ED
El Pretil
Parada de Autobuses UCR-Cartago
Ruta UCR-Alajuela
UCR-Heredia
Parada UCR-Tibás

XXI Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias | III LACSC: Congreso Internacional en Estadística Computacional



Programa y Resúmenes
27 de febrero al 2 de marzo, 2018,
San José, Costa Rica

DATE	TIME	ACTIVITY
Tuesday 27	9:30 a.m.	Opening Ceremony
Tuesday 27	6:10 p.m.	Welcome Toast
Thursday 01	4:35 p.m.	Departure Dinner
Thursday 01	7:00 p.m.	Dinner
Friday 02	5:45 pm	Closing Toast

WiFi Access
Visitante UCR

If you have EDUROAM account you can connect directly.
Si tiene una cuenta de EDUROAM, puede conectarse directamente.

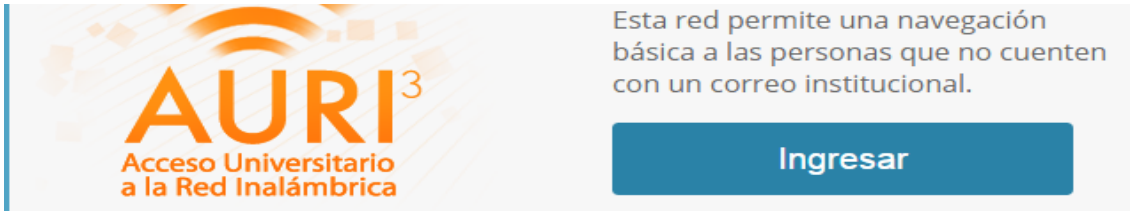
Otherwise:

Please follow the next steps:

1. Connect your laptop to the wireless connection "VISITANTE UCR".

2. Open your web browser and click "Proceed anyway":

3. Automatically, the browser will redirect you and click "Ingresar":



XXI SIMMAC / III LACSC:

Martes / Tuesday 27th

Auditorium

8:00	-->	∞	Inscripciones / Inscription
9:30	10:00	am	Inauguración / Opening ceremony
10:00	10:30	am	Café / Coffee break
10:30	11:15	am	Opening conference
			Sorin, S.
			Replicator dynamics: old and new
			Chair: Trejos, J
11:20	12:05	am	Opening conference LACSC
			Fung, W.
			Efficient Monte Carlo evaluation of resampling-based hypoth
			Chair: Canas, Paulo
12:05	01:30	pm	Almuerzo / Lunch

Auditorium

Session	Apl1-1			**Applications / Aplicaciones 1**	
Chair:					
1	01:30	01:50	pm	Arce, J.	Gravitational tests with a K
2	01:50	02:10	pm	Gómez, P.	Perturbing compact object
3	02:10	02:30	pm	Oliva, G.	Applications of GRT for cor
Session	Apl2-1			**Applications / Aplicaciones 2**	
Chair:					
13	02:35	02:55	pm	Fernández, J.	Metodología para la constr
14	02:55	03:15	pm	Caicedo, L.	Planteamiento de un méto
15	03:15	03:35	pm	Calderón-Arce,	PROE: Simulación computa
	03:35	04:00	pm	Café / Coffee break	
Session	Apl3-1			**Applications / Aplicaciones 3**	
Chair:					
25	04:00	04:20	pm	Martínez, I.	Grafo n-residual Módulo m
26	04:20	04:40	pm	González, R.	A generalized algorithm fo
27	04:40	05:00	pm	Reyes, A.	An automatic framework f
28	05:00	05:20	pm	Camacho, F.	A characterization for dom

Room 1

Session	L3-1-1			**LACSC-New developments in Genomic Selection and Prediction studies**	
Chair: Vanda Lourenço					
4	01:30	01:50	pm	De_los_Campos,	Prediction of Complex Hur
5	01:50	02:10	pm	Pérez, M.	Genomic Selection: Is it wo
6	02:10	02:30	pm	Posekany, A.	Outlier detection on micro
Session	L3-2-1			**LACSC-Recent advances in Statistical Computing 1**	
Chair: Veronica A. Gonzal					
16	02:35	02:55	pm	Carvajal, R.	Parallel Algorithm for a Tes
17	02:55	03:15	pm	Ponsot-Balaguer,	An R implementation for tl
18	03:15	03:35	pm	Infante, S.	Estimation of Stochastic Vo
Session	L4-1-1			**LACSC-Statistical Computing for Data Science**	
Chair: Alba Martinez-Ruiz					
29	04:00	04:20	pm	Martinez, A.	On Some Theoretical Aspe
30	04:20	04:40	pm	Firinguetti-Limon	Bayesian Estimation of the
31	04:40	05:00	pm	Carvajal-Schiaffin	A Parallel / Distributed Alg
32	05:00	05:20	pm		

Room 2

Session	Num2-1			**Numerical Analysis / Análisis Numérico 1**	
Chair:					
7	01:30	01:50	pm	Fonseca, J.	Scalable Subsurface Simula
8	01:50	02:10	pm	Cuervo, O.	A Monte Carlo Approach to
9	02:10	02:30	pm	Sequeira, F.	Análisis de una formulació
Session	Num3-1			**Numerical Analysis / Análisis Numérico 2**	
Chair:					
19	02:35	02:55	pm	Guevara, J.	A Mixed Mimetic Formulat
20	02:55	03:15	pm	Forero, W.	Vectorizacion Contours for
21	03:15	03:35	pm	Herrera, E.	Spatial Interpolation of Pre
Session	Num4-1			**Numerical Analysis / Análisis Numérico 3**	
Chair:					
33	04:00	04:20	pm	Reyes, A.	Energy-based and PDE me
34	04:20	04:40	pm	Segura, E.	Computation of Matrix p-t
35	04:40	05:00	pm	Castro, E.	Some Examples of Dense S
36	05:00	05:20	pm	Lobo, J.	Zeroes of functions of Fres

6:10 7:30 pm Brindis de Bienvenida / Welcome Toast

XXI SIMMAC / III LACSC:

Miércoles / Wednesday 28th

Auditorium

08:00 09:00 am TUT-1
Vicente, P.
Recent Developments in Analysis of Three-Way Data

09:00 10:00 am TUT-2
Nieto, A.
Recent Developments in Analysis of Three-Way Data in

10:00 10:30 am Café / Coffee break

10:30 11:15 am Conf-2
Gatica, G.
Mixed finite element and related methods for some no
Chair:

11:20 12:05 am Conf-3
Schubert, L.
Estimation of holding periods applied to the case of sho
Chair:

12:05 01:30 pm Almuerzo / Lunch

Room 1

08:00 09:00 am
Griewank, A.
Nonsmooth Analysis and Optimization via Piecewise Line

09:00 10:00 am
Griewank, A.
Nonsmooth Analysis and Optimization via Piecewise Line

Session LACSC-1-1 LACSC-Statistical Computing 4
Chair:

42	10:30	10:50 am	Choulakian, V.	Optimality of the mea
43	10:50	11:10 am	Qu, L.	Copula Density Estima
44	11:10	11:30 am	Tita, D.	Minimum quadratic d
45	11:30	11:50 am	González-Estrada	An R package for test
46	11:50	12:10 pm	Cevallos, H.	On Fast Computation

Room 2

8:00 10:00 am TUT-4
Gutiérrez, J.
Teoremas fundamentales del bienestar económico

09:00 10:00 am TUT-5
Richter, W.
On p-generalized elliptically contoured distributions

10:30 11:15 am

11:20 12:05 pm Conf-7
Griewank, A.
A new successive piecewise linearization algorithm for ne

Lab1

8:00 10:00 am TUT-3
Becerra, N.
The role of data visualization on statistical analysis

09:00 10:00 am TUT-6
Reinecke, J.
Multiple Imputation with R

Auditorium

Session Mod1-1 Modeling / Modelación 1
Chair:

39	01:30	01:50 pm	Baca, G.	Los modelos de Cour
40	01:50	02:10 pm	Sánchez, I.	Juegos dinámicos y el
41	02:10	02:30 pm	Gutiérrez, J.	The Results of the Du

Session Mod2-1 Modeling / Modelación 2
Chair:

56	02:35	02:55 pm	Sanchez, F.	A Partial Differential E
57	02:55	03:15 pm	Murillo, W.	Estimación de tasas d
58	03:15	03:35 pm	Bravo, C.	Modelamiento mater

03:35 04:00 pm Café / Coffee break

Session Mod5-1 Modeling / Modelación 5
Chair:

68	04:00	04:20 pm	Dumaresq, Y.	Flow of a Ferrofluid Ir
69	04:20	04:40 pm	Iturraran-Viveros	Waveform inversion c
70	04:40	05:00 pm	Ovalle, D.	An ANOVA test for fu
71	05:00	05:20 pm	Criado, R.	New perspectives on

Room 1

Session L3-3-1 LACSC-Application of classification techniques in the
Chair:

47	01:30	01:50 pm	Arguedas, S.	Application of classifi
48	01:50	02:10 pm	Jiménez, J.	A bayesian hierarchic
49	02:10	02:30 pm	Melnykov, I.	Formal implementatio

Session L3-4-1 LACSC-Statistics and Applications
Chair:

59	02:35	02:55 pm	Louzada, F.	Fraud detection in ba
60	02:55	03:15 pm	Granzotto, D.	The risk of amputatio
61	03:15	03:35 pm	Nascimento, D.	Statistics Without Bor

Session L4-2-1 LACSC-Forecasting using Simulation
Chair:

72	04:00	04:20 pm	Muñoz, D.	Estimation of expecta
73	04:20	04:40 pm	Quiroz-Cornejo, J	Bayesian spatial inflat
74	04:40	05:00 pm	Christen, J.	On optimal direction
75	05:00	05:20 pm	López, E.	A model to estimate t

05:25 06:25 pm LARS-IASC General Assembly

Room 2

Session OR1-1 Operations Research / Investigación de Operaciones
Chair:

50	01:30	01:50 pm	Soria, I.	Modelo de optimizaci
51	01:50	02:10 pm	Hernández, F.	Optimización y Aplica
52	02:10	02:30 pm		

Session Opt1-1 Optimization / Optimización 1
Chair:

62	02:35	02:55 pm	Morales, L.	Optimization of super
63	02:55	03:15 pm	Soto, J.	Frobenius Norm: Adva
64	03:15	03:35 pm	Vílchez, E.	Método moderno de

Session DS-1 Dynamical Systems / Sistemas Dinámicos
Chair:

76	04:00	04:20 pm	Solis-Sanchez, H.	A new cryptographic s
77	04:20	04:40 pm	Ferreira, F.	Bounded solutions of
78	04:40	05:00 pm	Herrera, D.	Mathematical models
79	05:00	05:20 pm		

Room 3

Session DA2-1 #N/A
Chair:

53	01:30	01:50 pm		
54	01:50	02:10 pm		
55	02:10	02:30 pm		

Session Clas1-1 Classification / Clasificación
Chair:

65	02:35	02:55 pm	Reyes, A.	A word-distance
66	02:55	03:15 pm	Arce, J.	Point cloud regist
67	03:15	03:35 pm		

Session FinMat-1 Financial Mathematics / Matemática Financiera
Chair:

80	04:00	04:20 pm	Campos, W.	Técnica de Geom
81	04:20	04:40 pm	Cuevas, C.	An application of

XXI SIMMAC / III LACSC:
Jueves / Thursday 1st

Auditorium	Room 1	Room 2	Lab1
08:00 09:00 am Tutorial 1 Vicente, P. Recent Developments in Analysis of Three-Way Data	08:00 09:00 am SC-LACSC Tutorial Cherubini/Mulinacci Copula Functions	08:00 09:00 am Tutorial 3 Gutiérrez, J. Teoremas fundamentales del bienestar económico	08:00 09:00 am Tutorial 4 Becerra, N. The role of data visualization on statistical analysis
09:00 10:00 am Tutorial 5 Nieto, A. Recent Developments in Analysis of Three-Way Data	09:00 10:00 am SC-LACSC Cherubini/Mulinacci Copula Functions	09:00 10:00 am Tutorial 7 Richter, W. On p-generalized elliptically contoured distributions	09:00 10:00 am Tutorial 8 Reinecke, J. Multiple Imputation with R
10:00 10:30 am Café / Coffee break			
10:30 11:15 am Conf-4 Chowell, G. Infectious Disease Modeling	10:30 11:15 am SC-LACSC Tutorial Cherubini/Mulinacci Copula Functions		
11:20 12:05 am Conf-5 Widlund, O. Recientes avances en M etodos de Descomposici on de	11:20 12:05 am C-LACSC-2 Conference Allen, G. Inference, Computation, and Visualization for Convex Cl		
12:05 01:30 pm Almuerzo / Lunch			

Auditorium	Room 1	Room 2	Room 3
Session **Dmin-1** **Data Mining / Minera de Datos** Chair:	Session **LACSC-3-1** **LACSC-Copulas** Chair:	Session **Mod3-1** **Modeling / Modelación 3** Chair:	Session **Opt2-1** **Optimization / Optimización 2** Chair:
82 01:30 01:50 pm Calvo, D. Aplicación de los moc	85 01:30 01:50 pm Kolev, N. Bivariate Teissier's Co	88 01:30 01:50 pm Amaya, J. Optimization Modelir	91 01:30 01:50 pm Bernábe, M. Determination of Bin
83 01:50 02:10 pm Nieto, A. A review of unsuperv	86 01:50 02:10 pm García, J. Foreign Exchange Dep	89 01:50 02:10 pm Villaseñor, J. Tests for the Inverse (	92 01:50 02:10 pm Lara-Velázquez, F A Classifier System Us
84 02:10 02:30 pm Ordoñez, F. Determinacion de los	87 02:10 02:30 pm Gonzalez-Lopez, Bayesian Sensitivity A	90 02:10 02:30 pm Urbano, C. Hypothesis testing fo	93 02:10 02:30 pm Mora, D. Simulated annealing f
Session **Aprox1-1** **Approximation / Aproximación** Chair:	Session **LACSC-4-1** **LACSC-Statistical Computing 7** Chair:	Session **Mod4-1** **Modeling / Modelación 4** Chair:	
94 02:35 02:55 pm Vicente, P. Multivariate characte	97 02:35 02:55 pm Centeno, O. Monitoring a survey v	100 02:35 02:55 pm Accinelli, E. A Classification of infi	
95 02:55 03:15 pm Chaverri, F. Analysis of the effecti	98 02:55 03:15 pm Corain, L. Scatter and Joint Dep	101 02:55 03:15 pm Shuklin, G. Methods of mathema	
96 03:15 03:35 pm	99 03:15 03:35 pm Muñoz, S. Development and Exe	102 03:15 03:35 pm Solís, M. Nonparametric estim	
03:35 04:35 pm	**Coffee & posters**		

04:35 pm Salida para el Paseo del Evento/Departure for Conference Tour

6:30 9:00 pm Cena del evento / Conference Dinner

XXI SIMMAC / III LACSC:
Viernes / Friday 2nd

Auditorium					Room 1					Room 2					Lab1				
08:00	09:00	am			08:00	09:00	am			08:00	09:00	am	Tutorial 3		08:00	09:00	am	Tutorial 4	
					Griewank, A.					Gutiérrez, J.					Becerra, N.				
					Nonsmooth Analysis and Optimization via Piecewise Line					Teoremas fundamentales del bienestar económico					The role of data visualization on statistical analysis				
09:00	10:00	am			09:00	10:00	am			09:00	10:00	am	Tutorial 7		09:00	10:00	am	Tutorial 8	
					Griewank, A.					Richter, W.					Reinecke, J.				
					Nonsmooth Analysis and Optimization via Piecewise Line					On p-generalized elliptically contoured distributions					Multiple Imputation with R				
10:00	10:30	am	Café / Coffee break																
Session	Bio2-1		Biomathematics / Biomatemáticas		Session	L4-3-1		LACSC-Recent advances in robust statistics and distribution theory		Session	DA1-1		Data Analysis / Análisis de datos		Session	R-1		Clustering and Visualization of Complex Data	
Chair:					Chair:	0				Chair:	0				Chair:	0			
105	10:30	10:50 am	Prendas, J.	Diagnóstico automátic	109	10:30	10:50 am	Milheiro, V.	Robust inference for t	113	10:30	10:50 am	Ponsot-Balaguer,	Aggregation of levels i	117	10:30	10:50 am	Cabanes, G.	dynamic topological cl
106	10:50	11:10 am	Lloret, M.	Cubrimiento e invariat	110	10:50	11:10 am	Rodrigues, P.	A robust DF-REML fra	114	10:50	11:10 am	Arce, J.	Best point symbolic pr	118	10:50	11:10 am	Rastin, P.	barycentric coordinate
107	11:10	11:30 am	Ferreira, W.	A Mathematical Mode	111	11:10	11:30 am	Mayo-Iscar, A.	Robust clustering app	115	11:10	11:30 am	Vargas, E.	Relative Risk of Cardio	119	11:10	11:30	Gallo, M.	A combined SWATLD-A
108	11:30	11:50 am	Arroyo, J.	Infection model for an	112	11:30	11:50 am	Richter, W.	Grouped p-generalize	116	11:30	11:50 am			120	11:30	11:50	Grozavu, N.	Topological co-clusteri
11:50	2:00	pm	Almuerzo / Lunch																
Auditorium					Room 1					Room 2					Room 3				
Session	Bio1-1		Biomathematics / Biomatemáticas		Session	LACSC-2-1		LACSC-Statistical Computing 5		Session	DJ-1		Simulación Multiagente: Herramien						
Chair:					Chair:					Chair:									
120	01:30	01:50 pm	Alpizar-Jara, R.	An overview on integr	123	01:30	01:50 pm	Bhowmik, A.	Computational Tools f	126	01:30	01:50 pm	Jiménez, D.	Simulación Multiagent					
121	01:50	02:10 pm	Fraguela, A.	Dengue fever models	124	01:50	02:10 pm	Soto, J.	Estimation of a mixed-	127	01:50	02:10 pm	Zúñiga-Rojas, R.	A Brief Survey of Higg					
122	02:10	02:30 pm	Velázquez-Castro	Explicit Spatial Contro	125	02:10	02:30 pm	Aguilar, V.	A New approach to Es	128	02:10	02:30 pm	Sánchez, J.	Operads					
02:30	02:45	pm	Breve descanso / Short break																
02:45	03:30	pm	Closing Conference LACSC																
Cribari-Neto, F.																			
A new log-linear bimodal birnbaum-saunders regression mode																			
Chair:																			
03:30	04:00	pm	Café / Coffee break																
Chair:																			
04:00	05:30	pm	Closing Conference SIMMAC																
Calvo, J.																			
Virtual coarse spaces for domain decomposition methods																			
Chair:																			
05:45	**06:00**	**pm**	**Clausura / Closing session**																

Programa / Program

Lunes/Monday, 26

8:00 – ∞: Inscripciones / Registration: `Faculty of Education`.

Martes/Tuesday, 27

8:00 – ∞: Inscripciones / Registration.

9:30 – 10:00 : Inauguración / Opening ceremony `Auditorium`.

10:00 – 10:30 : Café / Coffee break.

10:30 – 11:15: **Conferencia Inaugural / Opening Plenary Talk**: `Auditorium`.

SORIN, S.: Replicator dynamics: old and new (pág. 164).

11:20 **Conference LACSC 1 (C-LACSC)**: `Auditorium`. WING, K.: Efficient Monte Carlo evaluation of resampling-based hypothesis tests (pág. 207).

12:05 – 1:30 p.m.: Tiempo para almuerzo / Time for lunch.

1:30 **Session: Applications / Aplicaciones 1 (Apl1)**: `Auditorium`.

1:30 – 1:50 ARCE, J. & FRUTOS-ALFARO, F.: Gravitational tests with a Kerr-like metric (pág. 40).

1:50 – 2:10 GÓMEZ, P. & FRUTOS, F.: Perturbing compact objects with mass quadrupole moment (pág. 84).

2:10 – 2:30 OLIVA, G. & FRUTOS, F.: Applications of GRT for compact objects in General Relativity (pág. 131).

1:30 **Session: LACSC-New developments in Genomic Selection and Prediction studies** :`Room/aula 1`.

1:30 – 1:50 DE LOS CAMPOS, D.: Prediction of Complex Human Traits using Big Data from Biobank (pág. 204).

1:50 – 2:10 PEREZ, P.: Genomic Selection: Is it worth the sequence (pág. 229).

2:10 – 2:30 POSEKANY, A.POSEKANY, A.: Outlier detection on microarrays applying Mixtures of Gaussian and heavy-tailed distributions (pág. 232).

1:30 **Session: Numerical Analysis / Análisis Numérico 1 (Num2)**: `Room/aula 2`.

1:30 – 1:50 FONSECA, J.: Scalable Subsurface Simulations with ParFlow (pág. 75).

1:50 – 2:10 CUERVO, C. & GALVIS, J.: A Monte Carlo Approach to Computing Stiffness Matrices Arising (pág. 67).

2:10 – 2:30 SEQUEIRA, F. & GATICA, G. & GATICA, L.: Análisis de una formulación aumentada basada en pseudo-esfuerzo para un modelo Brinkman no lineal de flujo de medios porosos (pág. 157).

2:35 **Session: Numerical Analysis / Análisis Numérico 2 (Num3)**: `Room/aula 2`.

2:35 – 2:55 GUEVARA, J. & CHACON, C.: A Mixed Mimetic Formulation for Porous Media Flow (pág. 95).

2:55 – 3:15 FORERO, W. & HERRERA, E.: Vectorización de Contornos en Imágenes Rasterizadas (pág. 76).

3:15 – 3:35 HERRERA, E. & BARON, C. & CHAUSTRE, S.: Spatial Interpolation of Precipitation (pág. 101).

2:35 **Session: LACSC-Recent advances in Statistical Computing 1 (L3-2)**: `Room/aula 1`.

2:35 – 2:55 CARVAJAL, R.: Parallel Algorithm for a Test of Goodness of Fit of a Rash Type Models (pág. 193).

2:55 – 3:15 PONSOT, E.: An R implementation for the aggregation of factor levels in the binomial logit model (pág. 230).

3:15 – 3:35 INFANTE, S. & LUNA, C. & SÁNCHEZ, L. & HERNÁNDEZ, A.: Estimation of stochastic volatility models using optimized filtering algorithms (pág. 102).

2:35 **Session: Applications / Aplicaciones 2 (Apl2)**: `Auditorium`.

2:35 – 2:55 FERNANDEZ, J.: Metodología para la construcción de modelos de simulación discreta aplicados a la industria (pág. 70).

2:55 – 3:15 CAICEDO, L. & LAMADRID, M.A. & MANJARRES, M.D. & VARGAS, J.J.: Planteamiento de un método para la Optimización del agua utilizada en los lavaderos de autos en la ciudad de Santa, Colombia (pág. 53).

3:15 – 3:35 CALDERÓN, C. & SOLÍS, R. & BUSTILLOS,T. & ARIAS, F.: PROE: Simulación computacional para la determinación de rutas óptimas. (pág. 55).

3:35 – 4:00 pm: Café / Coffee break.

4:00 **Session: Applications / Aplicaciones 3 (Apl3)**: `Auditorium`.

4:00 – 4:20 MARTÍNEZ, I. & SOLÍS, R.: Grafo n-residual Módulo m y su aplicación en la estructuración de Residuos n-ádicos (pág. 119).

4:20 – 4:40 GONZÁLEZ, R. & CAMACHO, C.: A generalized algorithm for the enumeration of chiral and achiral isomers of a n-membered ring monocycloalkane m-polyheteroalkylsubstituted (pág. 87).

4:40 – 5:00 REYES, RA. & CAMACHO.J.A. & CERVANTES, S. & SALAZAR, I. & CRUZ,: An automatic framework for graph-based representation of coronary arteries (pág. 145).

5:00 – 5:20 CAMACHO, F. & GÓMEZ, B. & PINO-PÉREZ, R.: A characterization for dominace plausible rule through of qualitative probabilities (pág. 59).

4:00 **Session: LACSC-Statistical Computing for Data Science (L4-1)**: `Room/aula 1`.

4:00 – 4:20 MARTINEZ, A.: On Some Theoretical Aspects of SABSCOR, SSQCOR, SUMCOR and MAXVAR Criteria (pág. 217).

4:20 – 4:40 FIRINGUETTI, L. & PEREIRA-BARAHONA, M.: Bayesian Estimation of the Shrinkage Parameter in Ridge Regression (pág. 206).

4:40 – 5:00 CARVAJAL, R.: A Parallel / Distributed Algorithm for Bootstrapping (pág. 195).

4:00 **Session: Numerical Analysis / Análisis Numérico 3 (Num4)**: `Room/aula 2`.

4:00 – 4:20 REYES, A.: Energy-based and PDE methods for image restoration (pág. 146).

4:20 – 4:40 SEGURA, E.: Computation of Matrix p-th Root Using Solvents (pág. 155).

4:40 – 5:00 CASTRO, E. & ARGUEDAS, V.: Some Examples of Dense Sequences and its Relation to Some Spaces of Functions (pág. 61).

5:00 – 5:20 LOBO, J. & VILLALOBOS, M.A.: Zeroes of functions of Fresnel complementary integral type (pág. 118).

6:10 pm: 6:10 p.m. Welcome Toast

Miércoles/Wednesday, 28

8:00 **Session: Tutorial 1 (Tut)**: `Auditorium`.

VICENTE-GALINDO, P. & NIETO-LIBRERO, A.B. : Recent Developments in Analysis of Three-Way Data (pág. 176).

8:00 **Session: Tutorial 3 (Tut)**: `Laboratory 1`.

BECERRA, N. & QUINTAS, I.: the role of data visualization on statistical analysis (pág. 46).

8:00 **Session: Tutorial 4 (Tut)**: `Room/aula 2`.

GUTIÉRREZ, J.: Teoremas fundamentales del bienestar económico (pág. 96).

8:00 **Session: Tutorial 7 (Tut)**: `Room/aula 1`.

GRIEWANK, A. & WALTHER, A.: A new successive piecewise linearization method for nonsmooth optimization problems. (pág. 91).

9:00 **Session: Tutorial 2 (Tut)**: `Auditorium`.

NIETO, A.B. & VICENTE, P. : A review of unsupervised methods in data mining for pattern recognition (pág. 124).

9:00 **Session: Tutorial 5 (Tut)**: `Room/aula 2`.

RICHTER, W.: On p-generalized elliptically contoured distributions (pág. 148).

9:00 **Session: Tutorial 6 (Tut)**: `Laboratory 1`.

REINECKE, J.: Multiple Imputation with R (pág. 143).

10:30 **Session: Conference 2 (CONF)**: `Auditorium`.

GATICA, G.: Mixed finite element and related methods for some nonlinear problems in fluid mechanics (pág. 83).

10:30 **Session: LACSC-Statistical Computing 4 (LACSC-1)**: `Room/aula 1`.

10:30 CHOULAKIAN, V.: Optimality of the mean absolute deviation about the mean and its generalization to higher-way arrays (pág. 200).

10:50 QU, L.: Copula Density Estimation by Chebyshev Orthogonal Polynomials at the Padua Points (pág. 234).

11:10 TITA, D. & LOUZADA, F. & ANAYA-IZQUIERDO, K.: Minimum quadratic distance estimation based on the cumulative hazard function (pág. 240).

11:30 – 11:50 GONZÁLEZ, E. & VILLASEÑOR, J.A.: An R package for testing goodness of fit (pág. 85).

11:50 – 12:10 CEVALLOS, G. & CEVALLOS, H. & VAN, S. & SALIBIAN-BARRERA, M.: On Fast Computation of Robust Subspace Estimators (pág. 197).

11:20 **Session: Conference 3 (CONF)**: `Auditorium`.

SCHUBERT, L. & SCHUBERT, G.: Estimation of holding periods applied to the case of short and leveraged ETFs (pág. 153).

11:20 **Session: Conference 7 (CONF)**: `Room/aula 2`.

GRIEWANK, A. & WALTHER, A.: A new successive piecewise linearization method for nonsmooth optimization problems. (pág. 89).

12:05 1:30 p.m.: Tiempo para almuerzo / Time for lunch.

1:30 **Session: Modeling / Modelación 1 (Mod1)**: `Auditorium`.

1:30 – 1:50 BACA, G. & GUTIERREZ, J.D J.: Los modelos de Cournot y Stakelberg como equilibrios de Nash (pág. 45).

1:50 – 2:10 SÁNCHEZ, I. & BACA, G.: Juegos dinámicos y el comportamiento en las organizaciones industriales. (pág. 151).

2:10 – 2:30 GUTIÉRREZ, J. & BACA, G.I. & SÁNCHEZ, I.: The results of the duopolio in the cournot and stackelberg models as nash equilibria (pág. 97).

1:30 **Session: LACSC-Application of classification techniques in the study of student's success in propauedeutic college math courses (L3-3)**: `Room/aula 1`.

1:30 – 1:50 ARGUEDAS, S. & FERNÁNDEZ, C.: Application of classification techniques in the study of student's success in propauedeutic college math courses (pág. 189).

1:50 – 2:10 JIMÉNEZ, J. & PÉREZ, P. & LÓPEZ, M.: A bayesian hierarchical model for extreme concentration of carbon monoxide pollution in Mexico City (pág. 211).

2:10 – 2:30 MELNYKOV, I. & MELNYKOV, V.: Formal implementation of the semi-supervised K-means algorithm with hard constraints (pág. 219).

1:30 **Session: Operations Research / Investigación de Operaciones (OR1)**: `Room/aula 2`.

1:30 – 1:50 SORIA, I. & MEJÍA, C.: Modelo de optimización integral para programación de autobuses y tripulaciones de manera simultánea (pág. 163).

1:50 – 2:10 HERNÁNDEZ, F.J. & GUEVARA, A.P.: Optimización y Aplicaciones (pág. 98).

2:35 **Session: Optimization / Optimización 1 (Opt1)**: `Room/aula 2`.

2:35 – 2:55 MORALES, L.: Optimización de diseños sobresaturados con dos niveles a través de técnicas metaheurísticas (pág. 123).

2:55 – 3:15 SOTO, J.: Frobenius Norm: Advances and Trends in Optimization Problems in Signal Processing (pág. 166).

3:15 – 3:35 VÍLCHEZ, E. & LÓPEZ, E.: Método moderno de optimización combinatoria apoyado con el software Wolfram Mathematica (pág. 178).

2:35 **Session: Classification / Clasificación (Clas1)**: `Room/aula 3`.

2:35 – 2:55 REYES, A.: A word-distance based method for Maya language classification (pág. 144).

2:55 – 3:15 ARCE, J. E. & JIMÉNEZ, D.: Point cloud registration: matching a maximal common subset on 2D pointclouds with noise (pág. 41).

2:35 **Session: LACSC-Statistics and Applications (L3-4)**: `Room/aula 1`.

2:35 – 2:55 LOUZADA, F. & BAZÁN, J. & TORRES-AVILÉS, F. & SUZUKI, A.: Power and Reversal Power Links for Binary Classification Modelling (pág. 216).

2:55 – 3:15 GRANZOTTO, D.: The risk of amputation of the lower extremities of patients with diabetes mellitus: an application using the e-extended regression family of models (pág. 210).

3:15 – 3:35 NASCIMENTO, D. & TOOSTANI, I. & VAZIRI, Z. & COLACIQUE, M. & LOUZADA, F. & DELBEM, A. & BARROS, C. & OLIVEIRA, A. & HYPPOLITO, M. & LEITE, J.: Statistics Without Borders: Overpassing between conscious and brain response (pág. 227).

2:35 **Session: Modeling / Modelación 2 (Mod2)**: `Auditorium`.

2:35 – 2:55 SANCHEZ, F. & CALVO, J. & SEGURA, E.: A Partial Differential Equation Model With Age-Structure and Nonlinear Recidivism: Conditions for a Backward Bifurcation and a General Numerical Implementation (pág. 150).

2:55 – 3:15 MURILLO, W.: Estimación de tasas de transmisión de fiebre dengue con un modelo epidemiológico (pág. 127).

3:15 – 3:35 BRAVO, C. & SASA, M. & CORDOVEZ, J.M. & RENJIFO, C.: Modelamiento matemático de la incidencia del accidente ofídico en Costa Rica (pág. 50).

4:00 **Session: LACSC-Forecasting using Simulation (L4-2)**: `Room/aula 1`.

4:00 – 4:20 MUÑOZ, D.: Estimation of expectations in two-level nested simulation experiments (pág. 224).

4:20 – 4:40 QUIROZ, Z. & SAL Y ROSAS, V. & BAYES, C. & VALDIVIESO, L.: Bayesian spatial inflated beta regression model for assessment of reading level in school districts of Peru (pág. 235).

4:40 – 5:00 CHRISTEN, J. & MONTESINOS, C. & SANTANA-CIBRIAN, M.: On optimal direction gibbs sampling (pág. 201).

5:00 – 5:20 LÓPEZ, E. & MONTIEL, L.: A model to estimate the probability of wining the USA presidential elections (pág. 215).

4:00 **Session: Dynamical Systems / Sistemas Dinámicos (DS)**: `Room/aula 2`.

4:00 – 4:20 SOLIS, H. & BARRANTES, E. G.: A new cryptographic system using the Duffing map. (pág. 162).

4:20 – 4:40 FERREIRA, F. & CUEVAS, C. & SOTO, H.: Bounded solutions of Volterra functional difference equations (pág. 72).

4:40 – 5:00 HERRERA, D. & MACÍAS, F. & AHUATZI, J.G.: Mathematical models in topology (pág. 100).

4:00 **Session: Financial Mathematics / Matemática Financiera (FinMat)**: `Room/aula 3`.

4:00 – 4:20 CAMPOS, W.: Técnica de Geometría Fractal para el Cálculo del Riesgo Operacional (pág. 60).

4:20 – 4:40 CUEVAS, C.CUEVAS, C. & ÍÑIGO, J. & JIMÉNEZ, R.: Application of the normal distribution in finance: from statistical modelling to risk assessment.Application of the normal distribution in finance:from statistical modelling to risk assessment. (pág. 68).

4:00 **Session: Modeling / Modelación 5 (Mod5)**: `Auditorium`.

4:00 – 4:20 DUMARESQ, Y. & DE OLIVEIRA, C. & DA CUNHA, F.R.: FLOW OF A FERROFLUID IN A DRIVEN CAVITY (pág. 69).

4:20 – 4:40 ITURRARAN, U. & MUÑOZ-GARCÍA, A.M.: Waveform inversion of poststacked reflection seismic data using Artificial Neural Networks (pág. 103).

4:40 – 5:00 OVALLE, D. & OLAYA, J.: An ANOVA test for functional data applied to fine particulate matter measurements on air (pág. 134).

5:00 – 5:20 CRIADO, R. & ROMANCE, M.: New perspectives on multilayer and multiplex networks as structures to model real world systems (pág. 65).

Jueves/Thursday, 01

8:00 – 9:00 **Session: Tutorial 1 (Tut)**: `Auditorium`.

VICENTE-GALINDO, P. & NIETO-LIBRERO, A.B. : Recent Developments in Analysis of Three-Way Data (pág. 176).

9:00 – 10:00 **Session: Tutorial 1 (Tut)**: `Auditorium`.

NIETO, A.B. & VICENTE, P. : A review of unsupervised methods in data mining for pattern recognition (pág. 124).

8:00 – 9:00 **Session: Tutorial 3 (Tut)**: `Laboratory 1`.

BECERRA, N. & QUINTAS, I.: the role of data visualization on statistical analysis (pág. 46).

8:00 – 9:00 **Session: Tutorial 4 (Tut)**: `Room/aula 2`.

GUTIÉRREZ, J.: Teoremas fundamentales del bienestar económico (pág. 96).

8:00 – 9:00 **Session: Tutorial LACSC (SC-LACSC)**: `Room/aula 1`.

CHERUBINI, U. & MULINACCI, S.: Copula Functions (pág. 199).

9:00 – 10:00 **Session: Tutorial 5 (Tut)**: `Room/aula 2`.

RICHTER, W.: On p-generalized elliptically contoured distributions (pág. 148).

9:00 – 10:00 **Session: Tutorial 6 (Tut)**: `Laboratory 1`.

REINECKE, J.: Multiple Imputation with R (pág. 143).

10:30 – 11:15 **Session: Conference 4 (CONF)**: `Auditorium`.

CHOWELL, G.: Infectious Disease Modeling (pág. 63).

11:20 – 12:05 **Conference LACSC 2 (C-LACSC)**: `Room/aula 1`.

ALLEN, G.: Inference, Computation, and Visualization for Convex Clustering and Biclustering (pág. 35).

11:20 – 12:05 **Session: Conference 5 (CONF)**: `Auditorium`.

WIDLUND, O.: Recientes avances en M etodos de Descomposici on de Dominios (pág. 182).

12:05 1:30 p.m.: Tiempo para almuerzo / Time for lunch.

1:30 – 1:50 **Session: Data Mining / Minera de Datos (DMin)**: `Auditorium`.

1:30 – 1:50 CALVO, D.: Aplicación de los modelos: Bayes y K-nn, en un score para tarjetas de crédito (pág. 58).

1:50 – 2:10 NIETO, A. B. & GONZÁLEZ, N. & GALINDO, M. P.: Recent Developments in Analysis of Three-Way Data in R (pág. 125).

2:10 – 2:30 ORDOÑEZ, A.: Determinacion de los Factores Influyentes en Accidentes de Transporte de Carga Un analisis exploratorio para la identicacion de patrones (pág. 132).

1:30 **Session: LACSC-Copulas (LACSC-3)**: `Room/aula 1`.

1:30 – 1:50 KOLEV, N.: Dependence Analysis via Copulas - Nonstandard View (pág. 213).

1:50 – 2:10 GARCÍA, J.: Foreign Exchange Dependence through di erent Copula Models (pág. 208).

2:10 – 2:30 GONZALEZ, V.: Bayesian Sensitivity Analysis for Asymmetric Copulas with Cubic Sections (pág. 209).

1:30 **Session: Modeling / Modelación 3 (Mod3)**: `Room/aula 2`.

1:30 – 1:50 AMAYA, J.: Optimization Modeling for Resources Allocation in Public Education Systems (pág. 38).

1:50 – 2:10 VILLASEÑOR, J. & GONZÁLEZ-ESTRADA, E.: Tests for the Inverse Gaussian distribution hypothesis (pág. 180).

2:10 – 2:30 URBANO, C. & OLAYA, J.: Hypothesis testing for mean comparison of two non-independent functional populations (pág. 169).

1:30 **Session: Optimization / Optimización 2 (Opt2)**: `Room/aula 3`.

1:30 – 1:50 BERNÁBE, M. & GONZÁLEZ, R. & OCHOA, A. & ESTRADA, M. & SÁNCHEZ, A.: Determination of Bin Packing Algorithm for moving cattle of different size and age (pág. 48).

1:50 – 2:10 LARA, P.: A Classifier System Using Soft Graph Coloring (pág. 109).

2:10 – 2:30 MORA, D. & VILLALOBOS, M.A.: Simulated annealing for maximization of efficiency on dairy cattle attribute monitoring process. (pág. 121).

2:35 **Session: Modeling / Modelación 4 (Mod4)**: `Room/aula 2`.

2:35 – 2:55 ACCINELLI, E. & PUCHET, M.: A Classification of infinite dimensional Walrasian economies and the economic crisis (pág. 33).

2:55 – 3:15 SHUKLIN, G.: Methods of mathematical managing theory in the forecast of dynamic price share on a stock market (pág. 158).

3:15 – 3:35 SOLÍS, M. & LOUBES, J. & MARTEAU, C.: Nonparametric estimation of the first order Sobol indices with bootstrap bandwidth (pág. 160).

2:35 **Session: LACSC-Statistical Computing 7 (LACSC-4)**: `Room/aula 1`.

2:35 – 2:55 CENTENO, O.: Supervisando una encuesta con Google (pág. 196).

2:55 CORAIN, L. & PERUFFO, A. & SALMASO, L.: Scatter and Joint Dependence Ranking of Multivariate Populations with Applications to Brain Cytoarchitecture Complexity (pág. 202).

3:15 – 3:35 MUÑOZ, S.: Development and Execution of a technique to generate models of Credit Scoring (pág. 225).

2:35 **Session: Approximation / Aproximación (Aprox1)**: `Auditorium`.

2:35 – 2:55 VICENTE, M. P. & PATINO, M. C. & GALINDO, M. P.: Multivariate characterization of the profiles of rural women workers (pág. **??**).

2:55 CHAVERRI, F. & FRUTOS-ALFARO, F. & GÓMEZ-OVARES, P. & OLIVA-MERCADO, G.A.: Analysis of the effective potential and the stable orbits of a Kerr-like metric with quadrupole (pág. 62).

3:35 **Session: Poster/Carteles (Cart)**: `Posters`.

3:35 – 4:35 ACOSTA, A. & DEL CALLEJO–CANAL, D. & CANAL–MARTÍNEZ, M.: Comparación entre el análisis de correspondencias y el análisis de correspondencias no simétrico para variables ordinales (pág. 186).

3:35 – 4:35 AGUILAR, J. E.: Estimation of the attenuation coefficient of compacted selected fill used in the construction of a radiotherapy bunker (pág. 34).

3:35 – 4:35 ALONZO, I. & SALGADO J.: Modelo Estocástico para la Estimación del Coeficiente Semafórico (pág. 36). Salgado Salgado, Javier Alexander

3:35 – 4:35 ARCINIEGAS, S. & GARCÍA-PEÑA, M. & KRZANOWSKI, W.J.: An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: Some new aspects (pág. 43).

3:35 – 4:35 CORVALÁN, A. & CARDO, R.: Driving certain Impulsive Biological Systems with Maximal Operators (pág. 64).

3:35 – 4:35 DUARTE, C.: Análisis clásico y Bayesiano de series temporales para la predicción del precio del café (pág. 205).

3:35 – 4:35 GARCIA, J. & MUJICA, R.D.: Aplicación del MANOVA a Aplicación de MANOVA en Experimentos en Parcelas Divididas (pág. 81).

3:35 – 4:35 GARCÍA, M. & ARCINIEGAS, S. & BARBIN, D.: Climate data imputation using the singular value decomposition: An empirical comparison (pág. 82).

3:35 – 4:35 GONZÁLEZ, R. & HERNÁNDEZ, K. & WATERHOUSE, P. & HERRERA, Ó.A.: Design and construction of an experimental methodology for the detection of radical-pair reactions in butterflies (pág. 86).

3:35 – 4:35 LACAYO, R.: Exact queue length distribution in an equilibrium M/G/1 queue with Weibull Service (pág. 214).

3:35 – 4:35 LARA, P.: Solving the soft graph coloring problem using scatter search (pág. 106).

3:35 – 4:35 LARA, P.: Classical Coloring Problems Using the Soft Graph Coloring Model (pág. 108).

3:35 – 4:35 LARA, P.: A greedy algorithm for redistricting (pág. 110).

3:35 – 4:35 LARA, P.: Solving the portfolio optimization problem using swarms (pág. 112).

3:35 – 4:35 LARA, P.: Development of a hybrid method based on k-means, consensus and ant system for classifying 60 Mexican Universities (pág. 114).

3:35 – 4:35 LARA, P.: Development of a hybrid method based on Method of the Musical Composition, Ant System and dual simplex algorithm for solving the vehicle routing problem with windows of time (VRP-TW) (pág. 115).

3:35 – 4:35 MORALES, E. & DEL CALLEJO-CANAL, D. & CANAL-MARTÍNEZ, M.: Comparison of changes of scale for the simultaneous analysis of variables, qualitative and quantitative, using multivariate statistical methods. (pág. 221).

3:35 – 4:35 MUJICA, R. & GARCÍA, J.: Dominio afectivo de la estadística en estudiantes universitarios (pág. 223).

3:35 – 4:35 NASCIMIENTO, D. & LUIZ, P. & LOUZADA, F.: Generalized Gamma distribution: Different methods of estimations (pág. 228).

3:35 – 4:35 OJEDA, S. & BRITOS, G. & VALLEJOS, R.: An Image Quality Index Based on Coefficients of Spatial Association with an Application to Image Fusion (pág. 128).

3:35 – 4:35 .OLAYA, O. & CUEVAS, J.L. & MURILLO, J: A functional regression model for the prediction of PM2:5 from PM10 (pág. **??**).

3:35 – 4:35 QUINTAS, I. & BECERRA, N.: Algunas técnicas para la detección de la falta de homogeneidad en series de datos (pág. 139).

3:35 – 4:35 RIVAS, M.: Estudio fractal de la serie de tiempo de la variable CO2 en el volcan de San Salvador-año 2005 (pág. **??**).

3:35 – 4:35 SOTO, C. & CASTRO, J.C.: Hierarchical Bayesian Model for Diseases Frecuency in Costa Rica (pág. 167).

3:35 – 4:35 SUNJAY, S.: Bootstrap technique for hydrocarbon exploration and production (pág. 238).

3:35 – 4:35 VÍQUEZ, J.VÍQUEZ, J.: Estimation of the exchange rate between Costa Rica and the United States via Levy processes (pág. 181).

4:35 p.m. Salida para la Cena del evento / Departure for Conference Dinner : "Ram Luna" Restaurant

6:30 – 9:00 p.m.: Cena del evento / Conference Dinner.

Viernes/Friday, 02

8:00 – 9:00 **Session: Tutorial 1 (Tut)**: `Auditorium`.

VICENTE, P.VICENTE-GALINDO, P. & NIETO-LIBRERO, A.B. : Recent Developments in Analysis of Three-Way Data (pág. 176).

9:00 – 10:00 **Session: Tutorial 1 (Tut)**: `Auditorium`.

NIETO, A.B. & VICENTE, P. : A review of unsupervised methods in data mining for pattern recognition (pág. 124).

8:00 – 9:00 **Session: Tutorial 3 (Tut)**: `Laboratory 1`.

BECERRA, N. & QUINTAS, I.: the role of data visualization on statistical analysis (pág. 46).

8:00 – 9:00 **Session: Tutorial 4 (Tut)**: `Room/aula 2`.

GUTIÉRREZ, J.: Teoremas fundamentales del bienestar económico (pág. 96).

8:00 – 9:00 **Session: Tutorial 7 (Tut)**: `Room/aula 1`.

GRIEWANK, A. & WALTHER, A.: A new successive piecewise linearization method for nonsmooth optimization problems. (pág. 91).

9:00 – 10:00 **Session: Tutorial 5 (Tut)**: `Room/aula 2`.

RICHTER, W.: On p-generalized elliptically contoured distributions (pág. 148).

9:00 – 10:00 **Session: Tutorial 6 (Tut)**: `Laboratory 1`.

REINECKE, J.: Multiple Imputation with R (pág. 143).

10:30 – 10:50 **Session: Biomathematics / Biomatemáticas 2 (Bio2)**: `Auditorium`.

10:30 – 10:50 PRENDAS, J. & FIGUEROA, G. & RAMÍREZ, M. & CALDERÓN, R.A. & RAMÍREZ, M. & TRAVIESO, C.M.: Diagnóstico automático de infestación por Nosemiasis en abejas melíferas mediante procesado de imágenes. (pág. 137).

10:50 – 11:10 LLORET, M. & NESCOLARDE–SELVA, J.A. & GONZÁLEZ–FRANCO, L.: Cubrimiento e invariabilidad: una teoría sistémica en ecosistemas (pág. 117).

11:10 – 11:30 FERREIRA, W.: A Mathematical Model for Mimicry Dynamics in Evolutionary Ecology (pág. 74).

11:30 – 11:50 ARROYO, J. & SANCHEZ, F.: Infection model for analyzing biological control of coffee rust using bacterial antifungal compounds (pág. 44).

10:30 **Session: Data Analysis / Análisis de datos (DA1)**: `Room/aula 2`.

10:30 – 10:50 PONSOT, E. & PENALOZA, Y.: Aggregation of levels in the logit model: an application to the endotracheal extubation protocol of neonatal patients (pág. 135).

10:50 – 11:10 ARCE, J. & RODRÍGUEZ, O.: Best point symbolic principal components for interval-valued variables (pág. 42).

11:10 – 11:30 VARGAS, E. & CASTRO, M.: Relative risk of cardiovascular disease and type 2 diabetes in people with metabolic syndrome: systematic review and meta-analysis (pág. 170).

10:30 **Session: LACSC-Recent advances in robust statistics and distribution theory (L4-3)**: `Room/aula 1`.

10:30 – 10:50 MILHEIRO, V.: Robust inference for the ROC curve (pág. 220).

10:50 – 11:10 RODRIGUES, P.: A robust DF-REML framework for genetic association studies (pág. 237).

11:10 – 11:30 MAYO, A.: Robust clustering approaches based on the joint application of trimming and constrains (pág. 218).

11:30 – 11:50 RICHTER, W.: Grouped p-generalized spherical distributions (pág. 236).

10:30 **Session: Clustering and Visualization of Complex Data (R)**: `Room/aula 3`.

10:30 – 10:50 CABANES, C. & VERDE, R. & BENNANI, Y.: dynamic topological clustering of distributional data (pág. 51).

10:50 – 11:10 RASTIN, P. & MATEI. B: barycentric coordinates approach for complex data clustering (pág. 141).

11:10 – 11:30 GALLO, G.: A combined SWATLD-ALS algorithm for three-way compositional data (pág. 79).

11:30 – 11:50 GROZAVU, G. & ROGOVSCHI, N. & BENNANI, Y.: Topological co-clustering and vizualisation for heterogenoeous data (pág. 93).

12:00 1:30 p.m.: Almuerzo / Lunch.

1:30 **Session: Biomathematics / Biomatemáticas 1 (Bio1)**: `Auditorium`.

1:30 – 1:50 ALPIZAR, R.: An overview on integrated population dynamics models (pág. 37).

1:50 – 2:10 FRAGUELA, A. & PLIEGO, E. & VELAZQUEZ, J.: Dengue fever models and optimal control in the life-cycle of Aedes aegypti mosquitoes (pág. 77).

2:10 – 2:30 VELÁZQUEZ, J. & BONILLA-CAPILLA, B. & ANZO-HERNÁNDEZ, A. & SOTO-BAJO, M.: Explicit Spatial Control Strategies for Dengue, Zika and Chicunguya outbreaks (pág. 172).

1:30 **Session: LACSC-Statistical Computing 5 (LACSC-2)**: `Room/aula 1`.

1:30 – 1:50 BHOWMIK, A. & VARGHESE, E. & JAGGI, S.: Computational Tools for Generation of Cost-effective Run Orders Useful for Agricultural and Industrial Research (pág. 191).

1:50 – 2:10 SOTO, J. & INFANTE, S. & CAMACHO, F.: Estimation of a mixed-effects model using a partially observed diffusion process (pág. 165).

2:10 – 2:30 AGUILAR, V. & CÓRDOVA, K. & DELGADO, L. & ECHARRY, E. & MÁRQUEZ, I. & RAMOS, S.: A new approach to estimate the spatial impact of climate change on metaxenic disease transmission (pág. 188).

1:30 **Session: Simulación Multiagente: Herramientas y Aplicaciones (DJ)**: `Room/aula 2`.

1:30 – 1:50 JIMÉNEZ, D.: Simulación Multiagente: Herramientas y Aplicaciones (pág. 104).

1:50 – 2:10 ZÚÑIGA-ROJAS, R.A.: A Brief Survey of Higgs Bundles (pág. 183).

2:10 – 2:30 SÁNCHEZ, J.: Operads (pág. 152).

2:30 2:45 p.m.: Breve descanso / Short break

2:45 – 3:30 **Session: Conference LACSC (C-LACSC)**: `Auditorium`.

CRIBARI, F.: A new log-linear bimodal Birnbaum-Saunders regression model (pág. 203).

03:30 04:00 p.m.: Café / Coffee break.

4:00 – 4:45 **Conference 6 (CONF)**: `Auditorium`.

CALVO, J.G.: Virtual coarse spaces for domain decomposition methods (pág. 57).

5:45 – 6:00 p.m.: Clausura / Closing session.

6:00 p.m.: Closing Toast

Contributions List[1]

XXI Int. Symposium on Mathematical Methods Applied to the Sciences 32

[1] In strict alphabetic order according to the name of the author of the contribution.

Abstracts

XXI SIMMAC

XXI International Symposium on Mathematical Methods Applied to the Sciences

A Classification of infinite dimensional Walrasian economies and the economic crisis[I]

Invited Session / Sesion invitada

ACCINELLI, ELVIO[II] Puchet, Martín [III]

México

We consider pure exchange economies whose consumption spaces are Banach Lattices. Utility functions are strictly concave, Gateaux differentiable, and not necessarily separable. Following the Negishi approach and by using the excess utility function we introduce the set of social equilibria. We show that there exists a bijective correspondence between this set and the set of Walrasian equilibria. By transforming the infinite dimensional problem of finding Walrasian equilibria into an equivalent finite dimensional problem of finding social equilibria, we allow ourselves to use techniques of smooth functional analysis. We show that a suitable large subset of economies are regular and its equilibrium set is a Banach manifold. Finally, we focus on the complement of this set, i.e. the set of singular economies, and we analyze its main characteristics, among them those of being the causes of the economic crises.

Keywords: Negishi approach, Banach spaces, Banach manifolds, singular economies.

References

[1] Accinelli,E.; Covarrubias, E. (2014) "An extension of the Sard Samle Theorem to convex domain with an empty interior." *Journal of Matehmatical Economics* 55, 123-128

[2] Accinelli, E. (2013) "The equilibrium set of infinite dimensional Walrasian economies and the natural projection" *Journal of Mathematical economics* 49, 435-440.

[3] Balasko, Y. (2009) "The equilibrium manifold: postmodern developments in the theory of general economic equilibrium", Arne Ryde Memorial Lectures, MIT Press.

[I] Thursday/Jueves, 2:35p.m., Room/aula 2, Session: Mod4, Modeling / Modelacion 4

[II] Universidad Autónoma de San Luis Potosí, Economía, México, San Luis Potosí, elvio.accinelli@eco.uaslp.mx

[III] Facultad de Econom´ia, Divisi´on Posgrados, UNAM. E-mail: anyul@servidor.unam.mx

Estimation of the attenuation coefficient of compacted selected fill used in the construction of a radiotherapy bunker[I]

Poster / Cartel

AGUILAR ROMERO, JAVIER ENRIQUE[II] Martinez, Javier Enrique [III]

Honduras

Honduras manifests its characteristics of a third world country in many aspects, such as the construction of medical centers. The construction of radiotherapy bunkers usually involves the use of compacted select fill whose real attenuation coefficient is unknown; the magnitude of this coefficient is vital in determining the amount of radiation escaping from this bunkers. Considering that the cost of building a bunker from this material is generally low in comparison to others, an estimation of a good coefficient would lead to tranquility in the health system of Honduras, as cheap wouldn't imply inefficient. Both parametric and non parametric bayesian inference are used in the estimation of the coefficient.

Keywords: Radiation shielding, radiation protection, radiotherapy, bayesian inference.

References

[1] McGinley, Patton. (2002) *Shielding Techniques for Radiation Oncology Facilities*. Medical Physics Publishing Corporation, United States.

[2] Müller, P.; Quintana, F.A.; Jara, A.; Hanson, T. (2015) *Bayesian Non Parametric Data Analysis*. Springer International Publishing, Switzerland.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Nacional Autonoma de Honduras, Estadistica, Honduras, Tegucigalpa, javier.aguilar@unah.edu.hn

[III] same address

Fast Computation, and Dynamic Visualization for Convex Clustering[I]

Plenary Talk / Conferencia Plenaria

ALLEN, GENEVERA[II]

United States

Hierarchical clustering enjoys wide popularity because of its fast computation, ease of interpretation, and appealing visualizations via the dendogram and cluster heatmap. Recently, several have proposed and studied convex clustering and biclustering which, similar in spirit to hierarchical clustering, achieve cluster merges via convex fusion penalties. While these techniques enjoy superior statistical performance, they suffer from slower computation and are not generally conducive to representation as a dendogram. In this talk, we present new convex (bi)clustering methods and fast algorithms that inherit all of the advantages of hierarchical clustering. Specifically, we develop a new fast approximation and variation of the convex (bi)clustering solution path that can be represented as a dendogram or cluster heatmap. Also, as one tuning parameter indexes the sequence of convex (bi)clustering solutions, we can use these to develop interactive and dynamic visualization strategies that allow one to watch data form groups as the tuning parameter varies. We apply these techniques to examples from text mining and cancer genomics. This is joint work with John Nagorski and Michael Weylandt

Keywords: Clustering, Machine Learning, Statistical Computing, Dynamic Visualization.

[I] Thursday/Jueves, 11:20a.m., Room/aula 1, Session: C–LACSC, Conference LACSC 2

[II] United States, Rice University, Houston, gallen@rice.edu

Modelo Estocástico para la Estimación del Coeficiente Semafórico[I]

Poster / Cartel

ALONZO MATAMOROS, IZHAR ASAEL[II] Salgado Salgado, Javier Alexander[III]

Honduras

Debido a la necesidad de estimar el número óptimo de unidades que ofrecen un servicio de transporte público en las diferentes rutas de circulación en Honduras, se utiliza un modelo de regulación de vehículos, para el cual es necesario conocer el tiempo de recorrido de una unidad de transporte en una ruta. Dicho tiempo hace uso del coeficiente semafórico, este parámetro se define como la probabilidad de que una unidad de transporte no deba detenerse en alguna de las intersecciones reguladas por semáforos en una ruta previamente establecida. En este trabajo se propone la estimación del coeficiente semafórico mediante un enfoque probabilista, modelando el recorrido de una unidad de transporte a través de todos los semáforos de la ruta haciendo uso de un proceso estocástico.

Palabras clave: Transporte Público, Semáforo, Coeficiente Semafórico, Proceso Estocástico.

References

[1] Rabaza, J. (2009) *Tamaño Vehicular Óptimo para el Servicio de Transporte Público de Superficie*. Tesis de Especialidad, Infraestructura del Transporte y del Territorio, Universidad de Catalunya, Barcelona.

[2] Sminth, H.; Hemily, B.; Ivanovic, M. (2005) *Transit Signal Priority: A Planning and Implementation Handbook*, ITS América Department of Transportation.

[3] Shirayev, A. N. (1991).*Probability. (S. A. F. W. G. P. R. Halmos, Ed.)* (Second Edi). Moscow, Russia: Springer.

[4] Rincón, L. (UNAM). (2012). *Introducción a los procesos estocásticos*. Departamento de Matemáticas, Facultad de Ciencias UNAM (First Edit). Mexico DF.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Nacional Autónoma de Honduras, Escuela de Matemática, Departamento de Estadística, Tegucigalpa, Honduras, asael_am@hotmail.com

[III] Unidad de Normativa y Estudios Económicos, Instituto Hondureño de Trasnporte Público, Tegucigalpa M.D.C., Honduras

An overview on integrated population dynamics models[I]

Communication / Ponencia

ALPIZAR JARA, RUSSELL[II]

Portugal & Costa Rica

Integrated population dynamics models have become popularly used during the last decades [1]. These models jointly analyze data on population size and data on demographic parameters. Due to difficulties of incorporating data for parameter estimation in conventional population projection matrix-type models (i.e. Leslie and Lefkovitch), an integrated analyses with a state-space formulation has proven to be very useful [2]. This approach allows inferences about population dynamics accounting for parameter estimates and model uncertainties due to process variation, such as demographic and environmental stochasticity, and observational error. I will highlight some of the main features of these models and their utility for application to wildlife species.

Keywords: abundance, Bayesian inference, capture–recapture, dynamical systems, parameter estimation.

References

[1] Newman, K., Buckland, S.T., Morgan, B., King, R., Borchers, D.L., Cole, D., Besbeas, P., Gimenez, O., and Thomas, L. (2014) "Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-Space Methods". Methods in Statistical Ecology. Springer Science, New York, pp. 215.

[2] Buckland, S.T., Newman, K.B., Thomas, L., and Koesters, N.B. (2004). "State–space models for the dynamics of wild animal populations". Ecological Modelling, 171 (1–2), 157–175.

[I] Friday/Viernes, 1:30p.m., Auditorium, Session: Bio1, Biomathematics / Biomatematicas 1

[II] University of Évora and University of Costa Rica, Mathematics (UEvora) and Statistics (UCR), Portugal and Costa Rica, Évora and San José, alpizar@uevora.pt

Optimization Modeling for Resources Allocation in Public Education Systems[I]

Communication / Ponencia

AMAYA, JORGE[II]

Chile

This work presents an economic decision model for planning in public education, considering the satisfaction of the total demand for education in a municipality and the annual cost involved in the system. Given a set of established schools in a municipality and possible new locations, the aim of the model is to find the optimal supply in terms of which schools should be kept open, closed or consolidated, which grades (or year levels) should be made available and the number of classes that should be provided within each such grade to meet student demand.

The model is dynamic, in the sense that it works on a multiperiod context (several years) and includes specific information about curriculum requirements in terms of the number of hours per year/subject to be provided for each grade level. The objective function essentially seeks to minimize the total fixed and variable costs of the system. This concept can be expressed by a mixed-integer optimization model and in this talk we discuss the mathematical aspects and the solving strategies for practical resolution. We also present practical results on real cases and a computational implementation of the model in High Performance Computing.

This work presents an economic decision model for planning in public education, considering the satisfaction of the total demand for education in a municipality and the annual cost involved in the system. Given a set of established schools in a municipality and possible new locations, the aim of the model is to find the optimal supply in terms of which schools should be kept open, closed or consolidated, which grades (or year levels) should be made available and the number of classes that should be provided within each such grade to meet student demand.

The model is dynamic, in the sense that it works on a multiperiod context (several years) and includes specific information about curriculum requirements in terms of the number of hours per year/subject to be provided for each grade level. The objective function essentially seeks to minimize the total fixed and variable costs of the system. This concept can be expressed by a mixed-integer optimization model and in this talk we discuss the mathematical aspects and the solving strategies for practical resolution. We also present practical results on real cases and a computational implementation of the model in High Performance Computing.

Keywords: Education planning, optimization models

References

[1] Amaya, J. et al. Optimization Modeling for Resource Allocation in the Chilean Public Education System. International Regional Science Review, 39(2), 2016.

[I] Thursday/Jueves, 1:30p.m., Room/aula 2, Session: Mod3, Modeling / Modelacion 3

[II] University of Chile, Center for Mathematical Modeling and Department of Mathematical Engineering, Santiago, Chile, jamaya@dim.uchile.cl

[2] Antunes, A. et al. A Dynamic Optimization Model for School Network Planning. Socio-Economic Planning Sciences 34:101–20, 2000.

[3] Antunes, A. et al. On Solving Complex Multiperiod Location Models using Simulated Annealing. European Journal of Operational Research 130:190–201, 2001.

[4] Church, R. L. et al. Modeling School Utilization and Consolidation. Journal of Urban Planning and Development (ASCE) 119:23–38, 1993.

[5] Teixeira, J. et al. An Optimization–based Study on the Redeployment of a Secondary School Network. Environment and Planning B: Planning and Design 34:296–315, 2007.

Gravitational tests with a Kerr-like metric[I]

Communication / Ponencia

ARCE GAMBOA, JOSÉ RAFAEL[II] Frutos-Alfaro, Francisco[III]

Costa Rica

We revise the classical gravitational tests with a Kerr-like spacetime. This metric represents the spacetime of an object which posses mass, rotation and mass quadrupole moment. All these calculations were carried out until the second order in the parameters corresponding to the mass, rotation and mass quadrupole moment. Since this metric has mass quadrupole moment, it is not necessary to introduce this feature through the expansion of the mass potential which is common in the literature.

Keywords: general relativity, approximation procedures, weak fields.

References

[1] Frutos-Alfaro, F., Retana-Montenegro, E., Cordero-García, I., & González, J. B. (2012). Metric of a slow rotating body with quadrupole moment from the Erez-Rosen metric. arXiv preprint arXiv:1209.6126.

[2] Bodenner, J., & Will, C. M. (2003). Deflection of light to second order: A tool for illustrating principles of general relativity. American Journal of Physics, 71(8), 770-773.

[I] Tuesday/Martes, 1:30p.m., Auditorium, Session: Apl1, Applications / Aplicaciones 1

[II] Universidad de Costa Rica, Escuela de Física, Heredia, Costa Rica, jose.arcegamboa@ucr.ac.cr

[III] Universidad de Costa Rica, Escuela de Física, San José, Costa Rica, frutos62@gmail.com

Point cloud registration: matching a maximal common subset on 2D pointclouds with noise[I]

Communication / Ponencia

ARCE GARRO, JORGE EMMANUEL[II] Jiménez López, David[III]

Costa Rica

We analyze the problem of determining whether 2 given point clouds in 2D, with any distinct cardinality and any number of outliers, have subsets of the same size that can be matched via a rigid motion. This problem is important, for example, in the application of fingerprint matching with incomplete data. We propose an algorithm that, under assumptions on the tolerable noise, guarantees to find corresponding subclouds of the maximum possible size. Our procedure optimizes a potential energy function to do so, which was first inspired in the potential energy interaction that occurs between point charges in electrostatics.

Keywords: Computer vision, pointcloud registration, energy function, fingerprint.

References

[1] J.C.Bazin, Y.Seo, R.Hartley, and M.Pollefeys. Globally optimal inlier set maximization with unknown rotation and focal length. In European Conference on Computer Vision, pages 803–817

[2] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter. Efficient globally optimal consensus maximisation with tree search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2413–2421, 2015.

[3] V. Golyanik, S. Aziz Ali, and D. Stricker. Gravitational approach for point set registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5802–5810, 2016.

[4] M. T. Goodrich, J. S. Mitchell, and M. W. Orletsky. Approximate geometric pattern matching under rigid motions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(4):371–379, 1999.

[5] B. Jian and B. C. Vemuri. A robust algorithm for point set registration using mixture of gaussians. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 2, pages 1246–1251. IEEE, 2005.

[I] Wednesday/Miércoles, 2:55p.m., Room/aula 3, Session: Clas1, Classification / Clasificacion

[II] Universidad de Costa Rica, Escuela de Matemática, San José, Costa Rica, jorgeemmanuel.arce@ucr.ac.cr

[III] Universidad de Costa RIca, San José, Costa Rica, david.jimenezlopez@ucr.ac.cr

Best point symbolic principal components for interval–valued variables[I]

Communication / Ponencia

ARCE, JORGE[II] Rodríguez Rojas, Oldemar[III]

Costa Rica

In the last two decades, principal component analysis was adapted for symbolic data, first in the context of interval-valued data. A number of approaches have been proposed. In [2] and [3], the authors proposed the centers method and the vertices methods to extend the well known principal components analysis method to a particular kind of symbolic objects characterized by multi–valued variables of interval type. Two methods were proposed, a vertices method which uses all the vertices of the observation's hypercube, and a centers method which uses the centroid values.

This paper aims to improve the centers method applying an optimization algorithms in which instead of projecting the centroid value we look for the best point to project in supplementary the vertices. The best point in the sense that it minimizes the distance of the supplementary individuals to that point or the point that generates an principal components analysis with the best inertia in the first components and then from this projecting the vertices as supplementary elements. We obtain interval–valued symbolic principal components which recapture better the internal variation of the observations or maximizes the correlation measures between these principal components and the random variables and/or the observations themselves.

Besides, the reader may use all the methods presented herein and verify the results using the `RSDA` package written in `R` language, that can be downloaded and installed directly from `CRAN` , see [4].

Keywords: Interval-valued variables, Principal Curves, Principal Component Analysis, Symbolic Data Analysis, vertices principal components, centers principal components.

References

[1] Billard, L. & Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining* John Wiley & Sons Ltd, United Kingdom.

[2] Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales á des données de type intervalle. *Rev. Statistique Appliquée, Vol. XLV 3*, 5–24.

[3] Douzal-Chouakria A., Billard L., Diday E. (2011). Principal component analysis for interval-valued observations. *Statistical Analysis and Data Mining, Volume 4, Issue 2*, 229–246.

[4] Rodríguez, O. with contributions from Olger Calderon and Roberto Zuñiga (2017). RSDA – R to Symbolic Data Analysis. R package version 2.0. http://CRAN.R–project.org/package=RSDA.

[5] Rodríguez, O. (2000). *Classićation et Modéles Linéaires en Analyse des Données Symboliques.* Ph.D. Thesis, Paris IX-Dauphine University.

[I] Friday/Viernes, 10:50a.m., Room/aula 2, Session: DA1, Data Analysis / Analisis de datos

[II] Universidad de Costa Rica, CIMPA, Costa Rica, San Pedro, National Bank of Costa Rica, jarceg@bncr.fi.cr;

[III] Universidad de Costa Rica, CIMPA, Costa Rica, San Pedro, oldemar.rodriguez@ucr.ac.cr

An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: Some new aspects[I]

Poster / Cartel

ARCINIEGAS ALARCÓN, SERGIO[II] García-Peña, Marisol[III]
Krzanowski, Wojtek Janusz[IV]

Colombia

A common problem in multi-environment trials arises when some genotypeby-environment combinations are missing. In Arciniegas-Alarcón et al. (2010) we outlined a method of data imputation to estimate the missing values, the computational algorithm for which was a mixture of regression and lower-rank approximation of a matrix based on its singular value decomposition (SVD). In the present paper we provide two extensions to this methodology, by including weights chosen by cross-validation and allowing multiple as well as simple imputation, the considered methods are: GabrielEigen, WGabriel and MIWG(0.01). The three methods are assessed and compared in a simulation study, using a complete set of real data in which values are deleted randomly at different rates. Two data sets are used, a G×E trial with 26 wheat genotypes evaluated in 5 French environments and other, about 18 pea varieties evaluated in 9 different locations in Poland. The quality of the imputations is evaluated using three measures: the Procrustes statistic, the squared correlation between matrices and the normalised root mean squared error between these estimates and the true observed values. The WGabriel and MIWG(0.01) imputation procedures proposed here give the best results for the data matrix in the simulation study. None of the methods makes any distributional or structural assumptions, and all of them can be used for any pattern or mechanism of the missing values.

Keywords: cross-validation, singular value decomposition, imputation, genotype-by-environment interaction, weights.

References

[1] Arciniegas-Alarcón, S.; García-Peña, M.; Dias, C.T.S. and Krzanowski, W.J.(2010). "An alternative methodology for imputing missing data in trials with genotype-by-environment interaction", *Biometrical Letters, n^o 47*, p.1–14.

[2] Arciniegas-Alarcón, S.; García-Peña, M.; Krzanowski, W.J. and Dias, C.T.S.(2014). "An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: some new aspects", *Biometrical Letters, n^o 51*, p.75–88.

[3] Gabriel, K.R. (2002). "Le biplot - outil d´exploration de données multidimensionelles", *Journal de la Société Française de Statistique, v.143, n^o 3-4*, p. 5–55.

[4] Gauch H.G. (2013). "A simple protocol for AMMI analysis of yield trials", *Crop Science, n^o 53*, p. 1860–1869.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Militar Nueva Granada, Departamento de Matemáticas, Colombia, Cajicá, sergio.arciniegas@gmail.com

[III] Departamento de matemáticas, Pontificia Universidad Javeriana, Bogotá, Colombia, luzmara@gmail.com

[IV] College of Engineering, Mathematics and Physical Sciences, Harrison Building, University of Exeter, Exeter, W.J.Krzanowski@exeter.ac.uk

Infection model for analyzing biological control of coffee rust using bacterial antifungal compounds[I]

Communication / Ponencia

ARROYO ESQUIVEL, JORGE[II] Sanchez Peña, Fabio[III]

Costa Rica

Coffee rust is one of the main diseases that affect coffee plantations worldwide. This causes an important economic impact in the coffee production industry. A common method for combating this disease is through the use of copper hydroxide as a fungicide, which may have collateral effects both on the coffee tree and on human health. A novel method for biological control of coffee rust has been proven to be an effective alternative to copper hydroxide fungicides is anti-fungal compounds. In this paper, we develop a spatial stochastic model for the interaction of bacteria and coffee rust populationsin a coffee plantation, analyse equilibria for specific control strategies, compute conditions for stability under specific conditions, and perform numerical experiments under different control strategies.

Keywords: Infection model, biological control, epidemiology, spatial model, coffee rust.

References

[1] Cressey, D. (2013) Coffee rust regains foothold. Nature, 493, 587.

[2] Haddad F, Maffia L, Mizubuti E, and Texeira H. (2009). Biological control of coffee rust by antagonistic bacteria under field conditions in Brazil. Biological Control, 49, 114–119.

[3] Haddad F, Saraiva R, Mizubuti E, Romeiro R, and Maffia L. (2013). Antifungal compounds as a mechanism to control hemileia vastatrix by antagonistic bacteria. Tropical Plant Pathology, 38, 398–405.

[4] Pielaat A. and Bosch F. (1998). A model for dispersal of plant pathogens by rainsplash. Mathematical and Medical Biology, 15, 117–134

[I] Friday/Viernes, 11:30a.m., Auditorium, Session: Bio2, Biomathematics / Biomatematicas 2

[II] Universidad de Costa Rica, Escuela de Matemática, San Pedro de Montes de Oca, Costa Rica, jorgearroyo0731@gmail.com

[III] Universidad de Costa Rica,CIMPA y Escuela de Matemática, Costa Rica, San José, fabio.sanchez@ucr.ac.cr

Los modelos de Cournot y Stakelberg como equilibrios de Nash

Communication / Ponencia

BACA LOBERA, GLORIA IDALIA[I] Gutierrez Ramírez, José de Jesús[II]

México

A market in perfect competition is when there is a large number of companies, the production of each has an insignificant effect on the market. Such a company does not have to worry about the impact of its own production on the sale price of its product. On the contrary, in an oligopoly, a productive sector of n companies, each of which has an appreciable production, the situation is no longer so simple, companies have to consider the effect that their own decisions on production have on the price of the market. In the Cournot model, the decision variables of each company are the level of production, this decision is taken assuming that the level of production of its competitor is fixed. The level of production chosen by each company is "your best response" to what you think the other company will produce. The market price is the result of the interaction of the sum of the individual offers of each company and the market demand for the product. The price turns out to be, therefore, that in which any excess demand or supply is eliminated. This results in an interdependence between the profits of each company, which can be modeled using game theory. In this paper we present a demonstration about the equilibria in the models and their relation with the Nash equilibria for non-cooperative games.

Keywords: non-cooperative games, oligopolies.

Resumen

Un mercado en competencia perfecta es cuando existe un gran número de empresas, la producción de cada una tiene un efecto insignificante en el mercado. Una empresa así no tiene por qué preocuparse acerca del impacto de su propia producción sobre el precio de venta de su producto. Por el contrario, en un oligopolio, un sector productivo de n empresas, cada una de las cuales tiene una producción apreciable, la situación ya no es tan simple, las empresas tienen que considerar al efecto que sus propias decisiones sobre producción tienen sobre el precio de mercado. En el modelo de Cournot las variables de decisión de cada empresa son el nivel de producción, esta decisión se toma suponiendo que el nivel de producción de su competidora está fijo. El nivel de producción elegido por cada empresa es "su mejor respuesta" a lo que cree que producirá la otra empresa. El precio de mercado es resultado de la interacción de la suma de ofertas individuales de cada empresa y de la demanda de mercado por el producto. El precio resulta ser, por tanto, aquel en que se elimina cualquier exceso de demanda o de oferta. Esto da como resultado una interdependencia entre las utilidades de cada empresa, lo cual se puede modelar utilizando la teoría de juegos. En este trabajo se presenta una demostración sobre los equilibrios en los modelos y su relación con los equilibrios de Nash para juegos no cooperativos.

Palabras clave: juegos no cooperativos, oligopolios.

References

[1] Binmore, Ken. (1994). *Game Theory*. OXFORD.

[I]Universidad Autónoma Metropolitana, Producción Económica, México, Ciudad de México, gbaca@correo.xoc.uam.mx
[II]Universidad Autónoma Metropolitana, México Ciudad de México, jjgricb@hotmail.com

The role of data visualization on statistical analysis[I]

Short course / Curso corto

BECERRA RODRÍGUEZ, NOÉ[II] Quintas Pereira, Isabel[III]

México

Nowadays data visualization is considered as an accessory element within statistical analysis. Research reports, both from senior researchers and junior researchers, include data visualizations only as an auxiliary procedure to improve the understanding of the subject reviewed. However, it is necessary to recognize that the origin of the statistical analysis was initially based on the construction of graphs and infographics that allowed to make statistical inference and support the decision-making process.

Later on, the creation of statistical methods and techniques more developed enabled researchers to perform more robust data analysis. Nevertheless, the development of statistical procedures more sophisticated has biased the emphasis towards only to parameters estimations and building of models supported by numeric coefficients. This bias minimizes the role that data visualization plays in the data analysis.

The aim of this work is to rescue the significance of data visualization both as a statistical procedure by itself and as complementary tool for the statistical task. The argument unfolds by presenting and discussing the following topics:

1. ¿What is the meaning of data visualization?
2. Data visualization versus numeric statistical analysis.
3. Data visualization as procedure of exploratory data analysis.
4. Data visualization in the presentation of statistical research results.

Keywords: data visualization, data analysis, exploratory data analysis, statistical research results.

Resumen

Actualmente la visualización de datos se considera como un elemento accesorio del análisis estadístico. Los trabajos de investigación, tanto de académicos como de estudiantes universitarios los incluyen únicamente como un procedimiento auxiliar para mejorar la comprensión del tema estadístico revisado. Sin embargo, es necesario reconocer que el origen del análisis estadístico se basó inicialmente en la construcción de gráficas e infografías que permitieron hacer inferencia y apoyar en la toma de decisiones.

Posteriormente la construcción de técnicas y métodos estadísticos más elaborados permitió realizar análisis de datos más robustos. Empero, la evolución de procedimientos estadísticos cada vez más sofisticados ha sesgado el énfasis al uso casi exclusivo de estimación de parámetros y construcción de modelos respaldados por coeficientes numéricos. Este sesgo minimiza el rol que tiene la visualización de datos en el análisis de datos. El objetivo de este trabajo es resaltar la importancia de la visualización de datos tanto como procedimiento estadístico de análisis en sí mismo y como herramienta complementaria del trabajo estadístico. El argumento se construye con la presentación y discusión de los siguientes tópicos:

[I] Wednesday/Miércoles, 8:00a.m., Laboratory 1, Session: Tut, Tutorial 3

[II] Universidad Autónoma Metropolitana Xochimilco, Producción Económica, México, Ciudad de México, nbecerra@correo.xoc.uam.mx

[III] Universidad Autónoma Metropolitana Xochimilco, Ciudad de México, México, iquintas@correo.xoc.uam.mx

1. ¿Qué es la visualización de datos?
2. Visualización de datos vs análisis estadístico numérico
3. Visualización de datos como procedimiento de análisis exploratorio
4. Visualización de datos en la presentación de resultados estadísticos

Palabras clave:
visualización de datos, análisis de datos, análisis exploratorio de datos, resultados estadísticos de investigación.

References

[1] Chen, Ch., Hardle, W. and Unwin A. (Editors) (2008). Handbook of Data Visualization, Leipzig, Germany, Springer-Verlag

[2] Iliinsky, N. and Steele J. (2011). Designing Data Visualizations, Sebastopol, California, O´Reilly Media.

[3] Gohil, A. (2015). R Data Visualization Cookbook, Birmingham, UK, Packt Publishing.

Bin Packing for moving cattle[I]

Communication / Ponencia

BERNÁBE LORANCA, MARÍA BEATRIZ[II] González Velázquez, Rogelio[III]
Martínez Guzmán, Gerardo[IV] González Velázquez, Rogelio[V] Cruz, Edna[VI]
Ochoa Zezzatti, Alberto[VII]

México

In the state of Veracruz ´s main economic activities are agriculture and livestock, the existence of cattle is at the first place at national level, a very important process for their production and trade is the movement of live animals. To optimize the process of moving an optimization algorithm, implemented in solving the problem of optimal packing dimensional (2D -BPP), recommendations for the process to help reduce the losses to make the transfer of live cattle exposed road A container of sturdy straight truck is presented on a journey from 4 hours to 7 hours at most, several factors influence the choice of cattle to move depending on the order you have intended for animals, whether trade or in the state Veracruz occurs later grazing cattle market introduction, why two choices of accommodation livestock shows.

Keywords: Livestock, cattle, optimum dimensional packing problem, (2D -BPP), Transfer of live animals by road.

Resumen

En el estado de Veracruz, las actividades económicas principales son la agricultura y la ganadería, la existencia de ganado en el primer lugar a nivel nacional, el cual es un proceso muy importante para su producción y comercio en el movimiento de animales vivos. Para optimizar el proceso óptimo de mover el embalaje, se implementa un algoritmo de optimización, packing dimensional (2D - BPP) para atender el proceso de ayudar a reducir las pérdidas al hacer la transferencia de ganado vivo expuesto en la carretera. Un contenedor de camión recto robusto se presenta en un viaje de 4 a 7 horas a máximo, varios factores influyen en la elección del ganado que debe moverse según el orden de los animales destinados, si el comercio o en el estado Veracruz se produce después del pastoreo de ganado introducido al mercado en el mercado, entonces se muestran dos opciones de alojamiento del ganado.

Palabras clave: Ganado, borregos, Problema Packing problem.

References

[1] DE LA SOTA, D. M. D.,; SANITARIAS, D. D. L. (2005). Manual de Procedimientos en el transporte de animales. Dirección Nacional de Sanidad Animal, SENASA. Versión corregida.

[I] Thursday/Jueves, 1:30p.m., Room/aula 3, Session: Opt2, Optimization / Optimizacion 2

[II] Benemérita Universidad Autónoma de Puebla, México, beatriz.bernabe@gmail.com
[III] Maestría en Cómputo Aplicado, Universidad Autónoma de Ciudad Juárez, México,rogelio.gzzvzz@gmail.com
[IV] Benemérita Universidad Autónoma de Puebla, México, Benemérita Universidad Autónoma de Puebla, México
[V] Maestría en Cómputo Aplicado, Universidad Autónoma de Ciudad Juárez, México, rogelio.gzzvzz@gmail.com
[VI] Benemérita Universidad Autónoma de Puebla, México, cbr_ lad7@yahoo.com.mx
[VII] Benemérita Universidad Autónoma de Puebla, México, alberto.ochoa@uacj.mx

[2] Sánchez Gómez J. I., Flores Toro A. R.: Especificaciones para el Transporte de Ganado.Virbac Salud Animal. Pp. 1 (2006).

[3] INEGI. Cría y explotación de animales en Veracruz de Ignacio de la Llave. Censo Agropecuario. Pp. 7–8 (2007).

[4] Ros Piqueras J. M.: Bienestar Animal en el Transporte. Consejería de Agricultura y Agua. Pp. 1 (2008).

[5] Chambers, P. G., Grandin, T., Heinz, G., Srisuvan, T. (2001). Guidelines for humane handling, transport and slaughter of livestock. Rap publication, 4, 13–30.

[6] Guía de Buenas Prácticas para el Transporte de Ganado Porcino. Formación Agraria. Pp. 2-3. (2007).

[7] MÁXIMAS, S. E. P. Y. D., LOS, C., LAS VÍAS, G. D., FEDERAL, C. D. J. NORMA OFICIAL MEXICANA NOM-012-SCT-2-2008. NORMA, 1, 50.

[8] Ensminger, M. E. (1973). Producción bovina para carne. El Ateneo.

[9] Ramos Rama, J. M. (2002). Algunas observaciones de campo sobre cojeras traumáticas en ganado bovino. Producción Animal, 17(178), 26–32.

[10] Referencia web http://www.logisticamx.enfasis.com/articulos/67452-factores-determinar-los-costos-del-autotransporte

Modelamiento matemático de la incidencia del accidente ofídico en Costa Rica[I]

Invited Session / Sesion invitada

BRAVO VEGA, CARLOS ANDRÉS[II] Sasa Marin, Mahmood[III]
Cordovez, Juan Manuel[IV] Renjifo, Camila[V]

Colombia

Snakebite is a public health issue that affects mostly rural populations on tropical areas. Furthermore, in most those countries health system coverage is low, so not all patients reach medical attention: This causes an important underreporting on collected data. In central America and north of South America, the most relevant venomous snake in terms of epidemiology is the terciopelo viper (Bothrops asper, Garman 1884): In Costa Rica this species causes between 60 and 80% of the envenomings, and the total is of 504 events per year. Given the good coverage of health system that exist in Costa Rica, and a correct distribution of antivenom by the Instituto Clodomiro Picado, we evaluated the predictive capacity of a mathematical model in this country.

We did field work to estimate the frequency of encounters between humans and terciopelo vipers on 2 geographic areas in Costa Rica (Pacific and Caribbean), and at 3 different altitudes to extrapolate this variable to all the country. Moreover, we used a mathematical model by compartments to relate this frequency with rural human populations to estimate incidence at a district level. Finally, results were compared with incidence to evaluate model performance. The region with the higher incidence is the south of the pacific, and furthermore lowlands had highest frequency of encounters with venomous snakes. Highlands and northern pacific dry forest had the lowest incidence by Bothrops asper due to the low probability to find one of those snakes in those areas. Mathematical model did a nice work predicting the incidence, showing that it can be improved to be used in other countries to estimate total burden of snakebite and to improve decision making by government authorities.

Keywords: Snakebite, Mathematical model, Terciopelo (Bothrops asper), incidence.

[I] Wednesday/Miércoles, 3:15p.m., Auditorium, Session: Mod2, Modeling / Modelacion 2

[II] Universidad de los Andes, Ingeniería Biomédica, Colombia, Bogotá, ca.bravo955@uniandes.edu.co
[III] Instituto Clodomiro Picado, San Jose, Costa Rica, msasamarin@gmail.com
[IV] Universidad de los Andes, Bogotá, Colombia, jucordov@uniandes.edu.co
[V] CorpoICA, Bogotá, Colombia, mrenjifo@corpoica.org.co

Dynamic topological clustering of distributional data[I]

Invited Session / Sesion invitada

CABANES, GUÉNAEL[II] Verde, Rosanna[III] Bennani, Younes[IV]

France

Histogram data represents a useful tool for summarising sequences of numerical data. Some examples are financial data in economics applications, sensor data for environmental phenomena detection, or energy consumption data loaded by smart meters. Further sources of histogram data arise from Official Statistics Institutes, that make available results of surveys only in form of aggregated or summarised data for preserving the privacy of respondents. The main advantage of histogram data is in the possibility to take into account the shape of the distribution of the data with a smaller loss of information than a synthesis of the observed data by means of the position and scale indexes (for example, by means of an arithmetic mean and a standard deviation only). Formally, histogram data have been introduced in the context of Symbolic Data Analysis (SDA) by Bock and Diday [1]. In SDA, a histogram variable is defined as a particular kind of multi-valued variable (numerical modal), where each observation is represented by an empirical distribution, like a histogram. According to the given definition, a histogram is a sequence of disjoint intervals with associated a set of weights (e.g., relative frequencies). Since a histogram data arises as an empirical distribution, the techniques recently developed for such data refer to distributional data [4]. In the field of research on distributional data analysis, the L2 Wasserstein distance for comparing distributions has shown some interpretative properties. Irpino et al. in [4] proved that the L2 Wasserstein distance between two histograms can be computed by considering the distances between the corresponding centers and radii of the intervals (bins). This type of decomposition simplify some computational aspects and in the same time to use a consistent distance measure with the way to compute the average histograms (or prototype of the clusters). We are particularly interested in clustering methods adapted to the analysis of dynamic histogram data. Indeed, clustering task is a challenging problem. It aims at regrouping data into clusters based on their similarity, without any external or a-priori knowledge. Several difficulties are known for this task, including the selection of a suitable number of clusters to represent the data structure and the capability to detect non-convex clusters. In addition, in many cases, databases are characterized by a dynamic structure over time, new data constantly arriving. Sometimes, the evolution and the mass of data are so important that it is not possible to store it in a database and we need "on the fly" analyses. These processes have been the subject of numerous studies in recent years because of the important potential applications in many fields. However, it is a difficult problem because of the computation and memory cost associated with the volume of data involved. To deal with that type of data, it is essential to construct a condensed description of the dynamic properties of the data, and to be able to detect accurately the variations in the data structure. As this type of data can be seen as an infinite process constantly changing over time, the clustering of these data (i.e. the model of their structure) must also constantly evolve. An adapted clustering algorithm must therefore be able to compute the clustering of the dataset over different periods of time, and be able to compare the data structures over these different periods. The objective of this work is to develop a method of dynamic clustering algorithm for

[I] Friday/Viernes, 10:30a.m., Room/aula 3, Session: R, Clustering and Visualization of Complex Data

[II] Université Paris 13, , Villetaneuse, France, cabanes@lipn.univ-paris13.fr

[III] Dip. Matematica e Fisica, Universitá della Campania Luigi Vanvitelli, Italy

[IV] Université Paris 13, Villetaneuse, France

histogram data. The proposed algorithm is an adapted version of the Self Organizing Map (SOM) [5] algorithm. SOM is a popular non-linear technique for unsupervised learning and data visualization. The learning of a Self-Organizing Map is proposed as an efficient method to address the problem of clustering, especially for high-dimensional data having an input topological structure. The development of a SOM method to histogram data is suitable to analyse data that are already available in aggregated form (like some Official Statistics or confidential data) or they are generated as syntheses of huge amount of original data. The histograms, as empirical distributions, are able to preserve much more information of data than simple means and standard deviations that are usually used to synthesize dataset. In the analysis of histogram data a batch version of SOM has been proposed by [3] based on the L2 Wasserstein distance. Adaptive Wasserstein distances has been also developed in this context to find, automatically, weights for the variables for the whole dataset as well as for each node. However, these methods can provide a quantification and a visualization of symbolic data (intervals, histograms) but they are not able to partition the data in a set of homogeneous clusters. A two-level clustering algorithm based on SOM for interval data has been proposed by [2], that combines the dimension reduction, achieved by a SOM learning step, and groups the data in a reduced space in a certain number of homogeneous clusters. Here, we present an extension of this approach to histogram data. In the partitioning phase is used the L_2 Wasserstein distance, as in dynamic clustering algorithm for histogram data. Following the idea of [2], the number of cluster is not fixed in advance, as parameter of the clustering algorithm, but it is automatically induced by a criterion related to the estimation of a local density of the data in the original space. Tests on simulated data have been performed for showing the capability of the method in preserving the information about the distributions describing each data. In addition, an application on real data shows the usefulness the method. In addition, the proposed algorithm is capable of following the dynamic of the data. The model update itself according to the dynamic of the data structure and "forget" outdated information, using a modified update of the prototypes definition. I that way, the model converge to a stable state when the data structure is stationary, though it is able to change swiftly to adapt to a non–stationary structure. To perform the prototypes clustering, the idea is to learn dynamically an estimation of the local density of data for each prototype, then to use this density to detect clusters boundaries. The number of clusters is detected automatically and clusters of any type of structure can be detected.

Keywords: machine learning, dynamic dataset, histogram data, self-organizing map.

References

[1] H.-H. Bock and E. Diday, Eds., Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer Verlag, Heidelberg, 2000.

[2] G. Cabanes, Y. Bennani, R. Destenay, and A. Hardy, A new topological clustering algorithm for interval data, Pattern Recognition, vol. 46, no. 11, pp. 3030–3039, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320313001520

[3] F. d. A. De Carvalho, A. Irpino, and R. Verde, Batch self organizing maps for interval and histogram data, isi ed. Curran Associates, Inc. (2013), 2013, pp. 143–154.

[4] A. Irpino and E. Romano, Optimal histogram representation of large data sets: Fisher vs piecewise linear approximation, in EGC, ser. Revue des Nouvelles Technologies de l´Information, M. Noirhomme–Fraiture and G. Venturini, Eds., vol. RNTI–E–9. Cépaduès–Éditions, 2007, pp. 99–110.

[5] T. Kohonen, Self-Organizing Maps. Berlin: Springer-Verlag, 2001.

Planteamiento de un método para la Optimización del agua utilizada en los lavaderos de autos en la ciudad de Santa, Colombia[I]

Communication / Ponencia

CAICEDO VEGA, LINDA LUCÍA[II] Lamadrid Guerra, Maria Alejandra[III]
Manjarres Ballestas, Miggles Dayanna[IV] Vargas Sanchez, Jhon Jairo[V]

Colombia

Currently, Colombia is going through a crisis of scarcity of water, in spite of this, the society is not acting to address this problem and on the other hand, companies that use more water for their operations, such as the washing of cars, are those who are creating a negative impact by the excessive use of this resource. This research was carried out in order to propose a method for the auto wash that present significant savings in the consumption of water used in the lavautos of the city of Santa Marta.

Consecuente with the above, to develop research and meet the goal of it, initially assessed the process current that it is taking place by tracking and taking record of all the activities that make up the process. Similarly, were the process flow diagram, and analytical cursograma of the current method, then assessed the consumption of water used in this process, which settled a few meters digital foot water flow measures of water consumption a sample of 54 vehicles. We analysed the data collected in sampling, through the implementation of histograms, parameter estimation and goodness of fit tests. Parameter estimation is performed by the method of maximum likelihood. Then we calculated the cost of consumption.

Taking into account the information, cost-effective alternatives that would allow the optimization of the water used were designed, they were put to the test, where registration of the consumption of water used with its implementation under the same sample of 54 vehicles was taken. The data collected were subjected to evaluation and analysis with the application of techniques of flow analysis, the realization of histograms, tests of goodness of fit and estimation of parameters, as was done to evaluate the current process. Also, estimated the cost of water consumption which is annullable if they should implement.

The next step included statistical comparisons of the data collected from the current process with the data collected in the trial of the designed alternatives, in order to choose the most cost-effective alternative, which would be considered as the method to consider. These comparisons were made through the application of analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) to verify the existence of significant differences between the different methods. Expected that at least one present alternative proposals differences with average less than others. Then to collate what alternative had better averages, test postAnova LSD, and passed to the verification of the assumptions of normality, independence of errors and constant variance giving validity to the results of the ANOVA.

Finally, satisfaction surveys were conducted to measure the fulfilment of the objective of the investigation, whose results were analyzed through comparisons of vectors of proportions to find differences between the best

[I] Tuesday/Martes, 2:55p.m., Auditorium, Session: Apl2, Applications / Aplicaciones 2

[II] Universidad del Magdalena, Facultad de Ingeniería, Santa Marta, Colombia, jvargass@unimagdalena.edu.co
[III] Universidad del Magdalena, Santa Marta, Colombia, maleja.lamadridg@gmail.com
[IV] Universidad del Magdalena, Santa Marta, Colombia, miggles.manjarres@gmail.com
[V] Universidad del Magdalena, Santa Marta, Colombia, jvargass@unimagdalena.edu.co

method and the traditional regarding the quality of washing and customer satisfaction. Additionally, the technique of multivariate statistical analysis correspondences multiple ACM for the comparison of more variables that were presented in the survey designed was used. Were exploratory diagrams as a biplot, subnubes of categories by variables, graphics of inertia, etc. The exploratory analysis allowed to visualize associations and thus the proposed method was associated with good grades.

Keywords: Multivariate analysis of variance, anova, maximun likelihood, optimization water consumption.

Resumen

Actualmente, Colombia se encuentra pasando por una de las crisis más fuerte de escasez de agua, a pesar de ello, la sociedad no está actuando para enfrentar esta problemática y por el contrario, las empresas que más utilizan el agua para sus operaciones, como los lavaderos de autos, son quienes están generando un impacto negativo por el uso desmedido de este recurso. Ésta investigación se realizó con el propósito de plantear un método para el lavado de auto que presente ahorros significativos en el consumo de agua utilizada en los lavautos de la ciudad de Santa Marta. Consecuente con lo anterior, para desarrollar la investigación y cumplir con el objetivo de la misma, inicialmente se evaluó el proceso actual que se está llevando a cabo haciendo un seguimiento y tomando registro de todas las actividades que conforman el proceso. De igual forma, se realizaron el diagrama de flujo de proceso y cursograma analítico del método actual, luego se valoró el consumo de agua utilizada en este proceso, para lo cual se instalaron unos medidores digitales de flujo de agua pata tomar las medidas del consumo de agua a una muestra de 54 vehículos. Se analizaron los datos recolectados en la toma de muestra, mediante la realización de histogramas, pruebas de bondad de ajuste y estimación de parámetros. La estimación de parámetros se realizó por el método de máxima verosimilitud. Seguidamente se calculó el costo de dicho consumo. Teniendo en cuenta la información, se diseñaron alternativas rentables que permitieran la optimización del agua utilizada, fueron puestas a prueba, donde se tomó registro del consumo de agua utilizada con su implementación bajo la misma muestra de 54 vehículos. Los datos recolectados fueron sometidos a evaluación y análisis con la aplicación de técnicas de análisis de flujo, la realización de histogramas, pruebas de bondad de ajuste y estimación de parámetros, tal y como se hizo para evaluar el proceso actual. Asimismo, se calculó el costo de consumo de agua en el que se incurría si se llegasen a implementar. El paso siguiente comprendió las comparaciones estadísticas de los datos recolectados del proceso actual con los datos recolectados en el ensayo de las alternativas diseñadas, con la finalidad de escoger la alternativa más rentable, la cual sería considerada como el método a plantear. Estas comparaciones se realizaron a través de la aplicación de un análisis de varianza (ANOVA) y análisis de varianza multivariado (MANOVA) para verificar la existencia de diferencias significativas entre los diferentes métodos. Se esperaba que al menos uno de las alternativas propuestas presente diferencias con media menor a los demás. Seguidamente para cotejar cuál alternativa presentaba mejores promedios, se realizó la prueba postAnova LSD, y se pasó a la verificación de los supuestos de normalidad, independencia de errores y varianza constante dando validez a los resultados del ANOVA. Por último, se realizaron encuestas de satisfacción para medir el cumplimiento del objetivo de la investigación, cuyos resultados fueron analizados a través comparaciones de vectores de proporciones para encontrar las diferencias entre el mejor método y el tradicional con respecto a la calidad del lavado y satisfacción del cliente. Adicionalmente, se utilizó la técnica de estadística multivariada Análisis de correspondencias múltiples ACM para la comparación de más variables que se presentaron en la encuesta diseñada. Se obtuvieron diagramas exploratorios como biplot, subnubes de categorías por variables, gráficos de inercia, etc. El análisis exploratorio permitió visualizar asociaciones y de esta forma el método propuesto estuvo asociado a buenas calificaciones.

Palabras clave: Análisis multivariado de varianza, anova, máxima verosimilutud, optimización de consumo de agua.

References

[1] Cepeda-Cuervo, E. (2015). Estadística matemática, Bogotá, Colombia: Universidad Nacional de Colombia.

[2] Vargas, J. (2004). Estudio de factibilidad para el montaje de un lavadero de autos con servicios de valor agregado en la Pontificia Javeriana, Colombia: Pntificia Universidad Javeriana.

[3] Díaz, L. G. and Morales, A. M (2012). Análisis Estadístico de datos multivariados, Bogotá, Colombia: Universidad Nacional de Colombia.

[4] González, R. (2012). Consecuencias del efecto invernadero, Twenergy.

[5] Walpole, R. E. and Myers, R. H. and Myers, S. L. and Ye, K (2012). Probabilidad y Estadística para Ingeniería y ciencias, México: Pearson Education.

PROE: Computational simulation for the determination of optimal routes[I]

Communication / Ponencia

CALDERÓN ARCE, CINDY[II] Solís-Ortega, Rebeca[III] Bustillos-Lewis, Timothy[IV]
Arias-Núñez, Fernando[V]

Costa Rica

Planning optimal, efficient and safe routes contributes to a faster and better evacuation in static scenarios such as buildings, urban centers, forests, and other spaces with obstacles and danger areas. In that sense a simulation of agents representing a swarms of robots was implemented for the prior examination of static and unknown scenarios.

Thus, through a process of scanning the simulated swarm will crawl the space and collect data about the environment, for subsequent determination of optimal routes through optimization techniques.

We present the preliminary results of the performance of a cooperative algorithm that allows to explore the environment in which a simulated swarm are. Joint with an algorithm for processing and analysis of the information collected, in order to construct the mapping of the area, and another optimization algorithm for determining optimal routes, evacuation and/or access based on the map constructed.

Keywords: swarm, optimization, simulation, map generation optimal routes.

Resumen

La planificación de rutas óptimas, eficientes y seguras contribuye a una rápida y mejor evacuación en escenarios estáticos, con obstáculos y zonas de peligro fijas pero con ubicación desconocida, como edificios, centros urbanísticos, bosques, entre otros. En ese sentido, se implementó una simulación de agentes que representan enjambres de robots para la examinación previa de escenarios estáticos y desconocidos.

Y mediante un proceso de exploración, el enjambre simulado rastrea el espacio recolectando información sobre el entorno en que se encuentran, para la posterior determinación de rutas óptimas, de evacuación y/o acceso, por medio de técnicas de optimización.

Se presentarán los resultados preliminares de la ejecución de un algoritmo cooperativo que permite explorar de manera eficaz el entorno en el que se encuentra un enjambre simulado de agentes. En conjunto con un algoritmo que procesa y analiza la información recolectada para la construcción un mapa de la zona explorada y un algoritmo de optimización que determina rutas óptimas, de evacuación y/o acceso.

Palabras clave: enjambres, optimización, simulación, generación de mapas, rutas óptimas.

[I] Tuesday/Martes, 3:15p.m., Auditorium, Session: Apl2, Applications / Aplicaciones 2

[II] Instituto Tecnológico de Costa Rica, Escuela de Matemática, Cartago, Costa Rica, ccalderon@itcr.ac.cr
[III] Escuela de Matemática, Instituto Tecnológico de Costa Rica, Cartago, Costa Rica, rsolis@itcr.ac.cr
[IV] Ingeniería en Mecatrónica, ITCR, Cartago, Costa Rica, tbustillostec@gmail.com
[V] Ingeniería en Mecatrónica, ITCR, Cartago, Costa Rica, ariasnunez.fernando@gmail.com

References

[1] A. J. Sharkey and N. Sharkey, "The application of swarm intelligence to collective robots?, Advances in applied artificial intelligence, p. 157, 2006.

[2] S. M. LaValle, "Planning Algorithms", Cambridge University Press, 2006, ch. 5. [En línea]. Disponible: http://planning.cs.uiuc.edu/

[3] J. J. Kuffner and S. M. LaValle "RRT-Connect: An Efficient Approach to Single-Query Path Planning", IEEE Int. Conf. on Robotics & Automation, San Francisco, CA, April 2000.

[4] M. Rubenstein, C. Ahler, and R. Nagpal, "Kilobot: A low cost scalable robot system for collective behaviors", in Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012, pp. 3293?3298.

[5] Y. Tan and Z.-y. Zheng, "Research advance in swarm robotics", Defence Technology, vol. 9, no. 1, pp. 18?39, 2013.

Virtual coarse spaces for domain decomposition methods[I]

Plenary Talk / Conferencia Plenaria

CALVO, JUAN G[II]

Costa Rica

A new extension operator for a virtual coarse space will be presented which can be used in domain decomposition methods for nodal elliptic problems in two dimensions. In particular, a two-level overlapping Schwarz algorithm is considered and a bound for the condition number of the preconditioned system is obtained. The extension operator saves computational time compared to previous studies where discrete harmonic extensions are required and it is suitable for general polygonal meshes and irregular subdomains. Numerical experiments that verify the result are shown, including some with regular and irregular polygonal elements and with subdomains obtained by a mesh partitioner.

Keywords: Domain decomposition, virtual element methods, irregular subdomain boundaries, overlapping Schwarz algorithms, nodal elliptic problems.

References

[1] Calvo, J. (2017). “On the approximation of a virtual coarse space for domain decomposition methods in two dimensions”, *Tech. Rep. IM-2017-31, Institute of Mathematics, Czech Academy of Sciences*. To appear.

[2] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo (2013). “Basic principles of virtual element methods", *Math. Models Methods Appl. Sci.*, 23 (2013), pp. 199–214.

[3] A. Toselli and O. B. Widlund (2005). “Domain decomposition methods-algorithms and theory", *vol. 34 of Springer Ser. Comput. Math., Springer*.

[I] Friday/Viernes, 4:00p.m., Auditorium, Session: CONF, Conference 6

[II] Universidad de Costa Rica, CIMPA, San José, Costa Rica, juan.calvo@ucr.ac.cr

Aplicación de los modelos: Bayes y K-nn, en un score para tarjetas de crédito[I]

Communication / Ponencia

CALVO BRICEÑO, DANIEL[II]

Costa Rica

En los años recientes tanto en el ámbito local como internacional los reguladores financieros han expuesto la necesidad de que las entidades financieras establezcan metodologías adecuadas para la identificación y evaluación del riesgo. A raíz de esta necesidad se implementó un modelo para el segumiento de la cartera de tarjetas de credito, con el objetivo principal de alertar sobre clientes que dado una serie de atributos puedan llegar a caer en incumplimiento (default crediticio) de sus obligaciones con la entidad en un horizonte temporal menor o igual a 12 meses.

El modelo utiliza un consolidado de la probabilidad de default, que se genera como resultado de la aplicación a la cartera de tarjetas de creditos de los métodos de míneria de datos *Naive Bayes* y *K vecinos más cercanos* de los paquetes *e1071* y *kknn* respectivamente, programados en el software R. Los atributos o variables que represetan las condiciones de cada uno de los clientes fueron categorizados, esta segmentación se basó en los percetiles de su distribucion de frecuencias y el criterio experto. La versión definitiva se empezó a utilizar en abril del 2016, por lo cual se posee suficiente información para comparar y compartir los resultados obtenidos por el backtesting, los que reflejan un buen desempeño en la detección de clientes con alto riesgo para la entidad.

Palabras clave: Míneria de datos, Probabilidad, score de crédito, riesgo.

References

[1] Hechenbichler K, y Schliep, K (2004). “Weighted k Nearest Neighbor Techniques and Ordinal Classification, https://epub.ub.uni–muenchen.de/1769/1/paper_ 399.pdf

[2] Rodriguez, O (2015). “Apuntes programa experto en Míneria de Datos". Promidat

[3] Williams, G (2011). “Data Mining With Rattle and R: The art of excavating data for knowledge Discovery. Springer.

[I] Thursday/Jueves, 1:30p.m., Auditorium, Session: DMin, Data Mining / Minera de Datos

[II] Banco Promerica de Costa Rica, Riesgo Financiero, Costa Rica, San José, dacalvo@promerica.fi.cr

A characterization for dominace plausible rule through of qualitative probabilities[I]

Communication / Ponencia

CAMACHO, FRANKLIN[II] Gómez, Bian[III] Pino-Pérez, Ramón[IV]

Ecuador & Venezuela

In this work it's proposed a set qualitative postulates that characterize a, $\succeq$, relation on the set of policies, defined through dominance plausible rule. In Decision Theory a policy is a function from the set of states S to the set of consequence X. There are different forms to rank the policies, one of the most used is the expected utility, [1].

However, in the case that S is finite, this rule is sensitive to small changes in the parameters that define it.

A more robust rule, in this framework, is the dominance plausible rule. Its definition need two parameters: a $\sqsupseteq$ relation of plausibility on the events (subsets S) and a preference relation, on the set X. This rule has been characterized for some plausibility relations such as: possibilities relations and leximax relations, [2, 3].

Using similar techniques, it will be determined sufficient and necessaries qualitative properties, on a $\succeq$ that is defined through of dominance plausible rule having as the plausibility parameter a *probabilities qualitative* relation and a total preorder about the consecuences. A preference relation is a qualitative probability (satisfies the De Finetti's axioms) if it is a total preorder, no trivial, no negative and additive, [4].

For obtaining the properties that represent the decision rule, it is necessary to give a alternative characterization for the qualitative probabilities.

Keywords: Qualitative decision theory, Dominance plausible rule, Qualitative probabilities, De Finetti's axioms.

References

[1] Savage, L. (1972) "The Foundations of Statistics", *Second edition, Dover, New York*,

[2] Dubois, D and Fargier, H and Perny, P. (2003). "Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach", *Artif. Intell., Vol. 148 n^o 1-2* , p. 219–260.

[3] Camacho, F. and Pino Pérez, R. (2011). "Leximax Relations in Decision Making through the Dominance Plausible Rule", Symbolic and Quantitative Approaches to Reasoning with Uncertainty - 11th European Conference, ECSQARU 2011, Belfast, UK . *In Lecture Notes in Computer Science. Vol. 6717* p. 569-581.

[4] de Finetti, B.(1993)."La prévision: ses lois logiques, ses sources subjectives" *Annales de l'Institut Henri Poincaré", Vol. 7*, p. 1–68

[I] Tuesday/Martes, 5:00p.m., Auditorium, Session: Apl3, Applications / Aplicaciones 3

[II] Universidad de Investigación de Tecnología Experimental, Escuela de Ciencias Matemáticas y Tecnología Informática, Ecuador, Urcuquí, cfranklinj@gmail.com

[III] Escuela de Estadística, Universidad de Los Andes, Mérida, Venezuela, gomezbrian88@gmail.com

[IV] Departamento de Matemáticas, Universidad de Los Andes, Mérida, Venezuela, pino@ula.ve

Técnica de Geometría Fractal para el Cálculo del Riesgo Operacional[I]

Communication / Ponencia

CAMPOS GRANADOS, WALTER OTONIEL[II]

El Salvador

Since the New Basel Accord raised the requirement to incorporate operational risk into the solvency relationship in financial institutions, various methodological alternatives have emerged for its quantification. In this paper, it´s presented the called fractal estimation, based on Iterated Function Systems, is presented as an alternative for measuring risk. No specific comparison has been made with other methodologies, since it is an introduction to the method to show how fractal geometry is applied in the finance area

Keywords: Fractal geometry, fixed point, operational risk, distribution function.

References

[1] Kenneth,F. (2003), Fractal Geometry, Mathematical Foundations and Aplications,University of St Andrews, Wiley, UK.

[2] Cryer, J. & Chan, K,. (2010), Time Series Analysis: With Applications in R, Springer.

[3] Texts in StatisticsLacus, S. & La Torre, D., (2004), Approximating Distribution Functions by Iterated Function Systems, Department of Economics, Business and Statistics, Faculty of Political Science, University of Milan.

[I] Wednesday/Miércoles, 4:00p.m., Room/aula 3, Session: FinMat, Financial Mathematics / Matematica Financiera

[II] Universidad de El Salvador, Escuela de Matemática-Facultad de Ciencias Naturales y Matemática, El Salvador, San Salvador, walteroto@gmail.com

Some Examples of Dense Sequences and its Relation to Some Spaces of Functions[I]

Communication / Ponencia

CASTRO FERNÁNDEZ, EDWIN[II] Arguedas Troyo, Vernor[III]

Costa Rica

We analyze some methods for dense sequences for some functions classes. There are many well known examples : $\mathbb{Q}$ is a numerable dense set in $\mathbb{R}$, the Kronecker's and its consequences gives additional examples. In the paper of Tian-Xiao He , Zachariah Sinkala and Xiaoya Zha some other examples are shown for the building of numerable dense sets.

We present other results of dense sequences.

Keywords: Kronecker density theorem, dense sequences, generalized polyinomials.

Resumen

En este artículo analizaremos varios métodos de sucesiones densas para clases de funciones. Hay ejemplos muy conocidos: $\mathbb{Q}$ es un conjunto numerable denso en $\mathbb{R}$, el teorema de Kronecker y sus derivados brindan otros ejemplos. En el artículo de Tian-Xiao He , Zachariah Sinkala y Xiaoya Zha se muestran otros métodos para construir conjuntos densos numerables, los cuales nosotros derivamos de la teoría presentada.

Presentamos otros resultados de sucesiones densas .

Vamos a discutir este tipo de densidad para algunas clases de funciones en las que se pondrán ciertas condiciones

Palabras clave: Teorema de la densidad de Kronecker, sucesiones densas, polinomios generalizados.

References

[1] D.Andrica, S.Buzeteanu, Relatively dense universal sequences for the class of continous periodical functions of period T, Mathematica- L'Analyse Numerique et la Theorie de l'Approximation 16 (1987), 1-9 MR89i: 11084; Zbl: 642.26006

[2] Andrica Dorin, Buzeţeanu Şerban, *Relatively dense universal sequences for the class of continuous periodical functions of period T*, L'Analyse Numerique et la Theorie de l'Approximation, Tome 16, no.1, 1987, pp.1-9

[3] Arguedas Vernor, Castro Edwin, *Conjuntos de densidad y espacios lacunares*, Simposio de Matemáticas aplicadas a las ciencias, Universidad de Costa Rica, Escuela de Matemática, 2006, (por aparecer)

[4] Tian Xiao Hu, *Some Dense Sets of Real Numbers and Applications*,With P. J.-S. Shiue and X. Zha, J. Adv. Math. Studies, 4(2011), No. 2,pp 25

[I] Tuesday/Martes, 4:40p.m., Room/aula 2, Session: Num4, Numerical Analysis / Análisis Numérico 3

[II] Universidad de Costa Rica, CIMPA, Costa Rica, San José, hyperion32001@gmail.com

[III] Universidad de Costa Rica,San José,Costa Rica, vernor.arguedas@ucr.ac.cr

Analysis of the effective potential and the stable orbits of a Kerr-like metric with quadrupole[I]

Communication / Ponencia

CHAVERRI MIRANDA, FABIÁN[II] Frutos-Alfaro, Francisco[III] Gómez-Ovares, Pedro[IV]
Oliva-Mercado, G. Andre[V]

Costa Rica

We derived the effective potential of a Kerr-like metric for the radial coordinate by the Euler-Lagrange method. After this, we studied the stability of this potential to find the equations for null geodesics and the innermost stable circular orbit (ISCO). Finally, we reduced the found equations to the known limiting cases to verify the results.

We derived the effective potential of a Kerr-like metric for the radial coordinate by the Euler-Lagrange method. After this, we studied the stability of this potential to find the equations for null geodesics and the innermost stable circular orbit (ISCO). Finally, we reduced the found equations to the known limiting cases to verify the results.

Keywords: General Relativity, solutions of Einstein's equations, approximation procedures, weak fields.

Resumen

Se encontró el potencial efectivo de una métrica tipo Kerr con cuadrupolo para la componente radial por medio del método de Euler Lagrange. Con esto se analizó la estabilidad de dicho potencial y de esta forma se hallaron las ecuaciones para las geodésicas nulas y las órbitas estables circulares más cercanas (ISCO). Finalmente se compararon las ecuaciones encontradas con los casos conocidos más generales, y así verificar los resultados finales.

Palabras clave: Relatividad general, soluciones de las ecuaciones de Einstein, procedimientos de aproximación, campos débiles.

References

[1] F. Frutos-Alfaro (2016) "Approximate Kerr-like Metric with Quadrupole´´, *International Journal of Astronomy and Astrophysics*, 6, 334-345, .

[I] Thursday/Jueves, 2:55p.m., Auditorium, Session: Aprox1, Approximation / Aproximacion

[II] Universidad de Costa Rica, Física, San José, Costa Rica, fabian.chaverrimiranda@ucr.ac.cr
[III] Universidad de Costa Rica, Escuela de Física, San José, Costa Rica, frutos62@gmail.com
[IV] Universidad de Costa Rica, San José, Costa Rica, pedro.gomezovares@gmail.com
[V] Universidad de Costa Rica, San José, Costa Rica, gandreoliva@gmail.com

Epidemic Growth Scaling: Implications for Disease Forecasting and Estimation of the Reproduction Number[I]

Plenary Talk / Conferencia Plenaria

CHOWELL, GERARDO[II]

United States

There is a long tradition of using mathematical models to generate insights into the transmission dynamics of infectious diseases and assess the potential impact of different intervention strategies. The increasing use of mathematical models for epidemic forecasting has highlighted the importance of designing reliable models that capture the baseline transmission characteristics of specific pathogens and social contexts. More refined models are needed however, in particular to account for variation in the early growth dynamics of real epidemics and to gain a better understanding of the mechanisms at play. Here, we review recent progress on modeling and characterizing early epidemic growth patterns from infectious disease outbreak data, and survey the types of mathematical formulations that are most useful for capturing a diversity of early epidemic growth profiles, ranging from sub-exponential to exponential growth dynamics. Specifically, we review mathematical models that incorporate spatial details or realistic population mixing structures, including meta-population models, individual-based network models, and simple SIR-type models that incorporate the effects of reactive behavior changes or inhomogeneous mixing. In this process, we also analyze simulation data stemming from detailed large-scale agent-based models previously designed and calibrated to study how realistic social networks and disease transmission characteristics shape early epidemic growth patterns, general transmission dynamics, and control of international disease emergencies such as the 2009 A/H1N1 influenza pandemic and the 2014-2015 Ebola epidemic in West Africa.

Keywords: Mathematical modeling, epidemic, epidemic growth scaling, sub–exponential growth, polynomial growth.

References

[1] Perspectives on model forecasts of the 2014-2015 Ebola epidemic in West Africa: lessons and the way forward. Chowell G, Viboud C, Simonsen L, Merler S, Vespignani A. BMC Med. 2017 Mar 1;15(1):42. doi: 10.1186/s12916–017–0811–y.

[2] Forecasting the 2001 Foot and Mouth Disease Epidemic in the UK. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Ecohealth. 2017 Dec 13. doi: 10.1007/s10393-017-1293-2. [Epub ahead of print]

[3] Mathematical models to characterize early epidemic growth: A review. Chowell G, Sattenspiel L, Bansal S, Viboud C. Phys Life Rev. 2016 Sep;18:66-97. doi: 10.1016/j.plrev.2016.07.005. Epub 2016 Jul 11. Review.

[4] A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. Viboud C, Simonsen L, Chowell G. Epidemics. 2016 Jun;15:27-37. doi: 10.1016/j.epidem.2016.01.002. Epub 2016 Feb 1.

[5] Modeling household and community transmission of Ebola virus disease: Epidemic growth, spatial dynamics and insights for epidemic control. Kiskowski M, Chowell G. Virulence. 2016;7(2):163-73. doi: 10.1080/21505594.2015.1076613. Epub 2015 Aug 20.

[I] Thursday/Jueves, 10:30a.m., Auditorium, Session: CONF, Conference 4

[II] United States, Georgia, Georgia State University , gchowell@gsu.edu

Driving certain Impulsive Biological Systems with Maximal Operators[I]

Poster / Cartel

CORVALÁN, ALVARO[II] Cardo, Romina[III]

Argentina

In this work we study the dynamics of competing species in certain environments. In several lake habitats where there are frequent exogenous interventions localized in time the situation can be adequately represented by systems of impulsive differential equations. Such models deals with the behavior of variables generally governed by a differential system, except in a discrete sequence of localized instants where abrupt variations are introduced either by seeding alevins or by massive capture of specimens of one or more of the species.

$$\begin{cases} X'(t) = f(t, X(t)) t \neq \tau_k \\ X(\tau_k) = g(\tau_k, X(\tau_k^-)) t = \tau_k, \end{cases}$$

where $\tau_1, \tau_2, \cdots \tau_k, ...$ are the times for the impulses.

A key numerical question is usually the choice of the instant to introduce or allow the disturbance of the system - and the magnitude of the intervention - where in addition to satisfying certain criteria of economic interests it is desired that the system remain sustainable in the sense of avoiding non-desired extinctions.

Here we consider a strategy based on the values of the left one-sided maximal Hardy-Littlewood Operator and the Hardy Operator to control impulsive systems. A way to do that is by means the pulsating seeding of young fish in the case where many competitive species interact. This strategy may occur as a natural from the point of view of those who propose seeding and also allows an sustaintable control of the magnitudes resulting from competing species. We ilustrate some of this ideas with model examples of poblations of perch and trout fishes of many lakes in the south Patagonical region. To quantify the success of these controls we propose to measure the cross entropies of the time series of the total mass of the population of the involved species.

Some preliminar results can also be found in [1].

Keywords: Impulsive Differential Equations, Maximal Operators.

References

[1] "Control de Sistemas Biológicos Impulsivos mediante operadores maximales" Cardo R. & Corvalán A., MATUA, *Revista de Matemática de la Universidad del Atlántico*, Vol III N2 (2016).

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Nacional de General Sarmiento, IDH, Argentina, Malvinas Argentinas, acorvala@ungs.edu.ar

[III] Universidad Nacional de General Sarmiento, IDH, Argentina, Malvinas Argentinas, rcardo@ungs.edu.ar

New perspectives on multilayer and multiplex networks as structures to model real world systems[I]

Communication / Ponencia

CRIADO, REGINO[II] Romance, Miguel[III]

España

There is a broad variety of real-world systems where components cannot function independently, so these components interact with others through different routes of connectivity and dependencies. Since the publication of the two key papers, by D.J. Watts and S.H. Strogatz on *small world networks* [1] and by A.L. Barabási and R. Albert on *scale-free networks* [2], there is growing interest for complex networks reflected by the large number of papers that have been published on this topic, and the new problems, issues and research lines which have been tackled (see, for example, [3, 4, 5, 6]).

Multilayer networks [4, 5] explicitly incorporate diverse channels of connectivity in a system and constitute the natural mathematical environment to describe systems whose units are interconnected through different types and kinds of connections: each channel (relationship, category,...) is represented by a layer, and the same node or entity may have different kinds of interactions (coworker-ship, friendship, vicinity, etc.) with the rest of the nodes or entities.

Multiplex networks are a particular case of multilayer networks that provide a natural description for systems in which the same elementary units may interact amongst them through many kinds of links. Each type of link or relationship (e.g., a task, an activity, a category,...) is represented by a different layer. Thus, multiplex networks provide a natural description for systems in which entities have the same set of neighbours in each layer [4, 5].

Both concepts (multilayer and multiplex networks) have been introduced in order to embody some topological, dynamical and structural properties of heterogeneous-type complex systems which are not completely captured by the classical models of complex network's theory.

The aim of this talk is to present some examples, heuristics and applications related to these concepts. In the context of applications, some results related to an extension of the classical PageRank algorithm using a biplex approach [7] or a suitable personalization vector will be presented.

Keywords: Multilayer networks, multiplex networks, Centrality.

References

[1] Watts, D.J., Strogatz, S.H., " Collective dynamics of small-world networks", Nature 1998;393:440–442.

[2] Barabási, A. L. and Albert,R., "Emergence of scaling in random networks", Science 286, 509–512 (1999).

[I] Wednesday/Miércoles, 5:00p.m., Auditorium, Session: Mod5, Modeling / Modelacion 5

[II] Universidad Rey Juan Carlos, Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica, España, Madrid, regino.criado@urjc.es

[III] Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid, España, miguel.romance@urjc.es

[3] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D. U. : Complex Networks: Structure and Dynamics. Phys. Rep, **424**, 175 (2006).

[4] Boccaletti, S., Bianconi, G., Criado, R., Y., Del Genio, C.I., Gómez-Gardeñes, J. , Romance, M., Sendiña-Nadal, I. Wang, Z., Zanin, M.: The Structure and Dynamics of Multilayer Networks. Phys. Rep, **544**, 1, (2014).

[5] Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. J. of Complex Net. **2** 203–271, (2014).

[6] Newman, M.E.J.: Networks: An Introduction. Oxford Univ. Press, Oxford, 2010

[7] Pedroche, F. Romance M. and Criado,R., "A biplex approach to PageRank centrality: From classic to multiplex networks" Chaos **26**, (6), (2016) 06530.

A Monte Carlo Approach to Computing Stiffness Matrices Arising[I]

Communication / Ponencia

CUERVO FERNÁNDEZ, OMAR ANDRÉS[II] Galvis, Juan[III]

Colombia

We use a Monte Carlo method to assemble finite element matrices for polynomial Chaos approximations of elliptic equations with random coefficients. In this approach, all required expectations are approximated by a Monte Carlo method. The resulting methodology requires dealing with sparse block-diagonal matrices instead of block-full matrices. This leads to the solution of a coupled system of elliptic equations where the coupling is given by a Kronecker product matrix involving polynomial evaluation matrices. This generalizes the Classical Monte Carlo approximation and Collocation method for approximating functionals of solutions of these equations.

Keywords: Polynomial Chaos, Random Elliptic Partial Differential Equations, Monte Carlo Integration.

References

[1] I. Babuska, F. Nobile, and R. Tempone (2007)"A stochastic collocation method for elliptic partial differential equations with random input data"', *SIAM J. Numer. Anal.*, 45 , pp. 1005–1034.

[2] I. Babu?ska, R. Tempone, and G. E. Zouraris (2004) Galerkin finite element approximations of stochastic elliptic partial differential equations, *SIAM J. Numer. Anal.*, 42, pp. 800–825.

[3] M. Bachmayr, A. Cohen, and G. Migliorati(2017) "Sparse polynomial ap- proximation of parametric elliptic pdes. part i: affine coefficients", *ESAIM: Mathematical Modelling and Numerical Analysis*, 51 , pp. 321–339.

[I] Tuesday/Martes, 1:50p.m., Room/aula 2, Session: Num2, Numerical Analysis / Análisis Numérico 1

[II] Universidad Nacional de Colombia, Bogotá D.C., Colombia, Bogotá, oacuervof@unal.edu.co

[III] Universidad Nacional de Colombia, Bogotá, Colombia, jcgalvisa@unal.edu.co

An application of Gaussian mixtures in finance statistical modelling and risk assessment[I]

Communication / Ponencia

CUEVAS COVARRUBIAS, CARLOS[II] Íñigo Martínez, Jorge[III]
Jiménez Padilla, Raquel[IV]

México

Many important models in quantitative finance are based on the assumption that stock returns are independent and normally distributed. However, the empirical distributions of price changes are frequently skewed and leptokurtic. Therefore, flexible distributions and their potential in financial modeling constitute an important research topic in mathematical finance. We explore the potential of Gaussian mixtures as an alternative to the normal distribution. This article is not limited to the estimation of marginal distributions. Contrasting with some other papers in the literature, the application of multivariate Gaussian mixtures to estimate joint distributions of financial returns is also analyzed. This multivariate approach gives us the opportunity to illustrate the application of Gaussian mixtures in financial modelling for risk assessment and portfolio optimization. We illustrate our proposal with practical examples.

Keywords: Gaussian mixtures, Financial Returns, VaR, CVaR.

References

[1] Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Trans. Automatic Control*, Vol. AC-19, 716 - 723.

[2] Behr, A.; and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models, *Annals of Finance*. Vol. 5, 49 - 68.

[3] Cuevas-Covarrubias, C.; Iñigo-Martínez, J.; and Jiménez-Padilla, R. (2017). Gaussian Mixtures and Financial Returns, Aceptado para publicación en *Discussiones Mathematicae Probability and Statistics.*

[4] Fama, E.F. (1965). The behavior of Stock Market Prices, *The Journal of Business*, Vol. 38(1), 34 - 105.

[5] Jorion, Ph. (2007). Value at Risk: The New Benchmark for Managing Financial Risk 3rd Edition, *McGraw-Hill.*

[I] Wednesday/Miércoles, 4:20p.m., Room/aula 3, Session: FinMat, Financial Mathematics / Matematica Financiera

[II] Universidad Anáhuac, Centro de Investigación en Estadística y Matemáticas Aplicadas, México, Huixquilucan, ccuevas@anahuac.mx

[III] same address

[IV] same address

Flow of a ferrofluid a driven cavity[I]

Communication / Ponencia

DUMARESQ SOBRAL, YURI[II] de Oliveira Vieira, Camila[III]
da Cunha, Francisco Ricardo[IV]

Brazil

This work aims at studying the behavior of magnetic fluids in a 2D lid-driven cavity. In order to achieve this goal, a computer program was developed to solve the Navier-Stokes equation using finite differences. In addition to that, different equations for the evolution of the magnetism of the magnetic fluid were introduced. The governing equations are presented, as well as the dimensionless parameters governing the problem. The Navier-Stokes equations are solved using a projection method to decouple the velocity pressure problem. The solution of the Poisson equation is obtained by means of an implicit method with sparse matrices and Cholesky factorization. A few distinct Reynolds numbers are tested, as well as some different magnetic parameters. We observe that the magnetic field strongly influences the flow on the cavity.

Keywords: Fluid dynamics, Magnetic fluids, Finite differences, Driven Cavity.

References

[1] Tzirtzilakis E.E. & Xenos M.A., 2013. Biomagnetic fluid flow in a driven cavity. Meccanica, vol. 48, pp 187-200.

[2] Rosensweig R.E., 1987. Magnetic Fluids. Annual Review of Fluid Mechanics, vol. 19, pp 437-461.

[I] Wednesday/Miércoles, 4:00p.m., Auditorium, Session: Mod5, Modeling / Modelacion 5

[II] Universidade de Brasília, Departamento de Matemática, Brazil, Brasília, ydsobral@unb.br
[III] Universidade de Brasília, Departamento de Matemática, Brasília, Brasil, vieira_ camilla@yahoo.com.br
[IV] Universidade de Brasília, Departamento de Engenharia Mecânica, Brasília, Brasil, frcunha@unb.br

Metodología para la construcción de modelos de simulación discreta aplicados a la industria[I]

Communication / Ponencia

FERNANDEZ LEDESMA, JAVIER DARIO[II]

Colombia

The discrete simulation is a technique that contributes to the decision-making processes. There are many tools that have been developed for discrete simulation, fundamentally oriented in the modeling of the current situation of organizational processes for the analysis and impact of possible decisions. However, the development of the tools has not been accompanied by a development of the methodologies that contribute to a more comprehensive understanding of the simulation problems and that serve to make decisions in a faster and more efficient way, with the incorporation of other techniques that go from the modeling of the process to the validation and prospective of results. This article presents a methodology oriented to the construction of discrete simulation models from the previous phases of the analysis and the modeling of the process until the validation of impacts on the results of the model. The approach includes the definition of the problem, the construction of the process model, the construction and design of the discrete simulation model, and the generation and discussion of model results. The methodology has been successfully applied in different fields of industry, including services, metal mechanics, agro-industry, among others and has been validated with results that relate the proposed methodology and the impact of the results generated for decision making.

Keywords: Methodology, Models, Simulation.

Resumen

La simulación discreta constituye una técnica que aporta en los procesos de toma de decisiones. Son múltiples las herramientas que se han desarrollado para la simulación discreta orientadas fundamentalmente en el modelamiento de la situación actual de los procesos organizacionales para el análisis e impacto de posibles decisiones. Sin embargo, el desarrollo de las herramientas no se ha acompañado de un desarrollo de las metodologías que aporten a una comprensión más integral de los problemas de simulación y que sirvan a la toma de decisiones de una manera más rápida y eficiente, con la incorporación de otras técnicas que vayan desde el modelado del proceso hasta la validación y prospectiva de resultados. Este artículo presenta una metodología orientada a la construcción de modelos de simulación discreta desde las fases previas del análisis y el modelamiento del proceso hasta la validación de impactos sobre los resultados del modelo. El enfoque incluye la definición del problema, la construcción del modelo del proceso, la construcción y el diseño del modelo de simulación discreta y la generación y discusión de los resultados del modelo. La metodología se ha aplicado con éxito en diferentes campos de la industria, pasando por los servicios, la metalmécanica, la agroindustria, entre otros y se ha validado con resultados que relacionan la metodología propuesta y el impacto de los resultados generados para la toma de decisiones.

Palabras clave: Metodología, Modelos, Simulación.

[I] Tuesday/Martes, 2:35p.m., Auditorium, Session: Apl2, Applications / Aplicaciones 2

[II] Universidad Ponticificia Bolivariana, Antioquia, Colombia, Medellin, javier.fernandez@upb.edu.co

References

[1] Molano, A. (2014). Internet de las cosas: concepto y ecosistema. Colombia Digital. Retrieved from http://colombiadigital.net/actualidad/articulosinformativos/item/7821internet-de-las-cosas-concepto-y-ecosistema.html.

[2] Schwab, K. (2016). La cuarta revolución industrial. Editorial Debate. México. 224 p.

[3] Bauernhansl, T. (2014). Die vierte industrielle Revolution - Der Weg in ein wertschaffendes Produktionsparadigma. In Industrie 4.0 in Produktion, Automatisierung und Logistik (pp. 5– 35). http://doi.org/10.1007/978-3-658-04682-8.

[4] Bundesministerium für Bildung und Forschung. (2014a). Die neue Hightech-Strategie Innovationen für Deutschland. Retrieved May 20, 2015, from http://www.bmbf.de/de/26393.php.

[5] Aspray, William. (1990). John Von Neumann and the origins of modern computing. Massachusetts. The MIT Press. p. 110-113.

Bounded solutions of Volterra functional difference equations[I]

Communication / Ponencia

FERREIRA BERNARDO, FELIX[II] Cuevas, Claudio[III] Soto, Herme[IV]

Brasil

Introduction

We briefly present the relevance of **Dynamic systems** on the scope of *Input/output systems*, *Approximation (Peter Lax)*; *asymptotic behaviour of the solutions (stable or unstable)* and *boundedness*.
Following the work of - Cardoso, F., & Cuevas, C. (2009). *Exponential dichotomy and boundedness for retarded functional difference equations*. Journal of Difference Equations and Applications, 15(3), 261290. - we will characterize and give a sufficient condition to have **Exponential dichotomy**, consequences and robustness under several types of perturbations. To conclude this topic we will talk about the topological structure of the solution set: compactness and the open problem of connectivity.

Volterra Difference Equations

We present the Volterra difference equation

$$x(n+1) = \lambda \sum_{j=-\infty}^{n} a(n-j)x(j) + f(n, x_n),\ n \in \mathbb{Z}. \tag{1}$$

Examples and applications in [2], an integro-difference Turelli–Hoffmann Model [1] and control systems.

To deal whit such problem we present the concept of **Phase space**, $\mathcal{B} \subseteq \{\varphi : \mathbb{Z}^- \longrightarrow X\}$. Equipped whit the norm, $||\ .\ ||_{\mathcal{B}}$, $\mathcal{B}$ is a Banach space. To define the phase space we follow the standard axiomatic of Hale-Kato-Murakami (discrete case), adding an *fading memory* alike new axiom.

Results

To conclude the exposition we present some results from our next work.

Existence and stability of solutions

Theorem 1 *Let p and q be conjugated exponents. Assume that $\mathcal{B}$ is a phase space. Let $\lambda \in \Omega_s$ [2] and let $f : \mathbb{Z} \times \mathcal{B} \to X$ be a function that satisfies the following Lipschtz condition:*

$$||f(n, \varphi) - f(n, \psi)||_X \leq L(n)||\varphi(0) - \psi(0)||_X \tag{2}$$

for all $\varphi, \psi \in \mathcal{B}$ and each $n \in \mathbb{Z}$, where $L : \mathbb{Z} \to \mathbb{R}^+$ is ℓ^p-summable and $f(.,0) \in \ell^1(\mathbb{Z}; X)$.
If $\rho||s(\lambda, .)||_1||L||_{\ell^q(\mathbb{Z};\mathbb{R}^+)} < 1$ then the equation 1 has an unique $\ell^p - bounded$ solution.

[I] Wednesday/Miércoles, 4:20p.m., Room/aula 2, Session: DS, Dynamical Systems / Sistemas Dinamicos

[II] Universidade Federal de Pernambuco, Departamento de Matemática., Brasil, Recife, felix.pt@gmail.com

[III] Department of Mathematics, Federal University of Pernambuco, Recife, Brasil, cch@dmat.ufpe.br

[IV] Department of Mathematics and Statistics, University of La Frontera, Temuco, Chile, herme.soto@ufrontera.cl

[1] Wolbachia invasive wave fronts

[2] Ω_s is the region of $\mathbb{C}$ where the fundamental solution is summable.

Theorem 2 *Let $\lambda \in \Omega_s$. Assume that $\mathcal{B}$ is a phase space and the condition* $\mathbf{Hp}$ *holds. In addition suppose that:* ***(Lp-1)*** *For all $k \in \mathbb{Z}$ and every bounded subset, $K \subseteq \mathcal{B}$, the set $f(k, K)$ is relatively compact* [3] *in X.*
Then there is a $\ell^p - bounded$ solution of 1.

Topology of the solution set

Theorem 3 *Assume that $\mathcal{B}$ is a phase space and let $\lambda \in \Omega_s$. Let $f : \mathbb{Z} \times \mathcal{B} \to X$ be a function satisfying the* ***(Hp-2)****,* ***(Lp-1)*** *and* ***(TS2)*** *conditions whit $\gamma_R \in \ell^p(\mathbb{Z}, \mathbb{R}^+)$. Suppose that* ***(TS5)*** *and* ***(TS6)*** *holds, then the set S of all $\ell^p(\mathbb{Z}; X) - solutions$ of 1 is compact.*

Keywords: Boundedness, Volterra functional difference equation, Phase space, Infinite delay, topological properties.

References

[1] Agarwal, R. P., Cuevas, C., & Lizama, C. (2014). *Regularity of Difference Equations on Banach Spaces* (p. 202). Berlin: Springer.

[2] Cuevas, C., Choquehuanca, M., & Soto, H. (2014). *Asymptotic analysis for Volterra difference equations. Asymptotic Analysis*, 88(3), 125-164.

[3] Agarwal, R. P., Cuevas, C., & Dantas, F. (2013). *Almost automorphy profile of solutions for difference equations of Volterra type*. Journal of Applied Mathematics and Computing, 42(1-2), 1-18.

[4] de Andrade, B., Cuevas, C., Silva, C., & Soto, H. (2016). *Asymptotic Periodicity for Flexible Structural Systems and Applications*. Acta Applicandae Mathematicae, 143(1), 105-164.

[5] Andrade, F., Cuevas, C., & Henríquez, H. R. (2016). "Periodic solutions of abstract functional differential equations with state dependen–delay´´. *Mathematical Methods in the Applied Sciences*, 39(13), 3897–3909.

[3] Strongly non-linear situations, far common in problems from the real world, demand adicional compacticity hypothesis to achieve fixed point operators.

A Mathematical Model for Mimicry Dynamics in Evolutionary Ecology[I]

Communication / Ponencia

FERREIRA JR, WILSON[II]

Brazil

An extension of the mathematical model proposed in a previously published article (Ferreira & Marcon–[2014]) will be presented where Mimicry Dynamics between some toxic butterflies (known as Muller mimicry) is analysed. This peculiar phenomenon involves many aspects of animal behavior and perception besides being one of the most most singular example of evolutionary dynamics. Simulations will show how some striking features of wing pattern distributions widely observed in nature could be explained.

Keywords: Biomathematics, Evolution, Mimicry, Cellular Automata.

References

[1] W.C.Ferreira Jr.-D.Marcon-Revisiting the 1879 model of Evolutionary Mimicry by Fritz Muller: New Mathematical Approaches, Ecological Complexity 18 (2014), 25-38

[I] Friday/Viernes, 11:10a.m., Auditorium, Session: Bio2, Biomathematics / Biomatematicas 2

[II] UNICAMP-Campinas Brasil, Applied Mathematics, Brazil, Campinas, wilson@ime.unicamp.br

Scalable Subsurface Simulations with ParFlow[I]

Communication / Ponencia

FONSECA, JOSE A.[II]

Germany

The accurate simulation of variably saturated flow in a porous media is a valuable component in understanding physical processes ocurring in many water resources problems. Such simulations require expensive and extensive computations and efficient usage of the latest high performance parallel computing systems becomes a necesity. The simulation software ParFlow has been shown to have excellent solver scalability for up to 16,384 processes (MPI ranks).

To scale to the full size of current petascale systems, we propose a reorganization of ParFlow's mesh subsystem: we modify it to use state of the art mesh refinement and partition algorithms provided by the parallel software library p4est. Additionaly, we removed several overly strict assumptions on the parallel mesh layout in the processes.

We evaluate the scalability and performance of the modified version of ParFlow demonstrating weak and strong scaling to over 458.000 processes of the Juqueen supercomputer at the Research Center Juelich.

Keywords: Subsurface Flow, Numerical Simulation, High Performance Computing.

References

[1] Burstedde, C. Fonseca, J.A., Kollet, S. (2018) "Enhancing speed and scalability of the ParFlow simulation code". Computational Geosciences, 22(1), 347-361.

[I] Tuesday/Martes, 1:30p.m., Room/aula 2, Session: Num2, Numerical Analysis / Análisis Numérico 1

[II] Institute for Numerical Simulation and Hausdorff Center for Mathematics, University of Bonn, Germany, fonseca@ins.uni-bonn.de

Vectorización de Contornos en Imágenes Rasterizadas[I]

Communication / Ponencia

FORERO VELASCO, WILLIAM[II] Herrera Daza, Eddy[III]

Colombia

This work presents a method of rasterized contour vectorization, through the determination of the changes in the curvature of the contour, for the fitting of a Bézier curve to these points, minimizing the difference between the original contour and the resulting. The method was tested with different images for JPG, JPEG, BMP and PNG formats with or without blurred contours. The results show that the implemented algorithm reconstructs the contour of different forms, maintaining characteristics such as complex corners and curves. At the same time it significantly reduces the information needed to define the image and by extension the resulting size, which represent an alternative for vectorization of images and storage.

Keywords: Vectorization, image tracing, curve-fitting.

Resumen

Este trabajo presenta un método de vectorización de contornos rasterizados, a través de la determinación de los cambios en la curvatura del contorno, para el ajuste de una curva de Bézier a estos puntos, minimizando la diferencia entre el contorno original y el resultante. El método fue probado con diferentes imágenes para los formatos JPG, JPEG, BMP y PNG con o sin contornos borrosos. Los resultados muestran, que el algoritmo implementado logra reconstruir el contorno de diferentes formas, manteniendo características como las esquinas y curvas complejas, a la vez que reduce significativamente la información necesaria para definir la imagen y por extensión el tamaño resultante, lo cual representan una alternativa para la vectorización de imágenes y almacenamiento

Palabras clave: Vectorización, trazado de imágenes, ajuste de curvas.

References

[1] Chang, H. And Yan, H. (1998). Vectorization Of Hand–Drawn Image Using Piecewise Cubic Bézier Curves Fitting. Pattern Recognition, 31(11), pp.1747–1755.

[2] Doermann, D. (1998). An introduction to vectorization and segmentation. Graphics Recognition Algorithms and Systems, pp.1–8.

[3] Gameenginegems.net. (2017). Graphics and Game Gems Database - An Algorithm for Automatically Fitting Digitized Curves. [online] Available at: http://www.gameenginegems.net/gemsdb/article.php?id=780 [Accessed 24 Nov. 2017].

[4] Ren, M., Yang, J. and Sun, H. (2017). Tracing boundary contours in a binary image.

[I] Tuesday/Martes, 2:55p.m., Room/aula 2, Session: Num3, Numerical Analysis / Análisis Numérico 2

[II] Pontificia Universidad Javeriana, Cundinamarca, Colombia, Bogotá, forero.w@javeriana.edu.co

[III] Pontificia Universidad Javeriana, Bogotá, Colombia, eherrera@javeriana.edu.co

Dengue fever models and optimal control in the life-cycle of Aedes aegypti mosquitoes[I]

Communication / Ponencia

FRAGUELA COLLAR, ANDRES[II] Pliego Pliego, Emilene[III] Velazquez Castro, Jorge[IV]

México

A mathematical model with delay of SIR type (Susceptible, Infected and Recovered) is proposed to model the dynamics of the disease in a population of humans in an endemic state, coupled with a model type SI (Susceptible and Infected) for the dynamics of the cycle of life of the Aedes Aegypti mosquito, where the two stages of development in the life cycle of the vector have been considered; the aquatic phase (egg, larva, pupa) and the aerial phase or imago.

The model of the SIR-SI type is described by a system of seven ordinary differential equations with delays corresponding to the incubation periods of the virus in humans and mosquitoes respectively. This model for the endemic state can be reduced to a model of six differential equations if we assume the constant human population, to later decouple it in a system of two and four differential equations, respectively. The first system is associated with the life cycle dynamics of the Aedes Aegypti mosquito in its two phases of development and the second corresponds to the dynamics of the disease in humans, which considers, in a coupled form, population of humans and the population of infected vectors in their two stages of development; aquatic and aerial. We started by presenting the qualitative analysis of both systems and later we studied the model for the population of mosquitoes including seasonality and diapause, obtaining results related to the influence of the mobility of hatcheries in the state of diapause, in the emergence of outbreaks.

Subsequently, two models with optimal control for the life cycle of the Aedes Aegypti mosquito are studied through the use of larvicides and adulticides (chemical control) as well as the elimination of certain breeding sites of the vector (mechanical control), presenting several strategies for the implementation of the two types of control through a cost-benefit analysis.

Finally, applications of these results are presented to different states of the Mexican territory.

Keywords: dengue dynamics, outbreaks, seasonality, diapause, control.

Resumen

Se propone un modelo matemático con retraso de tipo SIR (Susceptible, Infectado y Recuperado) para modelar la dinámica de la enfermedad en una población de humanos en un estado endémico, acoplado con un modelo tipo SI (Susceptible e Infectado) para la dinámica del ciclo de vida del mosquito Aedes Aegypti, donde se han considerado las dos etapas de desarrollo en el ciclo de vida del vector; la fase acuática (huevo, larva, pupa) y la fase aérea o imago.

El modelo del tipo SIR-SI se describe mediante un sistema de siete ecuaciones diferenciales ordinarias con retardos correspondientes a los periodos de incubación del virus en los humanos y los mosquitos respectivamente. Este modelo para el estado endémico se puede reducir a un modelo de seis ecuaciones diferenciales

[I] Friday/Viernes, 1:50p.m., Auditorium, Session: Bio1, Biomathematics / Biomatematicas 1

[II] Benemerita Universidad Autonoma de Puebla, Facultad de Ciencias Físico Matemáticas, Mexico, Puebla, fraguela@fcfm.buap.mx

[III]

[IV]

si suponemos la población humana constante, para posteriormente desacoplarlo en un sistema de dos y cuatro ecuaciones diferenciales, respectivamente. El primer sistema está asociado a la dinámica del ciclo de vida del mosquito Aedes Aegypti en sus dos fases de desarrollo y el segundo corresponde a la dinámica de la enfermedad en los humanos, el cual considera en forma acoplada a la población de humanos y a la población de vectores infectados en sus dos etapas de desarrollo; acuática y aérea. Comenzamos presentando el análisis cualitativo de ambos sistemas y posteriormente estudiamos el modelo para la población de mosquitos incluyendo estacionalidad y diapausa, obteniéndose resultados relativos a la influencia de la movilidad de criaderos en estado de diapausa, en el surgimiento de brotes.

Posteriormente se estudian dos modelos con control óptimo para el ciclo de vida del mosquito Aedes Aegypti mediante el uso de larvicidas y adulticidas (control químico) así como también la eliminación de ciertos criaderos del vector (control mecánico), presentándose diversas estrategias para la implementación de los dos tipos de control na través de un análisis de costo-beneficio.

Finalmente se presentan aplicaciones de estos resultados a diferentes estados del territorio mexicano.

Palabras clave: dinámica del dengue, brotes, estacionalidad, diapausa, control.

References

[1] Pliego Pliego, E., Velásquez-Castro, J., and Fraguela-Collar, A. (2017) Seasonality on the life cycle of aedes aegypti mosquito and its statistical relation with dengue outbreaks. Applied Mathematical Modelling 50, 484–496.

[2] Pliego Pliego, E., Velásquez-Castro, J., EICHHORN, M. P., and Fraguela-Collar, A. (2017) Increased eficiency in the second-hand tire trade provides opportunity for dengue control. Journal of Theoretical Biology.

[3] Pliego Pliego, E., Velásquez-Castro, J., Vasilieva, O., and Fraguela-Collar, A.(2017) Optimal control in a life-cycle model for Aedes aegypti mosquitoes with seasonality. Submitted to Applied Mathematical Modelling.

A combined SWATLD-ALS algorithm for three-way compositional data[I]

Invited Session / Sesion invitada

GALLO, MICHELE[II]

Italy

Compositional Data (CoDa) consist of vectors of positive values summing to a unit or, in general, to some fixed constant for all vectors. These data are quantitative descriptions of the parts of a whole and only carry relative information, expressed as ratios between components. As a consequence of the special constant-sum structure, compositional vectors belong in a specific sample space referred to as simplex. Usually, these data are organized in a two-way array where each row vector is known as composition or in general as CoDa (Compositional Data). When studying complex phenomena, it is possible to run into repeated observations collected for the same variables on the same occasions (such as conditions, times, locations). In particular, if the data represent proportions of a total, recorded on several occasion along a third dimension, they define compositional vectors of strictly positive parts (or components) summing to some fixed constant and organized in a three-way array. The main interest when modelling these vectors lays in the relative changes among components across time or space. Modelling these data entails many difficulties because of their bounded nature. Since standard statistical methods are designed to operate in a Euclidean framework they cannot be directly applied to compositions. Consequently, to avoid the complications of rethinking these tools within the linear vector space structure of the simplex, two different strategies to analyse compositions are suggested in statistical literature (stay-in-simplex or move from simplex to real space). Following the second approach, multilinear techniques such as Tucker3 or the Candecomp/Parafac (CP) model are proposed to analyse compositions arranged in three-way array. These techniques allow to extract hidden entities and capture the individual differences in a variety of conditions, preventing the loss of information that occurs when two-way analysis, like PCA (Principal Component Analysis), are carried out. In this work, we focus on the CP model.

Usually, the CP model is defined as an optimal representation of a trilinear model with respect to the criteria of the best low-rank approximation in a least squares sense. In fact the most widely used algorithm for fitting the model is the ALS-Alternating Least Square due to its appealing properties: a bounded loss function; guaranteed convergence; uniqueness of the solution under rather mild conditions; extendibility to higher order arrays; and possibility of imposing several constraints to the loading vectors. In recent years, however, different fitting alternatives to solve the trilinear decomposition problem were compared and discussed extensively, especially after the shortcomings of the iterative ALS algorithm were pointed out.

Several algorithms based on iterative procedures were introduced and tested in literature but only few of them reached reasonable results if compared to the ALS procedure. In particular the Self-Weighted Alternating Tri-Linear Decomposition SWATLD has turned out to be an acceptable alternative, presenting several advantages. Specifically, it appears less sensitive to over-factoring; relatively stable in presence of collinearity; less affected by the error model; capable of providing a correct estimation of the underlying factors; and also computationally fast. However, its inability to regularly find a solution in a least squares sense.

[I] Friday/Viernes, 11:10a.m., Room/aula 3, Session: R, Clustering and Visualization of Complex Data

[II] University of Naples L´Orientale, Department of Human and Social Sciences, Italy, mgallo@unior.it

The main focus of this work is to combine the merits of Parafac-ALS and Parafac-SWATLD in a contaminated compositional framework. The ALS algorithm is known to be particularly sensible to the presence of anomalous observations which might result in biased estimates of the parameters. Essentially this algorithm lacks resistance to extreme points due to the fact that it minimizes the objective function of the errors. On the other hand, the iterative principle at the base of SWATLD, in which a non-least square objective function is employed, could be an interesting starting point to obtain less affected estimates with an alternative fitting criteria instead of making use of a robust version of ALS . The complementary performance of SWATLD and ALS are thus combined and adapted to compositional data.

Keywords: Compositional Data, Log-ratios, CP model, Parafac-ALS, SWATLD.

References

[1] Aitchison, J. The statistical analysis of compositional data, 1986, Chapman and Hall London

[2] Kroonenberg P.M. Apply Multi-way Data Analysis,Wiley, Chichester, UK, 2008.

[3] Faber, N. K. M., Bro, R. and Hopke, P.K. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. Chemometrics and Intelligent Laboratory Systems, 65(1), 119-137, 2003.

[4] Tomasi, G. and Bro,R. A comparison of algorithms for fitting the PARAFAC model. Computational Statistics & Data Analysis, 50(7), 1700-1734, 2006.

[5] Zhang, S.R., Wu, H.L. and Yu R.Q. A study on the differential strategy of some iterative trilinear decomposition algorithms: PARAFAC-ALS, ATLD, SWATLD and APTLD. Journal of Chemometrics, 29, 179-192, 2015.

Aplicación del MANOVA a Aplicación de MANOVA en Experimentos en Parcelas Divididas[I]

Poster / Cartel

GARCIA BOLIVAR, JUDITH JOSEFINA[II] Mujica Betancourt, Ruben Darío[III]

Ecuador

This investigation has analyzed MANOVA for Split plot experiments. The general rationale is identical to that of the completely randomized factorial multivariate analysis of variance and fulfillments of MANOVA suppositions are assumed. There are techniques for afterwards interpretation of significant MANOVA that were not considered. It was assessed the effects of two factors where one of them has restrictions on its application. It provides four tests to prove the effects of each main effect and each interaction. For the development of this experimental situation, it was included a practical example that valued the effect of three levels of irrigation and four doses of N on the response variables measured in the leaves: %N, %P, %K and %Mg, in a crop nutrition investigation, with random blocks in a Split Plot arrangement. In this example, it was established that existed statistical differences related to the doses of nitrogen applied to the soil on the response variables.

Keywords: Split plot, multivariate analysis, practical example.

Resumen

En el presente estudio se realizó el cálculo e interpretación del MANOVA en diseños en parcelas divididas, la teoría general es idéntica a la de un análisis factorial multivariado. Aquí se asume cumplimiento de los supuestos del MANOVA. Existen técnicas para la interpretación posterior de los resultados de un MANOVA significativo las cuales no fueron contempladas. Se valoró el efecto de dos factores donde uno de ellos tiene restricciones para su aplicación. Se utilizaron cuatro pruebas multivariadas para determinar la significación de cada efecto principal y sus interacciones. Para el desarrollo de ese tipo de situación experimental, se incluyó un ejemplo práctico que valoró el efecto de tres láminas de riego y cuatro dosis de N en las variable respuesta medidas en las hojas: %N, %P, %K y %Mg, en un ensayo de nutrición en un cultivo, realizado en bloques al azar en un arreglo en Parcelas Divididas. En este ejemplo se estableció que existían diferencias estadísticamente significativas en las dosis de nitrógeno aplicadas al suelo sobre las variables respuesta evaluadas.

Palabras clave: Parcelas Divididas, análisis multivariado, ejemplo aplicativo.

References

[1] Cuadras C. M. 1991. Métodos de Análisis Multivariante. PPU. Barcelona. España. 641.

[2] Infostat. 2004. Manual del Usuario. Grupo InfoStat, FCA. Universidad Nacional de Córdoba. Primera edición. Editorial Brujas. Argentina.

[3] Johnson, R.A. and D. W. Wichern. 2007. Applied Multivariate Statistical Analysis, 6th edition Pearson International.

[4] Rencher A. 1998. Multivariate Statistical Inference and its applications. John Wiley & Sons, USA. 559.

[5] Stevens J. P. 1979. Comment on Olson: Choosing a Test statistic in multivariate analysis of variance. Psychological Bulletin 86, 335-360.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad politecnica estatal del Carchi,provincia del Carchi, Ecuador, Tulcan, garciaj66@gmail.com

[III] Universidad Politécnica Estatal del Carchi, Ecuador,Tulcan, ruben.mujica@upec.edu.ec

Climate data imputation using the singular value decomposition: An empirical comparison[I]

Poster / Cartel

GARCÍA PEÑA, MARISOL[II] Arciniegas Alarcón, Sergio[III] Barbin, Décio[IV]

Colombia

Often, on weather studies complete observations are necessary (no missing information) to perform appropriate analysis of the data that are collected during a certain period of time, at different stations of one or more regions of interest. It is common to find climatological data sets with missing data due to several reasons, such as failure of measuring instruments, extreme weather conditions and errors in typing. A very common way to analyse data from studies with missing information is imputed the missing observations and subsequently apply classical procedures on the completed data (observed + imputed). A widely used method in the literature is to use the mean as imputation.

Recently, four methods have been developed which are based in the singular value decomposition of a matrix (SVD). The imputation methods used in this study will be the Mean, EM-SVD, biplot imputation, imputation by an approximation of weighted lower rank (or EMSJ) and GabrielEigen imputation. The aim of this paper is to evaluate the new developments making a comparison by means of a simulation study based on two complete matrices of real data. One corresponds to the historical precipitation of Piracicaba/SP - Brazil and the other matrix corresponds to multivariate meteorological characteristics in the same city during the year 2012. In the study, values were deleted randomly at different percentages with subsequent imputation, comparing the methodologies by three criteria: the normalised root mean squared error, the similarity statistic of Procrustes and the Spearman correlation coefficient. It was concluded that the SVD should be used only when multivariate matrices are analysed and when matrices of precipitation are used, the monthly mean overcome the performance of other methods based on the SVD.

Keywords: imputation, singular value decomposition, missing values.

References

[1] Arciniegas-Alarcón, S.; García-Peña, M.; Dias, C.T.S. and Krzanowski, W.J.(2010). "An alternative methodology for imputing missing data in trials with genotype-by-environment interaction", *Biometrical Letters, n^o 47*, p.1–14.

[2] Gabriel, K.R. (2002). "Le biplot - outil d´exploration de données multidimensionelles", *Journal de la Société Française de Statistique, v.143, n^o 3-4*, p. 5–55.

[3] García-Peña, M.; Arciniegas-Alarcón, S. and Barbin, D. (2014). "Climate data imputation using the singular value decomposition: an empirical comparison", *Revista Brasileira de Meteorologia, n^o 29* , p. 527–536.

[4] Krzanowski, W.J. (1988). "Missing value imputation in multivariate data using the singular value decomposition of a matrix", *Biometrical Letters, v.25*, p. 31–39.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Pontificia Universidad Javeriana, Departamento de Matemáticas, Colombia, Bogotá, luzmara@gmail.com

[III] Departamento de matemáticas, Universidad Militar Nueva Granada, Cajicá, Colombia, sergio.arciniegas@gmail.com

[IV] Departamento de Ciências Exatas, Universidade de São Paulo, Piracicaba, Brasil, decio.barbin@usp.br

Mixed finite element and related methods for some nonlinear problems in fluid mechanics[I]

Plenary Talk / Conferencia Plenaria

GATICA, GABRIEL[II]

Chile

In this talk we describe some recent results concerning the application of mixed finite element and mixed virtual element methods to numerically solve diverse nonlinear boundary value problems in fluid mechanics. The models to be considered include: coupled flow-transport problems; Navier-Stokes equations; stationary Boussinesq problems; stress-assisted diffusion problems; and quasi-Newtonian flows. Our methods are based, on one hand, on the introduction of modified pseudostress tensors and, on the other hand, on the augmentation of the original variational formulations with suitable Galerkin type terms arising from the constitutive and equilibrium equations, the relations defining the additional unknowns, and the boundary conditions involved. In turn, the main tools yielding the solvability of the continuous and discrete formulations include the classical Schauder, Banach, and Brouwer theorems for fixed-point equations, the Lax-Milgram theorem, the Babuska-Brezzi theory, the Sobolev embedding and Rellich-Kondrachov compactness theorems, and well-known results on monotone operators. Some numerical results illustrating the good performance of the methods and confirming the theoretical rates of convergence are reported.

Keywords: Mixed FEM, Mixed VEM, Navier-Stokes, Boussinesq.

References

[1] Gatica, G.N. and Sequeira, F.: Analysis of the HDG method for the Stokes- Darcy coupling. Numerical Methods for Partial Differential Equations, vol. 33, 3, pp. 885-917, (2017)

[2] Cáceres, E. and Gatica, G.N.: A mixed virtual element method for a pseudo- stress-velocity formulation of the Stokes problem. IMA Journal of Numerical Analysis, vol. 37, 1, pp. 296-331, (2017).

[3] Cáceres, E., Gatica, G.N. and Sequeira, F., A mixed virtual element met- hod for the Brinkman problem. Mathematical Models and Methods in Applied Sciences, vol. 27, 4, pp. 707-743, (2017).

[I] Wednesday/Miércoles, 10:30a.m., Auditorium, Session: CONF, Conference 2

[II] Universidad de Concepción, CI2MA, Chile, Concepción, ggatica@ci2ma.udec.cl

Perturbing compact objects with mass quadrupole moment[I]

Communication / Ponencia

GÓMEZ OVARES, PEDRO ANTONIO[II] Frutos Alfaro, Francisco[III]

Costa Rica

We derive a new partial differential equation, the so called master equation, describing the perturbations of scalar, gravitational and electromagnetic fields of compact objets with mass quadrupole moment and rotation.

To this goal, the Newman-Penrose formalism is applied. To include the mass quadrupole moment into this equation, a Kerr-like metric is employed. The advantage of this procedure is that our Master equation can be applied to compact objects, such as neutron stars, otherwise this would not be possible. We derive a new partial differential equation, the so called master equation, describing the perturbations of scalar, gravitational and electromagnetic fields of compact objets with mass quadrupole moment and rotation.

Keywords: general relativity, solutions of Einstein's equations, approximation procedures, weak fields.

References

[1] F. Frutos-Alfaro, Approximate Kerr-like Metric with Quadrupole, International Journal of Astronomy and Astrophysics, 6, 334-345, 2016.

[2] Teukolsky, S. A. (1973). Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations. The Astrophysical Journal, 185, 635-648.

[I] Tuesday/Martes, 1:50p.m., Auditorium, Session: Apl1, Applications / Aplicaciones 1

[II] Universidad de Costa Rica, , San José, Costa Rica, pedro.gomezovares@gmail.com

[III] Universidad de Costa Rica, Escuela de Física, San José, Costa Rica, frutos62@gmail.com

An R package for testing goodness of fit[I]

Communication / Ponencia

GONZÁLEZ ESTRADA, ELIZABETH[II] Villaseñor, J. A.[III]

México

The R package "goft" is presented in this talk. This package implements three types of goodness-of-fit tests for some widely used probability distributions when the parameters are unknown. The tests are either based on data transformations or on the ratio of two estimators of a dispersion parameter. Some correlation tests are also implemented. The package´s functionality is illustrated with two examples by using data sets from the areas of finance and reliability.

Keywords: hypothesis testing, probability distributions, R.

References

[1] Villaseñor, J.A., González, E. 2009. "A Generalization of Shapiro-Wilk's Test for Multivariate Normality", Communications in Statistics - Theory and Methods 38: 1870-1883.

[I] Wednesday/Miércoles, 11:30a.m., Room/aula 1, Session: LACSC–1, LACSC–Statistical Computing 4

[II] Colegio de Postgraduados, Posgrado en Estadística, México, Texcoco, egonzalez@colpos.mx

[III] Colegio de Postgraduados, Mexico jvillasr@colpos.mx

Design and construction of an experimental methodology for the detection of radical-pair reactions in butterflies[I]

Poster / Cartel

GONZÁLEZ LEÓN, ROBERTO ANDRÉS[II] Hernández Jiménez, Kevin[III]

Waterhouse Morales, Priscilla[IV] Herrera Sancho, Óscar A.[V]

Costa Rica

The migratory behaviour of the monarch butterflies ($Danaus\ plexippus$) has been extensively studied for the multiple mechanisms that it involves, including temperature, blossoming of different plants, position of the sun and circadian clock queues. One of the compasses that had been proposed is a magnetic compass that guides the butterflies using the magnetic field of the Earth.

Here a device and experimental setting consisting of a double 1 m^3 Helmholtz coil is built and characterized, designed for making different behavioural experiments upon demonstrating and exploring the magnetic compass of $Danausplexippus$ and other migrating butterflies. Also, a protocol is proposed for making such an experiment and quantify the effect of a magnetic field on the orientation of the specimen.

Keywords: quantum biology, butterfly, migration mechanism, magnetic field.

References

[1] Gegear, R. J., Casselman, A., Waddell, S., & Reppert, S. M. (2008). Cryptochrome mediates light–dependent magnetosensitivity in Drosophila. Nature, 454(7207), 1014–1018. https://doi.org/10.1038/nature07183

[2] Guerra, P. A., Gegear, R. J., & Reppert, S. M. (2014). A magnetic compass aids monarch butterfly migration. Nature Communications, 5. https://doi.org/10.1038/ncomms5164

[3] Liedvogel, M., & Mouritsen, H. (2010). Cryptochromes–a potential magnetoreceptor: what do we know and what do we want to know? Journal of The Royal Society Interface, 7($Suppl_2$),S147–S162. https://doi.org/10.1098/rsif.2009.0411.focus

[4] Reppert, S. M., Gegear, R. J., & Merlin, C. (2010). Navigational mechanisms of migrating monarch butterflies. Trends in Neurosciences, 33(9), 399–406. https://doi.org/10.1016/j.tins.2010.04.004

[5] Reppert, S. M., Guerra, P. A., & Merlin, C. (2016). Neurobiology of Monarch Butterfly Migration. Annual Review of Entomology, 61(1), 25–42. https://doi.org/10.1146/annurev-ento-010814-020855

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad de Costa Rica, Escuela de Física, Costa Rica, San José, roberto9695@gmail.com

[III] Universidad de Costa Rica, San José, Costa Rica, kevin.hernandezjimenez@ucr.ac.cr

[IV] Universidad de Costa Rica, San José, Costa Rica, priscilla.waterhouse@ucr.ac.cr

[V] Universidad de Costa Rica, San José, Costa Rica, oscar.herrerasancho@ucr.ac.cr

A generalized algorithm for the enumeration of chiral and achiral isomers of a n–membered ring monocycloalkane m–polyheteroalkylsubstituted[I]

Communication / Ponencia

GONZÁLEZ LEÓN, ROBERTO ANDRÉS[II] Camacho, Cristopher[III]

Costa Rica

A common problem in combinatorial chemistry is the enumeration of possible compounds that can be produced in a chemical synthesis. Particularly, differentiating whether the possible products present a given stereochemistry, as they can be in a racemic mixture.

Here, a generalized algorithm for the enumeration of chiral and achiral isomers of a m-polyheterosubstituted monocycloalcane with ring size n is presented, derived from combinatorial formulae. The formulae are derived for r kind of substituents of k_r carbons each. This algorithm is also applicable for homomorphically substituted monocycloalkane and generalizes both cases into one.

Keywords: Combinatorial Chemistry, Group Theory, Graph Theory, Stereochemistry, Topology.

Resumen

Un problema común en la química combinatoria es la enumeración de posibles productos en una síntesis química. Particularmente, diferenciar si los posibles productos presentan una estereoquímica dada, ya que pueden estar en una mezcla racémica.

Aquí se presenta un algoritmo generalizado para la enumeración de isómeros quirales y aquirales de un monocicloalcano m-poliheteroalquilsustituido de tamaño de anillo n, derivado a partir de fórmulas combinatorias. Las formulas son derivadas para r tipos de sustituyentes de kr carbonos cada uno. Este algoritmo es también applicable para un monocicloalcano sustituido homomórficamente, generalizando ambos casos en uno solo.

Palabras clave: Química Combinatoria, Teoría de Grupos, Teoría de Grafos, Estereoquímica, Topología.

References

[1] Harris, J., Hirst, J. L., & Mossinghoff, M. (2008). Combinatorics and Graph Theory. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-79711-3

[2] Nemba, Robert M., & Balaban, A. T. (1998). Algorithm for the Direct Enumeration of Chiral and Achiral Skeletons of a Homosubstituted Derivative of a Monocyclic Cycloalkane with a Large and Factorizable Ring Size n. Journal of Chemical Information and Computer Sciences, 38(6), 1145–1150. https://doi.org/10.1021/ci980079f

[I] Tuesday/Martes, 4:20p.m., Auditorium, Session: Apl3, Applications / Aplicaciones 3

[II] University of Costa Rica, School of Chemistry, Costa Rica, San José, roberto9695@gmail.com

[III] University of Costa Rica, Costa Rica, cristopher.camacho@ucr.ac.cr

[3] Nemba, Robert M., & Balaban, A. T. (2002). Enumeration of Chiral and Achiral Isomers of an n-Membered Ring with n Homomorphic Alkyl Groups. MATCH. Communications in Mathematical and in Computer Chemistry, 46, 235—250.

[4] Nemba, Robert Martin. (1996). Solution générale du problème de dénombrement des stéréoisomères d´un cycloalcane monocyclique homosubstitué. In Comptes Rendus de l´ Académie des Sciences (Vol. 323, pp. 773–779). Paris.

[5] Parks, C. A., & Hendrickson, J. B. (1991). Enumeration of monocyclic and bicyclic carbon skeletons. Journal of Chemical Information and Computer Sciences, 31(2), 334–339.

A new successive piecewise linearization algorithm for nonsmooth optimization[I]

Plenary Talk / Conferencia Plenaria

GRIEWANK, ANDREAS[II] Walther, Andrea[III]

Ecuador & Germany

As indicated by the title this will first focus on nonsmooth optimization but later also discuss applications of the piecewise linearization approach [1] to nonsmooth equation solving and other fundamental computational tasks.

Many problems in machine learning involve objective functions that are piecewise smooth [7] due to the occurrence of absolute values mins and maxes in their evaluation procedures. See e.g. [8]. For such function we derived in [5] first order (KKT) and second order (SSC) optimality conditions, which can be checked on the basis of a local piecewise linearization [1] that can be computed in an AD like fashion, e.g. using ADOL-C or Tapenade.

In that analysis, a key assumption on the local piecewise linearization was the Linear Independence Kink Qualification (LIKQ), a generalization of the Linear Independence Constraint Qualification (LICQ) known from smooth Nonlinear Optimization. A rather surprising consequence is that checking the optimality conditions is not at all combinatorial but can be done with a cubic effort like in the classical smooth case. Moreover, as we show here first under LIKQ with SSC the natural algorithm of successive piecewise linear optimization with a proximal term (SPLOP) achieves a linear rate of convergence. A version of SPLOP has already been implemented and tested in [4, 1].

Secondly, we observe that, even without any kink qualifications, local optimality of the nonlinear objective always requires local optimality of its piecewise linearization, and strict minimality of the latter is in fact equivalent to sharp minimality of the former. Moreover, we show that SPLOP will converge quadratically to such sharp minimizers, where the function exhibits linear growth. These results are independent of the particular function representation, and allow in particular duplications of switching variables and other intermediates.

We note that the classical theory for subgradient [9] , proximal [6] and bundle [5] methods usually only yields convergence rates like $1/sqrt(k)$ or $\log(k)/k$, where k is the iteration counter. Only for strongly convex functions a linear convergence rate can sometimes be established. Our assumptions LIKQ and SSC are certainly quite strong, but they do not require convexity, even locally, near a minimizer. In case of the Lasso problem $\min ||x||_1 + \rho||Ax - b||$ our method coincides with ISTA as described in [6].

Our current implementation of SPLOPT allows the verification of the theoretical results mentioned above on the usual set of academic test problems. The number of outer iterations is usually extremely low compared to more established approaches. However, the setting up and solving the local piecewise linear problem is not yet adapted to large structured problems.

In effect we have to solve a sequence of closely related, convex Quadratic Optimization Problems (QOP), while marching through a polyhedral decomposition of the variable domain. For several aspects, like the selection of the next polyhedron, the handling of the many locally redundant constraints, and the exploitation of

[I] Wednesday/Miércoles, 11:20a.m., Room/aula 2, Session: CONF, Conference 7

[II] Yachay Tech, matemática, Ecuador, Urcuquí, griewank@yachaytech.edu.ec

[III] Institut für Mathematik, Universität Paderborn, Paderborn, Germany, andrea.walther@uni-paderborn.de

sparsity there are obvious improvements, which we are currently exploring and implementing. We expect to present results at least on the Lasso problem [6] and fuzzy pattern trees as described in [8].

Piecewise linear approximations to piecewise smooth vector functions can be used to efficiently solve nonlinear equations and integrate Lipschitzian dynamical systems without loss of accuracy due to kinks and jumps. In the case of equations, successive piecewise linearization can achieve quadratic convergence to roots even when the vector functions is only open but not necessarily injective in its neighborhood. It is also possible to compute generalized Jacobians for use in the well-known semi-smooth Newton methods. On initial value problems in ODEs generalizations of the mid point and the trapezoidal rule maintain a local third order error truncation error and thus achieve uniform global second order convergence. Under a week form of transversality, Richardson extrapolation increases the convergence order to three. Finally, we discuss other possible applications like the quadratures for piecewise smooth functions.

Keywords: Piecewise smooth, KKT conditions, Linear or Quadratic Convergence.

References

[1] S. Fiege, A. Walther, and A. Griewank. An algorithm for nonsmooth optimization by succes- sive piecewise linearization. Technical report, Universität Paderborn, 2016. available at optimization-online.

[2] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. Optimization Methods and Software, 28(6):1139–1178, 2013.

[3] A. Griewank and A. Walther. First and second order optimality conditions for piecewise smooth objective functions. Optimization Methods and Software, 31(5):904–930, 2016.

[4] A. Griewank, A. Walther, S. Fiege, and T. Bosse. On Lipschitz optimization based on gray-box piecewise linearization. Mathematical Programming Series A, 158(1–2):383–415, 2016.

[5] N. Karmitsa and M. Mäkelä. Limited memory bundle method for large bound constrained nons mooth optimization: convergence analysis. Optimization Methods and Software, 25(6):895–916, 2010.

[6] Juan Peypouquet. Convex Optimization in Normed Spaces. Number ISBN 978–3–319–13710–0. Springer, 2015.

[7] S. Scholtes. Introduction to Piecewise Differentiable Functions. Springer, 2012.

[8] R. Senge and E. Hüllermeier. Top-down induction of fuzzy pattern trees. IEEE Transactions on Fuzzy Systems,, 19(2):241–252, 2011.

[9] N.Z. Shor. Nondifferentiable optimization and polynomial problems. Kluwer, 1998.

Nonsmooth Analysis and Optimization via Piecewise Linearization/Differentiation[I]

Short course / Curso corto

GRIEWANK, ANDREAS[II] Walther, Andrea[III]

Ecuador & Germany

In the plenary talk on a new method for nonsmooth optimization we demonstrate how successive piecewise linearization can be used to rapidly attain accurate solutions of piecewise smooth optimization [7] problems. The key assumption is that nonsmoothness can be cast in terms of the absolute value function, min, and max. Under the same assumption we cover in this tutorial the wider context and the application of piecewise linearization to other fundamental computational tasks. In particular we explain and elaborate.

- The automatic derivation of piecewise linearizations by an extension of Algorithmic Differentiation [8].
- The approximation and continuity properties of these piecewise linear (PL) local model [1].
- The constructive characterization of local properties like stationarity and minimality in terms of the PL models [5].
- The exact computation of generalized gradients and Jacobians without any regularity assumptions [6].
- The near equivalence between differential regularity and local convexity, which is generally co-NP complete [9].
- The reformulation of PL systems of equations as linear complementarity problems and the related solvability conditions and algorithms [2].
- The quadratic convergence of generalized Newton to roots where the vector function is open but not necessarily injective[3] .
- The generalization of the trapezoidal and midpoint rule to maintain global second order convergence [4] on Lipschitzian IVPs. .
- The generalization of the PL concept to discontinuos functions, such as right hand sides in dynamical systems.

The theoretical explanations are complemented by computational demonstrations and numerical experiments.

[I] Wednesday/Miércoles, 8:00a.m., Room/aula 1, Session: Tut, Tutorial 7

[II] School of Mathematical Sciences and Information Technology, Yachay Tech, matemática, Ecuador, Urcuquí, griewank@yachaytech.edu.ec

[III] Institut für Mathematik, Universität Paderborn, Paderborn, Germany, andrea.walther@uni-paderborn.de

References

[1] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013.

[2] A. Griewank, J.-U. Bernt, M. Randons, and T. Streubel. Solving piecewise linear equations in abs-normal form. *Lin. Algebra and its Appl.*, 471:500–530, 2015.

[3] A. Griewank, R. Hasenfelder, L. Lehmann, M. Radons, and T. Streubel. Piecewise linear secant approximation via algorithmic piecewise differentiation. *Optimization Methods and Software*, pages online pp1–20, 2017.

[4] A. Griewank, R. Hasenfelder, M. Radons, and T. Streubel. Integrating lipschitzian dynamical systems using piecewise algorithmic differentiation. *Optimization Methods and Software*, pages online pp1–19, 2017.

[5] A. Griewank and A. Walther. First and second order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5):904–930, 2016.

[6] B. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Springer, 2006.

[7] S. Scholtes. *Introduction to Piecewise Differentiable Functions*. Springer, 2012.

[8] A. Walther and A. Griewank. *Combinatorial Scientific Computing*, chapter Getting Started with ADOL-C, pages 181–202. Chapman-Hall CRC Computational Science, 2012.

[9] A. Walther and A. Griewank. Characterizing and testing subdifferential regularity for piecewise smoothobjective functions. Technical report, Universität Paderborn, 2017. available at optimization-online.

Topological co-clustering and visualization for heterogenoeous data[I]

Communication / Ponencia

GROZAVU, NISTOR[II] Rogovschi, Nicoleta[III] Bennani, Younès[IV]

France

Topological learning is a recent direction in Machine Learning which aims to develop methods grounded on statistics to recover the topological invariants from the observed data points. Most of the existed topological learning approaches are based on graph theory or graph-based clustering methods. The topological learning is one of the most known technique which allows clustering and visualization simultaneously. At the end of the topographic learning, the "similar" data will be collect in clusters, which correspond to the sets of similar observations. These clusters can be represented by more concise information than the brutal listing of their patterns, such as their gravity center or different statistical moments. As expected, this information is easier to manipulate than the original data points. The neural networks based techniques are the most adapted to topological learning as these approaches represent already a network (graph).

Unsupervised feature learning algorithms aim to find good representations for data, which can be used for different tasks i.e. classification, clustering, reconstruction, visualization,...
Recently, the SNE and t-SNE [van der Maaten and Hinton, 2008] methods have shown high feature learning performance used for dimensionality reduction and visualization [Kitazono et al., 2016].

This work introduces a new topological co-clustering approach to cluster and to visualize high heterogeneous dimensional datasets based on Topological Spectral Clustering (TSC) and the SNE (Stochastic Neighbor Embedding). The SNE method which performs good results for visulaization allows a projection of the dataset in low dimensional spaces that make it easy to use for very large datasets, even if the data is heterogeneous. Using SNE and spectral clustering during the learning process will allow to reduce the dimensionality and to preserve the topology of the dataset by increasing the clustering accuracy.

Given a data matrix represented as vectors of variables (p observations and n features), the goal of the unsupervised transformation of feature space is to produce another data matrix of dimension (p, n') (the transformed representation of n' new latent variables) or a similarity matrix between the data of size (p, p). Applying a model on the transformed matrix should provide better results compared to the original dataset.
The transformation of the feature space is done in two steps. First, we decompose the sparse data matrix using the SNE approach, then the matrix of latent variables obtained after this decomposition is used to co-cluster the dataset.

Spectral clustering ([Ng et al., 2001]; [Jordan et al., 2003]) has been well studied in the literature. It is easy to efficiently implement and often outperforms traditional clustering methods such as K-Means. The spectral clustering methods based on the graph partitioning theory focus on finding the best cuts of a graph that optimize certain predefined criterion functions. The optimization of the criterion functions usually leads to the

[I] Friday/Viernes, 11:30a.m., Room/aula 3, Session: R, Clustering and Visualization of Complex Data
[II] Université de Paris 13, , Sorbonne Paris Cité, France, nistor.grozavu@lipn.univ-paris13.fr
[III] LIPADE University of Paris 5, France nicoleta.rogovschi@parisdescartes.fr
[IV] LIPN-UMR CNRS 7030, Paris 13 University, France, younes.bennani@lipn.univ-paris13.fr

computation of singular vectors or eigenvectors of certain graph affinity matrices. One of the disadvantages of this type of methods is the computing of the similarity matrix which is very expensive computationally.

At the end of learning the SNE, the "similar" data will be separated from the dissimilar data. As expected, this information is easier to manipulate than the original data points.

Algorithm 1 Proposed topological co-clustering algorithm

Input: n data points $x_1, x_2, ..., x_n \in \mathbb{R}^m$; Cluster number k ;
Output: k clusters;

1. Compute pairwise affinities $p_{j|i}$
2. **for** i=1 to Iter **do**
   compute low-dimensional affinities q_{ij}
   compute the low dimensional data (Y) using the gradient descend
3. **end for**
4. Compute the prototypes matrix W using the TSC algorithm on the low dimensions Y algorithm

The topological spectral co-clustering (TSC) proposed in [Rogovschi et al., 2012] consists of two parts :

(1) Spectral decomposition of the data matrix: This step consists of a spectral decomposition of a weighted data matrix using an SVD and the construction of a new data matrix $\mathbf{D}$ adapted to the problem of co-clustering.

(2) Co-clustering of original data A by applying the Self-Organizing Maps (SOM) algorithm on $\mathbf{D}$.

We tested the proposed model on five datasets (waveform, wdbc, madelon, spambase, mnist) and we can note that our method outperforms the classical k-means and the spectral clustering. We increased the accuracy index by 0.5% for wdbc, 1.3% for waveform, 0.3% for madelon, 1,4% for spambase and 3.4% for mnist dataset. But we have to note here that the goal of this apparoahc is also to preserve the topological structure of the data for visualization.

Keywords: Visualization, spectral clustering, Stochastic Neighbor Embedding.

References

[Jordan et al., 2003] Jordan, M. I., Bach, F. R., and Bach, F. R. (2003). Learning spectral clustering. In *Advances in Neural Information Processing Systems 16*. MIT Press.

[Kitazono et al., 2016] Kitazono, J., Grozavu, N., Rogovschi, N., Omori, T., and Ozawa, S. (2016). t-distributed stochastic neighbor embedding with inhomogeneous degrees of freedom. In *Neural Information Processing - 23rd International Conference, ICONIP 2016, Kyoto, Japan, 2016, Proceedings, Part III*, pages 119–128.

[Ng et al., 2001] Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press.

[Rogovschi et al., 2012] Rogovschi, N., Labiod, L., and Nadif, M. (2012). A spectral algorithm for topographical co-clustering. In *The 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, June 10-15, 2012*, pages 1–6.

[van der Maaten and Hinton, 2008] van der Maaten, L. and Hinton, G. (2008). Visualizing high-dimensional data using t-sne.

A Mixed Mimetic Formulation for Porous Media Flow[I]

Communication / Ponencia

GUEVARA JORDAN, JUAN MANUEL[II] Chacon, Chanel[III]

Venezuela

A new mixed mimetic formulation for the approximation and simultaneous solution of the continuity equation and Darcy's law is presented. This system describes the pressure and velocities of the fluid flow in porous media. The mimetic approximation is based on a novel second order tensor representation of mimetic operators, which generates a large linear system for pressure and velocity with a strong block structure. This system is solved by the Schur's complement for the pressure or the velocity. Numerical experiments give evidence that solution by the Schur's complement for the velocity produces linear systems with a better condition number and lower dimension, so its solution by iterative techniques is faster than the traditional approach based on the Schur's complement for the pressure.

Keywords: mimetic, Darcy's law, Schur's complement, velocity, tensor.

Resumen

Se presenta una novedosa formulación mimética mixta para aproximar la solución del sistema de ecuaciones formado por la ecuación de continuidad y la ley de Darcy para obtener la presión y la velocidad que describen el flujo de fluidos en medios porosos. En ella se utilizan una novedosa representacion tensorial de los operadores miméticos de segundo orden generando un sistema lineal con estructura de bloques que se resuelve utilizando el complemento de Schur para la presión o la velocidad. Se evidencia, mediante ejemplos numéricos, que el sistema lineal con menor dimensión y mejor condicionamiento es generado mediante el complemento de Schur de la velocidad, lo cual acelera la convergencia de los métodos iterativos en comparación con el tradicional complemento de Schur para la presión.

Palabras clave: mimetico, ley de Darcy, complemento de Schur, velocity, tensor.

References

[1] Chacón Suescún, C.C. (2017). "Una Formulación Mimética Mixta para Ecuaciones Elípticas", *Tesis de Maestria en Ciencias de la Computacion, Universidad Central de Venezuela*.

[2] Castillo, J.E., and Miranda G. (2013). *Mimetic Discretizations Method*, Boca Raton, Florida: CRC Press.

[I] Tuesday/Martes, 2:35p.m., Room/aula 2, Session: Num3, Numerical Analysis / Análisis Numérico 2

[II] Universidad Central de Venezuela, Matematicas, Venezuela, Caracas, jmguevarajordan@gmail.com

[III] Universidad Nacional Abierta, Caracas, Venezuela,chchcon@hotmail.com

Teoremas fundamentales del bienestar económico[I]

Short course / Curso corto

GUTIÉRREZ RAMÍREZ, JOSÉ DE JESÚS[II]

México

This course is an introduction to the themes of the general equilibrium, where the existence of an economy is supposed to be n Consumers who have unsatiated preferences locally, a set of m Firms that own production sets other than vacuum and closed, and there is also an assumption that there is an initial endowment of k Goods; One of the results Main To expose is about that, Given the previous assumptions, If you have a balance with transfers then the allocation to companies and consumers will represent an optimal Pareto. Another of the results to be studied, has to do with that, If you have an economy with signatures, consumers and initial endowment, with convex preferences, not locally satiated, and also convex production sets, you will have to, For any allocation for consumers and firms that are optimal Pareto, there will be a price vector that will form, together with the allocation to companies and consumers, a cuasiequilibrium with transfers.

Also, some other results will be exhibited on the set of points that make up the Pareto border.

This course represents an interesting application to that of the microeconomics that uses elements of the analysis of R^n.

Keywords: Optimal, Pareto, cuasiequilibrium, transfers.

References

[1] Mas-Colell, Andreu,(1995). *Microeconomic Theory*

[I] Wednesday/Miércoles, 8:00a.m., Room/aula 2, Session: Tut, Tutorial 4

[II] Universidad Autónoma Metropolitana, Producción Económica, México, Ciudad de México, jgramirez@correo.xoc.uam.mx

The results of the duopolio in the Cournot and stackelberg models as Nash equilibria[I]

Communication / Ponencia

GUTIÉRREZ RAMÍREZ, JOSÉ DE JESÚS[II] Baca Lobera, Gloria Idalia[III]
Sánchez Guevara, Irene[IV]

México

The Theory of games is a set of concepts and formal mathematical methods to study human games, military conflicts, situations of conflict and cooperation between countries and between companies, and which is becoming an important instrument of study in all branches of the economy. The main objective of the Game Theory is to find the optimal strategies for each player, assuming that all the others play in turn with their optimal strategies. There are two major themes within game theory, one that corresponds to non-cooperative games, and the other is cooperative games. This separation is clear in terms of the methods of solution and in terms of interpretations. For non-cooperative games a complete description of the rules of the game is required, so that the strategies available to the players can be studied in detail. The objective is to find an appropriate set of strategies that will be called the solution of the game. In this paper we present a mathematical demonstration of the optimal results for the duopoly models of Cournot and Stakelberg that are effectively Nash equilibria.

Keywords: Continuous games, Nash equilibrium.

Resumen

La Teoría de juegos es un la Teoría de Juegos es encontrar las estrategias óptimas para cada jugador, suponiendo que todos los demás juegan a su vez con sus estrategias óptimas. Hay dos grandes temas dentro de la teoría de juegos, una que corresponde a juegos no cooperativos, y el otro tema es el de juegos cooperativos. Esta separación es tajante en cuanto los métodos de solución y en cuanto a las interpretaciones. Para los juegos no cooperativos se requiere una descripción completa de las reglas del juego, de manera que las estrategias disponibles a los jugadores puedan ser estudiadas en detalle. El objetivo es encontrar un conjunto adecuado de estrategias a las que se le llamarán la solución del juego. En este trabajo se presenta una demostración matemática de los resultados óptimos para los modelos de duopolio de Cournot y de Stakelberg que son efectivamente equilibrios de Nash.

Palabras clave: juegos continuos, equilibros de Nash.

References

[1] Harrington, J.E. (2015) Games, Strategies, and Decision Making. Segunda edición. Worth Publishers. EU.

[2] Gibbons, R. (1992), Game Theory for Applied Economists. Princenton University Press.

[I] Wednesday/Miércoles, 2:10p.m., Auditorium, Session: Mod1, Modeling / Modelacion 1

[II] Universidad Autónoma Metropolitana, Producción Económica, México, Ciudad de México, jjgribc@hotmail.com
[III] Universidad Autónoma Metropolitana, México, gbaca@correo.xoc.uam.mx
[IV] Universidad Autónoma Metropolitana, México, isabiro@yahoo.com.mx

Heuristic algorithms of coincidence for the estimation of movements in compression of images[I]

(Algoritmos heurísticos de coincidencia para la estimación de movimientos en compresión de imágenes)

Communication / Ponencia

HERNÁNDEZ GÓMEZ, FERNANDO J.[II] Guevara, Antonio Parajón[III]

Nicaragua

Combinatorial optimization problems appear very often in everyday life. In a combinatorial optimization problem we want to minimize or maximize a function defined on a discrete set. Exact methods for combinatorial optimization problems provide always the optimal solution but in some cases there are not able to do so in an efficient way. That is, their execution times grow exponentially with the size of the problem. There are lots of combinatorial optimization problems that belongs to the NP-hard class for which no polynomial algorithm are known. In these cases the goal is to design heuristics that provide good, approximate, solutions in an affordable computational time.

Motion estimation is very hard from the computational complexity point of view.It requires a big amount of memory and running time. It represents $\frac{2}{3}$ of the cost of encoding static or dynamic image sequences or video. The task is then, to minimize the *distortion rate* and to improve the visual quality. So, the focus of the research within the field of encoding is to look for efficient algorithms for carrying out the motion estimation.

In this work we propose heuristic algorithms, based on the frequency domain, that are applied on the coefficients of the discrete cosine transform or the wavelets transform. We also propose temporal domain algorithms as those of block matching which intend to maximize the coincidence of the present image with the reference one.

We implemented the algorithms with the software MATLAB. We also review basic concepts of image and video processing and of the compression and motion estimation algorithms most frequently used.

If the list of images to be analyzed has n elements there are 2^n feasible solutions. So, an exhaustive search is too slow and then, from the practical point of view, exhaustive search has to be avoided.

The evaluation of the algorithms was done with images provided by an acquisition system. We showed the improvement of the visual quality, the amount of information compressed or reconstructed and the behavior of the methods when looking for similarities between pixels or images.

We would like to contribute to the diffusion of new research in Nicaragua. This area is pretty new in the country.

Keywords: Motion estimation, heuristics, discrete cosine transformation. wavelets transformation, local search algorithms.

[I] Wednesday/Miércoles, 1:50p.m., Room/aula 2, Session: OR1, Operations Research / Investigacion de Operaciones

[II] Departamento de Matemática, UNAN-Managua, Nicaragua, fhernandez@dc.uba.ar

[III] Departamento de Matemática, UNAN-Managua, Nicaragua, Ramon.A.Parajon@uv.es

Resumen

La teoría de NP-Completitud plantea que los algoritmos exactos y eficientes son poco probable que exista para la clase de problemas NP −difíciles. Una forma de hacer frente a la dureza NP es relajar el requisito de optimalidad y en su lugar buscar soluciones que estén probablemente *cerca del óptimo*. Esta es la idea principal detrás de los algoritmos de aproximación, que tienen por nombre heurísticos o metaheurísticos.

El problema de estimación de movimiento es un proceso con alto grado de complejidad computacional, requiere suficiente espacio de memoria y tiempo de ejecución. Representa $\frac{2}{3}$ del costo de codificación de secuencia de imágenes estática, dinámicas y de vídeo. La principal tarea consiste en minimizar la *tasa de distorsión* y mejorar la calidad visual. Esto hace que las investigaciones en el campo de la codificación, compresión de imágenes y vídeo se centre en buscar algoritmos eficientes para llevar a cabo estimación de movimiento en un tiempo razonable.

Si se analiza una lista de imágenes de n elementos, hay 2^n soluciones factibles. Entonces, una búsqueda exhaustiva es demasiado lenta, incluso para pequeños valores del espacio de soluciones. Por lo tanto, desde un punto de vista práctico, es crucial contar con algoritmos heurísticos eficientes y rápidos que eviten la búsqueda exhaustiva.

En esta investigación diseñamos e implementamos algoritmos heurísticos, basados en el dominio de frecuencias, que se aplican sobre los coeficientes de la transformada discreta del coseno, wavelets. También, proponemos los algoritmos de dominio temporal como los de coincidencia por bloque, que centran su búsqueda en la máxima coincidencia de la imagen actual con la de referencia. Los algoritmos utilizados durante la implementación de este trabajo de investigación fueron escritos con el lenguaje de programación matemático MATLAB. Además, revisamos los conceptos básicos del procesamiento de imágenes, vídeo, algoritmos de compresión y estimación de movimiento frecuentemente utilizados.

La evaluación de los algoritmos se realizó con un conjunto de imágenes proporcionadas por un sistema de adquisición previo. Mostramos la mejora de la calidad visual, la cantidad de información comprimida o reconstruida y el comportamientos de los métodos en la búsqueda de similitudes entre píxeles o imágenes.

Por último, contribuimos a la difusión de nuevas investigación científica que conllevan a la ampliación y mejora del estudio, la generación de nuevos conocimientos, ya es un área joven dentro de la disciplina de la Educación de Nicaragua.

Palabras clave: Estimación de movimiento, Heurísticos, Transformada discreta del coseno. Wavelets, Algoritmos de búsqueda.

References

[1] Chen, W. H., Smith, C. H., and Fralick, S. C. (1977). *A fast computational algorithm for the discrete cosine transform"*. IEEE Transactions on communications, 25(9), 1004-1009.

[2] Cuevas, E., Zaldívar, D., Pérez-Cisneros, M., Sossa, H., Osuna, V. (2013). *Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)"*, Applied Soft Computing Journal 13 (6), pp. 3047-3059.

[3] DUFAUX, F. (2010). *Motion Estimation Techniques for Digital TV: A Review and a New Contribution"*.

[4] Evan Herbst, Xiaofeng Ren, and Dieter Fox. (2013) *Rgb-d flow: Dense 3-d motion estimation using color and depth"*. In Robotics and Automation (ICRA), IEEE International Conference on, pages 2276-2282. IEEE, 2013.

[5] Jagalingam, P., and Hegde, A. V. (2015). *A review of quality metrics for fused image"*. Aquatic Procedia, 4, 133-142.

Mathematical models in topology[I]

Communication / Ponencia

HERRERA CARRASCO, DAVID[II] Macías Romero, Fernando[III]
Ahuatzi Reyes, José Gerardo[IV]

México

A *continuum* is a compact connected metric space, with more than one point. The set of positive integers is denoted by $\mathbb{N}$. Given a continuum X and $n \in \mathbb{N}$, we consider the following hyperspaces of X: $2^X = \{A \subset X : A \text{ is a nonempty closed subset of } X\}$, $C_n(X) = \{A \in 2^X : A \text{ has at most } n \text{ components}\}$, $C(X) = C_1(X)$; and $F_n(X) = \{A \in 2^X : A \text{ has at most } n \text{ points}\}$. All these hyperspaces are metrized by the Hausdorff metric. The hyperspaces $F_n(X)$ and $C_n(X)$ are called the *n-th symmetric product of X* and the *n-fold hyperspace* of X, respectively.

In the theory of hyperspaces it is very useful to have geometric ideas of how they look. Since they are defined as certain classes of subsets of a given space, this task is not easy. For this reason, we try to construct models for them. A model for a given hyperspace K(X) is a topologically equivalent space, where the elements are points instead of subsets.

From the geometric point of view, models of hyperspaces is a very attractive subject. Moreover, models are a very powerful tool to suggest properties and results on hyperspaces. In this talk we present some hyperspace models of continuum.

Keywords: Continuum, Euclidean space, hyperspace, model.

References

[1] K. Borsuk, On the third symmetric of the circunference, Fund. math. 36 (1949), 235-244.

[2] David Herrera carrasdco, maría de J. López, fernando macías Romero, Almost meshed locally connected continua have unique second symmetric products, Topology Appl. 209 (2016), 1-13.

[3] Alejandro Illanes, A model for the hyperspaces C_2(¨1), Question Answers Gen. Topology 22(2004), 117-130.

[4] Alejandro Illanes, Models of hyperspaces, Topology Proc. 41(2013). 1-26

[I] Wednesday/Miércoles, 4:40p.m., Room/aula 2, Session: DS, Dynamical Systems / Sistemas Dinamicos

[II] Benemérita Universidad Autónoma de Puebla, Posgrado de Matemáticas, Puebla, México, dherrera@fcfm.buap.mx

[III] Benemérita Universidad Autónoma de Puebla, Puebla, México, fmacias@fcfm.buap.mx

[IV] Benemérita Universidad Autónoma de Puebla, Puebla, México, Puebla, jgahuatzi@gmail.com

Spatial Interpolation of Precipitation[I]

Communication / Ponencia

HERRERA DAZA, EDDY[II] Baron, Carlos[III] Chaustre, Santiago[IV]

Colombia

The availability of rainfall data is fundamental for most hydrological analyses .There are many methods of spatial interpolation of rainfall despite their effectiveness and ease of use variables, the choice of one or the other depends t also on the territorial context of the area ,surface, network density, etc.

The purpose of this study is to determine the most adequate rainfall interpolation technique to Bogotá city using radar data. Different techniques of numerical interpolation were tested on a set of neighborhoods to determine the most efficient interpolator

Keywords: Spatial Interpolation, Precipitation, Numerical interpolation.

Resumen

La disponibilidad de datos de lluvia es fundamental para la mayoría de los análisis hidrológicos. Existen muchos métodos de interpolación espacial de las precipitaciones a pesar de su efectividad y facilidad de uso. La elección de uno u otro depende también del contexto territorial del área, superficie, densidad de red, etc.

El propósito de este estudio es determinar la técnica de interpolación de lluvia más adecuada para la ciudad de Bogotá utilizando datos de radar. Se probaron diferentes técnicas de interpolación numérica en un conjunto de vecindarios para determinar el interpolador más eficiente.

Palabras clave: Interpolación Espacial, Precipitación, Métodos numéricos.

References

[1] Bárdossy A, Pegram G (2013) Interpolation of precipitation under topographic influence at different time scales. Water Resour Res 49:4545–4565

[2] Syed, K. H., D. C. Goodrich, D. E. Myers, and S. Sorooshian (2003), Spatial characteristics of thunderstorm rainfall fields and their relation to runoff, J. Hydrol., 271, 1 – 21

[3] Caruso C. & Quarta F., 1998. Interpolation methods comparison. Comput. Math. Appl., 35(12), 109–126.

[I] Tuesday/Martes, 3:15p.m., Room/aula 2, Session: Num3, Numerical Analysis / Análisis Numérico 2

[II] Pontificia Universidad Javeriana, Cundinamarca, Colombia, Bogotá, eherrera@javeriana.edu.co

[III] Pontificia Universidad Javeriana,Bogotá,Colombia,baron_ carlos@javeriana.edu.co

[IV] Pontificia Universidad Javeriana,Bogotá,Colombia, santiago-chaustre@javeriana.edu.co

Estimation of stochastic volatility models using optimized filtering algorithms[I]

Communication / Ponencia

INFANTE QUÍRPA, SABA RAFAEL[II] Luna, Cesar[III] Sánchez, Luis[IV]
Hernández, Aracelis[V]

Ecuador

We describe and implement two recursive filtering algorithms, the optimized particle filter, and the Viterbi algorithm, which allow the joint estimation of states and parameters of continuous-time stochastic volatility models, such as the Cox Ingersoll Ross and Heston model. In practice, good parameter estimates are required so that the models are able to generate accurate forecasts. To achieve the objectives the proposed algorithms were implemented using daily empirical data from the time series of the $S\&500$ returns of the stock exchange index. The proposed methodology facilitates computational calculations of the marginal likelihood of states and allows the reconstruction of unknown states in a suitable way, and reliable estimation of the parameters. To measure the quality of estimation of the algorithms, we used the square root of the mean square error as a measure of goodness of fit. The estimated errors are insignificant for the analyzed data and the two models considered. We also calculated the execution times of the algorithms, demonstrating that the Viterbi algorithm has less execution time than the optimized particle filter.

Keywords: Stochastic volatility models, The optimized particle filter, The Viterbi algorithm filter.

References

[1] Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica, 70:223–262.

[2] Aït-Sahalia, Y., and Kimmel, R. (2007) Maximum likelihood estimation of stochastic volatility models. Financ. Econ. 134:507–551.

[3] Andrieu, C., Doucet, A., and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods. J. of the Royal Statistical Society B, 72(3):269–342.

[4] Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. J. of political economy, 81(3): 637–654.

[5] Carvalho, C., Johannes, M., Lopez, H., and Nicholas, P. (2010). Particle Learning and Smoothing. Statistical Science, 25(1):88–106.

[I] Tuesday/Martes, 3:15p.m., Room/aula 1, Session: L3–2, LACSC–Recent advances in Statistical Computing 1

[II] Yachay Tech, Escuela de Matemática y Tecnologías de Información, Ecuador, Urcuquí, sinfante64@gmail.com
[III] Universidad de Carabobo. Valencia; Venezuela, cvluna.m@gmail.com
[IV] Universidad de Carabobo. Valencia; Venezuela,arhernan65@gmail.com
[V] Universidad de Carabobo. Valencia; Venezuela, arhernan65@gmail.com

Waveform inversion of poststacked reflection seismic data using Artificial Neural Networks[I]

Communication / Ponencia

ITURRARAN VIVEROS, URSULA[II] Muñoz-García, Andrés M.[III]

México

We apply instantaneous seismic attributes to a stacked P-wave reflected seismic section in the Tenerife field located in the Middle Magdalena Valley Basin (MMVB) in Colombia to estimate effective volume of clay V_{clay} and density ρ at seismic scale. The well-logs and the seismic attributes associated to the seismic trace closer to one of the available wells is the information used to train some multi-layered Artificial Neural Networks (ANN). We perform data analysis via the Gamma test, a mathematically non-parametric nonlinear smooth modeling tool, to choose the best input combination of seismic attributes to train an artificial neural network (ANN) for estimating V_{clay} and ρ. Once the ANN is trained it is applied to predict these parameters along the seismic line. This is a significant result that shows for the first time a petrophysical characterization of this field at seismic scale. From the continuous estimations of V_{clay} we distinguish two facies: sands and shales, these estimations confirm the production of the Mugrosa C-Arenas zone and we draw brown clay that correlate with the high amplitude attributes and the yellow sand correlate with the low amplitude attributes. From the facies we build a model for the velocity V_P at seismic scale using polynomial fitting. These is our initial velocity model which is used together with the estimation for ρ in a forward model to produce synthetic seismograms. The inversion process consists of minimizing the error between the synthetic seismograms and the real acquired seismic traces obtaining a modified velocity model which in the eyes of interpreters seems physically feasible and enhance the contrast of the productive Mugrosa C-Arenas zone.

Keywords: Artificial Neural Networks, Seismic Attributes, Facies Inversion.

Mathematics Subject Classification (2010): 86A15,86A22, 68U20, 68T05

References

[1] Iturrarán-Viveros, U., 2012, Smooth regression to estimate effective porosity using seismic attributes: J. Appl. Geophys., **76**, 1–12.

[2] Iturrarán-Viveros, U., and Parra, J. O., 2014, Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data: J. Appl. Geophys., **107**, 45–54.

[3] Parra, J. O., Parra, J. S., & Iturrarán-Viveros, U., 2017, Waveform inversion of poststacked reflection seismic data with well log constrained using nonlinear optimization methods: The Leading Edge, **36**, no. 12, 306–309.

[I] Wednesday/Miércoles, 4:20p.m., Auditorium, Session: Mod5, Modeling / Modelacion 5

[II] Universidad Nacional Autonoma de Mexico, Mathematics, Mexico, Mexico, ursula@ciencias.unam.mx

[III] Instituto Tecnológico Metropolitano de Medellín, Colombia FDcorreo

Simulación Multiagente: Herramientas y Aplicaciones[I]

Communication / Ponencia

JIMÉNEZ LÓPEZ, DAVID[II]

Costa Rica

An Agent-Based Simulation (ABS) is a computational system where a given number of virtual autonomous entities within a specified environment. This technology is used as a tool to validate hypothesis in research areas as diverse as population genetics, behavioral science and fluid dynamics; when other models are not as appropriate. We present a proposed domain-specific interpreted programming language to design ABS models, developed under the philosophy of ease of use by novice programmers. Current state of the project can be found at: `https://github.com/djimenez81/onei`

Keywords: Computer Simulation, Agent-Based Models.

References

[1] Schelling, Thomas C. (1971). Dynamic Models of Segregation (PDF). Journal of Mathematical Sociology. 1 (2): 143–186.

[2] Axelrod, Robert (1997). The Complexity of Cooperation: Agent–Based Models of Competition and Collaboration. Princeton: Princeton University Press.

[3] Bonabeau, Eric (2002). Agent-based modeling: methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America. 99 (3): 7280–7287.

[4] Sun, Ron (2006). Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation. Cambridge University Press.

[5] Aditya Kurve; Khashayar Kotobi; George Kesidis (2013). An agent-based framework for performance modeling of an optimistic parallel discrete event simulator. Complex Adaptive Systems Modeling. 1: 12.

[I] Friday/Viernes, 1:30p.m., Room/aula 2, Session: DJ, Simulacion Multiagente: Herramientas y Aplicaciones

[II] Universidad de Costa Rica, Matemática, Costa Rica, San José, david.jimenezlopez@ucr.ac.cr

Dependence Analysis via Copulas - Nonstandard View[I]

Short course / Curso corto

KOLEV, NIKOLAI[II]

Brazil

Objectives:

The copula approach for dependence analysis of random variables will be introduced. In fact, a mix of recent probability, multivariate statistical modeling and computing techniques will be presented along with corresponding R codes.

We will outline not only the profit of copula methodology for data analysis (being advocated by Basel III and Solvency II accords), but copula pitfalls will be discussed as well. Based on our current investigations, a new copula alternative - the Sibuya's dependence function will be introduced.

Several non-standard measures of dependence will be presented and their impact on data analysis will be demonstrated. Applications in Insurance and Finance will be provided.

Content:

1. Copula theory: basic properties and main copula families;
2. Dependence concepts: concordance, perfect dependence;
3. Measures of monotone dependence (Spearman, Kendall, Gini, Blets);
4. Tail copula approach for modeling of extreme events;
5. Sibuya's dependence function - a copula alternative;
6. Statistical inference using copulas;
7. Applications using R.

Observation: Master level in Statistics is a must to understand the exposition.

Keywords: Copulas, Dependence measures, Extreme events, Inference, Sibuya's dependence function.

References

[1] Joe, H. (2015). *Dependence Modeling with Copulas*. CRC Press, Boca Raton.

[2] Mai, J-F. and Scherer, M. (2014). *Finacial Engineering with Copulas Explained*. Palgrave, MacMilan.

[3] Pinto, J. and Kolev, N. (2015). Sibuya-type bivariate lack of memory property. *Journal of Multivariate Analysis* 134, 119-128.

[I] Thursday/Jueves, 1:30p.m., Room/aula 1, Session: LACSC–3, LACSC–Copulas

[II] Institute of Mathematics and Statistics, University of Sao Paulo, Statistics, Brazil, Sao Paulo, kolev.ime@gmail.com

Solving the soft graph coloring problem using scatter search[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

The fundamental concepts for the scatter search algorithm were proposed by Fred Glover in the 1970s and throughout the years it has been constantly improve and has consolidated as a suitable metaheuristic method for solving optimization problems [5] because it obtains good quality solutions at reasonably short runtimes. Due to these qualities and to the fact than the algorithm has been previously tested to solve the robust graph coloring problem [3] (a special case of the soft graph coloring problem), this algorithm is also suitable for our particular problem.

In this paper a scatter search algorithm is proposed as a solution tool for the soft graph coloring problem. This is a metaheuristic consists of 5 methods, the flexibility of this methods is shown and the best strategies for each method are analyzed.

The soft graph coloring model is a special case of the graph coloring problem where given a complete graph with weights in the edges, finding a coloring than minimizes the sum of the edges with the same color in both ends is the fitness function. The soft graph coloring problem has demonstrated that can solve other graph coloring problems such as minimal, equitable, weak and robust graph coloring [2] and also has shown its application as an unsupervised classifier [1]. Being a generalization of the robust graph coloring problem is known that it is a NP-Hard problem, and it is necessary to use metaheuristics for instances higher than 20 vertexes.

Keywords: Scatter search, graph coloring, optimization.

References

[1] Flores-Cruz, J., Lara–Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., and Rincón-García, E.A. (2016). "Un Sistema Clasificador Utilizando Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones*. http://revistas.ucr.ac.cr/index.php/matematica/article/view/20838.

[2] Glover, F. (1998). A Template for Scatter Search and Path Relinking. In Artificial Evolution, edited by Jin–Kao Hao, Evelyne Lutton, Edmund Ronald, Marc Schoenauer, and Dominique Snyers, 1363:1351. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/BFb0026589.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx
[III] Universidad Autónoma Metropolitana, Ciudad de México, México, gamma@xanum.uam.mx
[IV] Universidad Autónoma Metropolitana, Ciudad de México, México, cobos@xanum.uam.mx
[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx
[VI] Universidad Autónoma Metropolitana, Ciudad de México, México, mgra@correo.azc.uam.mx

[3] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G. and Rincón-García, E.A. (2015). "Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones* 22 (2): 311. doi:10.15517/rmta.v22i2.20838.

[4] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., Ramírez-Rodríguez, J. and López-Bracho. R. (2012). "Un Algoritmo Evolutivo Para Resolver El Problema de Coloración Robusta." *Revista de Matemática: Teoría Y Aplicaciones* 12 (1-2): 111. doi:10.15517/rmta.v12i1-2.255.

[5] Martí, R., Laguna, M., and Glover. F. (2006). "Principles of Scatter Search." *European Journal of Operational Research* 169 (2): 359?72. doi:10.1016/j.ejor.2004.08.004.

Classical Coloring Problems Using the Soft Graph Coloring Model[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

There is a wide variety of coloring problems, which model real-world problems such as assigning radio frequencies, registrer allocation, coloring maps, timetable scheduling, pattern recognition, and so on. Graph coloring consist of painting the graph vertices taking certain restrictions into account such as not painting the same color those vertexes joined by an edge.

In this work we describe a unifying model which allows us to solve classic coloring instances such as: minimal coloring, weak graph coloring, equitable coloring and robust coloring using the Soft Graph Coloring problem; additionally, it is detailed the particularities each of the problems that allow mapping them into the Soft Graph Coloring Model.

Keywords: Scatter search graph coloring optimization.

References

[1] Flores-Cruz, J., Lara-Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., and Rincón-García, E.A. (2016). "Un Sistema Clasificador Utilizando Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones*. http://revistas.ucr.ac.cr/index.php/matematica/article/view/20838.

[2] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G. and Rincón-García, E.A. (2015). "Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones* 22 (2): 311. doi:10.15517/rmta.v22i2.20838.

[3] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., Ramírez-Rodríguez, J. and López-Bracho. R. (2012). "Un Algoritmo Evolutivo Para resolver el Problema de Coloración Robusta." *Revista de Matemática: Teoría Y Aplicaciones* 12 (1-2): 111. doi:10.15517/rmta.v12i1-2.255.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx
[III] Universidad Autónoma Metropolitana, Ciudad de México, México, gamma@xanum.uam.mx
[IV] Universidad Autónoma Metropolitana, Ciudad de México, México, cobos@xanum.uam.mx
[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx
[VI] Universidad Autónoma Metropolitana, Ciudad de México, México, mgra@correo.azc.uam.mx

A classifier system using soft graph coloring[I]

Communication / Ponencia

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI] Flores Cruz, Jorge [VII]

México

One of the applications of Artificial Intelligence is recognition of patterns, which can be understood as classifying large quantities of physical or abstract objects with the purpose of extracting useful information, which allows to establish properties between clusters of objects.

The challenge of a pattern recognition model is to teach a machine, in this case a computer, to classify efficiently a set of items with few parameters. In a classifier system, the label can be determined previously, then it has to decide for a given object, in which particular class it is more appropriate to place it given the existing options. For example, a server must decide whether incoming mail is "spam" or goes to the inbox. There are times when classes are not defined previously and it is necessary to find a way to group the elements in some optimal way.

Unsupervised classifiers allow clustering methods with next to none or no human intervention. In this work an unsupervised classifier system using the model of soft graph coloring is described and tested with some classic instances from the literature and the obtained results were compared with classifications made with suppervised classifiers, yielding our method as good or better results, sometimes providing alternative classifications that brings additional information that humans did not consider.

Keywords: Scatter search, graph coloring optimization.

References

[1] Flores-Cruz, J., Lara-Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., and Rincón-García, E.A. (2016). "Un Sistema Clasificador Utilizando Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones*. http://revistas.ucr.ac.cr/index.php/matematica/article/view/20838.

[2] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G. and Rincón-García, E.A. (2015). "Coloración de Gráficas Suaves." *Revista de Matemática: Teoría Y Aplicaciones* 22 (2): 311. doi:10.15517/rmta.v22i2.20838.

[3] Lara-Velázquez, P., Gutiérrez-Andrade, M.A., Ramírez-Rodríguez, J. and López-Bracho. R. (2012). "Un Algoritmo Evolutivo Para resolver el Problema de Coloración Robusta." *Revista de Matemática: Teoría Y Aplicaciones* 12 (1-2): 111. doi:10.15517/rmta.v12i1-2.255.

[I] Thursday/Jueves, 1:50p.m., Room/aula 3, Session: Opt2, Optimization / Optimizacion 2

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx

[III] Universidad Autónoma Metropolitana, Ciudad de México, México, gamma@xanum.uam.mx

[IV] Universidad Autónoma Metropolitana, Ciudad de México, México, cobos@xanum.uam.mx

[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx

[VI] Universidad Autónoma Metropolitana, Ciudad de México, México, mgra@correo.azc.uam.mx

[VII] Posgrado en Ciencias y Tecnologías de la Información, Universidad Autónoma Metropolitana, same address as Lara-Velázquez

A greedy algorithm for redistricting[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

Typically, there does not exist a unique optimal solution for redistricting problems, a solution can be a small set of good alternatives, or a group of alternatives into different preference sets, and it is necessary to use an appropriate technique to propose only one solution. "In engineering and management practice there is a need to select a final solution to be implemented" [4].

Redistricting is the redrawing of the boundaries of legislative districts for electoral purposes in such a way that the generated districts full fill federal and state requirements such as contiguity, population equality and compactness. Redistricting is a multi-objective problem which has been proved to be NP-hard. In Mexico, the redistricting process has been done using an aggregation function, considering a weighted sum of the objectives. However, if different weighting factors are used then a set of diverse, high quality solutions can be generated and a new problem arises: which solution should be implemented?

In this paper we propose a novel alternative, called FuGA, to select the best solution for the redistricting problem using a fuzzyfication of the objective function. The proposed algorithm was applied in a real case, and its solutions were compared with those produced by VIKOR, a well-known algorithm for decision making. FuGA showed a better performance since it was able to avoid the selection of dominated solutions.

Keywords: Fuzzy Ranking, Redistricting Problem, Multiobjective Programming.

References

[1] Rincón-García, E.A., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., Lara-Velázquez, P. (2012), "Nuevas medidas de compacidad para el diseño de zonas", *Revista de Matemática: Teoría y Aplicaciones*, Vol. 19 No. 2, pp.183-199.

[2] Rincón-García, E.A., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., Lara-Velázquez, P., Mora-Gutiérrez, R.A. and Ponsich, A.S. (2012), "A discrete particle swarm optimization algorithm for designing electoral zones", *Book Series World Scientific*, Vol. 6, pp. 174-187.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx
[III] gamma@xanum.uam.mx
[IV] cobos@xanum.uam.mx
[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx
[VI] mgra@correo.azc.uam.mx

[3] Rincón-García, E.A., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., Lara-Velázquez, P., Ponsich, A.S. and Mora-Gutiérrez, R.A. (2013), "A multiobjective algorithm for redistricting", *Journal of Applied Research and Technology*, Vol. 11 No. 3, pp. 324-330.

[4] Opricovic, S. and Tzeng, G.H. (2007), "Extended VIKOR Method in Comparison with Outranking Methods", *European Journal of Operational Research*, Vol. 178 No. 2, pp. 514-529.

[5] Razmak, J. and Aouni, B. (2015), "Decision Support System and Multi-Criteria Decision Aid: A State of the Art and Perspectives", *Journal of Multi-Criteria Decision Analysis*, Vol. 22 No. 101-117.

Solving the portfolio optimization problem using swarms[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

The constrained portfolio optimization problem with multi-objective functions cannot be efficiently solved using exact techniques. Thus, heuristics approaches seem to be the best option to find high quality solutions in a limited amount of time. Selecting investment portfolios nowadays represents a very common practice and a difficult task, covering a wide range of applications. Because of its importance, the solution of this problem has attracted considerable attention in different research areas such as economics, finance, computer science, operation research and mathematics. The construction of an investment portfolio seeks to balance gains and losses, distributing and offsetting the risk and return of different assets.

However, the problem becomes much more difficult when real-world requirements are imposed, such as cardinality, which forces the construction of a portfolio having a specific number of assets, or constraints on the ratios of capital to invest in a selected asset. When such constraints are included, a nonlinear mixed integer programming model must be solved and classical exact solution techniques become inadequate. This higher complexity has encouraged the development and application of more robust heuristic strategies, such as genetic algorithms tabu search, simulated annealing, neural networks, evolutionary algorithms and multi-objective techniques.

For solving this problem, we used a variant of Particle Swarm Optimization named PSO-3P [1], [2]. PSO-3P incorporates two strategies of diversification and intensification to improve the performance of the classical PSO algorithm. The proposed algorithm was tested over five well-known benchmark data sets and the obtained results prove to be highly competitive since they outperform those reported in the specialized literature in almost all tackled instances.

Keywords: PSO-3P, Portfolio Optimization, Markowitz Model.

References

[1] De-Los-Cobos-Silva S. G., Gutiérrez-Andrade M. A., Mora-Gutiérrez R. A., Lara-Velázquez P., Rincón-García E. A., Ponsich A. (2015). “An Efficient Algorithm For Unconstrained Optimization”. Mathematical Problems In Engineering. Available Inhttp://Dx.Doi.Org/10.1155/2015/178545.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx
[III] gamma@xanum.uam.mx
[IV] cobos@xanum.uam.mx
[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx
[VI] mgra@correo.azc.uam.mx

[2] De-Los-Cobos-Silva S. G., Mora-Gutiérrez R. A., Gutiérrez-Andrade M. A., Rincón-García E. A., Ponsich A.,Lara-Velázquez P. (2016). "Development Of Seven Hybrid Methods Based On Collective Intelligence For Solving Nonlinear Constrained Optimization Problems". Artificial Intelligence Review. Available In Http://Dx.Doi.Org/10.1007/S10462-016-9524-4.

[3] Markowitz H. (1952)."Portfolio Selection". The Journal Of Finance. 7(1), 77-91.

[4] Ponsich A., López Jaimes A., Coello Coello C.A. (2013)."A Survey On Multiobjective Evolutionary Algorithms For The Solution Of The Portfolio Optimization Problem And Other Finance And Economics Applications". IEEE Transactions On Evolutionary Computation. 17(3), 321-344.

Development of a hybrid method based on k–means, consensus and ant system for classifying 60 Mexican Universities[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

In this work, a hybrid method , based on ideas of the k-mean [2], the consensus [3] and ant system [4] is presented, called ASKC. The ASKC is used to classify 60 Mexican universities. This method is self-adaptative since it is able to regulate the number of clusters employed. For this reason, there is no need that user determine the number of cluster.

The information used in work was obtained of [1]. Initially, we determined if there not likeness between the $i-th$ and the $j-th$ universities, based on numerical results of non parametric statistical test. After, the ASKC was run 20 times on the set of universities. Numerical results shown that the set of universities could be divided in 6 or 13 subset. Each subset is integrated by element strongly similar.

Keywords: k-means, ant colony optimization, classification problem.

References

[1] Ordorika, Imanol and Rodríguez-Gómez, Roberto and Lozano, Francisco Javier. (2016). "Estudio Comparativo de las Universidades Mexicanas" - Explorador de datos (ExECUM). http://www.execum.unam.mx. Web; accedido el 08-06-2017

[2] MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281?297.

[3] Ruggiero MA, Gordon DP, Orrell TM, et al. "A Higher Level Classification of All Living Organisms". Thuesen EV, ed. PLoS ONE. 2015;10(4):e0119248. doi:10.1371/journal.pone.0119248.

[4] Dorigo, M., Birattari, M., & Stutzle, T. (2006). "Ant colony optimization". IEEE computational intelligence magazine, 1(4), 28-39.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx

[III] gamma@xanum.uam.mx

[IV] cobos@xanum.uam.mx

[V] Departamento de Sistemas, Universidad Autónoma Metropolitana, Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx

[VI] mgra@correo.azc.uam.mx

Development of a hybrid method based on Method of the Musical Composition, Ant System and dual simplex algorithm for solving the vehicle routing problem with windows of time (VRP-TW)[I]

Poster / Cartel

LARA VELÁZQUEZ, PEDRO[II] Gutiérrez Andrade, Miguel Angel [III]
De los Cobos Silva, Sergio Gerardo [IV] Rincón García, Eric Alfredo [V]
Mora Gutiérrez Roman Anselmo [VI]

México

We present a metaheuristic hybrid method, called *DSM-MMC-AS*, which is used to solve the vehicle routing problem with windows of time *(VRP-TW)* [2]. This method is based on *Method of the musical composition (MMC)* [3, 4] and *ant system (AS)* [1] and the dual simplex algorithm (DSM) [5] . Note that in this hybrid two metaheuristc method and one exact algorithm of linear programming are combining. Initially, the *DSM* is used to solve two relaxations of the original model, this activity is made in the *Gurobi* program. After, solutions generated by *Gurobi* are input by *AS* for producing a set initial solution for the *MMC*. Subsequently, in each iteration of the algorithm the information of solutions found by *MMC* is use update pheromone matrix into *AS*.

The numerical results, show that *DSM-MMC-AS* obtain the best results shown in the literature, using less computational resources, since the *DSM-MMC-AS* involves a smaller number of evaluation of the objetive function.

Keywords: Hybridization, Combination of exact with heuritic methods, optimization.

References

[1] Dorigo, M., Maniezzo, V., & Colorni, A. (1996).“ Ant system: optimization by a colony of cooperating agents”. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 26(1), 29-41.

[2] El-Sherbeny, N. A. (2010).“ Vehicle routing with time windows: An overview of exact, heuristic and metaheuristic methods”. Journal of King Saud University-Science, 22(3), 123-131.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Autónoma Metropolitana, Departamento de Ingeniería Eléctrica, México, Ciudad de México, plara@xanum.uam.mx
[III] gamma@xanum.uam.mx
[IV] cobos@xanum.uam.mx
[V] Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco, Ciudad de México, México, rigaeral@correo.azc.uam.mx
[VI] mgra@correo.azc.uam.mx

[3] Mora-Gutiérrez, R., Ramírez-Rodríguez, J.,Rincón-García, E. (2012). "An optimization algorithm inspired by musical composition". Artificial Intelligence Review. Springer Netherlands. Url: http://dx.doi.org/10.1007/s10462–011–9309–8. Doi: 10.1007/s10462–011–9309–8

[4] Mora-Gutiérrez, R. A., Ramírez-Rodríguez, J., Rincón-García, E. A., Ponsich, A., Herrera, O., & Lara-Velázquez, P. (2013). "Un algoritmo de optimización inspirado en composición musical para el problema de optimización con restricciones". Revista de Matemática: Teoría y Aplicaciones, 20(2), 183-202.

[5] Wagner, H. M. (1958). "The dual simplex algorithm for bounded variables". Naval Research Logistics (NRL), 5(3), 257-261.

Cubrimiento e invariabilidad: una teoría sistémica en ecosistemas[I]

Communication / Ponencia

LLORET CLIMENT, MIGUEL[II] Nescolarde–Selva, J. A.[III] González–Franco, L.[IV]

España

A major problem related to the treatment of ecosystems is that they have no available mathematical formalization. This implies that many of their properties are not presented as short, rigorous modalities, but rather as long expressions which, from a biological standpoint, totally capture the significance of the property, but which have the disadvantage of not being sufficiently manageable, from a mathematical standpoint. The interpretation of ecosystems through networks allows us to employ the concepts of coverage and invariance alongside other related concepts. The latter will allow us to present the two most important relations in an ecosystem – predator–prey and competition – in a different way. In this article, we apply the developed concepts of structural functions, coverage, invariant sets, etc. (Lloret et al., 1998; Esteve and Lloret, 2006,a,b, 2007) to ecosystems.

Keywords: Ecosystem, structural function, coverage, invariability.

Resumen

Un problema importante en el tratamiento de ecosistemas es no disponer de una formalización matemática de estos. Esto implica que muchas de sus propiedades no se presentan bajo apariencias cortas y rigurosas sino sobre desarrollos largos, que aunque desde un punto de vista biológico, capturan totalmente el significado de la propiedad, presentan el inconveniente de no ser suficientemente manejables. La interpretación de ecosistemas mediante redes nos permitirá usar los conceptos de cubrimiento e invarianza junto con otros conceptos afines. Estos conceptos nos permitirán presentar de forma diferente las dos relaciones más importantes en un ecosistema, las de depredador-presa y la de competición.

En este articulo aplicamos los conceptos desarrollados de funciones estructurales, cubrimientos, conjuntos invariantes, etc.;(Lloret et al., 1998; Esteve and Lloret, 2005) a ecosistemas.

Palabras clave: Ecosistemas, función estructural, cubrimiento, invariabilidad.

References

[1] Esteve, P.F., and Lloret, M.(2005) Coverage and invariability by structural functions. International Journal of General Systems (submitted to referee).

[2] Lloret, M., Villacampa, Y. and Usó, J.L. (1998). System-linkage: structural functions and hierarchies. Cybernetics and Systems: An International Journal 29, 29–39

[I] Friday/Viernes, 10:50a.m., Auditorium, Session: Bio2, Biomathematics / Biomatematicas 2

[II] Universidad de Alicante, Matemática Aplicada, España, Alicante, miguel.lloret@ua.es

[III] Universidad de Alicante, Matemática Aplicada, España, Alicante, josue.selva@ua.es

[IV] Universidad de Alicante, Matemática Aplicada, España, Alicante, luciadhara@hotmail.com

Zeroes of functions of Fresnel complementary integral type[I]

Communication / Ponencia

LOBO SEGURA, JAIME[II] Villalobos Arias, Mario Alberto[III][IV]

Costa Rica

A study of the real zeros of some families of real variable functions dependent on a real parameter α belonging to a line interval is presented below. In general terms, these families are defined by the Cartesian coordinates of certain parametric flat curves.

This study extends the one carried out on the same problem for the generalized Fresnel functions, whose results appear in [1]. In this paper, we introduced the generalized integral Fresnel complement (FCI) functions, denoted by c_a, s, where $\alpha > 0$. We sought here to reframe and extend this study thanks to the use of a theory of the fixed point for parametric functions (PF theory) that will appear in a special section. The main achievement of Lobo-Villalobos in [1] is the Theorem about the real zeros of the functions c_α, s_α,

Keywords: Zeroes, special functions, Fresnel integral.

Resumen

Se presentaun estudio de los ceros reales de algunas familias de funciones de variable real dependientes de un parámetro real α perteneciente a un intervalo de la recta. En términos generales dichas familias están definidas por las coordenadas cartesianas de ciertas curvas planas paramétricas.

Este estudio extiende el realizado sobre el mismo problema para las funciones de Fresnel generalizadas, cuyos resultados aparecen en [1]. En este trabajo introdujimos las funciones complemento integral de Fresnel generalizadas (FCI), denotas por c_α, s_α, donde $\alpha > 0$. Buscamos aquí replantear y extender este estudio gracias al recurso de una teoría del punto fijo para funciones paramétricas (teoría PF) que aparecerá en una sección especial. El principal logro de Lobo-Villalobos en [1] es el teorema acerca de los ceros reales de las funciones c_α, s_α,

Palabras clave: ceros, funciones especiales, integral de Fresnel.

References

[1] Lobo, J.; Villalobos, M.A. (2016) "Ceros de funciones integrales complementarias de Fresnel generalizadas", *Revista de Matemótica: Teoróa y Aplicaciones*, 23(**2**), 321-338

[2] Koumandos, S.; Lamprecht, M. (2012) "The zeros of certain Lommel functions", *Proceedings of The American Mathematical Society*, **140** (9), pp. 3091–3100.

[3] Kreyszig, E. (1957) "On the complementary functions of the Fresnel integrals", *Canadian Journal of Mathematics*, **9**, pp. 500–510.

[4] Slavik, A. (2013) "Intrinsecally defined curves and special functions", *Mathematics Magazine*, **86** (3), pp. 211–218.

[5] Livingstone, A. E. (1954) "The Zeros of a Certain Class of Indefinite Integrals", *Proceedings of the American Mathematical Society*, **5**(2),pp. 296-300

[I] Tuesday/Martes, 5:00p.m., Room/aula 2, Session: Num4, Numerical Analysis / Análisis Numérico 3

[II] Universidad de Costa Rica, CIMPA, San José, Costa Rica, jaimelobosegura@hotmail.com

[III] same address, mario.villalobos@ucr.ac.cr

[IV] Instituto tecnologico de Costa Rica, Escuela de Matemática, Cartago, Costa Rica, marvillalobos@itcr.ac.cr

Grafo n-residual Módulo m y su aplicación en la estructuración de Residuos n-ádicos[I]

Communication / Ponencia

MARTÍNEZ, IVETH V.[II] Solís, René[III]

Panamá

The reciprocity laws allowed to characterize the residues in prime modules, where some results were established for residues of quadratic, cubic powers, among others. We define the sets n -residuales module m in order to characterize the resulting residues given the power and the module, to then characterize the conditions the residues. We characterize the sets by restricting the modules that support primitive roots, and then determine the behavior of the residues based on the relationship that exists between the exponent and the modules with said restriction. Once defined, using graphs, we can observe the behavior of the residuals as we assign values, where we observe special cases where cycles are generated that compose related graphs.

The reciprocity laws allowed to characterize the residues in prime modules, where some results were established for residues of quadratic, cubic powers, among others. We define the sets n -residuales module m in order to characterize the resulting residues given the power and the module, to then characterize the conditions the residues.

We characterize the sets by restricting the modules that support primitive roots, and then determine the behavior of the residues based on the relationship that exists between the exponent and the modules with said restriction. Once defined, using graphs, we can observe the behavior of the residuals as we assign values, where we observe special cases where cycles are generated that compose related graphs.

Keywords: Reciprocity laws, n-residuales module, prime modules, primitive roots.

Resumen

Las leyes de reciprocidad permitieron caracterizar los residuos en módulos primos, donde se pudo establecer algunos resultados para residuos de potencias cuadráticas, cúbicas, entre otras. Definimos los conjuntos n-residuales módulo m con el propósito de caracterizar los residuos resultantes dado la potencia y el módulo, para luego caracterizar las condiciones los residuos.

Caracterizamos los conjuntos restringiendo los módulos que admitan raíces primitivas, para luego determinar el comportamiento de los residuos en base a la relación que exista entre el exponente y los módulos con dicha restricción. Una vez definido, mediante grafos, es posible realizar diversas pruebas conforme asignamos valores, con algunas rutinas construidas en $MATHEMATICA$, observando que los residuos que en casos especiales generan ciclos que componen grafos conexos.

Palabras clave: Leyes de reciprocidad, módulo n-residuales, módulos principales, raíces primitivas.

[I] Tuesday/Martes, 4:00p.m., Auditorium, Session: Apl3, Applications / Aplicaciones 3

[II] Universidad de Panamá, Matemática, Panamá, Panamá, iveth.martinez@up.ac.pa

[III] Universidad Tecnológica de Panamá, ciudad de Panamá, rene.solis@utp.ac.pa

References

[1] D.M. Burton. Elementary Number Theory. Allyn and Bacon Inc., 1980.

[2] F. Lemmermeyer. Introduction to number theory. November 2000.

[3] F. Lemmermeyer. Reciprocity Laws. From Euler to Eisenstein. Springer-Verlag, 2000.

[4] Primitive roots and exponential iterations. http://www.mathpages.com/home/kmath148/kmath148.htm.

[5] F. Toranzos. Introducion a la Teoria de Grafos. OEA, 1976.

[6] D.R. Wilkins. http://www.maths.tecd.ie/dwilkins/Courses/311/311NumTh.pdf,1997.

Simulated annealing for maximization of efficiency on dairy cattle attribute monitoring process.[I]

Communication / Ponencia

MORA VALVERDE, DAVID[II] Villalobos Arias, Mario Alberto[IV]

Costa Rica

Having a viable and reliable design mechanism for dairy cattle diets, but not being able to deploy it to practical reality, represents a dilemma for the animal nutrition professional. Much of the design effort that seeks to fill the physiological and nutritional requirement in the cow's diet loses precision as soon as the requirements of a group of animals are averaged to meet the practical conditions of canoe supply. It aroused the interest to evaluate the operative imprecision of the cow-feeding milk process. Part of the causes of the variability lie in the low or null intensity of monitoring of productive variables that allow to update rations of the animals according to their nutritional requirements, where there is no clear structure to define the most convenient frequency of updating information on farm attributes and animals as it results in a process of high economic cost and intellectual resources. So, performance and viability in the use of metaheuristic algorithms, specifically the simulated annealing, were evaluated on the processes of evaluation, design and supply of dairy feeding practices in different search spaces within them, evaluating their methodological performance as problem solving engines. An algorithm was designed which was subjected to three different levels of precision nutrition intensity (low, medium and simulated annealing). Measurements of animal and food attributes at frequencies that are highly distanced in time, as is customary in the dairy sector, can respond to a more traditional practice that contributes little robustness to the process.

Likewise, the results allow us to affirm that it is practically impossible to predict compositions without measuring with higher frequencies than those traditionally used. The application of the simulated annealing algorithm allows to indicate two profiles of its use: mainly as an evidence tool and secondly as a planning tool, with which the traditional monitoring scheme is contrasted, versus another based on the intrinsic variability itself combined with combinatorial optimization, where it is possible to differentiate, through a truncated heuristic mechanism, that the frequencies of traditional measurements highly distanced in time do not adjust to the prevailing variability. Likewise, the tool allows to outline the minimum frequencies necessary to adjust to the global overfeeding according to the farm's own variability. The proposed algorithm has the potential to define minimum sampling ranges to approach the actual inefficiency value. In descriptive terms, it is necessary to maximize the adjustment of the frequencies to the global calculation of imprecision of the feeding process for the total data set, with which the difference between the real imprecision value and the one calculated by the frequencies obtained from the configuration of each run, described by the following formula:

Min[Differential of megacalories between real waste and waste of the run]

Where the real waste is defined by the following equation:

[I] Thursday/Jueves, 2:10p.m., Room/aula 3, Session: Opt2, Optimization / Optimizacion 2

[II] Universidad de Costa Rica, Estacion de ganado Lechero, San José, Costa Rica, david.mora@ucr.ac.cr

[III] Universidad de Costa Rica, CIMPA y Escuela de Matemática, San José, Costa Rica, mario.villalobos@ucr.ac.cr

[IV] Instituto tecnologico de Costa Rica, Escuela de Matemática, Cartago, Costa Rica, marvillalobos@itcr.ac.cr

Megacalories of overfeeding $= \sum_{i=1}^{n} |R_i - Si| -$ Fixed requirement

Where: R_i is the nutrient requirement of an animal on a specific day
S_i is the real supply to the animal on a specific day
Thus:

$$f(I) = |SUMA_MCAL(I) - SUMA_MCAL0|$$

Keywords: Simulated annealing, dairy cattle, dairy nutrition.

References

[1] D´Agaro, E. 2012.Use of Simulated Annealing Algorithms for Optimizing Selection Schemes in Farm Animal Populations.Simulated Annealing Advances, Applications and Hybridizations.Chapter 9. Italy

[2] Casado, S.; Pacheco, J.. 2003. Estudio comparativo de diferentes estrategias metaheurísticas para la resolución del labor schedulingproblem Estudios de Economía Aplicada, vol. 21, núm. 3, diciembre, 2003, pp. 537–557, Asociación Internacional de Economía Aplicada.

[3] Fox, D. G., Tedeschi, L.O. y Guiroy, P. J. 20011. Determining feed intake and feed efficiency of individual cattle fed in groups. Animal Science Dept. Cornell University

[4] Hills JL, Wales WJ, Dunshea FR, Garcia SC, Roche JR (2015) Invited review: An evaluation of the likely effectsof individualized feeding of concentrate supplements to pasture–based dairy cows. Journal of Dairy Science 98,1363–1401.

[5] Kirckpatrick, S.; Gelatt, J. Vecchi, P. 1983. Optimization by simulated annealing.Revista Science. Volumen 220, número 4598.

Optimización de diseños sobresaturados con dos niveles a través de técnicas metaheurísticas[I]

Communication / Ponencia

MORALES, LUIS B[II]

México

Supersaturated designs are factorial designs in which the number of experimental runs is less than or equal to the number of factors. Such designs are used in various fields including computer experiments, software testing, medical, industrial and engineering experiments, as well as in biometric applications. Several computational methods have been used to construct $E(s^2)$-optimal supersaturated designs. In particular, metaheuristic methods have also been used, such as genetic algorithms, simulated annealing, tabu search and swarm techniques particles In this paper we present optimization algorithms based on the tabu search to construct $E(s^2)$-optimal supersaturated designs.

Keywords: $E(s^2)$-optimality, Supersaturated designs, Tabu search algorithm.

References

[1] K.H.V.Booth & D.R.Cox. Some systematic supersaturated designs. Technometrics 4 (1962) 489-495.

[2] D.A. Bulutoglu & C.S. Cheng, Construction of $E(s^2)$-optimal supersaturated designs. Ann. Statist. 32 (2004) 1662–1678.

[3] S.Gupta & L.B. Morales, Constructing $E(s^2)$-optimal and minimax-optimal k-circulant supersaturated designs via multi-objective tabu search. J. Statist. Plann. Inference 142 (2012) 1415–1420.

[I] Wednesday/Miércoles, 2:35p.m., Room/aula 2, Session: Opt1, Optimization / Optimizacion 1

[II] IIMAS Universidad Nacional Autónoma de México, México, México, lbm@unam.mx

Recent Developments in Analysis of Three-Way Data in R[I]

Short course / Curso corto

NIETO LIBRERO, ANA BELÉN [III] Vicente Galindo, M. Purificación[IV]

España

This course is a continuation of the course named Recent Developments in Analysis of Three-Way Data. The libraries available in R of the methods seen above will be shown.

Starting with the analysis of two-way tables, the biplotbootGUI package will be shown with which the Biplot methods can be executed through a graphical user interface and which also provides measures of stability of the results through the Bootstrap methods. Among the libraries related to the STATIS methods and their different versions are ade4, in which the STATIS, STATIS dual, PTA, STATIC and COSTATIS methods can be executed. In the DistatisR package, the researcher will be able to execute the COVSTATIS and DISTATIS methods. STATIS, dual STATIS, PTA, K + 1 STATIS, CANOSTATIS and ANISOSTATIS can be found in the MExPosition package.

Recent developments in the MFA line are the HMFA (Hierarchical Multiple Factorial Analysis), proposed as an extension of the MFA in which the variables are structured according to a hierarchy; when a set of individuals is described by a set of variables that

can be continuous and/or categorical, the proposed analysis is a particular case of the MFA called Factorial analysis of mixed data; if what is intended to be analyzed are multiple contingency tables, the proposed method is called MFACT. These methods can be used through the FactomineR package among others.

Keywords: biplot, three-way data, STATIS, MFA, multibiplot.

Mathematics Subject Classification (2010): 62H25, 62H86

References

[1] Gabriel, K.R. (1971). "The biplot graphic display of matrices with application to principal component analysis", *Biometrika n^o 58(3)*, p. 453–467.

[2] Galindo-Villardón, P. (1986). "Una alternativa de representacion simultanea: HJ-Biplot", *Qüestiió n^o 10(1)*, p. 13–23.

[3] Lê, S. and Josse, J. and Husson, F. (2008). "FactomineR: an R package for multivariate analysis", *Journal of Statistical Software n^o 25(1)*, p. 1–18.

[4] Thioulouse, J. and Dray, S. (2007). "Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages", *Journal of Statistical Software n^o 22(5)*, p. 1–14.

[I] Wednesday/Miércoles, 9:00a.m., Auditorium, Session: Tut, Tutorial 1

[II] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, ananieto@usal.es

[III] EE.UU. de Relaciones Laborales, Zamora, España, ananieto@usal.es

[IV] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, purivg@usal.es

A review of unsupervised methods in data mining for pattern recognition[I]

Communication / Ponencia

NIETO LIBRERO, ANA B.[II] González García, Nerea[III]
Galindo Villardón, M. Purificación[IV]

España

The latest advances in technology, the universalization of Internet, data mining and the reduction of storage and costs of large volumes are causing a paradigm shift in society. It defines the knows as 'Data Revolution'. Currently there are wide and complex databases increasingly frequent in disciplines such as genetics or finance, where the collected databases are formed by miles of variables. Its analysis is not simple or direct and therefore, the univariate descriptive statistics is far from providing optimal results. It is necessary to use multivariate statistical techniques that use all available information. Therefore, one of the current research lines focuses on the development of mathematical models for the study of high-dimensional data.

In machine learning and data mining it is necessary the use of automatic pattern recognition techniques to detect a structure of groups underlying the data. The main purpose of these classification techniques is to assign an object to an underlying group within the original multidimensional data. These clusters exist due to existing relationships and they are useful to understand some behavior patterns in multiple disciplines. These methods are grouped into supervised methods and unsupervised methods.

The supervised classification techniques are characterized because the information of the groups is known a priori. Among them are decision trees such as CHAID (Chi-square Automatic Interaction Detection), TAID (Tao Automatic Interaction Detection), CART or ID3, used in artificial intelligence.

In the case of unsupervised methods, the information corresponding to the groups is unknown. The unsupervised algorithms include association rules and clustering methods. The most commonly used clustering method is the K-MEANS. However, this algorithm has several disadvantages: the number of clusters is unknown, it is sensitive to outliers and initial points too. Moreover, K-MEANS usually generates clusters with relatively uniform size and spherical shape. Due to K-Means is the king of clustering algorithms it has got a lot of variants because of its disadvantages. Some of them are better for large data set analysis and another ones give better results for finding cluster with arbitrary shapes. Some alternatives are K-MEDOIDS which is less sensitive to outliers, CLARA and CLARANS. These methods can be applied to high-dimensional data in the same way as QT clustering. Another weakness of the use of K-MEANS is the shape of the clusters, but this can be solved using DBSCAN.

[I] Thursday/Jueves, 1:50p.m., Auditorium, Session: DMin, Data Mining / Minera de Datos

[II] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, ananieto@usal.es
[III] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, nerea_gonzalez_garcia@usal.es
[IV] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, pgalindo@usal.es

Other novel algorithms of the K-means family are K-means++ and K-Means-Quantile (which improves the K-means seeds), competitive K-means (that can be applied to large data sets), Fuzzy C-Means (the soft version of K-Means), Kernel K-Means (for clusters with shapes that are not hyper-eliptical) or the Spherical version (based on angular distance instead of euclidean distance).

In this work we describe the application of some of these alternative methods to the analysis of leukemia datasets. Later the results obtained are compared by different measures as Silhouette coefficient, purity or entropy.

Keywords: data mining, high-dimensional data, K-means.

References

[1] Kaufman, L. and Rousseeuw, P.J (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley.

[2] Ng, R. T. and Han, J. (1994). "Efficient and Effective Clustering Methods for Spatial Data Mining", *In Proceedings of VLDB*, p. 144–155.

[3] Ester, M., Kriegel, H.P., Sander, J. and Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise", *In Proceedings of Second International Conference on Knowledge Discovery and Data Mining*, p. 226–231.

[4] McQueen, J.B.(1967). "Some methods for classification and analysis of multivariate observations", *In Procedings of 5-th Berkeley Symposium on mathematical statistics and probablity, 1*, p. 281–297.

[5] Esteves, R.M., Thomas, H. and Chunming, R.(2014). "A new approach for accurate distributed cluster analysis for Big Data: competitive K-Means", *International Journal of Big Data intelligence, 5, n^o 1 (1-2)*, p. 50–64.

Estimación de tasas de transmisión de fiebre dengue con un modelo epidemiológico[I]

Communication / Ponencia

MURILLO LOPEZ, WILFREDO[II]

Honduras

Dengue fever is an infectious disease spread by infected female Aedes Aegypti mosquitoes. The virus consists of four serotypes which confers lifelong immunity to the infecting serotype and partial cross-immunity against other serotypes. In this work, we study a mathematical epidemiological model of non-linear ordinary di?erential equations for vector-borne infectious diseases, The model is built assuming total cross-immunity in the human population , that is, epidemiological classes of secondary infections are not considered. The main objective of the study is to estimate the transmission rates from mosquitoes to humans . The statistical estimation is performed using ordinary and generalized least squares techniques , and compare these results with Bayesian estimation. Residuals plots are used as diagnostic tools of the underlying assumptions of the estimation techniques, Finally, based on these estimations, numerical simulations are performed for the proposed model and the basic reproductive number R0 will be calculated.

Keywords: Fiebre Dengue, temporary cross-immunity, Basic reproduction number.

References

[1] Brauer, F., Castillo-Chavez, C., (2010). Mathematical Models in Population Biology and Epidemiology - Second Edition. Springer.

[2] Aguiar, M., Kooi, B., Stollenwerk, N., (2008). Epidemiology of Dengue Fever: A Model with Temporary Cross-Immunity and Possible Secondary Infection Shows Bifurcations and Chaotic Behaviour in Wide Parameter Regions. Mathematical Modelling of Natural Phenomena, Vol 3, No 4, pp 48-70.

[3] van den Driessche , P., Watmough ,J.(2001).Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission

[4] Nuraini,N.,Soewono E.,Sidarto K A.(2007)Mathematical Model of Dengue Disease Transmission with Severe DHF Compartment

[I] Wednesday/Miércoles, 2:55p.m., Auditorium, Session: Mod2, Modeling / Modelacion 2

[II] Universidad Nacional Autonoma de Honduras(UNAH), Estadistica, Honduras, Tegucigalpa, wilfredo.murillo@unah.edu.hn

An Image Quality Index Based on Coefficients of Spatial Association with an Application to Image Fusion[I]

Poster / Cartel

OJEDA, SILVIA MARÍA[II] Britos, Grisel[III] Vallejos, Ronny[IV]

Argentina

In the last decade, image quality indices have received considerable attention to quantify the dissimilarity between two images. The codispersion coefficient, commonly used in spatial statistics to address the association between two processes has also been used for this aim. Here we introduce an image quality index (CQmax) that is based on codispersion. This new coefficient is a directional evaluation of the spatial association, and consists on computing the maximum codispersion for a finite set of spatial lags on the plane, which also allows to obtain the direction associated with the maximum codispersion. From the CQmax index, a pseudo-metric that can be used as a cost functional for related optimization problems is defined. We carry out Monte Carlo simulations to explore the performance of the proposed index and its capability to detect directional contaminations. Additionally, we introduce a novel algorithm to restore directionally contaminated images and present an application with real data in the context of image fusion.

Keywords: Image similarity, SSIM index, CQmax index, Spatial lag, Image fusion.

References

[1] Brunet, D., Vrscay, E.R.,Wanng, Z., 2012. On the mathematical properties of the structural similarity index. IEEE Trans. Image Process 21, 1488–1498.

[2] Ojeda S. M., Vallejos R. O., Lamberti P. W., 2012. Measure of similarity between images based on the codispersion coefficient. Journal of Electronic Imaging 21, 023019.

[3] Pistonesi, S., Martinez, J., Ojeda, S., Vallejos, R., 2015. A novel quality image fusion assessment based on maximum codispersion. Lecture Notes in Computer Science 9423, 383–390.

[4] Vallejos, R., Mancilla, D., Acosta, J., 2016. Image similarity assessment based on coefficients of spatial association. Journal of Mathematical Imaging and Vision 56, 77–98.

[5] Wang Z., Bovik A., 2002. A universal image quality index. IEEE Signal Processing Letters 9, 81–84.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] FaMAF - Universidad Nacional de Córdoba, Grupo de Probabilidad y Estadística, Argentina, Córdoba, ojeda@famaf.unc.edu.ar

[III] Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba. Córdoba, Argentina, griselbritos@gmail.com

[IV] Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile, ronny.vallejos@usm.cl

Modelo de regresión funcional para predecir $PM_{2.5}$ a partir PM_{10}[I]

Poster / Cartel

OLAYA OCHOA, JAVIER[II] Cuevas Muñoz, José Luis [III] Murillo Alvarado, Jennifer [IV]

Colombia

Particulate matter (PM) is widely recognized as a harsh air pollutant. In fact, as an indicator of how contaminated is a place, for instance a city, people use to mention the daily average level of fine PM. Actually, fine ($PM_{2.5}$) and coarse (PM_{10}) particulate matter are two of the six criteria air pollutants, according to EPA. However, although air quality surveillance systems around the world measure both pollutants regularly, few stations measure both. This fact has led to the search of methods to estimate one of them from the other. And, since measuring $PM_{2.5}$ is more difficult, measurements of PM_{10} are more common. Also, there are many cases on which the presence of missing points is quite usual. Then, prediction of $PM_{2.5}$ from PM_{10} has become an important issue on environmental studies. Formally, PM_{10} is defined as the particulate matter whose aerodynamic diameter is less than 10 but greater than 2.5 μm. $PM_{2.5}$ refers to particles having an aerodynamic diameter under 2.5 μm. Some authors have worked on estimating the ratio $PM_{2.5}$ / PM_{10}. To reach this goal, most authors have used a simple linear regression model, but some other models, like generalized additive models (GAM), have also been fitted. In this paper we assume that hourly data can be viewed as coming from an unknown continuous function, and so we suggest using a functional data approach to model these data. Under this way of thinking, we would have as many as 24 observations of each pollutant per day, and we would treat them as functional data. Afterwards, we fit a concurrent functional linear model of functional response and functional predictor. We smooth the daily data using B-splines. A first model shows that the functional intercept may no be needed, so we moved towards a second model on which we kept only the functional slope. The R^2 is close to 91%. This model compares very well to some other models in the literature and allows one to estimate the whole unknown mean generating function of the response, as well as the model variance. This method could even be used for imputation purposes. All computing work was conducted using R.

Keywords: Functional data, Smoothing, B-splines, Air pollution, Particulate Matter.

References

[1] Faraway, J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*, 2nd. edn, Chapman & Hall/CRC.

[2] Ferraty, F. (2013). *Recent Advances in Functional Data Analysis and Related Topics*, New York, NY: Physica-Verlag.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad del Valle, Escuela de Estadística, Cali, Colombia, javier.olaya@correounivalle.edu.co

[III] Escuela de Estadística, Universidad del Valle, Cali, Colombia, jose.cuevas@correounivalle.edu.co

[IV] Escuela de Estadística, Universidad del Valle, Cali, Colombia, jennifer.murillo@cor

[3] Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd. edn, Springer.

[4] Kokoszka, P. & Reimherr, M. (2017). *Introduction to Functional Data Analysis* , New York, NY: Chapman & Hall/CRC.

[5] Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd. edn, Springer.

Applications of GRT for compact objects in General Relativity[I]

Communication / Ponencia

OLIVA MERCADO, GUILLERMO ANDRÉ[II] Frutos Alfaro, Francisco[III]

Costa Rica

We present a new geodesic ray-tracing (GRT) program that also calculates the gravitational redshift and gravitational lenses effects, given an arbitrary spacetime metric in spherical coordinates. An analytic module calculates the geodesic equations that are subsequently evaluated numerically. In general relativity, an open problem is the determination of a spacetime model for the exterior and interior of a neutron star; this problem has no unique solution. Many approximations exist, but so far, comparisons have focused only on analytical reductions to special cases. This program allows for an easy comparison of effects with observational connections for different metrics, which is applied in here to newly developed metrics with rotation and quadrupole moment and the Hartle-Thorne exterior metric.

Keywords: general relativity, solutions of Einstein's equations, approximation procedures, weak fields.

Resumen

Presentamos un nuevo programa de trazado de geodésicas (*geodesic ray-tracing*, GRT) que calcula también los efectos de corrimiento al rojo y lentes gravitacionales, dada una métrica arbitraria en coordenadas esféricas. Un módulo analítico calcula las ecuaciones de las geodésicas, que son luego evaluadas numéricamente. En relatividad general, un problema abierto es la determinación de un modelo de espaciotiempo para el exterior e interior de una estrella de neutrones; este problema no tiene solución única. Existen muchas aproximaciones, pero hasta ahora, las comparaciones entre ellas se han enfocado solamente en reducciones analíticas a casos especiales. El presente programa permite la comparación fácil de efectos con conexiones observacionales para diferentes métricas, la cual se aplica en este trabajo para métricas desarrolladas recientemente con rotación y momento de cuadrupolo, y la métrica exterior de Hartle-Thorne.

Palabras clave: relatividad general, soluciones de las ecuaciones de Einstein, procedimientos de aproximación, campos débiles.

References

[1] Frutos-Alfaro, F. (2016). "Approximate Kerr-like Metric with Quadrupole", International Journal of Astronomy and Astrophysics, 6, 334-345.

[2] Oliva-Mercado, G.A., Bonatti-González, J., Cordero-García, I., Frutos-Alfaro, F. (2015). "A Visualization of Null Geodesics for the Bonnor massive dipole", Revista de Matemática: Teoría y Aplicaciones, 22(2): 255–264

[I] Tuesday/Martes, 2:10p.m., Auditorium, Session: Apl1, Applications / Aplicaciones 1

[II] Instituto Tecnológico de Costa Rica, Escuela de Física, San José, Costa Rica, gandreoliva@gmail.com

[III] Centro de Investigaciones Espaciales,Universidad de Costa Rica, San Pedro, San José, Costa Rica, francisco.frutos@ucr.ac.cr

Determinacion de los Factores Influyentes en Accidentes de Transporte de Carga Un analisis exploratorio para la identicacion de patrones[I]

Communication / Ponencia

ORDOÑEZ ORDOÑEZ, FREDY ANTONIO[II]

Honduras

Según la Organización Panamericana de la Salud (OPS) las muertes asociadas con accidentes de tránsito representan una de las primeras causas de mortalidad en la región de América Latina. El observatorio de la violencia de la Universidad Nacional Autónoma de Honduras reporto en 2006, que las lesiones causadas por eventos de transito representaron la segunda causa de lesión externa. Como iniciativa a esta problemática, se propone un análisis estadístico exploratorio que determine los principales factores que influyen en dichos accidentes, en particular los relacionados a transporte de carga reportados por la Dirección Nacional de Transito durante los años 2004– 2012.

Introducción

El presente estudio se enfoca en el análisis de datos suministrados por la dirección nacional de vialidad y transporte, relacionados con la cuantía de siniestros ocurridos en las principales autopistas que componen la red vial pavimentada de la región, dentro del proceso de levantamiento de información, en el sitio del accidente intervienen de forma directa las unidades especializadas de la agencia de tránsito y de forma indirecta en menor grado la dirección general de carreteras y el instituto de transporte terrestre. Dicho proceso es controlado mediante los criterios y especificaciones de la ingeniería de caminos, haciendo uso de la relación entre factores de diseño, curvatura, pendientes, distancias visuales para paradas, carga viva vehicular, carriles de diseño y carriles de carga, Las clasificaciones de tamaños y cargas de vehículos de motor deben su importancia al diseño de caminos puesto que las vías de tráficos se adaptan para el vehículo más ancho, las cargas de ejes afectan la elección del espesor del pavimento, la base de ruedas incluye en la elección del radio mínimo, en caminos que hacen intersección, la altura de los vehículos afectan la decisión sobre la altura libre, en los pasos inferiores. Un camino planeado para que pasen camiones se adapta a cualquier vehículo de pasajeros considerándose unidades sencillas de camión de carga o autobús El objetivo principal de esta investigación consiste en moderar la proporción de accidentes en las vías terrestres, mediante la toma de disposiciones y criterios de vialidad, específicamente donde se ve involucrado el transporte de carga, para lo cual se hizo uso de herramientas de modelos multi-lineal y técnicas de minería de datos (data mining). La automatización de los procesos se basó en la ayuda de un ordenador acompañado con el paquete estadístico de libre distribución R, más un conjunto de programas (funciones en la nomenclatura de R) que se ejecutan bajo dicho sistemas. La metodología de este trabajo se basa principalmente en procesos de análisis exploratorios, buscando patrones de comportamiento, clasificaciones en grupo de datos dado el gran volumen de los mismos, relaciones entre las p variables para descubrir posibles estructuras a través de componentes principales, o un escalamiento multi-dimensional, o clasificaciones con

[I] Thursday/Jueves, 2:10p.m., Auditorium, Session: DMin, Data Mining / Minera de Datos

[II] IHTT instituto Hondureño Transporte Terrestre, Normativa y estudios, Honduras, Tegucigalpa, estructmate@yahoo.es / estructmate@gmail.com

un análisis de conglomerados, también de forma alternativa clasificamos grupos de datos usando métodos de estadística supervisada como el análisis de discriminante y modelos lineales, cabe mencionar que fue necesario la utilización de la estadística robusta por sus características de insensibilidad en las inferencias realizadas y sobre todo por la presencia de datos anómalos y posibles desviaciones en la distribución modelo supuesta. Con ayuda de minería de datos se logró diseñar nuevas métricas para el proceso e Identificación de accidentes de transporte de carga, lo que permite llegar a disposiciones más Exactas en los factores de incidencia, y aplicar estrategias diferenciadas para la actividad del transporte de carga en específico.

Palabras clave: Accidentes de Transito, Análisis Exploratorio, Transporte de Carga.

References

[1] Garcia, Alfonso, Advanced methods applied statistics

[2] N.H.Bingham ,Regression linears models in statistics

[3] Tellez Martinez, Carlos Statistical Control of the lenguate Process R

An ANOVA test for functional data applied to fine particulate matter measurements on air[I]

Communication / Ponencia

OVALLE MUÑOZ, DIANA PAOLA[II] Olaya Ochoa, Javier[III]

Colombia

Environmental authorities have defined fine particulate matter as particles suspended in air whose aerodynamic diameter is less than 2.5 μm (usually denoted as $PM_{2.5}$). We have daily information on this air pollutant coming from three surveillance stations and we want to check whether the levels of $PM_{2.5}$ are the same at the three places or not.

Datasets consist of daily records of as much as 24 observations per day, and so we have the typical framework on which Functional Data Analysis plays a key role. The reason is that $PM_{2.5}$ levels are originated from a continuous phenomenon and that we collect discrete observations from it. Then, using those discrete observations from a continuous phenomenon, we get a curve using smoothing techniques. This way, we have one curve, rather than one real number, per day.

We conducted the analysis using a Functional Analysis of Variance. Statistical comparison of the means from more than two populations is a very well-known problem if we are dealing with scalar values. However, it is not the case as soon as we move toward the observation of variables whose values are curves, rather than scalars.

Finally, since data from these stations were likely not to be independent, we needed to get estimations of the functional correlation between stations. Then we introduced such correlation structure into the analysis. Final results indicate statistically significant differences among the three stations.

Keywords: Functional analysis of variance, Smoothing, Air pollution, $PM_{2.5}$.

References

[1] Ferraty, F. (2013). Recent Advances in Functional Data Analysis and Related Topics, New York, NY: Physica-Verlag

[2] Hastie, T., Tibshirani, R. & Friedman, J. (2009). The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 2nd. edn, Springer

[3] Kokoszka, P. & Reimherr, M. (2017). Introduction to Functional Data Analysis , New York, NY: Chapman & Hall/CRC

[4] Ramsay, J. O. & Silverman, B. W. (2005). Functional Data Analysis, 2nd. edn, Springer]

[5] Zhang, J-T. (2014). Analysis of Variance for Functional Data, 2nd. edn, Chapman & Hall/CRC

[I] Wednesday/Miércoles, 4:40p.m., Auditorium, Session: Mod5, Modeling / Modelacion 5

[II] Universidad del Valle, Escuela Estadística, Cali, Colombia, diana.ovalle@correounivalle.edu.co

[III] Universidad del Valle, Cali, Colombia, javier.olaya@correounivalle.edu.co

Aggregation of levels in the logit model: an application to the endotracheal extubation protocol of neonatal patients[I]

Communication / Ponencia

PONSOT BALAGUER, ERNESTO[II] Penaloza, Yonel[III]

Ecuador

The binomial logit model Agresti (2015) is one of the most widely used statistical techniques for the analysis of categorical data. When after a logit adjustment, it is decided to group levels of the explanatory factors and repeat the analysis, Ponsot et al (2009) and Ponsot et al (2012) show that a distributional error is incurred. The distribution is binomial if and only if the probabilities of success are the same. The authors also suggest an alternative course of action that decreases the variance of the estimators, improves the accuracy of the estimates and increases the power of hypothesis tests on the parameters.

The methodology is tested for the first time with real data of 183 neonatal patients who needed to be intubated to a mechanical ventilator at the Dr. Miguel Perez Carreno Hospital in the city of Caracas, Venezuela, between 2012 and 2013. Postulating a logit model whose response is the success or failure of the process of extubation of patients and whose explanatory factors include clinical, hemodynamic, gasometric and ventilatory variables, the methodology shows its effectiveness, contributing to establish a protocol for the extubation procedure, which maximizes the chances of success. This application is important, since the failure of endotracheal extubation in neonates is a public health problem (Balcells, 2003) that seriously compromises the integrity of the newborn. Among other problems, it may be necessary to reintubate the patient in much more precarious conditions, also occupying again a scarce and highly demanded ventilation equipment. The main results were:

* A premature limit between 34 and 36 weeks of gestational age has 2.95 times more chances of having a successful extubation than a premature intermediate, with less than 34 weeks. With more than 36 weeks of gestational age, she has 9.33 times more likely to have a successful extubation than an intermediate preterm with less than 34 weeks.

* A neonate with a normal peak inspiratory pressure between 12 and 16 cm H2O has 3.08 times more chance of having successful extubation than one with a low peak inspiratory pressure less than 12 cm H2O.

* A neonate with three or more days on mechanical ventilation has 2.3 times more possibility of having a failed extubation of which one has less than three days. While one with seven days or more, has 4 times more chance of have a failed extubation of those who have one with less than 24 hours ventilated mechanically.

Keywords: logit model, aggregation of factor levels, endotracheal extubation protocol.

References

[1] Agresti, A. (2015) Foundations of Linear and Generalized Linear Models, John Wiley & Sons. Wiley series in probability and statistics, p. 165–201.

[I] Friday/Viernes, 10:30a.m., Room/aula 2, Session: DA1, Data Analysis / Analisis de datos

[II] Yachay Tech, School of Mathematical Sciences and Information Technology, Ecuador, Urcuqui, eponsot@yachaytech.edu.ec

[III] School of Bioanalysis, University of Los Andes, Merida, Venezuela, yonelp@gmail.com

[2] Ponsot Balaguer, E.; Sinha, S. and Goitia, A. (2009) On the grouping of levels of the explanatory factor in the binary logit model, Colombian Journal of Statistics, v 32, no 2, p. 157–187.

[3] Ponsot Balaguer, E.; Sinha, S. and Goitia, A (2012) Aggregation of explanatory factor levels in a binomial logit model: Generalization to the multifactorial unsaturated case, Colombian Journal of Statistics, v 35, no 1, p. 139–166.

[4] Balcells, J. (2003) Withdrawal of ventilation, complications and other types of ventilation, Annals of Pediatrics, no 2, p. 155–159.

Diagnóstico automático de infestación por Nosemiasis en abejas melíferas mediante procesado de imágenes.[I]

Communication / Ponencia

PRENDAS ROJAS, JUAN PABLO[II] Figueroa Mata, Geovanni [III]

Ramírez Montero, Marianyela[IV] Calderón Fallas, Rafael Ángel[V]
Ramírez Bogantes, Melvin[VI] Travieso González, Carlos Manuel[VII]

Costa Rica

Nosemiasis infestation is one of the leading causes of bee hive loss worldwide. The laboratory methods for the diagnosis of the level of infestation by this microsporidium are slow, expensive and require the presence of an expert for spore count. It is proposed the creation of an automatic, reliable and economical system of quantification of Nosema infestation from digital image processing. Using the techniques of image segmentation, object characterization and shape counting, the Cantwell and Hemocytometer techniques have been automatically reproduced. We worked with a total of 375 photographs grouped in folders of 5, which were previously labeled according to the level of infestation (very mild, mild, moderate, semi-strong and strong). The correct diagnosis rate was 84

Keywords: Nosema, image segmentation, object count, image processing.

Resumen

La infestación por Nosemiasis es una de las principales causas de la pérdida de colmenas a nivel mundial. Los métodos de laboratorio para el diagnóstico del nivel de infestación por este microsporidio son lentos, caros y demandan la presencia de un experto para el conteo de esporas. Se propone la creación de un sistema automático, confiable y económico de cuantificación de infestación por Nosema a partir del procesamiento digital de imágenes.

Con el uso de técnicas de segmentación de imágenes, caracterización de objetos y conteo de formas se ha reproducido la técnica de Cantwell y Hemocitómetro de manera automática. Se trabajó con un total de 375 fotografías agrupadas en carpetas de 5, las cuales estaban previamente etiquetadas según el nivel de infestación (muy leve, leve, moderado, semifuerte y fuerte). El porcentaje de diagnóstico correcto fue del 84% .

Palabras clave: Nosema, segmentación de imágenes, conteo de objetos, procesamiento de imágenes.

[I] Friday/Viernes, 10:30a.m., Auditorium, Session: Bio2, Biomathematics / Biomatematicas 2

[II] Instituto Tecnológico de Costa Rica, Escuela de Matemáticas, Costa Rica, Cartago, jpprendas@itcr.ac.cr
[III] Instituto Tecnológico de Costa Rica, Cartago, Costa Rica, gfigueroa@itcr.ac.cr
[IV] Universidad Nacional, Heredia, Costa Rica marianyela.ramirez.montero@una.cr
[V] Universidad Nacional, Heredia, Costa Rica rafael.calderon.fallas@una.cr
[VI] Instituto Tecnológico de Costa Rica, Cartago, Costa Rica, meramirez@itcr.ac.cr
[VII] Universidad de Las Palmas de Gran Canaria, España, carlos.travieso@ulpgc.es

References

[1] Bravo, J; Carbonell, V; Valdebenedito, JT; Figueroa, C; Valdovinos, CE; Martín-Hérnandez, R; Higes, M; Delporte, C. 2014. Identification of Nosema ceranae in the Valparaíso District, Chile. Archivos de Medicina Veterinaria. 46:487-491.

[2] Nabian, S; Ahmadi, K; Nazem-Shirazi, M.H; Gerami-Sadeghian, A. 2011. First detection of Nosema ceranae, a microsporidian protozoa of european honeybees (Apis mellifera) in Iran. Iran Journal Parasitology. 6:89-95.

[3] Prendas, J.P; Ramirez, M; Aguilar, I; Figueroa, G; Travieso, C.M. & Herrera, E. Automatic discrimination of Costa Rican stingless bees based on modified SIFT of its wings. IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI), 2016. Electronic ISBN: 978-1-4673-9578-6

[4] Ramírez, M., Prendas, J.P, Figueroa, G., Calderón, RA., Travieso, C.M & Salas, O. Cognitive Modelin of the Natural Behavior of the Varroa Destructor mite on Video. Springer. Cognitive Computation, mayo 2017. DOI 10.1007/s12559-017-9471-7

[5] Alvarez-Ramos, C.M., Niño, E., & Santos, M. Automatic Classification of Nosema Pathogenic Agents through Machine Vision techniques and Kernel-based Vector Machines. IEEE 8th Computing Colombian Conference (8CCC). Electronic ISBN: 978-1-4799-1056-4

Algunas técnicas para la detección de la falta de homogeneidad en series de datos[I]

Poster / Cartel

QUINTAS, ISABEL[II] Becerra Rodríguez, Noe[III]

México

The statistical models used in the natural and social sciences presuppose some hypotheses that do not always meet the available data series. There are models that require the data to have a normal or binomial distribution , or compliance with the homoce-dasticidad of data. In the case of the whether data series , these have been collected over the years by different institutions, even in different countries, so it does not ensure its homogeneity (as example we can mention the change occurred in 2015-2016 in Mexico for the measurement of poverty index). In economics, the deflation is used to correct the series expressed in monetary units due to effects such as inflation or devaluation; but this may not correct the effects of changes in the methods for measurement. In the case of analyzing data about employment, the number of occurrences of birth or particular disease, the inhomogeneity of the instrument with which these data have been measured may invalidate any conclusion. The bibliography has several tests to detect the inhomogeneity of the data series as part of the exploratory data analysis, EDA. This paper explain several statistical tests to apply to the data sets to determine where can be inhomogeneities such as changes in the average, variance, or periods with spurious trends. In particular some parametric tests are presented as the Buishand, the standard normal homogeneity test, from Alexandersson and Morbeg, the nonparametric test of Pettit and the von Newmann test, when testing against a series of reference and some argument about them. This test are the more selected in the hydrometeorology data analysis: in Spain (Martínez et al. 2009), in Switzerland (Alexandersson y M., 1997), in Finland and the Artic (Tourmenvirte, 2001) o Australia (Peterson, 1998). In social science are used some variance test, but are not common the use of test to detect changes in the average due to changes in the methods of determine the poverty index, feminicides, etc. in temporal or transversal series. To display the goodness of the test a randon serie is built and three systematic errors are induced: a break in the mean, a trend and a step. The tests are applied to tat series to show if they can find the errors. This test are already implemented in the R packages.

Keywords: series, series homogeneity, inhomogeneity, EDA, preliminary analysis.

Resumen

Los modelos estadísticos utilizados en las ciencias las naturales, presuponen algunas hipótesis que no siempre cumplen las series de datos disponibles. Hay modelos como los modelos de regresión simple que requieren que los datos tengan una distribución normal o binomial por ejemplo, o que se cumpla con la homocedasticidad de los datos. En el caso de las series de datos climatológicos, éstos han sido recopilados a lo largo de los años durante los cuales ocurrieron cambios en los instrumentos, los procedimientos e inclusive la localización del punto de medición. Estas modificaciones de origen no climático producen saltos, tendencias o información espúrea que puede invalidar los resultados del modelo climático. En la bibliografía aparecen numerosas

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Uninversidad Autónoma Metropolitana, Producción Económica, México, Ciudad de México, iquintas@correo.xoc.uam.mx

[III] UAM, Mexico, berono@correo.xoc.uam.mx

pruebas para detectar la inhomogeneidad de las series como parte del análisis exploratorio de datos, EDA. En este trabajo se explican varias de estas pruebas para encontrar cambios en la media, la varianza, o periodos con tendencias espúreas. En particular se presentan algunas pruebas paramétricas como la de Buishand, la prueba normal estándar de homogeneidad de Alexandersson y Morbeg, la prueba no paramétrica de Pettit y la prueba contra una serie de referencia como la von Newmann y alguna discusión sobre la bondad de ellas. Estas pruebas están entre las más seleccionadas para hacer el análisis de datos hidroclimatológicos en diversos países como España (Martínez et al, 2009), Suiza (Alexandersson y Moberg, 1997), Finlandia y el Ártico, (Tourmenvirte, 2001) o Australia (Peterson, 1998). En las ciencias sociales se utilizan más las pruebas sobre la varianza o la varianza relativa de la muestra pero no parecen interesarse por la variación no deseada de la media debido a por ejemplo cambios en el método de medición de indicadores como los de pobreza, feminicidios, etc. tantos en series temporales como transversales. Para ejemplificar la bondad de las pruebas se genera una serie de datos aleatorios simulados a la que se le introducen tres tipos de errores sistemáticos: un salto en la media, un periodo con tendencia creciente y un escalón en la media y se muestra como cada una de las pruebas es capaz de reconocer a los diferentes errores introducidos. La mayoría de estas pruebas ya están en paquetería de R por lo que se pueden hacer con unas pocas líneas de código.

Palabras clave: análisis de datos (data analysis), análisis exploratorio de datos (EDA), homogeneidad de los datos (data homogeneity), estadística en R (R programming).

Keywords: data analysis, EDA, data homogeneity, R programming.

References

[1] Toreti,A., 2010, J. of Climatology, Martinez, (2009), Aguilar, E. (2003)

Barycentric coordinates approach for complex data clustering[I]

Invited Session / Sesion invitada

RASTIN, PARISA[II] Matei, Basarab [III]

France

The purpose of Machine Learning approaches is to extract and interpret information from raw data representing a set of objects (or a set of observations). Supervised Learning methods are presented with "labelled" training data and their desired outputs, they try to obtain a function that maps inputs to outputs. On the other hand, the task of Unsupervised Learning approaches is to discover the underlying structure from "unlabelled" data. Most algorithms try to detect "clusters" in the data, i.e. groups of similar objects. The idea is that objects in the same clusters must be more similar than objects in different clusters. This task of clustering is very important in many fields, including bio-informatics, medical imaging, marketing, climatology and fraud detection.

Among the variety of algorithms that have been developed for clustering, we are interested in prototypebased approaches. The idea of such approaches is to represent the data-set by a (usually small) set of prototypes.

These prototypes are new objects in the representation space. Each prototype usually represents a cluster of objects. The main advantages of prototype-based methods are that they provide an intuitive summarize of the given data in few prototypical instances and thus lead to plausible and interpretative cluster structures. In addition, they have a low computational complexity, usually in $O(NxK)$, with N the numbers of objects in the data set and K the number of prototypes. This low complexity alone explains the popularity of prototype-based approach in real-life applications. The most used prototype-based algorithms are the K-means algorithm and its variations (e.g. K-means++, K-medoids, fuzzy C-means), as well as the family of Unsupervised Neural Network approaches such as Self-Organizing Map, Neural Gas or Boltzmann Machines.

If the objects of a data set are described in vectorial forms, the definition of cluster´s prototypes is straightforward.

In that case a prototype is a vector defined in the same vectorial space, usually defined as the vectorial barycentre of the objects (vectors) belonging to its cluster. However, in many case the objects cannot be easily defined in a vectorial space without a loss of information and/or a costly preprocessing (e.g. images, networks, sequences, texts). To analyze such non-vectorial data sets, it is common to describe the data using the relations or the similarities between the objects, using a dissimilarity or distance matrix. For this reason, they are sometimes called Relational Data. Relational clustering algorithms form a family of methods adapted to relational data. Some clustering algorithms are naturally adapted to deal with dissimilarity matrix and can be used to analyse relational datasets, such as HDBSCAN, Spectral Clustering, Affinity Propagation or Hierarchical Clustering.

None of these algorithms use prototypes and they does not benefit from the associated advantages. In particular, they all have a non-linear computational complexity.

Few work have been done yet in relational prototype-based clustering, but some authors have worked on using K-means with relational data [1, 2, 4]. The Partitioning Around Medoids (PAM) algorithm proposes to choose the K prototypes among the objects of the dataset, allowing a distortion in the cluster representation.

[I] Friday/Viernes, 10:50a.m., Room/aula 3, Session: R, Clustering and Visualization of Complex Data

[II] Université Paris 13, Villetaneuse, France, rastin@lipn.univ–paris13.fr

[III] Université Paris 13, Villetaneuse, France, matei@lipn.univ–paris13.fr

In order to propose a better representation of the clusters, in [2] the authors represent the prototypes as a linear combination of the input data. But, as far as processing power and memory usage are concerned, this implementation is very expensive (O(N2) for N objects), making it unusable for large sets of data. In [4] the authors proposed to enforce sparser prototyping by considering the prototypes as a linear combinations of a subset from the data set (called support points), succeeding in improving the computational complexity. However, the support points are specific to the cluster represented by the prototype, and the complete dissimilarity matrix (size N2) is still needed to be kept in memory.

In this work, we propose a new formalism for relational data clustering, allowing a low complexity and memory cost. The main idea is to use a set of fixed support points chosen among the objects of the data set, independently from the clusters, and use these support points as a basis for the definition of a representation space using the Barycentric Coordinates formalism [3]. All relational objects in the data set can be defined in the barycentric space. We show that it is also possible to define and update the prototypes in the same space.

This representation allows a low computational and memory cost with an accurate description of the clusters prototypes. We propose two algorithms for our proposed approach: a batch version, were the data set is kept in memory during the whole learning process and a stochastic version were objects are presented one by one. The batch version is faster than the stochastic version, at the cost of an increase in memory usage. Both are much faster and need much less memory than other approaches for relational data and are a significant improvement in terms of computational and memory complexity compared to the state-of-the-art algorithms. We demonstrate these properties theoretically and experimentally on a set of artificial and real relational data. The experiments also show the high quality of the algorithm in terms of clustering results.

For our future works, we aim to combine this approach with a dynamic artificial neural network algorithm such as Growing Neural Gas (GNG). The aim will be to deal with dynamic non-vectorial data-sets. The neural network is able to create or delete prototypes dynamically and adapt it´s structure. By defining these prototypes in the Barycentric space, it would be possible to follow the changes in the clusters structure and update the model accordingly.

Keywords: clustering, relational data, barycentric coordinates.

References

[1] 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016, Athens, Greece, December 6-9, 2016. IEEE, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7840087

[2] R. J. Hathaway, J. W. Davenport, and J. C. Bezdek, Relational duals of the c-means clustering algorithms, Pattern Recognition, vol. 22, no. 2, pp. 205–212, 1989.

[3] E. Hille, Analytic Function Theory, ser. AMS Chelsea Publishing Series. AMS Chelsea Publishing, 2005, no. vol. 2. [Online]. Available: https://books.google.com/books?id=rY8_ WxS2zsC

[4] F. Rossi, A. Hasenfuss, and B. Hammer, Accelerating relational clustering algorithms with sparse prototype representation, in Proceedings of the 6th International Workshop on Self–Organizing Maps, Bielefeld (Germany), 9 2007.

Multiple Imputation with R[I]

Short course / Curso corto

REINECKE, JOST[II]

Germany

Missing values often occur in empirical data sets across different disciplines (e. g., Economy, Educational and Social Science). A missing value means that no data value is stored for variable j of observational unit i. Possible reasons are general refusals (a possible participant of a study chooses not to take part in a study), partial refusals (participant refuses to answer a certain question), noncontacts (participant could not be contacted) or other reasons during data collection. Schafer (1997) discusses three factors that determine whether or not missing data pose a threat to the validity of statistical inferences: the missing data percentage, the selectiveness of the missing data process and the applied missing data procedure. Currently, there are state-of-the-art procedures to cope with the problem of missing data: multiple imputation, direct (or full information) maximum likelihood estimation and weighting. The short course will focus on techniques of multiple imputation and the purpose will be

- to familiarize participants with the problems and pitfalls, missing data pose to data analysts,
- to discuss the pros and cons of various approaches to analyze incomplete empirical data (including multilevel data), and
- to demonstrate the application of the most important and widely used missing data imputation tools in R (e.g. `norm2` package, `mice` package).

latex and pdf File

Keywords: Multiple Imputation, R, Norm2, mice.

References

[1] Kleinke, K., Reinecke, J., Salfrán, D. and Spiess, M. (2018). *Applied Multiple Imputation: Advantages, New Developments and Pitfalls*, Heidelberg: Springer.

[2] Schafer, J. L (1997). *Analysis of incomplete multivariate data*, London: Chapman & Hall.

[3] Schafer, J. L (2016). *norm2: Analysis of incomplete multivariate data under a normal model*, Retrieved from https://CRAN.R-project.org/package=norm2.

[4] van Buuren, S. and Groothuis-Oudshoorn, K. (2011). "MICE: Multivariate imputation by chained equations in R", *J. Stat. Software, nº 45*, p. 1–67.

[I] Wednesday/Miércoles, 9:00a.m., Laboratory 1, Session: Tut, Tutorial 6

[II] University of Bielefeld, Faculty of Sociology, Germany, Bielefeld, jost.reinecke@uni-bielefeld.de

A word-distance based method for Maya language classification[I]

Communication / Ponencia

REYES FIGUEROA, ALAN GERARDO[II]

Guatemala

A methodology for linguistic classification based on estimate distances between languages is developed. These language distances are constructed based on word-metrics, which are widely used in the computer science community. On this framework, we apply hierarchical classifications to a subfamily of 20 languages of the maya linguistic group, and later we extend to the whole family of 31 maya languages. For each of these languages, we compute distance or similitude matrices, attending to different metric methodologies, specifically by using the Levenstein metric and other similar word-metric approximations. This computation is based on morphological principles of writing in each language. For the construction of the distance matrices, we chose a fixed list of 65 words and compare one by one its writing form in each language of study. From these matrices, we then compute clusters by several hierarchical methods. We explore also some other classifications methodologies, such as K-means, SVM clustering and graph-based methods. Finally, the obtained results are compared to accepted classification schemes in maya linguistics, specifically the Campbell- Kaufman classification and the Brown-Holman-Wichmann-Velupillai classification. As a result we obtain high degree of compatibility with the accepted schemes.

Keywords: Quantitative linguistics, computational linguistics, clustering, hierarchical classification.

References

[1] Cecil H. Brown, Eric W. Holman, Viveka Velupillai and Sren Wichmann (2008). Automated classification of the worlds languages: a description of the method and preliminary results. Language Typology and Universals 61(4), 285–308.

[2] Lyle Campbell and Terrence Kaufman (1985). Mayan linguistics: Where are we now?. Annual Review of Anthropology 14:187–198.

[3] W. Cohen, P. Ravikumar, y F. Fienberg (2003). A comparison of string distance metrics for name–matching tasks. In Proceedings of the IJCAJ–03 workshop on Information Integration on the Web, 73–78.

[4] C Cab, Carlos (2003). Maya Choltzij: Vocabulario comparativo de los idiomas mayas de Gautemala. (OKMA: Antigua Guatemala), 343pp.

[5] Gwang-Yoon Goh (2000). Probabilistic meaning of multiple matchings for language relationship. Journal of Quantitative Linguistics 7, 53–64.

[I] Wednesday/Miércoles, 2:35p.m., Room/aula 3, Session: Clas1, Classification / Clasificacion

[II] CIMAT, Universidad del Valle de Guatemala, Usac, Departamento de Matemática, Guatemala, Guatemala, agreyes@uvg.edu.gt

An automatic framework for graph-based representation of coronary arteries[I]

Communication / Ponencia

REYES FIGUEROA, ALAN GERARDO[II] Camacho Gutiérrez, José Ariel [III]
Cervantes Sánchez, Fernando [IV] Salazar Colores, Sebastián [V] Cruz Aceves, Ivan[VI]

México

We present an automatic framework for the segmentation and graph-based representation of coronary arteries in X-ray angiograms. The first stage of vessel segmentation is carried out by using Gaussian matched filters, estimated by using a Univariate Marginal Distribution Algorithm (UMDA) heuristics. This stage is evaluated in terms of the area (Az) under the receiver operating characteristic curve and by the accuracy metric. In the second step, the line simplification method of Ramer-Douglas-Peucker is used to represent the coronary artery skeleton as a graph, where the nodes are the dominant pixels for shape preservation. This step is evaluated in terms of the data compression ratio. In the experimental results, the proposed framework obtained an Az = 0.9135, with a segmentation accuracy of 0.9215, and a data compression ratio of 95.7328 using a test set of 30 angiograms.

In addition, the experimental results has demonstrated that the proposed method can be suitable for computer-aided diagnosis in clinical practice and with useful results for further purposes such as image registration.

Keywords: Image segmentation, estimation of distribution algorithms, signal filters, medical imaging.

References

[1] Urs Ramer (1972). An iterative procedure for the polygonal approximation of plane curves. Computer Graphics and Image Processing, 1(3), 244-256.

[2] David Douglas and Thomas Peucker (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer,10(2), 112–122.

[3] A. Salhi, B. Minaoui and M. Fakir (2014). Robust Automatic Traffic Signs Detection using fast polygonal approximation of digital curves. International Conference on Multimedia Computing and Systems Proceedings 1, 433-437.

[4] A. Saalfeld (1999). Topologically Consistent Line Simplification with the Douglas–Peucker Algorithm. Cartogr. Geogr. Inf. Sci. 26, 7-18.

[5] P. Larrañaga and J.A. Lozano (2002). Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer, Boston, MA.

[I] Tuesday/Martes, 4:40p.m., Auditorium, Session: Apl3, Applications / Aplicaciones 3

[II] Centro de Investigación en Matemáticas (CIMAT), Ciencias de la Computación, México, Guanajuato, agreyes@cimat.mx
[III] Matemáticas Aplicadas, Centro de Invetigación en Matemáticas, CIMAT, jose.camacho@cimat.mx
[IV] Ciencias de la Computación, Centro de Invetigación en Matemáticas, CIMAT, fernando.cervantes@cimat.mx
[V] Departamento de Ciencias de la Computación, Universidad Autónoma de Querétaro, s.salazarcolores@gmail.com
[VI] Ciencias de la Computación. CONACYT - Centro de Invetigación en Matemáticas, CIMAT, ivan.cruz@cimat.mx

Energy-based and PDE methods for image restoration[I]

Communication / Ponencia

REYES FIGUEROA, ALAN GERARDO[II]

Guatemala

Digital inpainting is the process of modifying and interpolate a digital image in an undetectable form. This process is used in the restoration of damaged paintings and photographs or the removal/replacement of selected objects. In this paper we review some classical techniques for image inpainting based on minimization of energy functionals, namely: a simple method involving the harmonic diffusion equation, a second method involving a advection-diffusion equation, based mainly on works of Perona and Malik [1], and Bertalmio et al. [2], and a third method based on the Mumford-Shah functional [3], [4]. We also explore some relaxed versions of the previous methodologies, such as the Total Variation method and the Ambrosio-Tortorelli functional. For each of these approaches, we develop a partial differential equation scheme, which we solve numerically in terms of a discrete finite difference process. We test the developed algorithms in a set of several images differing in relation to texture and structure diversity. We compare the obtained results in terms of visual appreciation, error and convergence of the methods.

Keywords: Image inpainting, diffusion equation, transport equation, Mumford-Shah functional.

References

[1] P. Perona and J. Malik. (1990). Scale-space and Edge Detection Using Anisotropic Diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):629-639.

[2] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester. (2000). Image Inpainting. Proceedings of SIGGRAPH 2000, New Orleans, USA.

[3] D. Mumford and J. Shah. (1985). Boundary detection by minimizing functionals I. Proc. IEEE Conf. CVPR 1985, San Francisco, pp. 2226.

[4] D. Mumford and J. Shah. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. 42, 577-685.

[5] M. Bertalmio, A.L. Bertozzi, and G. Sapiro. (2001). Navier-Stokes, fluid dynamics, and image and video inpainting. Proceedings of the 2001 IEEE Conference on CVPR 1, pp. 355-362.

[I] Tuesday/Martes, 4:00p.m., Room/aula 2, Session: Num4, Numerical Analysis / Análisis Numérico 3

[II] CIMAT, Universidad del Valle de Guatemala, Usac, Guatemala, Departamento de Matemática, Guatemala, Guatemala, agreyes@uvg.edu.gt

Grouped p-generalized spherical distributions[I]

Communication / Ponencia

RICHTER, WOLF-DIETER[II]

Germany

The Gauss-exponential distribution plays an important role as a high risk limit law, see the lectures on high risk scenarios and extremes in [1]. The Gauss-exponential distribution can be considered as a particular asymmetric derivation of the Gauss-Laplace law. In particular, Gauss-exponential probabilities of arbitrary events can be dealt with by considering the corresponding Gauss-Laplace probabilities. Density level sets of the standard Gauss-Laplace distribution are topological boundaries of star bodies centered at the origin. The Minkowski functionals of the corresponding star bodies, however, are not homogeneous functions of order one as it is often assumed in the literature on star-shaped distributions. Instead, the bodies corresponding to different density levels reflect different geometric properties and are typically directed to different directions. The aim of the talk is to first recall (p, q)-spherical generalizations of the two-dimensional Gauss-Laplace distribution from [2] and to secondly extend this consideration to the grouped $(p_1, ..., p_k)$-spherical case. We prove geometric and stochastic representations which can be considered as standard tools for dealing with the present distributions in a way being similar to that one has already successfully been dealing for a long time with spherical and, since more recently, even with more general homogeneous star-shaped distributions. This will be demonstrated by generalizing the Box-Muller simulation method following [3] and [4] and discussing further statistical properties following [5].

Keywords:

Gauss-exponential distribution, Gauss-Laplace distribution, grouped $(p_1, ..., p_k)$-spherical uniform distribution, stochastic vector representation, dynamic geometric disintegration.

References

[1] Balkema, G., Embrechts, P. (2007). High risk scenarios and extremes. In: Zurich Lectures in advanced mathematics. Zurich: European Mathematical Society.

[2] Richter, W.-D. (2017). The class of (p,q)-spherical distributions with an extension of the sector and circle number functions. Risks 2017, 3. Doi: 103390/risks5030040

[3] Kalke, S., Richter, W.-D. (2013). Simulation of the p-generalized Gaussian distribution. J. Stat.Comput. Simul. 83 (4), 641-667. Doi: 10.1080/00949655.2011.631187

[4] Richter, W.-D., Schicker, K. (2017). Simulation of polyhedral convex contoured distributions. Journal of Statistical Distributions and Applications 4:1. Doi: 10.1186/s40488-017-0055-6

[5] Richter, W.-D. (2017). Statistical reasoning in dependent p-generalized elliptically contoured distributions and beyond. Journal of Statistical Distributions and Applications 4:21. Doi: 10.1186/s40488-017-0074-3

[I] Friday/Viernes, 11:30a.m., Room/aula 1, Session: L4–3, LACSC–Recent advances in robust statistics and distribution theory

[II] University of Rostock, Institute of Mathematics, Germany, Germany, wolf-dieter.richter@uni-rostock.de

On p-generalized elliptically contoured distributions[I]

Short course / Curso corto

RICHTER, WOLF-DIETER[II]

Germany

Flexible statistical modeling is needed in many areas of applied scientific and related work. Multivariate distribution theory involves both big mathematical challenges and many chances for new applications. The short course gives an introduction to a p-generalization, $p > 0$, of the classical theory of elliptically contoured distributions (where $p = 2$) and some of its probabilistic and statistical applications.

A starting point of this far reaching generalization is the stochastic representation $X \stackrel{d}{=} R \cdot U$ holding for every p –generalized spherical random vector X with independent R and U. Here, $R = |X|_p = (\sum_{i=1}^{n} |X_i|^p)^{1/p}$ and $U = \frac{1}{R}X$ are the p–generalized radius (or p–functional) and the p-generalized uniform basis of the random vector X, respectively. Although the distribution ω of U is singular with respect to the Lebesgue measure in the sample space R^n many authors, as in [1], describe it by discussing the density of its $n-1$-dimensional marginal distribution. Here, we fully explain the geometric nature of ω as follows. Based on the notion of the p-generalized surface content measure O_p on the Borel σ-field of the unit p-sphere $S_p = \{x \in R^n : |x|_p = 1\}$, according to [2], $\omega(A) = O_p(A)/O_p(S_p), A \in \mathcal{B}^n \cap S_p$. Special emphasis, when explaining this in detail, will be on dimension two where the notion of generalized surface content is reduced to that of generalized arc-length which can (and will) be explained much easier.

Additionally scaling and rotating p-spherical random vectors, together with further generalizing the notion of the uniform distribution to that on the Borel σ-field on p-generalized ellipsoids, allow then to introduce the more general class of random vectors following a p-generalized elliptically contoured distribution according to [3].

The subfamilies where the p-functional $|.|_p$ is a norm ($p \geq 1$) or an antinorm ($0 < p \leq 1$) are particularly discussed. Stochastic vector and geometric measure representations for these cases will be extended to those for the general families of convex and radially concave contoured distributions being important subfamilies of the again much more general family of star-shaped distributions.

Probabilistic properties as moments and characteristic functions will be discussed as well as statistical aspects as parameter estimation and hypothesis testing. Particular emphasis will be on the case that a density exists. Modeling heavy or light distribution tails and centers by choosing a suitable density generator is then of particular interest for various applications.

First applications of stochastic vector and geometric measure representations deal with the derivation of the distributions of certain statistics (such as generalized Chisquare-, Student- and Fisher-statistics, products, ratios and order statistics) under non-standard model assumptions and with simulation, more advanced ones with the construction of both skewed p-generalized elliptically contoured distributions extending the class considered in [4] and with p-generalized von Mises distributions as in [5].

Keywords: convex and radially concave contoured distributions, stochastic vector and geometric measure representations, simulation, skewed $l_{n,p}$-symmetric distributions, p-generalized von Mises distributions.

[I] Wednesday/Miércoles, 9:00a.m., Room/aula 2, Session: Tut, Tutorial 5

[II] University of Rostock, Institute of Methematics, Germany, Rostock, wolf-dieter.richter@uni-rostock.de

References

[1] Gupta, A. K., Song, D. (1997). l_P-norm spherical distributions. J Stat Plann Inference 60(2),241-260.

[2] Richter, W.-D. (2009). Continuous $l_{n,p}$-symmetric distributions. Lith Math J 49(1), 93-108. Doi: 10.1007/ s10986-009-9030-3.

[3] Richter, W.-D. (2014). Geometric disintegration and star-shaped distributions. J Stat Distrib Appl 1:20, 1-24. Doi: 10.1186/ s40488-014-0020-6.

[4] Arellano-Valle, R., Richter, W.-D. (2012). On skewed continuous $l_{n,p}$-symmetric distributions. Chilean Journal of Statistics v.3, no. 2, 193-212.

[5] Dietrich,T., Richter (2017). Classes of geometrically generalized von Mises distributions. Sankhya v. 79-B, part 1, 21-59. Doi: 10.1007/ s13571-016-0118-6.

A Partial Differential Equation Model With Age-Structure and Nonlinear Recidivism: Conditions for a Backward Bifurcation and a General Numerical Implementation[I]

Communication / Ponencia

SANCHEZ, FABIO[II] Calvo, Juan Gabriel [III] Segura, Esteban [IV]

Costa Rica

We formulate an age–structured three stage nonlinear partial differential equation model that features nonlinear recidivism to the infected (infectious) class (I) from the temporarily recovered class (R). Equilibria are computed, as well as, stability analysis of the infection–free equilibrium. A generalized numerical framework is established and numerical experiments are explored where necessary conditions are needed for two positive solutions to exist in the "infected" class. As a result a backward-bifurcation exists under those necessary conditions.

Keywords: epidemic models, backward bifurcation, modeling, math models, partial differential equations.

References

[1] Brauer, F. (2004). Backward bifurcations in simple vaccination models, Journal of Mathematical Analysis and Applications, 298, pp. 418–431.

[2] Castillo—havez, C. and Song, B. (2004). Dynamical models of tuberculosis and their applications, Mathematical Biosciences and Engineering, 1, pp. 361–404.

[3] Kot, M. (2001). Elements of Mathematical Ecology, Elements of Mathematical Ecology, Cambridge University Press.

[4] F. Sanchez, F., Wang, X., Castillo–Chavez, C., Gorman, D. and Gruenewald, P. (2007). Drinking as an Epidemic–A Simple Mathematical Model with Recovery and Relapse, Elsevier Inc., pp. 353–368.

[5] Song, B. Du, W. and Lou, J. (2013). Different types of backward bifurcations due to density–dependent treatments, Mathematical biosciences and engineering : MBE, 10, p. 1651–1668.

[I] Wednesday/Miércoles, 2:35p.m., Auditorium, Session: Mod2, Modeling / Modelacion 2

[II] Universidad de Costa Rica,CIMPA y Escuela de Matemática, Costa Rica, San José, fabio.sanchez@ucr.ac.cr

[III] Universidad de Costa Rica, San José, Costa Rica, juangabriel10@gmail.com

[IV] Universidad de Costa Rica, CIMPA, San José, Costa Rica, estebansu@gmail.com

Juegos dinámicos y el comportamiento en las organizaciones industriales.[I]

Communication / Ponencia

SÁNCHEZ GUEVARA, IRENE[II] Baca Lobera, Gloria [III]

México

Dynamic games are a powerful tool for decision making in organizations, particularly in industrial organizations. In Mexico there are certain sectors of the food industry that have the characteristics of oligopolies such as the dairy industry and the producer of corn flour, which, because there are few, are likely to avoid competition by raising prices. In this paper, the study of two companies is presented, using the dynamic models proposed by Cournot and Stakelberg. The mathematical modeling of games is a tool that allows analyzing the structure of the market. On the one hand, analyze the conditions that allow cooperation between them, which means that supply can be restricted and prices raised, or the conditions of competition for the benefit of society can be analyzed.

Keywords: industrial organization, dynamic games.

Resumen

Los juegos dinámicos constituyen una herramienta poderosa para la toma de decisiones en las organizaciones, en particular en las organizaciones industriales. En México hay ciertos sectores de la industria alimentaria que presentan características de oligopolios tales como la industria lechera y la productora de harina de maíz, que por ser pocas existe la posibilidad de que eviten la competencia, elevando los precios. En este trabajo se presenta el estudio de dos empresas, utilizando los modelos dinámicos que proponen Cournot y Stakelberg. La modelación matemática de los juegos es una herramienta que permite analizar la estructura del mercado. Por un lado analizar las condiciones que permiten la cooperación entre ellas, lo que hace que se pueda restringir la oferta y elevar los precios, o bien analizar las condiciones de competencia para beneficio de la sociedad.

Palabras clave: organización industrial, juegos dinámicos.

References

[1] Gibbons, R. (1992) Game theory for Applied Economies. Princeton, University Press. Princeton, New Jersey.

[I] Wednesday/Miércoles, 1:50p.m., Auditorium, Session: Mod1, Modeling / Modelacion 1

[II] Universidad Autónoma Metropolitana, política y Cultura, México, Ciudad de México, isabiro@yahoo.com.mx

[III] Universidad Autónoma Metropolitana, Ciudad de México, México, gbaca@correo.xoc.uam.mx

Operads[I]

Communication / Ponencia

SÁNCHEZ GUEVARA, JESÚS[II]

Costa Rica

La teoría de operads se presenta como una herramienta para clasificar el tipo de homotopía de espacios topológicos al caracterizar las estructuras algebraicas que los determinan. En esta breve exposición vamos a explicar algunos de sus alcances y una posible aplicación en el estudio de espacios moduli.

Palabras clave: operad, homotopias, espacios moduli.

References

[1] Markl, S. Shnider, and J. D. Stasheff. Operads in algebra, topology and physics. Mathematical Surveys and Monographs. American Mathematical Society, 2007.

[2] P. May. The geometry of iterated loop space. Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 1972.

[3] A. Prouté. Sur la transformation d´Eilenberg–Maclane. C. R. Acad. Sc. Paris, 297:193–194, 1983.

[4] Sánchez, Jesús 2016. About E-infinity structures in L-algebras, PhD thesis. Université Paris-Diderot. Paris, France.

[I] Friday/Viernes, 2:10p.m., Room/aula 2, Session: DJ, Simulacion Multiagente: Herramientas y Aplicaciones

[II] Escuela de Matemáticas, Universidad de Costa Rica, CIMPA, jesus.sanchez_g@ucr.ac.cr

Estimation of holding periods applied to the case of short and leveraged ETFs[I]

Plenary Talk / Conferencia Plenaria

SCHUBERT, LEO[II] Schubert, David [III]

Germany

The estimation of the holding periods of financial products has to be done in a dynamic process in which the size of the observation time interval influences the result. Small intervals will produce smaller average holding periods than bigger ones. The approach developed in this paper offers the possibility of estimating this average independently of the size of this time interval. This method is demonstrated on the example of two distributions, based on the exponential and the geometric probability functions. The estimation will be found by maximizing the likelihood function. The two examples will finally be applied to the financial instrument Exchange Traded Fund (ETF). The analysis contains ETFs with leverage factors of -2, -1, $+1$ and $+2$. Although different ETFs are treated, the majority of the data is concerned with the "db x-tracker ShortDAX ETF", "db x-trackers DAX ETF", "iShares DAX (DE)" and the "Lyxor ETF LevDAX". By the application of the proposed estimation approaches, the average holding periods of ETFs increase by 4%–29%. This increase depends on the time interval T of observation, the leverage factor, and the average holding period.

Keywords: Holding periods, Exchange Traded Funds (ETF), Maximum Likelihood estimation, Exponential distribution.

Resumen

La estimación de los períodos de tenencia de una inversión en instrumentos financieros debe realizarse mediante un proceso dinámico, en el cual el tamaño del intervalo de observación influye en los resultados. Pequeños intervalos de observación producirán en promedio períodos de tenencia más pequeños a los obtenidos con intervalos grandes de observación. El enfoque desarrollado en este artículo ofrece la posibilidad de estimar estos promedios, independientemente del tamaño del intervalo de observación. El método es ilustrado con el ejemplo de dos distribuciones, basadas en las leyes de probabilidad exponencial y geométrica. La estimación óptima se encuentra al maximizar la función de verosimilitud. Los dos ejemplos examinados fueron aplicados al instrumento financiero Exchange Traded Fund (ETF). Estos fondos, negociables en la bolsa de valores, tienen factores de apalancamiento -2, -1, +1, y +2. Aunque más de 30 ETFs fueron estudiados, la mayoría de los datos están orientados al "db x-tracker ShortDAX ETF", "db x-trackers DAX ETF", "iShares DAX (DE)" y el "Lyxor ETF LevDAX". La aplicación del método propuesto a las ETFs aumenta la amplitud del período de tenencia entre un 4% y 29%. El aumento es dependiente del tamaño de la ventana de observación, del factor de apalancamiento y del promedio de los períodos de tenencia.

Palabras clave: Períodos de tenencia, distribución exponencial, estimación de máxima verosimilitud, Exchange-Traded Funds (ETFs).

[I] Wednesday/Miércoles, 11:20a.m., Auditorium, Session: CONF, Conference 3

[II] Constance University of applied Sciences, WISO, Germany, Constance, schubert@htwg-konstanz.de

[III] Munich Germany, David.Schubert1@gmx.de

References

[1] Funke C., Gebken T., Johanning L. (2012). Investitionsverhalten bei Short- und Leveraged ETFs. WHU-Otto Beisheim School of Management. Retrieved from http://www.fondsprofessionell.de/upload/attach/1355396407.pdf

[2] Michalik T., Schubert L. (2009). Hedging with Short ETFs. Economics Analysis working Papers, 8(9).

[3] Schubert L., (2011). Investitionsverhalten bei Short- und Leveraged ETFs. HTWG-Konstanz. Retrieved from http://www.fondsprofessionell.de/upload/attach/8911.pdf

Computation of Matrix p-th Root Using Solvents[I]

Communication / Ponencia

SEGURA UGALDE, ESTEBAN[II]

Costa Rica

We study the matrix equation:

$$X^p - A = 0, \tag{3}$$

where $A \in \mathbb{C}^{n\times n}$. A matrix X satisfying (3) is called a p-th root of A (see, e.g., [4]). It is known that if A is nonsingular, it always has a p-th root. This is a classic problem in mathematics and still an active research topic. A well studied particular case of (3) is when $p = 2$, i.e.:

$$X^2 - A = 0. \tag{4}$$

Here a solution X of (4) is called a matrix square root of A (see, e.g., [3]). An interesting fact in this case, is that if A is singular, the existence of a square root depends on the Jordan structure of the zero eigenvalues (see, e.g., [1]). For instance, the matrix:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has no square root, while the matrix:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

does have a square root, specifically:

$$X = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The method of reference for computing matrix square roots is to compute a Schur decomposition, then compute a square root of the triangular factor, and finally transform back (see [2]). Iterative methods are alternatives for computing p-th roots. In [3], [4], for instance, several variations of the Newton's method to approximate the (square roots) p-th roots are presented. The new approach we present here, to compute a p-th root of A, is based on matrix solvents and its computation using a (block) moment method. Matrix solvents are studied in detail in [5] and are defined as follows. Let $P(\lambda)$ be an $n \times n$ matrix polynomial, i.e., $P(\lambda) = \sum_{i=0}^{\ell} A_i \lambda^i$, for $A_i \in \mathbb{C}^{n\times n}$. Then, a matrix $S \in \mathbb{C}^{n\times n}$ is called a (right) solvent for $P(S)$ if satisfies the relation:

$$P(S) := A_\ell S^\ell + \cdots + A_2 S^2 + A_1 S + A_0 = 0. \tag{5}$$

Note that equation (3) is a particular case of the monic matrix solvent problem of degree p:

$$P(S) := S^p + A_{p-1} S^{p-1} + \cdots + A_2 S^2 + A_1 S + A_0 = 0$$

[I] Tuesday/Martes, 4:20p.m., Room/aula 2, Session: Num4, Numerical Analysis / Análisis Numérico 3

[II] Universidad de Costa Rica, Matemática, Costa Rica, San Pedro, estebansu@gmail.com

when the matrices A_i are zero for $i = 1, \ldots, p-1$.

Keywords: Matrix p-th Root, Solvents, Matrix Polynomials, Moment Method. Matrix p-th Root, Solvents, Matrix Polynomials, Moment Method.

References

[1] G. Cross and P. Lancaster. Square roots of complex matrices. *Linear and Multilinear Algebra*, 1(4):289–293, 1974.

[2] N. J. Higham. Computing real square roots of a real matrix. *Linear Algebra Appl.*, 88:405–430, 1987.

[3] N. J. Higham. Newton's method for the matrix square root. *Mathematics of Computation*, 46(174):537–549, 1986.

[4] B. Iannazzo. On the Newton method for the matrix pth root. *SIAM J. Matrix Anal. Appl.*, 28(2):503–523, 2006.

[5] E. Segura Ugalde, M. Barkatou and P. Boito. A contour integral approach to the computation of invariant pairs. *TCS, Elsevier*, 681:3–26, 2017.

Análisis de una formulación aumentada basada en pseudo-esfuerzo para un modelo Brinkman no lineal de flujo de medios porosos[I]

Communication / Ponencia

SEQUEIRA CHAVARRÍA, FILÁNDER[II] Gatica, Gabriel[III] Gatica, Luis[IV]

Costa Rica

We introduce and analyze an augmented mixed finite element method for the two-dimensional nonlinear Brinkman model of porous media flow with mixed boundary conditions. More precisely, we extend a previous approach for the respective linear model to the present nonlinear case, and employ a dual-mixed formulation in which the main unknowns are given by the gradient of the velocity and the pseudostress. In this way, and similarly as before, the original velocity and pressure unknowns are easily recovered through a simple postprocessing. In addition, since the Neumann boundary condition becomes essential, we impose it in a weak sense, which yields the introduction of the trace of the fluid velocity over the Neumann boundary as the associated Lagrange multiplier. We apply known results from nonlinear functional analysis to prove that the corresponding continuous and discrete schemes are well-posed. In particular, a feasible choice of finite element subspaces is given by Raviart-Thomas elements of order $k \geq 0$ for the pseudostress, piecewise polynomials of degree $\leq k$ for the gradient, and continuous piecewise polynomials of degree $\leq k+1$ for the Lagrange multiplier. We also derive a reliable and efficient residual-based a posteriori error estimator for this problem. Finally, several numerical results illustrating the performance and the robustness of the method, confirming the theoretical properties of the estimator, and showing the behaviour of the associated adaptive algorithm, are provided.

Keywords: nonlinear Brinkman model, mixed finite element method, augmented formulation, high-order approximations.

References

[1] Gatica G.N., Gatica L.F., Sequeira F.A., Analysis of an augmented pseudostress-based mixed formulation for a nonlinear Brinkman model of porous media flow. Computer Methods in Applied Mechanics and Engineering, vol. 289, 1, pp. 104-130, (2015).

[I] Tuesday/Martes, 2:10p.m., Room/aula 2, Session: Num2, Numerical Analysis / Análisis Numérico 1

[II] Universidad Nacional, Escuela de Matemática, Costa Rica, Heredia, filander.sequeira@una.cr

[III] CI2MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile, ggatica@ci2ma.udec.cl

[IV] CI2MA and Departamento de Matemática y Física Aplicadas, Facultad de Ingeniería, Universidad Católica de la Santísima Concepción, Casilla 297, Concepción, Chile, lgatica@ucsc.cl

Methods of mathematical managing theory in the forecast of dynamic price share on a stock market[I]

Communication / Ponencia

SHUKLIN, GERMAN[II]

Ukrain

The purpose of this work is to create a mathematical model of price share management on the stock exchange using the non linear differential equation of the pendulum, which swings are caused by a force with a delay. Using of this method makes it possible to analyse the impact of external factors on the dynamic stock market price share, creating managing mechanisms against negative influences, as well as establishing correlative contacts between the portfolios of securities., we consider the price of the asset sold in the stock market as some variable C(t), which is the solution of equation,

$$\dot{c}(t) = Ac(t) + Bc(t-\tau) \tag{6}$$

$x(t) \in R^n, t \geq 0, \tau = const.$
Here $x(t) \equiv \varphi(t)$,$-\tau \leq t \leq 0$,and $\varphi(t)$ is arbitrary continuously differentiable initial vector function. A,B - are quadratic matrices with constant coefficients.

DEFINITION 1. Delay exponentional of matrix B is matrix function e_τ^{Bt} that has a form

$$e_\tau^{Bt} = \begin{cases} \theta, & \text{at} -infty < t < -\tau; \\ I, & \text{at} -\tau \leq t < 0; \\ I + B\frac{t}{1!} + B^2\frac{(t-\tau)^2}{2!} + \ldots + B^k\frac{(t-(k-1)\tau)^k}{k!}, & \text{at}(k-1)\tau \leq t < k\tau. \end{cases}$$

$k = 0, 1, 2, \ldots$,where θ is zero matrix.
THEOREM 1.Let $AB = BA$ of the system (1) is true.Then the solution Cauchy problem for system (1)has a form
$x(t) = e^{A(t-\tau)}e_\tau^{B_1(t-\tau)}\varphi(-\tau) + \int_{-\tau}^{0} e^{A(t-\tau-s)}e_\tau^{B_1(t-\tau-s)}e^{A\tau}[\varphi^{/}(s) - A\varphi(s)]ds$ Will consider linear nonhomogeneous system with constant delay

$$\dot{x}(t) = Ax(t) + B(t-\tau) + f(t) \tag{7}$$

THEOREM 2.The solution $\bar{x}(t)$ of nonhomogeneous system (2)with zero initial conditions has a form
$\bar{x}(t) = \int_0^t e^{A(t-\tau-s)}e_\tau^{B_1(t-\tau-s)}f(s)ds$,$t \geq 0$.

Keywords: Share,system,control function,stock market, delayed argument

[I]Thursday/Jueves, 2:55p.m., Room/aula 2, Session: Mod4, Modeling / Modelacion 4

[II]College of Information System and Technology named after Vadym Hetman, department of mathematic, Kiyv, Ukrain, redwaveplus@ukr.net

References

[1] D. YA. KHUSAINOV and G.V.SHUKLIN. LINEAR AUTONOMOUS TIME-DELAY SYSTEM WITH PERMUTATION MATRICES SOLVING// Studies of the University of Zilina. Mathematical Series. Vol.17 (2003), 101-108.

[2] D.Ya. Khusainov, A.F. Ivanov, and G.V. Shuklin. On a Representation of Solutions of Linear Delay Systems.//Differential Equations, Vol.41,No.7,2005,pp.1054-1058.

Nonparametric estimation of the first order Sobol indices with bootstrap bandwidth[I]

Communication / Ponencia

SOLÍS CHACÓN, MAIKOL[II] Loubes, Jean-Michel[III] Marteau, Clement[IV]

Costa Rica

Implement policies or take decisions in real-world problems is often supported by complex models.

The scientist has to process, analyze and interpret those complex systems with numerous variables and interactions. One option to overcome these issues is selecting the most relevant variables of the system.

In this way, we will gain insight on the model and we will discover the principal characteristics of it.

Also, we have to produce a stable approximation of the model to avoid large variations on the input given by small perturbations on the input. In any case, the analyst has to validate, check and correct the model if it is necessary.

We assume a set of inputs variables $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ producing an output $Y \in \mathbb{R}$ related by the model

$$Y = m(X_1, \ldots, X_p). \tag{8}$$

The function m is, generally, an unknown and complex function. However, a computer code can gauge the function m in some cases. Also, we can replace the original model by a low fidelity approximation called a *meta-model*. Given a set of inputs $(X_1, \ldots, X_p)$ in the model (8), we can rank them according different criteria. The work of [sobol1993sensitivity], inspired by a ANOVA (or Hoeffding) decomposition, split down the variance of the model in partial variances generated by the conditionals expectations of Y giving each input X_i for $i = 1, \ldots, p$. The partial variances represent the uncertainty created by each input or its interactions. Dividing each partial variance by the model total variance, we obtain a normalized index of importance. Specifically, we call the first-order Sobol indices to the quantities,

$$S_i = \frac{V(\, E[Y|X_i])}{V(Y)} \quad \text{for} \quad i = 1, \ldots, p.$$

Notice that $E[Y|X_i]$ is the best approximation of Y given the information of X_i. Thus, if the variance of $E[Y|X_i]$ is large, it means a large influence of X_i into Y.

The main endeavor with the Sobol indices relays in its computation. Some methods propose the use of multiple samples (of the order of hundreds or thousands) for the evaluation of the model outputs. From applied problems in engineering, biology, oceanography and others; the scientists have developed Monte-Carlo or quasi Monte-Carlo methods.

For instance, the Fourier amplitude sensitivity test (FAST) or the Sobol pick-freeze (SPF). The Monte-Carlo methods suffer of the high-computational stress in its implementation. For example, the FAST method requires estimate a set of suitable transformation functions and integer angular frequencies for each variable.

[I] Thursday/Jueves, 3:15p.m., Room/aula 2, Session: mod4, Modeling / Modelacion 4

[II] Universidad de Costa Rica, Escuela de Matemática, San Pedro, Costa Rica, maikol.solis@ucr.ac.cr

[III] Institut de Mathématiques de Toulouse, Toulouse, France, loubes@math.univ-toulouse.fr

[IV] Institut Camille Jordan, Lyon, France., marteau@math.univ-lyon1.fr

The SPF scheme creates a new copy of the variable in each iteration. For complex and high-dimensional models, those techniques could be expensive in computational time. This work proposes an alternative way to compute the Sobol indices. In particular, we will take the ideas of [zhu1996asymptotics] and we shall apply a nonparametric Nadaraya-Watson to estimate the value S_i for $i = 1, \ldots, p$.

With this estimator, we avoid the stochastic techniques and we use the structure of the data to fit the nonparametric model. We consider only the indices with simple interaction between one variable with respect the output. We leave out the sensitivity indices with interactions for a further study. Furthermore, we will show that if the joint distribution of (X_i, Y) is twice differentiable, the nonparametric estimator of S_i, has a parametric rate of convergence. Otherwise, we will get a nonparametric rate of convergence depending on the regularity of the density. To complete the analysis, we have to estimate the optimal bandwidth. Even the cross-validation method is asymptotically unbiased, it possess a relatively large finite-sample bias. This behavior has been observed studying nonparametric estimators for the density, quantiles and the mode respectively. We propose correct the bias part using a bootstrapping procedure, following the ideas in [Racine2001].

Keywords: Nonparametric, Sobol indices, bootstrap

References

[Racine2001] Racine, Jeff (2001). "Bias-Corrected Kernel Regression´´ In: *Journal of Quantitative Economics* 17.1, pp. 25–42.

[sobol1993sensitivity] Sobol', I.M. (1993). "Sensitivity Estimates for Nonlinear Mathematical Models´´ In: *Mathematical Modeling and Computational experiment* 1.4, pp. 407–414.

[zhu1996asymptotics] Zhu, Li-Xing and Kai-Tai Fang (1996). "Asymptotics for kernel estimate of sliced inverse regression´´ In: *The Annals of Statistics* 24.3, pp. 1053–1068.

A new cryptographic system using the Duffing map[I]

Communication / Ponencia

SOLIS SANCHEZ, HUGO[II] Barrantes, E. Gabriela[III]

Costa Rica

There exists a necessity to review and construct new cryptosystems due to the RSA compromising by the Quantum Architectures and nature represents a good source of inspiration to propose new schemes. The Duffing map is an example of a dynamical system that exhibits chaotic behavior obtained from the modeling of a simple physical system, a spring pendulum whose spring's stiffness does not exactly obey Hooke's law[1]. The Distributed Dynamics Encryption (DDE) represents a promissory mathematical method to generate a chaotic public key crypto-system[2]. Under DDE we have developed a new cryptographic scheme using the Duffing map. In this work, we have revisited the cryptographic requirements [3] for the new scheme and we found how the use of chaotic maps extracted from the modelling of the nature represents an advantage in the developing of chaotic cryptosystems over the traditional.

Keywords: Public key cryptosystems, duffing map, chaos, dynamical system, RSA.

References

[1] Gonzalo Alvarez and Shujun Li. Some basic cryptographic requirements for chaos–based cryptosystems. International Journal of Bifurcation and Chaos, 16(08):2129–2151, 2006.

[2] J Koz lowski, U Parlitz, and W Lauterborn. Bifurcation analysis of two coupled periodically driven duffing oscillators. Physical Review E, 51(3):1861, 1995.

[3] J Koz lowski, U Parlitz, and W Lauterborn. Bifurcation analysis of two coupled periodically driven duffing oscillators. Physical Review E, 51(3):1861, 1995.

[I] Wednesday/Miércoles, 4:00p.m., Room/aula 2, Session: DS, Dynamical Systems / Sistemas Dinamicos

[II] Universidad de Costa Rica, Física, Costa Rica, San Jose, hugo.solis@ucr.ac.cr

[III] Universidad de Costa Rica, San Jose, Costa Rica, gabriela.barrantes@ecci.ucr.ac.cr

Modelo de optimización integral para programación de autobuses y tripulaciones de manera simultánea[I]

Communication / Ponencia

SORIA ARGUELLO, ISIDRO[II] Mejía Argueta, Christopher[III]

México

El exponencial crecimiento poblacional provoca problemas de movilidad dentro de las ciudades y entre ellas, generando grandes retos operativos al transporte público en México, comúnmente etiquetado de ineficiente y caro. Actualmente, existen brechas metodológicas y prácticas en el problema de asignación de vehículos a corridas, tripulaciones y programación de mantenimiento; ya que no hay modelos integrales y los existentes están respaldados en supuestos no operativos para todos los países al enfocarse en problemas con restricciones específicas de operación como jornadas de trabajo de un día y terminales dentro de la misma ciudad. La presente investigación propone un modelo genérico de optimización que permite programar autobuses, tripulaciones y esquemas de mantenimiento en forma simultánea, minimizando costos operativos. El modelo es validado en una empresa mexicana que tiene una flota heterogénea de 1,000 autobuses, 2100 operadores, atendiendo más de 500 corridas diarias en 70 terminales con 15 talleres. Los resultados muestran que el modelo integral mejora el desempeño global comparado con la operación actual de la empresa; atendiendo las corridas programadas en la oferta comercial de la compañía con un menor número de autobuses.

Palabras clave: Transporte público, programación de vehículos y tripulaciones, mantenimiento, optimización.

References

[1] Huisman, D., & Wagelmans, A. P. (2006). A solution approach for dynamic vehicle and crew scheduling. European Journal of Operational Research, 172(2), 453-471.

[2] Mesquita, M., Paias, A., & Respício, A. (2009). Branching approaches for integrated vehicle and crew scheduling. Public Transport, 1(1), 21-37.

[3] Jütte, S., & Thonemann, U. W. (2015). A graph partitioning strategy for solving large-scale crew scheduling problems. OR Spectrum, 37(1), 137-170.

[4] Maher, S. J., Desaulniers, G., & Soumis, F. (2014). Recoverable robust single day aircraft maintenance routing problem. Computers & Operations Research, 51, 130-145.

[I] Wednesday/Miércoles, 1:30p.m., Room/aula 2, Session: OR1, Operations Research / Investigacion de Operaciones

[II] Universidad Anáhuac México, Ingeniería Industrial, México, México, soriaarguelloisidro@hotmail.com

[III] Eindhoven University of Technology, Eindhoven, The Netherlands

Replicator dynamics: old and new[I]

Plenary Talk / Conferencia Plenaria

SORIN, SYLVAIN[II]

France

We will describe the links between some agorithms for convex optimization, equilibria in games and on line learning. Starting with the repicator dynamics we will describe the associated unilateral version and its connection to no-regret learning procedures. We will survey recent results on extensions of this dynamics: regularization functions and variable weights, links with classical gradient algorithms in discrete and continuous time, equilibria and variational inequalities, potential and dissipative games.

Keywords:
optimization, learning, repeated games.

[I] Tuesday/Martes, 10:30a.m., Auditorium, Session: CONF, Conference 1

[II] Université Pierre et Marie Curie, Combinatoire et Optimisation Institut de Mathématiques de Jussieu-PRG , Fance , sylvain.sorin@imj-prg.fr

Estimation of a mixed–effects model using a partially observed diffusion process[I]

Communication / Ponencia

SOTO, JOSÉ[II] Infante, Saba[III] Camacho, Franklin[IV] Amaro, Isidro[V]

Ecuador

The mixed-effects models can be a extension of regression models for longitudinal data adding a component of random effects. This component explains the variations, between and inside, of experimental units. In this work, we propose a mixed effects diffusion model defined in terms of a stochastic differential equation, which describes the behavior repeated measures taken to a time discrete and with errors on a series observations. Since the estimate of parameters in these models is not clear, except for simples cases ([1, 2]). A methodology, based on Monte Carlo's methods, is proposed to estimate parameters. This depends of type of solution of the stochastic differential equation. If the solution is explicit, we proposed a Monte Carlo algorithm by chains of Markov to approximate the posteriori distribution, and then it estimates the parameters. If the solution is an implicit form, it does a Euler-Maruyama discretization, and it uses a Sequential Monte Carlo Algorithm for the approximations of posteriori distribution. For determinate if the model is appropriate, it is made a convergence analysis (diagnostic analysis). Finally, the methodology is illustrated using synthetic data and real data.

Keywords: Misted effects models, Stochastic Differential Equation, Monte Carlo algorithm by chains of Markov.

References

[1] Aït-Sahalia, Y. (2002) "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach´´, *Econometrica, Vol. 70, Blackwell Publishers Ltd.* p. 223–262. DOI: 10.1111/1468-0262.00274

[2] Aït-Sahalia, Y. (2008) "Closed-form likelihood expansions for multivariate diffusions". *Ann. Statist. 36, No. 2*, p. 906–937. DOI:10.1214/009053607000000622.

[3] Picchini,U; Ditlevsen, S. (2011) "Practical estimation of high dimensional stochastic differential mixed-effects models". *Computational Statistics and Data Analysis, Vol. 55, No 3.* p. 1426 - 1444. DOI: https://doi.org/10.1016/j.csda.2010.10.003

[4] Whitaker, G.; Golightly, A.; Boys, R.; Sherlock, C. (2017) "Bayesian Inference for Diffusion-Driven Mixed-Effects Models". *Bayesian Anal. 12, No. 2*, p. 435–463. DOI:10.1214/16-BA1009

[I] Friday/Viernes, 1:50p.m., Room/aula 1, Session: LACSC–2, LACSC–Statistical Computing 5

[II] Universidad de Investigación de Tecnología Experimental, Escuela de Ciencias Matemáticas y Tecnología Informática, Ecuador, Urcuquí, josemayns@gmail.com

[III] School of Mathematical Sciences and Information Technology, Yachay Tech University, Urcuquí, Ecuador, sinfante64@gmail.com

[IV] School of Mathematical Sciences and Information Technology – Yachay Tech University, Urcuquí, E

[V] School of Mathematical Sciences and Information Technology – Yachay Tech University, Urcuquí, E

Frobenius Norm: Advances and Trends in Optimization Problems in Signal Processing[I]

Communication / Ponencia

SOTO QUIRÓS, JUAN PABLO[II]

Costa Rica

The Frobenius norm is a matrix norm defined as the square root of the sum of the absolute squares of the elements of $A \in \mathbb{R}^{m\times n}$, i.e.,

$$\|A\|_f = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2}.$$

This matrix norm has been applied in many minimization problems in signal processing such that, to name a few, nonnegative matrix factorization for dimensionality reduction of data [4], reduced-rank transforms and reconstruction of noisy signals [2], shadow/light removal of images [3] and sparse representation of signals using an over complete dictionary [**?**]. A generative model for these minimization problems is represented by

$$\min_{X_1,...,X_p} \|\mathcal{F}(X_1,...,X_p)\|_f^2 + \mathcal{G}(X_1,...,X_p), \tag{9}$$

where $X_j \in \mathbb{S}_j^{m_j\times n_j} \subseteq \mathbb{R}^{m_j\times n_j}$, for $j = 1,..,p$, $\mathcal{G} : \mathbb{R}^{m_1\times n_1} \times ... \times \mathbb{R}^{m_p\times n_p} \to \mathbb{R}$ is a regularization transform and $\mathcal{F} : \mathbb{R}^{m_1\times n_1} \times ... \times \mathbb{R}^{m_p\times n_p} \to \mathbb{R}^{m\times n}$ is a matrix transform. This talk mainly focuses on some relevant optimization problems in signal processing related with equation (9), where the principles, properties and algorithms are summarized. Relevant application areas of each specific problem are also briefly described.

Keywords: Frobenius norm, Optimization, Signal processing.

References

[1] Y. X. Wang and J. Y. Zhang, Nonnegative matrix factorization: A comprehensive review, IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 6, pp. 1336–1353, June 2013.

[2] P. Soto Quiros and A. Torokhti, Optimal transforms of random vectors: The case of successive optimizations, Signal Processing, vol. 132, pp. 183 – 196, 2017.

[3] T. Zhou and D. Tao, Godec: Randomized low-rank & sparse matrix decomposition in noisy case, in in International Conference on Machine Learning, 2011.

[4] M. Aharon, M. Elad, and A. Bruckstein, rmk–svd: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[I] Wednesday/Miércoles, 2:55p.m., Room/aula 2, Session: Opt1, Optimization / Optimizacion 1

[II] Instituto Tecnológico de Costa Rica, Matemáticas, Costa Rica, Cartago, jusoto@itcr.ac.cr

Hierarchical Bayesian Model for Diseases Frecuency in Costa Rica[I]

Poster / Cartel

SOTO ROJAS, CRISTINA[II] Castro Porras, Jose Carlos[III]

Costa Rica

In Costa Rica exist the C.C.S.S.(Caja Costarricense del Seguro Social), the institution in charge of the social security. Per year attends millions of people and has the record of that, but what happens with all of those records. There's a lot of information that can be very useful hide in that data. This project starts from the idea of giving an use to all of that information. Specifically, the objective is to analyze the spatial frequency of different diseases. If that frequency was known by the C.C.S.S, better decisions can be made, and a lot of resources can be better distributed.

In the literature, we find that in similar situations it has been used a Bayesian hierarchical model, as [1] mention, this kind of models are used to try to describe these situations, and work very well when spatial variables are involved. [2] explain in their book the different Bayesian hierarchical models that can be used in the solution of similar problems, also explain their advantages and disadvantages. Also, the implementation of this models is show and some examples. There are two thesis related to the subject. First, the one of [4], about a spatial Bayesian model for the incidence of dengue in Puerto Rico, which is a problem similar to the ours, and also have some improves that we plan to implement in future advances. And the other one, of [5] is about a spatial Bayesian hierarchical model in agricole epidemiology. This one explains in a complete way all the different models that can be used and gives an interesting application. In [2] we found an implementation of an A Multivariate Hierarchical Bayesian Framework for Healthcare Predictions which explains how the Bayesian hierarchical model works and how we can include more predictors in the basic model and study the effect of more parameters. Also, it allows them to analyze the impact of some factors in the service that is provided. There is an atlas [1] from Costa Rica that shows the disease lethality per canton in the period of 2000-2007 and also use data form the C.C.S.S..

The data consist in 3 years(2014,2015,2016) of hospital discharges records distributed by canton, from C.C.S.S. and a spatial hierarchical Poisson-Gamma models is used in order to observe the incidence in each canton of the 21 big groups of diseases. As a result, we can observe the cantons with highest incidence of each disease in the country . Also, the results obtained are compared with the frequentist model.

Keywords: Hierarchical Model, Disease Incidence, Poisson-Gamma Model.

References

[1] Amanda Aparicio Llanos y Melvin Morera Salas. Atlas de mortalidad por todas las causas Costa Rica 2000-2007. San José, C.R. :EDNASSS, 2009.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad de Costa Rica, San José, Costa Rica, anacris_ 28-94@hotmail.com

[III] Univeresidad de Costa Rica, San José , Costa Rica, josecarloscast@hotmail.com

[2] Sudipto Banerjee, Bradley P. Carlin y Alan E. Gelfand. Hierarchical Modeling and Analysis for Spatial Data . Chapman y Hall/CRC, 2004.

[3] Hall Chapman. Spatio-Temporal Methods in Environmental Epidemiology. CRC Press,2015.

[4] Greivin Hernández Gónzalez. Modelo Espacial Bayesiano para la Incidencia del Dengue en la Isla Principal de Puerto Rico para el año 2014. Sistema de Estudios de Posgrado de la Universidad de Costa Rica, 2017.

[5] Nora Coromoto Monsalve Graterol.Modelos Jerárquicos Bayesianos Espaciales en Epidemología Agrícola. Departamento de Estadística e Investigación Operativa de la Universitat de Valencia, ene. de 2013.

Hypothesis testing for mean comparison of two non-independent functional populations[I]

Communication / Ponencia

URBANO LEON, CRISTHIAN LEONARDO[II] Olaya Ochoa, Javier[III]

Colombia

Functional Data Analysis (FDA) is a branch of Statistics intended to deal with objects which are not scalars but function curves. Such curves are obtained through smoothing techniques, from discrete measurements of a phenomenon of continuous origin. Given that objects are functions, it is possible to use the properties of a Hilbert subspace of the infinite functional space of integrable square functions (L_2 space). This fact allows using the common definitions on this vector space of functions, such as the norm and the internal product.

FDA theory is still under heavy construction. One reason is that functions do not follow a specific known distribution, and so it becomes hard to make sound statistical comparisons of, for instance, two or more populations functional mean curves. This is true since it requires the use of test statistics considering the functional nature of the data, added to other phenomena such as non-independence between curves of two distinct populations and between curves within each population. Until now, as far as we know, functional test statistics are extensions of classic test statistics. This work sticks to this idea of extending known test statistics based on real numbers to the functional data setting. We discuss the problem of statistical comparison of curves and implement a test statistic for the comparison of two populations functional means. We consider the case of two non-independent populations and then we illustrate the proposed test with an application to air pollution data.

Keywords: Functional data analysis, Smoothing, Fourier basis, Air pollution, Hilbert space.

References

[1] Faraway, J. (2016). Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models, 2nd. edn, Chapman & Hall/CRC.

[2] Ferraty, F. (2013). Recent Advances in Functional Data Analysis and Related Topics, New York, NY: Physica-Verlag.

[3] Hastie, T., Tibshirani, R. & Friedman, J. (2009). The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 2nd. edn, Springer.

[4] Kokoszka, P. & Reimherr, M. (2017). Introduction to Functional Data Analysis , New York, NY: Chapman & Hall/CRC.

[5] Ramsay, J. O. & Silverman, B. W. (2005). Functional Data Analysis, 2nd. edn, Springer.

[I] Thursday/Jueves, 2:10p.m., Room/aula 2, Session: Mod3, Modeling / Modelacion 3

[II] Universidad del Valle, Escuela de Estadística, Cali, Colombia, cristhian.urbano@correounivalle.edu.co

[III] Universidad del Valle, Santiago de Cali, Colombia, javier.olaya@correounivalle.edu.co

Relative risk of cardiovascular disease and type 2 diabetes in people with metabolic syndrome: systematic review and meta-analysis[I]

Communication / Ponencia

VARGAS PANIAGUA, ESTEBAN[II] Castro, Milena[III]

Costa Rica

Background

The purpose of this study is to synthesize the existing evidence related to the associated risk of developing cardiovascular diseases or type II diabetes, by carrying out a systematic review followed by a meta-analysis; in order to be able to objectively integrate the "state of the art" on this phenomenon. It is known that people with metabolic syndrome (MetS) are at increased risk of developing cardiovascular diseases and/or type II diabetes. We quantified the risk of cardiovascular disease (CVD), coronary heart disease (CHD), stroke, T2DM and all-cause mortality in individuals with known MetS using NCEP criteria.

Methods

We identified, from EMBASE and Medline (consulted on September 11st, 2017), 110 prospective cohort studies in which the diagnosis of MetS was evaluated at the beginning of the observation and individuals were followed from 0.8 up to 20 years. Studies included had a mean sample size of 8847 individuals (Range: 154 to 275 867). Random-e?ects meta-analysis models were used to calculate a pooled relative risk estimate (RR) for each outcome. Some extracted estimates were adjusted by age, gender, smoking, physical activity or low-density lipoprotein cholesterol. The national cholesterol education program (NCEP) definition for MetS was used as inclusion criteria. Assessment of publication bias was performed with previously evaluated methods.

Findings

Individuals with MetS were at higher risk compared to those without for all-cause mortality with a RR of 1.23 and 95% CI 1.14-1.33, CHD (RR 1.60, 95% CI 1.48-1.73), stroke (RR 1.52, 95% CI 1.35-1.70), CVD (RR 1.49, 95% CI 1.37-1.62) and DM (RR 4.28, 95% CI 3.27-5.59). **Interpretation**

Metabolic syndrome diagnosis can be used as a tool to identify people at higher risk of developing type 2 diabetes mellitus, cardiovascular outcomes and/or Mortality. The estimation of the risk can be used to develop different strategies in order to treat individuals with MetS. The prevention of these diseases is a major public health issue.

Keywords: Relative Risk, Meta-Analysis, CVD, Diabetes Mellitus, All cause mortality and Evidence synthesis

[I] Friday/Viernes, 11:10a.m., Room/aula 2, Session: DA1, Data Analysis / Analisis de datos

[II] Universidad de Costa Rica, Escuela de Estadística, Alajuela, Costa Rica, esvapa8@gmail.com

[III] University of Costa Rica, San José, Costa Rica milena.argentina@gmail.com

References

[1] Andreadis, E. A., Tsourous, G. I., Tzavara, C. K., Georgiopoulos, D. X., Katsanou, P. M., Marakomichelakis, G. E., & Diamantopoulos, E. J. (2007). Metabolic syndrome and incident cardiovascular morbidity and mortality in a Mediterranean hypertensive population. American Journal of Hypertension, 20(5), 558–564. https://doi.org/https://dx.doi.org/10.1016/j.amjhyper.2006.12.001

[2] Arnlov, J., Sundstrom, J., Ingelsson, E., & Lind, L. (2011). Impact of BMI and the metabolic syndrome on the risk of diabetes in middle–aged men. Diabetes Care, 34(1), 61–65. https://doi.org/https://dx.doi.org/10.2337/dc10-0955

[3] Asociación latinoamericana de diabetes ALAD (2009). Guías ALAD de diagnóstico, control y tratamiento de la Diabetes Mellitus Tipo 2. Obtenido de : https://www.google.com/url?sa=t& rct=j& q=& esrc=s& source=web& cd=2& ved=0ahUKEwjHq4 C139zVAhVilVQKHcJpBwAQFggxMAE& url=http% 3A% 2F% 2Fwww1.paho.org% 2Fhq%2Fd mdocuments% 2F2010% 2FGuias_ ALAD_ 2009.pdf& usg=AFQjCNHyG6PR4hbKyQ0iaoHAAjG_ 4jA6eQ

[4] Assmann, G., Schulte, H., & Seedorf, U. (2008). Cardiovascular risk assessment in the metabolic syndrome: Results from the Prospective Cardiovascular Munster (PROCAM) study. International Journal of Obesity, 32(SUPPL. 2), S11–S16. https://doi.org/10.1038/ijo.2008.29

[5] Benetos, A., Thomas, F., Pannier, B., Bean, K., Jego, B., & Guize, L. (2008). All-cause and cardiovascular mortality using the different definitions of metabolic syndrome. The American Journal of Cardiology, 102(2), 188–191. https://doi.org/https://dx.doi.org/10.1016/j.amjcard.2008.03.037

Explicit Spatial Control Strategies for Dengue, Zika and Chicunguya outbreaks[I]

Communication / Ponencia

VELÁZQUEZ CASTRO, JORGE[II] Bonilla Capilla, Beatriz[III]
Anzo Hernández, Andrés[IV] Soto Bajo, Moises[V] Fraguela Collar, Andrés [VI]

México

Dengue, Zika and Chikungunya are deseases of mayor concern in many countries. In particular, Dengue is of major importance due to its epidemiological magnitude. On the other hand, in many countries, diseases like Zika and Chikungunya can be considered as new diseases thus, a great portion of the population is susceptible. This characteristic make of them a mayor epidemiological threat. All of this are vector-borne diseases transmitted by the bite of the mosquito A. Aegypti. On the other hand, when applying control strategies of urban vector borne diseases, there are normally economic and physical limitations. In particular, there are increasing concerns and interest in designing the most efficient strategies and regulations that the health agencies can follow in order to reduce the imminent impact of virus like Denge, Zika and Chikungunya. A basic question that arise when designing control strategies is about where the resources should be allocated. This include ether, geographical and methods allocation. This means that government agencies would like to know where to implement the control strategies in order to make the greatest impact. In addition to this there is the question of when to apply this control strategies. In this work, we proposed dynamical indexes based on a mathematical model that can guide the decisions or when and where is best to invest in control strategies. This indexes takes into account the evolution of the epidemic where urban mobility and the differences in population density play and important role. For each neighborhood in a city, we found two different dynamical indexes. One that guide the control measures focused on reducing the disease spread cause by human habits and interactions. One example of this kind of control is the immediate hospitalization of possible infected individuals. The other index can guide control strategies focused in the reduction of the vector population. This dynamical indexes can be used to stratify a city in different priority areas. This indexes takes into account the evolution of the epidemic and thus they may change in order to propose the best strategy at a specific stage of the outbreak. One of these indexes can be use to give priority to certain city areas to applied education campaigns about symptoms and held care with focus in avoiding the geographical disease spread. The other one can be use to give priority to areas where vector control like fumigation and abatization is most needed. Both kind of control strategies must not coincide neither in space nor in time.

Keywords: dinamical indexes, control, human mobility, A. aegypty, metapopulations.

[I] Friday/Viernes, 2:10p.m., Auditorium, Session: Bio1, Biomathematics / Biomatematicas 1

[II] Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, México, Puebla, jorge.velazquezcastro@correo.buap.mx

[III] Facultad de Ciencias Físico-Matemáticas, BUAP, Puebla, México, beatriz.bonillac@correo.buap.mx

[IV] Facultad de Ciencias Físico-Matemáticas, BUAP, Puebla, México, andres.anzo@fcfm.buap.mx

[V] Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, Puebla, México, moises.soto@fcfm.buap.mx

[VI] Facultad de Ciencias Físico-Matemáticas, BUAP, Puebla, México, fraguela@fcfm.buap.mx

References

[1] Bichara, D., Holechek, S. A., Velázquez-Castro, J., Murillo, A. L., & Castillo-Chavez, C. (2016). On the dynamics of dengue virus type 2 with residence times and vertical transmission. Letters in Biomathematics, 3(1), 140–160.

[2] Bowman, L. R., Runge-Ranzinger, S., & McCall, P. J. (2014). Assessing the Relationship between Vector Indices and Dengue Transmission: A Systematic Review of the Evidence. PLoS Neglected Tropical Diseases, 8(5), e2848.

[3] Honório, N. A., Nogueira, R. M. R., Codeço, C. T., Carvalho, M. S., Cruz, O. G., de Avelar Figueiredo Mafra Magalhaes, M., Lourenço–de–Oliveira, R. (2009). Spatial Evaluation and Modeling of Dengue Seroprevalence and Vector Density in Rio de Janeiro, Brazil. PLoS Neglected Tropical Diseases, 3(11), e545.

[4] Stolerman, L. M., Coombs, D., & Boatto, S. (2015). SIR-Network Model and Its Application to Dengue Fever. SIAM Journal on Applied Mathematics, 75(6), 2581–2609.

Multivariate characterization of the profiles of rural women workers[I]

Communication / Ponencia

VICENTE GALINDO, M^A PURIFICACIÓN[II] Patino Alonso, M^a Carmen[III]
Galindo Villardón, M^a Purificación[IV]

España

The reality of rural Spain is diverse, from small villages that populate the northeast interior of the country, to villas on the Mediterranean coast, to the outskirts of the major urban centres of the country. There is, therefore, no single definition of the rural in Spain. Different authors have advanced several criteria to characterize an area as rural, based on cultural, economic, demographic, geographic etc. connotations. According to García (1994), the definition of what is rural encounters important difficulties owing to the complexity of rural society as such and to the impossibility of attaining a certain degree of generalization for different rural areas.

In the 1980s, the studies done from a gender perspective introduced a new approach to investigations on feminine uprootedness in the rural world (Camarero et al. 2009). According to Gómez and Rico (2005), in the rural setting the female population is ageing progressively because the women have a longer life expectancy than the men and because most young women migrate to the cities in search of educational and job opportunities. Women leave the villages in greater numbers than men, leaving behind a very male-dominated territory. However, many of them would like to return to a rural setting if they were offered the personal and professional development denied them there.

The objetive has been to perform a study aimed at determining the job situation of women working in rural areas.

To analyze the problems of rural women, field work was performed in the province of Salamanca (Spain). The data formed part of a questionnaire given to women working in a rural municipality of the province, employed regularly or irregularly. The overall sample comprised 772 women working in the rural setting, with ages between 16 and 64 years, employed regularly or irregularly. According to the Economic and Social Council (1999), the concept of irregular employment is based on behaviours reflecting a decline in the collection of social contributions (Social Security) or taxes.

The distribution observed with respect to the type of municipality was as follows: in the sample considered, 23.2 per cent belonged to municipalities with fewer than 500 inhabitants (1.4 per cent of the questionnaires given in villages with fewer than 100 inhabitants and 21.6 per cent between 100 and 500 inhabitants), that is, strictly rural; 34.88 per cent of the sample belonged to municipalities with fewer than 2,000 inhabitants - intermediate rurality- and 42 per cent had between 2,000 and 10,000 inhabitants, lax rurality. Groups of women with similar characteristics were separated using Multidimensional Scaling (MDS) (Green and Carmone, 1969). The coordinates obtained concerning the first two dimensions of the MDS were used in a cluster analysis (K-Means) to distribute each of the women in the group to which she belonged.

[I] Thursday/Jueves, 2:35p.m., Auditorium, Session: Aprox1, Approximation / Aproximacion

[II] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, purivg@usal.es
[III] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, carpatino@@usal.es
[IV] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, pgalindo@usal.es

Four well differentiated groups were discerned: two comprising rural women working in regular work and the other two with women in irregular job situations; both types of groups were seen to be differentiated by their marital status and the activity sector where they work .

Keywords: rural women, cluster, MDS.

Mathematics Subject Classification (2010): 62Hxx

References

[1] Camarero L.A., Cruz F., González M.., Del Pino J., Oliva J. and Sampedro R. (2009). "The rural population in Spain. From disequilibrium to social sustainability", *Social Studies Collection , nº 27*.

[2] Gómez J.M and Rico M. (2005). "La mujer en el medio rural de Castilla y León: Diversificación sectorial y proceso de dinamización económica", *Estudios de Economía Aplicada, nº 23* , p. 465–490.

[3] Economic and Social Council (1999). *La Economía Sumergida en Relación a la Quinta Recomendación del Pacto de Toledo Sesión del Pleno de 21 de abril*, Madrid: Informe CES.

[4] Green and Carmone (1969). "Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques", *Journal of Marketing Research, nº 6*, p. 330–341.

Recent Developments in Analysis of Three-Way Data[I]

Short course / Curso corto

VICENTE GALINDO, PURIFICACIÓN [II] Nieto-Librero, Ana Belén [III][IV]

España

In standard multivariate analysis the data are placed in a two-way matrix (individuals $\times$ variables). In this case, the data can be indexed by two indices, one referring to the individual and the other to identify the variable. However, many data have a three-way structure, for example, each individual is considered at different times or conditions. In this case, the data is indexed with three indexes (Three-way data). Multi-way analysis is the natural extension of standard multivariate analysis when the data are placed in matrices of three or more ways.

STATIS, an extension of Principal Component Analysis, is a method to analyze tables of three entries that consider several sets of variables on the same individuals, or, the same variables on several sets of individuals (STATIS DUAL), at different times or conditions.

STATIS has had a great impact on the Data Analysis and is very broad literature that contemplates different developments of related techniques and hundreds of applications in different fields of the science. This method, and similarly the AFM try to find a consensus configuration over time (or conditions) and therefore tries to evaluate the stable part in the dynamics of the relationships. Among the different methods to study three-way data, the STATIS is undoubtedly the most used.

The STATIS methods (STATIS DUAL)

We start from a set of T studies that we will identify (X_t, M_t, D_t) where X_t represents the data matrix with I individuals and J variables, M_t is the metric in the hyperspace of the individuals and D_t is the metric in the hyperspace of the variables. In general, the metrics used match the identity matrix. We try to simultaneously explore several X_t data matrices where each of them collects information about J variables in I individuals, in T occasions, or different experimental situations.

Each of the studies (X_t, M_t, D_t) can be identified by two structures which are associated to scalar product matrices: $W_t = X_t M_t X_t^T$ which emphasizes the positions of the individuals and $C_t = X_t^T D_t X_t$ which emphasizes the structures of covariation of the variables. We will name W_t and C_t objects or configurations representing each study. In both cases, it is a question of studying the inter-structure between the different studies and if this evidence shows a similar behavior, a commitment structure is defined and the different studies on the structure of commitment are analyzed in order to analyze the intra structures through the trajectories.

Comparing inter-structures means having some way of comparing the configurations: it is necessary to define a distance between them in order to assess the similarity or differences. The STATIS (dual STATIS) method uses the Hilber-Schmidt product to induce a distance between the objects W_t (C_t) representative of the studies.

The objectives of the STATIS are:

- to compare and analyze the relationships between the matrices of starting data

[I] Wednesday/Miércoles, 8:00a.m., Auditorium, Session: Tut, Tutorial 1

[II] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, purivg@usal.es

[III] Departamento de Estadística, Universidad de Salamanca, Salamanca, España, ananieto@usal.es

[IV] EE.UU. de Relaciones Laborales, Zamora, España, ananieto@usal.es

- to integrate these matrices in a single one that will be a weighted average of the starting matrices and that we will call consensus matrix
- to project each one of the original matrices on the consensus and analyze what they have in common and in what they disagree

The applications to different fields of science are counted by dozens: in chemometrics, in ecology, in medicine, in statistical quality control, in molecular biology, etc. etc.

Recent developments of the STATIS

The developments of the STATIS can be classified into four sections:

- According to the starting data: When all the data tables contain information on the same variables considered on the same individuals, in different times or situations, the X-STATIS or Partial Triadic Analysis (PTA) is used. In this case we work with data matrices and with operators. If the operators that we integrate are matrices of covariances taken on the same individuals, it is known as COVSTATIS, and if it starts from several matrices of defined distances on the same individuals, it is called DISTATIS and can be understood as a three-way MDS.
- According to the weights assigned when creating the consensus matrix: BENASSENI and BENNANI-DOSSE, 2012 call Power-Statis a version of STATIS in which the weight given to each matrix differs from the proposal of L' HERMIER in which the weight is related to the components of the first eigenvector of the vector correlation matrix.
- If external information is taken into account: $t + 1$ STATIS consider the matrices that are integrated in the STATIS and another matrix with external information about the individuals. In 2008, SABATIER and VIVEN, extend the $t + 1$ STATIS and propose the STATIS-4. For the case in which the individuals have structure of groups, VALLEJO-ARBOLEDA et al in 2007 develop the CANOSTATIS.
- If pairs of tables are available in t different situations or times: When pairs of tables are available, STATICO (STATIS and COINERTIA) and more recently COSTATIS (COINERTIA Y STATIS), are interesting alternatives that allow the study of co-structures.

STATIS limitations

The STATIS declares highly related those matrices that have a common first principal address. It does not capture multidimensional structures, or similar structures in which the first main address does not match. The STATIS and related techniques are only able to capture the stable part of the structure. In fact, the matrices must be similar so that it makes sense to create the consensus matrix. To solve the limitations of STATIS, there are alternative techniques in the literature. To capture multidimensional structures you can use the AFM.

Keywords: biplot, three-way data, STATIS, MFA, multibiplot.

Mathematics Subject Classification (2010): 62H25, 62H86

References

[1] Escoufier, B and Pagès, J (1988). *Analyses factorielles multiples*, Dunod, Paris.

[2] Gabriel, K.R. (1971). "The biplot graphic display of matrices with application to principal component analysis", *Biometrika n^o 58(3)*, p. 453–467.

[3] Galindo-Villardón, P. (1986). "Una alternativa de representacion simultanea: HJ-Biplot", *Qüestiió n^o 10(1)*, p. 13–23.

[4] L'Hermier des Plantes, H. (1976). "Structuration des Tableaux à Trois Indices de la Statistique", *These de Troisieme cycle. University of Montpellier*.

Método moderno de optimización combinatoria apoyado con el software Wolfram Mathematica[I]

Communication / Ponencia

VÍLCHEZ QUESADA, ENRIQUE[II] López, Erasmo[III]

Costa Rica

We present the results obtained from an algorithm based on the taboo search that was programmed using the commercial software Wolfram Mathematica. In Wolfram Language, different implementations of random instances and others available in the TSPLIB were made, comparing them later, with the results provided with the same algorithm, in the Visual Basic 6.0 programming environment. The improvements that were obtained are due to the structuring of predesigned functions that allowed specifically analyzing two aspects: the optimization of the solution and its exploration in the neighborhoods where the presence of the optimum was already known. For this we focus on developing an oscillation in the taboo matrix analogously to what is applied to solutions where it is perceived that the global optimum could be. Finally, conclusive results are shown that allow observing the good performance of the Wolfram Mathematica program to deal with this type of problems, through the proper structuring of its internal functions.

Keywords: Commercial software, tabu search, oscillation, internal functions, traveling salesman problem.

Resumen

En este trabajo se presentan los resultados obtenidos de un algoritmo basado en la búsqueda tabú que fue programado utilizando el software comercial Wolfram Mathematica. En Wolfram Language se realizaron distintas implementaciones de instancias aleatorias y otras disponibles en la TSPLIB, comparándolas posteriormente, con los resultados provistos del mismo algoritmo, en el ambiente de programación Visual Basic 6.0. Las mejoras que se obtuvieron obedecen a la estructuración de funciones prediseñadas que permitieron analizar específicamente dos aspectos: la optimización de la solución y su exploración en las vecindades donde ya se conocía la presencia del óptimo. Para ello nos centramos, en desarrollar una oscilación en la matriz tabú de manera análoga a lo que se aplica a las soluciones en donde se percibe que podría estar el óptimo global. Finalmente, se muestran resultados concluyentes que permiten observar el buen desempeño del programa Wolfram Mathematica para tratar este tipo de problemas, mediante la estructuración adecuada de sus funciones internas.

Palabras clave: Software comercial, búsqueda tabú, oscilación, funciones internas, problema del agente viajero.

References

[1] De los Cobos, S., Goddard, J., Gutiérrez, M. y Martínez, A. (2010). Búsqueda y Exploración Estocástica. Mexico: Universidad Autónoma Metropolitana.

[I] Wednesday/Miércoles, 3:15p.m., Room/aula 2, Session: Opt1, Optimization / Optimizacion 1

[II] Universidad Nacional de Costa Rica, Escuela de Informática, Costa Rica, Heredia, evq1529@gmail.com

[III] Colegio Técnico Profesional de Upala, Upala, Costa Rica, erasmolopezlopez@gmail.com

[2] López, E., Salas, O. y Murillo, A. (2014). El Agente Viajero: Un Algoritmo Determinístico, Revista de Matemática: Teoría y Aplicaciones, 21(1), 127-144.

[3] Vílchez, E. (2012). Álgebra lineal apoyada con mathematica. Costa Rica: Editorial Técnológica de Costa Rica.

[4] Wolfram Mathematica 11: Documentation Center. (2017). Mathematica functions and tutorials. Recuperado de: http://reference.wolfram.com/mathematica/guide/Mathematica.html

Tests for the Inverse Gaussian distribution hypothesis[I]

Communication / Ponencia

VILLASEÑOR, JOSÉ A.[II] González-Estrada, Elizabeth[III]

México

The family of Inverse Gaussian (IG) distributions has applications in areas such as hydrology, lifetime testing and reliability, among others. In this talk, a new characterization for this family of distributions is presented, and is used to propose a bootstrap test of fit for the IG distribution hypothesis with unknown parameters. As a second test, observations are transformed to normal variables and then Shapiro-Wilk test is used to test for normality. Simulation results show that the proposed tests preserve the nominal test size and are competitive against some existing tests for the same problem. Real datasets are used to illustrate the application of these tests.

Keywords: goodness of fit, hypothesis testing, data transformations, parametric bootstrap.

References

[1] Villaseñor, J.A., GonzálezEstrada, E. 2015. Tests of fit for inverse gaussian distributions. Statist Probab Lett. 105:189-194.

[I] Thursday/Jueves, 1:50p.m., Room/aula 2, Session: Mod3, Modeling / Modelacion 3

[II] Colegio de Postgraduados, Posgrado en Estadística, México, Texcoco, jvillasr@colpos.mx

[III] Colegio de Postgraduados, México, egonzalez@colpos.mx

Estimation of the exchange rate between Costa Rica and the United States via Levy processes[I]

Poster / Cartel

VÍQUEZ BOLAÑOS, JORGE[II]

Costa Rica

According to Tankov & Cont financial series posses characteristics that are better explained by models with jumps than those who only have a diffusion component[1]. In this work we apply the Kou and temperated stable models in order to predict the behavior of the exchange rate between Costa Rican Colones and USA Dollars from july 1st 2015 to june 30th 2017. To do so, robust estimators are used to verify that the studied series has no diffusion component [2], then, using least squares, the fourier transform wil be fitted to the Levy exponent determined by the models. All this while keeping the Martingale property expected by the theoretical behavior of the series [3]. Finally, the models' goodness of fit and predictive capabilities are tested agains an ARIMA fitted from the same data. [4].

Keywords: Probability, Stochastic Processes, Time Series, Exchange Rate.

Resumen

Palabras clave:

References

[1] Edward W. Frees. Regression Modeling with Actuarial and Financial Applications. International Series on Actuarial Science. Cambridge University Press, 2009.

[2] Helyette Geman, Nicole E. Karoui, and Jean C. Rochet. Changes of Numeraire, Changes of probability measure and Option Pricing. Journal of Applied Probaility, 32:443–458, 1995.

[3] Zhaoxia Ren. Estimation and Testing of the Jump Component in Levy Processes. PhD thesis, 2013.

[4] P. Tankov and R. Cont. Financial Modelling with Jump Processes, Second Edition. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis, 2015.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad de Costa Rica, Matemáticas, San Joaquin, Costa Rica, javiquez42@gmail.com

Algunos algoritmos de descomposición de dominios BDDC Deluxe[I]

Plenary Talk / Conferencia Plenaria

WIDLUND, OLOF[II]

United States

When designing domain decomposition algorithms, approximate inverses, also known as preconditioners, are constructed for very large matrices by using solvers for many smaller linear systems often obtained from much smaller instances of the given problem. The problems considered often arise in continuum mechanics, e.g., in linear elasticity or electro-magnetics. The preconditioners are used to increase the rate of convergence of conjugate gradient iterations. For fast convergence with a rate of convergence independent of the number of local problems, a coarse component of the preconditioner will be needed and these components are at the heart of most domain decomposition algorithms. While the local components relatively easily can be distributed to individual processors, the coarse, global problem require careful considerations in order to minimize the communication and to provide rapid convergence of the iterations.

The BDDC algorithms, first developed by Clark Dohrmann, have proven to be very successful domain decomposition algorithms for a variety of elliptic problems. For any particular application, the success of such an algorithm depends on the choice of a set of primal constraints and the choice of an averaging operator, which is used to restore the continuity of certain intermediate vectors in each iteration.

In the deluxe version, a new averaging procedure is used; it was first developed in joint work with Dohrmann on H(curl) problems. A theory will be outlined and a variety of successful applications will be discussed in particular to problems formulated in H(div) and H(curl). This work has been developed jointly with Clark Dohrmann and Duk-Soon Oh. These algorithms have been carefully implemented by Stefano Zampini for massively parallel computing systems. This work also involves the adaptive selection of the primal constraints and exploring three-level BDDC algorithms for problems with billions of degrees of freedom. These programs are made available to all thorugh the PETSc system headquartered at the Argonne National Laboratory.

The BDDC deluxe algorithms have also proven very effective for solving the large and often very ill-conditioned system of algebraic equations arising in isogeometric analysis. Current work, with Pavarino, Scacchi, and Zampini is focused on compressible and almost incompressible elasticity.

Keywords:

Domain decomposition, BDDC methods, deluxe scaling, adaptive primal constraints, isogeometric analysis.

[I] Thursday/Jueves, 11:20a.m., Auditorium, Session: CONF, Conference 5

[II] Courant Institute, Nueva York, Courant Institute of Mathematical Sciences, widlund@cims.nyu.edu

A Brief Survey of Higgs Bundles[I]por favor llenar esto

(Un estudio conciso de fibrados de Higgs)

Communication / Ponencia

ZÚÑIGA ROJAS, RONALD ALBERTO[II]

Costa Rica

Considering a compact Riemann surface of genus greater than two, a Higgs bundle is a pair composed of a holomorphic bundle over the Riemann surface, joint with an auxiliar vector field, so-called Higgs field. This theory started around thirty years ago, with Hitchin's work, when he reduced the self-duality equations from dimension four to dimension two, and so, studied those equations over Riemann surfaces. Hitchin baptized those fields as *"Higgs fields"* beacuse in the context of physics and gauge theory, they describe similar particles to those described by the Higgs bosson. Later, Simpson used the name *"Higgs bundle"* for a holomorphic bundle together with a Higgs field. Today, Higgs bundles are the subject of research in several areas such as non-abelian Hodge theory, Langlands, mirror symmetry, integrable systems, quantum field theory (QFT), among others. The main purposes here are to introduce these objects, and to present a brief but complete construction of the moduli space of Higgs bundles, and some of its stratifications.

Keywords: Moduli of Higgs Bundles, Variations of Hodge Structures, Vector Bundles.

Mathematics Subject Classification (2010): 14H60, 14D07, 55Q52.

Resumen

Considerando una superficie compacta de Riemann de género mayor que dos, un fibrado de Higgs es un par compuesto por un fibrado holomorfo sobre la superficie de Riemann, junto con un campo vectorial auxiliar, llamado campo de Higgs. Esta teoría inició hace unos treinta años, con el trabajo de Hitchin, cuando él reduce las ecuaciones de autodualidad de dimensión cuatro a dimensión dos, y así, estudiar esas ecuaciones sobre superficies de Riemann. Hitchin bautizó esos campos como *"campos de Higgs"* pues en el contexto de la física y de la teoría de gauge, describen partículas similares a las descritas por el bozón de Higgs. Más tarde, Simpson usó el nombre *"fibrado de Higgs"* para un fibrado holomorfo junto con un campo de Higgs. Hoy, los fibrados de Higgs son objeto de investigación en varias áreas tales como la teoría de Hodge no abeliana, Langlands, simetría de espejo, sistemas integrables, teoría cuántica de campos (QFT), entre otros. Los propósitos principales aquí son introducir estos objetos y presentar una breve pero completa construcción del espacio móduli de los fibrados de Higgs y algunas de sus estratificaciones.

Palabras clave: Móduli de fibrados de Higgs, Variaciones de estructura de Hodge, Fibrados vectoriales.

[I] Friday/Viernes, 1:50p.m., Room/aula 2, Session: DJ, Simulacion Multiagente: Herramientas y Aplicaciones

[II] Universidad de Costa Rica, Escuela de Matemática, , Costa Rica, ronald.zunigarojas@ucr.ac.cr

References

[1] M. F. Atiyah and R. Bott, *"Yang–Mills equations over Riemann surfaces"*, Phil. Trans. Roy. Soc. London A **308** (1982), 523–615.

[2] S.B. Bradlow, O. García-Prada and P.B. Gothen, *"What is a Higgs Bundle?"*, Notices of the AMS **54**, No. 8 (2007), 980–981.

[3] P. B. Gothen and R. A. Zúñiga-Rojas, *"Stratifications on the moduli space of Higgs bundles"*, Portugaliae Mathematica, **74**, (2017), 127–148.

[4] N. J. Hitchin, *"The self-duality equations on a Riemann surface"*, Proc. London Math. Soc. **55** (1987), 59–126.

[5] C. T. Simpson, *"Higgs bundles and local systems"*, Publ. Math. IHÉS **?** (1992), 5–95.

Abstracts

III LACSC

III Latin American Conference on Statistical Computing

Comparación entre el análisis de correspondencias y el análisis de correspondencias no simétrico para variables ordinales

Poster / Cartel

ACOSTA BARRADAS, ADRIANA[I] Del Callejo–Canal, Diana[II]
Canal-Martínez, Margarita[III]

México

The multivariate statistical analysis is an important branch of statistics, whose main function is to analyze, describe and interpret the observations that come from many variables, which act simultaneously to represent a phenomenon through a set of statistical and mathematical methods.

Multivariate analysis is a necessary tool for current research and especially studies related to certain social areas.

One of the best known techniques used to study the relationship of categorical variables of multivariate statistical methods is the correspondence analysis, "is an effective method to analyze the contingency tables with numerical frequency data, since it provides a graphic representation elegant and simple that allows a quick interpretation and understanding of the study data" (Greenacre, 2008).

A disadvantage of the correspondence analysis technique is that it does not distinguish symmetric from non-symmetric relationships, therefore it does not take into account the dependency ratio (if it exists) between the categories of a response variable and a predictor variable, another important disadvantage. that presents the technique is that the categorical variables are nominal or ordinal works with the same process, and does not emphasize for response variables that have ordered categories and leaves aside the efficient information that these variables provide (Agresti, 1996).

The technique of non–symmetric correspondence analysis doubly ordinal has as a gain that takes into account the dependency ratio and adds the importance of the structure of the ordinal variables in a table of two inputs and helps the identification of dependence between the variables (Lombardo, Beh, & D´Ambra , 2007).

So the research question to be solved is the following: Is the methodology of the DONSCA technique a better tool than the classic CA technique for the additional information it presents?

Therefore, the main objective is to compare the efficiency and effectiveness of the non-symmetric correspondence analysis method for ordinal categorical variables and to define the scope thereof in comparison with the classical correspondence analysis technique. The comparison was made with practically performing both analyzes with the statistical software R-Project version 3.4.2, using 3 real databases obtained from the page of the National Institute of Statistics and Geography (INEGI for its acronym in Spanish).

When put into practice with the application of 3 real databases, it is concluded that in 2 of the 3 of the applications, the DONSCA technique proved to be more efficient than the orthodox CA methodology in terms of the representation of reality.

For these reasons, it is concluded that according to what has been studied in this work; both the theory and practice of DONSCA is functional for the characteristics of a cross-table with nonsymmetric ordinal variables,

[I]Universidad Veracruzana, Facultad de Estadística e Informática, México, Xalapa, aab_1993@hotmail.com
[II]Instituto de Investigaciones de Estudios Superiores, Económicos y Sociales. Universidad Veracruzana, Xalapa, Veracruz, México, ddelcallejo@uv.mx
[III]Instituto de Investigaciones de Estudios Superiores, Económicos y Sociales. Universidad Veracruzana, Xalapa, Veracruz, México

that is, that the DONSCA methodology is a better tool than the classical technique of correspondence analysis when there is additional information on symmetry. of the variables and both variables are in ordinal scale.

Keywords:
Non-symmetric correspondence analysis, Ordinal variables, Bivariate moment decomposition, Orthogonal polynomial.

Resumen

El análisis estadístico multivariante es una rama importante de la estadística, cuya principal función es analizar, describir e interpretar las observaciones que provienen de muchas variables, las cuales actúan simultáneamente para representar un fenómeno a través de un conjunto de métodos estadísticos y matemáticos. El análisis multivariante es una herramienta necesaria para la investigación actual y sobre todo estudios referentes a ciertas áreas sociales.

Una de las técnicas más conocidas que sirve para estudiar la relación de variables categóricas de los métodos estadísticos multivariantes es el análisis de correspondencias, "es un método eficaz para analizar las tablas de contingencia con datos de frecuencias numéricas, ya que nos proporciona una representación gráfica elegante y simple que permite una rápida interpretación y comprensión de los datos de estudio´´ (Greenacre, 2008).

Una desventaja de la técnica de análisis de correspondencias es que no distingue relaciones simétricas de las no simétricas, por consiguiente, no toma en cuenta la relación de dependencia (si existe) entre las categorías de una variable respuesta y una variable predictora, otra desventaja importante que presenta la técnica es que las variables categóricas ya sean nominales u ordinales las trabaja con el mismo proceso, y no hace énfasis para las variables de respuesta que tienen categorías ordenadas y deja de lado la información eficiente que estas variables proporcionan (Agresti, 1996).

La técnica análisis de correspondencias no simétrico doblemente ordinal tiene como ganancia que toma en cuenta la relación de dependencia y agrega la importancia de la estructura de las variables ordinales en una tabla de dos entradas y ayuda a la identificación de dependencia entre las variables (Lombardo, Beh, y D'Ambra , 2007).

Por lo que la pregunta de investigación a resolver es la siguiente ¿La metodología de la técnica DONSCA es mejor herramienta que la técnica clásica de CA por la información adicional que presenta?

Por tanto, el objetivo principal es comparar la eficiencia y eficacia del método de análisis de correspondencias no simétrico para variables categóricas ordinales y definir los alcances de la misma en comparación de la técnica clásica del análisis de correspondencias.

La comparación se realizó con Realizar de manera práctica ambos análisis con el software estadístico R-Project versión 3.4.2, utilizando 3 bases de datos reales obtenidos de la página de El Instituto Nacional de Estadística y Geografía (INEGI).

Al llevarlo a la práctica con la aplicación de 3 bases de datos reales se concluye que en 2 de las 3 de las aplicaciones, la técnica DONSCA resulto ser más eficiente que la metodología ortodoxa del CA en cuestión de la representación de la realidad.

Por tales motivos se concluye que de acuerdo con lo estudiado en este trabajo; tanto la teoría como la práctica del DONSCA resulta funcional para las características de una tabla cruzada con variables ordinales no simétricas, es decir, que la metodología DONSCA es una mejor herramienta que la técnica clásica de análisis de correspondencias cuando se tiene información adicional sobre la simetría de las variables y ambas variables se encuentran en escala ordinal.

Palabras clave: Análisis de correspondencia no simétrica, variables ordinales, descomposición de momento bivariado, polinomio ortogonal.

References

[1] Sociology (Second edi, Vol. 22). Gainesville, Florida. https://doi.org/10.1198/jasa.2008.s251.

[2] Greenacre, M. J. (2008). La práctica del análisis de correspondencias (Primera ed). España: Rubes Editorial, Fundación BBVA. Retrieved from www.fbbva.es.

[3] Lombardo, R., Beh, E. J., & D´Ambra, L. (2007). Non–symmetric correspondence analysis with ordinal variables using orthogonal polynomials. Computational Statistics & Data Analysis, 52(1), 566–577. https://doi.org/10.1016/j.csda.2006.12.040

A new approach to estimate the spatial impact of climate change on metaxenic disease transmission[I]

Communication / Ponencia

AGUILAR, VÍCTOR HUGO[II] Córdova, Karenia[III] Delgado, Petrocelli Laura [IV]
Echarry, Exyeleth [V] Márquez, Ingrid [VI] Ramos, Santiago[VII]

Venezuela

Summary Climate change should be considered as emergent phenomena with an inequitable distribution on Earth. It expresses progressively and in a long-term fashion. It makes evident itself by means of a statistically significant change of weather average conditions, changes in the distribution of climate around its annual average values, as well as the occurrence of extreme and severe climatic events, like hurricanes, droughts, and inundations. To achieve minimization of its negative impacts, in particular, those due to widespread of metaxenic diseases. We developed a spatial model to represent malaria widespread, starting by the use of climatic data from public WorldClim Database with a spatial resolution grille of about 1 km2 surface, corresponding to the output from models Hadgem-2 and Miroc-5. Then, the Maxent program was used to obtain a valid baseline of the presence of malaria disease, derived from historical malaria records in municipalities of Sucre and Bolivar states, the Venezuelan's two mayor malaria western focus. From there, we estimated probability values as malaria risk forecasting for a medium RCP4.5and severe RCP 8.5 climatic scenarios for years 2050 and 2070. Results were expressed in a thematic map series discussed in terms of transmission chance for malaria in western Venezuela region.

Keywords: Keywords: Climatic change, Metaxenic disease, Disease transmission, Spatial Model, risk forecasting.

References

[1] INAMEH (2017) Escenarios climáticos. Proyecto: Segunda Comunicación Nacional en Cambio Climático de la República Bolivariana de Venezuela. Caracas, Venezuela. 29pp.

[2] Rodríguez. D., Delgado. L., Ramos. S., Weinberger. V. & Rangel. Y. (2013). A model for the dynamics of malaria in Paria Peninsula. Sucre State, Venezuela. Ecol. Modelling 259: 1-9.

[3] Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259.

[I] Friday/Viernes, 2:10p.m., Room/aula 1, Session: LACSC–2, LACSC–Statistical Computing 5

[II] Universidad Central de Venezuela, Facultad de Humanidades,Inst. de Geografía, Venezuela, Caracas, vha1999@gmail.com
[III] same address
[IV] same address
[V] same address
[VI] same address
[VII] same address

Application of classification techniques in the study of student's success in propauedeutic college math courses[I]

Communication / Ponencia

ARGUEDAS MÉNDEZ, SILVIA MARÍA[II] Fernández Corrales, Carla[III]

Costa Rica

High failure rates in the propaedeutic mathematics courses that students take during their first year of college have a significant impact in many areas such as student satisfaction, student retention and even university finances due to the economic expense incurred by the opening of courses for repeaters (Gómez, OP´T Eynde and De Corte, 2006). This phenomenon has been studied extensively by the scientific community, which has identified several factors associated with academic performance in these courses. Despite this vast research, we still have much to learn about the determinants of success in first-year college math courses. The high complexity of the phenomenon, the presence of various contingency factors and the entangled interaction of factors supported by many different theories are some of the challenges scholars face while trying to understand it. At the same time, the growing availability of data, the accessibility of algorithms to analyze them and the availability of computational power to carry out the calculations, make data mining a suitable method to discover patterns that give us better insights about the success and failure in first-year college courses. In this paper, we use decision trees (Quinlan, 1986), a machine learning technique for classification, in the exploration of relationships between several factors and their influence on the success or failure of a propaedeutic mathematics course. Some of the advantages of decision trees over other classification techniques are their adaptability to various problems, high automation that allows handling numerical, categorical variables and even lost data, the possibility of excluding non-relevant factors, adaptability to large and small datasets and ease of interpretation (Han, Kamber, & Pei, 2011). The selection of the factors included in the model was supported by a strong theoretical basis. Both cognitive and non-cognitive variables were included in our analysis. Cognitive variables are mainly related to intelligence and variables such as socio-cognitive, demographic, gender, personality, and so on, are included among non-cognitive. Fluid intelligence is a factor associated with achievement in propaedeutic courses. Fluid intelligence is the innate ability that a person has to think and reason in an abstract way. Fluid intelligence relates to the ability to cope with new ingenuity and unusual problems; whereas the ability to apply previously acquired knowledge to different problems is linked to crystallized intelligence (Hunt, 2011). According to Garbanzo (2007), pre-college academic training is considered a clear indicator of the academic performance of college students. Some of the measures of previous academic training correspond to the grades obtained in secondary education, and even the comprehensive tests done by the Ministry of Public Education at the end of high school. For the purposes of this study, we included the score in the comprehensive math test as a measure of crystallized intelligence. There are also institutional determinants (Garbanzo, 2007, Montero, Villalobos and Valverde, 2007) that are associated with academic performance, these are non-personal components that intervene in the educational process and that may influence the success in the initial courses of mathematics. Therefore, we include the type of school (private, public or semi-public) as a proxy of these determinants. We

[I] Wednesday/Miércoles, 1:30p.m., Room/aula 1, Session: L3–3, LACSC–Application of classification techniques in the study of student's success in propauedeutic college math courses

[II] Universidad de Costa Rica, Escuela de Ingeniería Industrial, Costa Rica, Heredia, smarguedas@gmail.com

[III] Universidad de Costa Rica, Costa Rica, carla.fernandez@ucr.ac.cr

also consider gender and verbal reasoning score at a college admission standardized test. Once we select our variables, for tree induction we use the Weka implementation of algorithm C4.5 (Quinlan, 1993). We perform post-pruning and pre-pruning by limiting the number of instances per leaf, in order to avoid overfitting. To evaluate the performance of the models obtained by adjusting the different parameters, we use stratified cross-validation, which has been shown to reduce the bias in the calculation of performance metrics (Kohavi, 1995). Our preliminary results (Figure 1) discover interesting relationships between the factors such as the role of gender in academic performance in the introductory math course. According to Montero and Villalobos (2004) there is empirical evidence of a significant relationship between gender and GPA; additionally, there are studies concluding that women present a superior performance than men (Garbanzo, 2007). Similarly, Barahona (2014) states that there are studies that indicate a statistically positive relationship between the variable female sex with academic performance, explained by women showing a greater commitment than men and spending more hours studying. The application of machine learning techniques has both practical and theoretical contributions. First, a better understanding of the factors influencing success in math courses can generate new research focused on the patterns uncovered. Second, those patterns could also provide insights to university authorities and college professors about initiatives to reduce failure in these courses.

Keywords: Techniques of classification, data mining, academic performance, machine learning.

References

[1] Barahona, P. (2014). Factores determinantes del rendimiento académico de los estudiantes de la Universidad de Atacama. Revista Estudios pedagógicos, 40 (1).

[2] Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública Educación. Revista Educación, 31 (1), pp. 43-63.

[3] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques, Third Edition. Morgan Kaufmann; 3 Edition.

[4] Kohavi, R. (1995). A study of cross–validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th Joint International Conference on Artificial Intelligence (IJCAI´95) (pp. 1137–1143). Montreal.

[5] Montero, E., Villalobos, J. y Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: un análisis multinivel, RELIEVE, 13 (2).

Computational Tools for Generation of Cost-effective Run Orders Useful for Agricultural and Industrial Research[I]

Communication / Ponencia

BHOWMIK, ARPAN[II] Varghese, Eldho[III] Jaggi, Seema[IV] Varghese, Cini [V]
Lall, Shwetank [VI]

India

Factorial experiments, wherein two or more factors each at two or more levels are used simultaneously, have profound applications in many fields of agricultural and allied sciences. These experiments allow studying the effect of each individual factor as well as the effects of interactions between factors on the response variable. In order to avoid any kind of bias in the estimation of these effects, it is always advisable that the order of execution of runs in a factorial design is random. However, experimentation under factorial setup may become expensive, time-consuming and difficult due to a large number of changes in factor levels induced by randomization. The number of level changes is a serious issue to experimenters in many agricultural and allied experiments that involve hard-to-change factors. In such situations, an effective solution is to adopt factorial run orders with minimum number of changes in the factor levels which will make the experiment less expensive both in terms of time and cost. Here, methodologies for obtaining minimally changed run sequences for factorial designs have been discussed. For potential use of these minimally changed run orders, an online software for generation of factorial designs with minimum number of changes in the factor levels has been developed. This paper describes the online computational tool developed for generation of such cost-effective minimally changed run orders called webFMC (http://webfmc.iasri.res.in) using client–server architecture along with an online catalogue of the designs within a permissible range of parameters. webFMC is accessible any time from arbitrary platforms through internet. This software provides a freely available solution for the researchers and students working in this area. This paper also describes an alternative solution for generation of such cost-effective minimally changed run orders in the form of an R-package known as FMC: Factorial Experiments with Minimum Level Changes (https://cran.r-project.org/package=FMC). Methodology has been highlighted to develop half replicate of a full factorial designs with 2-factors along with the generation of such run orders through a SAS macro which has been developed for the purpose. In experiments with multiple factors, sometime we may come across some situations where experiments may involve quantitative factors and the purpose of the experimenter is to identify the combinations of such factors which can optimize the response. Response surface designs can be used for such situations. Keeping in view the importance of hard-to-change factors, different computational tools like an R-package namely minimalRSD: Minimally Changed CCD and BBD (https://cran.r-project.org/package=minimalRSD) along with SAS macros which were developed for the generation of minimally changed run orders for response surface designs based on certain methodologies have

[I] Friday/Viernes, 1:30p.m., Room/aula 1, Session: LACSC–2, LACSC–Statistical Computing 5

[II] ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India, Design of Experiments, India, New Delhi, arpan.stat@gmail.com

[III] ICAR-Central Marine Fisheries Research Institute, Kochi-682 018, India, eldhoiasri@gmail.com

[IV] ICAR-Indian Agricultural Statistics Research institute, New Delhi-110 012, India, seema.jaggi@icar.gov.in

[V] ICAR-Indian Agricultural Statistics Research institute, New Delhi-110 012, India, cini2204@gmail.com (

[VI] ICAR-Indian Agricultural Statistics Research institute, New Delhi-110 012, India, shwetanklall@gmail.com

also been highlighted.

Keywords: Hard-to-change factors, Minimum level change, webFMC, FMC, minimalRSD.

References

[1] Bhowmik, A., Varghese, E., Jaggi, S., Varghese, C (2015). Factorial experiments with minimum changes in run sequences, Journal of the Indian Society of Agricultural Statistics, no 69 (3), p. 243-255.

[2] Bhowmik, A., Varghese, E., Jaggi, S., Varghese, C (2016). Minimally changed run sequences in factorial experiments, Communications in Statistics – Theory and Methods, no 46(15), p. 7444–7459.

[3] Varghese, E., Bhowmik, A., Jaggi, S., Varghese, C., Lall, S (2016). On the construction of response surface designs with minimum level changes. Utilitas Mathematica (in press).

[4] Varghese, E., Bhowmik, A., Jaggi, S., Varghese, C. and Kaur, C (2017). On the generation of cost effective response surface designs, Computers and Electronics in Agriculture, no 133, p. 37–45.

Parallel Algorithm for a Test of Goodness of Fit of a Rash Type Models[I]

Communication / Ponencia

CARVAJAL SCHIAFFINO, RUBEN[II]

Chile

Parallel computing is a programming technique where instructions can be executed simultaneously by different processors. This technique is based in the principle that big problems can be divided into smaller parts that can resolve concurrently.

On the other hand, the Item Response Theory (IRT) it refers to the class of latent trait models for discrete multivariate data obtained from the responses to a set of items of questionnaire /tests intended to measure educational attainment, attitudes, etc. (Reise S. and Revicki D., 2015). Rash-type models are a subset of IRT models, named after the pioneering work of Rash (1960).

Once an IRT model has been adjusted, it is necessary to quantify the discrepancy between the model and the data, that is, the absolute goodness of the fit of the Maydeu-Olivares model (2013); Koehler, K and Larntz, K. (1980).

An index of goodness of fit (GOF) summarizes the discrepancy between the values observed in the data and the expected values in a statistical model.A goodness-of-fit statistic is a GOF index with a known sampling distribution, the difficulty of these indices is that gigantic contingency tables are often applied (for example, 20 items are tested $2^{20} = 1048576$ cells, to comply with the general rule > 5, you need at least one sample size 1048576×5).

The R_1 test was proposed by Glas (1988) to evaluate the goodness of fit of the Rasch models, it is based on the comparison of the observed and expected first order frequencies, that is, the frequencies of the correct answers in subgroups which are homogeneous with respect to the level of punctuation. This is based mainly on a matrix $\mathbf{X}$ of diagonal blocks of dimension $k(k-2) + 2 \times 2^k$, with k equal to the number of items in the questionnaire, so that the columns of each block contain all the patterns of possible answers whose total score is r, so each block corresponds to a submatrix $\mathbf{X_r}$ of dimension $k \times \begin{pmatrix} k \\ r \end{pmatrix}$ with $r = 1, \ldots, k-1$ where x_0 and x_k take value 1.

The statistic in matrix form is expressed by

$$R_1 = N(\mathbf{p} - \pi(\widehat{\mathbf{\Theta}}))' \mathbf{X}' (\mathbf{X} \widehat{\mathbf{D}} \mathbf{X}')^{-1} \mathbf{X} (p - \pi(\widehat{\mathbf{\Theta}}))$$

[I] Tuesday/Martes, 2:35p.m., Room/aula 1, Session: L3–2, LACSC–Recent advances in Statistical Computing 1

[II] Universidad de Santiago de Chile, Departamento de Matematica y Ciencia de la Computacion, Chile, Santiago de Chile, ruben.carvajal@usach.cl

Or through blocks

$$R_1 \;=\; N\left(\frac{(p_0 \;-\; \pi(\widehat{\Theta_0}))^2}{\pi(\widehat{\Theta_0})} \;+\; \sum_{r=1}^{k-1} \widehat{\mathbf{e}}'\,\mathbf{X}_\mathbf{r}'\,(\mathbf{X}_\mathbf{r}\,\widehat{\mathbf{D}}\,\mathbf{X}_\mathbf{r}')^{-\mathbf{1}}\,\mathbf{X}_\mathbf{r}\,\widehat{\mathbf{e}} \;+\; \frac{(p_k \;-\; \pi\,(\widehat{\theta_k}))^2}{\pi\,(\widehat{\theta_k})}\right)$$

This work compares the efficiency in time and capacity of different ways of estimating the R_1 statistic, as will be shown in the results. As it is observed the calculation of the statistic R_1 contemplates the inverse of a matrix corresponding to the quadratic form. This matrix has the property of being formed by a certain number of blocks that form square submatrices. The result R_1 can be obtained by inverting the complete matrix of order $k(k-1)+2$ with cost of order $\mathcal{O}(k^6)$ or inverting each of the submatrices with a cost of order $\mathcal{O}(k^4)$. In addition, the calculation of the inverse of the submatrices can be done in parallel, which further decreases the execution time and increases the number of items for which this statistic can be calculated.

Thanks: Proyecto Dicyt, código 041733ME, Vicerrectoría de Investigación, Desarrollo e Innovación de la Universidad de Santiago de Chile.

Keywords: Item Response Theory, Parallel Computing.

References

[1] Glas, C.A.W. (1988). "The Derivation of Some Tests for the Rash Model From the Multinomial Distribution", *Psychometrika 53*, p. 525-546.

[2] Maydeu-Olivares, A. (2013). "Goodness-of-fit Assesment of Item Response Theory Models". *Measurement: Interdisciplinary Research and Perspectives 11*, p. 71-101.

[3] Rash, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Test*. Paedagogiske Institut, Copenhagen.

[4] Koehler, K. and Larntz, K. (1980). "An Empirical Investigation of Goodness of Fit Statistics for Sparse Multidimensional Tables". *Journal of the American Statistical Association 75*, p. 336-344.

[5] Reise S. and Revicki D. (2015). *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*. Routledge Editors.

A Parallel / Distributed Algorithm for Bootstrapping[I]

Invited Session / Sesion invitada

CARVAJAL SCHIAFFINO, RUBÉN[II] Firinguetti, Luis [III]

Chile

Bootstrapping is a statistical technique that is very expensive from a computational point of view. For this reason, parallel versions have been developed, tipically implemented on multicore architectures. In the parallel version of the algorithm, the iterations for random sample selection are performed by the processors. We present a technique in which the random sample generation is parallelized, allowing the implementation of a distributed version, where the data are distributed among several computers with many processing cores. The distributed version, also allows to perform bootstraping with data sizes that exceed the memory storage capacity of one computer.

Keywords: Parallel algorithms, bootstrapping.

[I] Tuesday/Martes, 4:40p.m., Room/aula 1, Session: L4–1, LACSC–Statistical Computing for Data Science

[II] Universidad de Santiago de Chile, , Chile, Santiago, ruben.carvajal@usach.cl

[III] Departamento de Estadística, Universidad del Bío-Bío, Concepción, Chile, lfiringu@ubiobio.cl

Supervisando una encuesta con Google[I]

Communication / Ponencia

CENTENO MORA, OSCAR[II]

Costa Rica

The application of a telephone survey requires effective control of both the interviewers and the multiple responses, in order to ensure the quality of the information. The Google applications allows everything. This summary exposes the power of Google´s applications in the direct supervision of a survey, emphasizing the execution and administration of the information: data collection and storage, statistics of the general and specific performance of the interviewers, descriptive analysis of the answers, among others. It´s exposed the application for the National Corruption Prevention Survey of 2017 (ENPC – 2017), made by the Government Accountability Office of Costa Rica. The ENPC – 2017 consulted 3 types of populations, applying two questionnaires for citizenship, a questionnaire for public workers, and a questionnaire for the public sector providers. Using the Google sites, the entire administration and monitoring was focused on the ENPC – 2017: interviewers' access to telephone numbers and the different questionnaires, answer sheets, descriptive statistics of the answers and the interviewers' performance. The entire application process was online and with immediate analysis responses.

Keywords: survey, Google applications, statistics, supervision.

Resumen

La aplicación de una encuesta telefónica requiere el efectivo control tanto de los encuestadores como de las múltiples respuestas, con el fin de asegurar la calidad de la información. Las diversas aplicaciones de Google permiten todo lo anterior.

El presente resumen expone la potencia de las aplicaciones de Google en la supervisión directa de una encuesta, enfatizando la ejecución y administración de la información: herramienta de recolección, almacenamiento de los datos, estadísticas del rendimiento general y específico de los entrevistadores, análisis descriptivos de las respuestas, nivel de alcance, entre otros. Se expone la aplicación para la Encuesta Nacional de Prevención de la Corrupción del 2017 (ENPC – 2017), realizada por la Contraloría General de la República. La ENPC –2017 consultó a un total de 3 poblaciones, aplicando dos cuestionarios para la ciudadanía, un cuestionario para los funcionarios públicos, y un cuestionario para los oferentes de la administración. Mediante la utilización de un Google sites se concentró toda la administración la ENPC – 2017: acceso de los entrevistadores tanto a los números telefónicos como a los diferentes cuestionarios, las hojas de respuestas, estadísticas descriptivas de las respuestas como del rendimiento de los entrevistadores. Todo el proceso de aplicación fue en línea y con respuestas de análisis inmediatas.

Palabras clave: encuesta, aplicaciones de Google, estadística, supervisión.

[I] Thursday/Jueves, 2:35p.m., Room/aula 1, Session: LACSC–4, LACSC–Statistical Computing 7

[II] Universidad de Costa Rica, Escuela de Estadística, Costa Rica, San José, oscarcenteno86@gmail.com

On Fast Computation of Robust Subspace Estimators[I]

Communication / Ponencia

CEVALLOS VALDIVIEZO, HOLGER GEOVANNY[II] Van Aelst, Stefan[III]
Salibian-Barrera, Matias[IV]

Ecuador

Principal component analysis (PCA) is a popular exploratory tool for multivariate data. Especially for high-dimensional data, PCA is a popular technique to find a low-dimensional representation of the data that yields the best possible approximation to the original data. However, classical PCA is very sensitive to atypical data due to the use of quadratic loss. Several competitive approaches for robust subspace estimation are available nowadays, whose estimation algorithms are based on computing robust covariance matrices (e.g. [2, 4]). Other approaches characterize their solutions in the orthogonal complement subspace (see e.g. [1, 5]). However, for high-dimensional data these methods can become computationally expensive or even unfeasible. In this work we propose an alternative algorithm for the S-L and the S-M estimators of [4] that is able to compute the solution in high-dimensional problems. The main ingredients of our new algorithm are twofold. First, we use the estimating equations corresponding to the estimators to iteratively update the principal directions of the subspace. This approach only requires solving vector operations rather than working with high-dimensional covariance matrices. Second, instead of using random starting values, similar to [3] we propose five robust deterministic values to initialize the algorithm. These starting values yield robust fits that are usually close to the sought after robust solution so that convergence occurs quickly. Several experiments were conducted to compare our new algorithms to the original algorithms of Maronna and other robust estimators such as Spherical PCA and projection pursuit (PP) which have been shown to be computationally efficient for high-dimensional data. These experiments comprised low and high-dimensional data with complex structures and outlying configurations. When estimating the best 2-dimensional subspace our algorithms in general showed a better performance in terms of prediction error, especially when outliers were close to the regular data. This result held in low-dimensional as well as high-dimensional settings. Our algorithms also showed tremendously smaller computation times as we let the dimension grow, even on settings with 10000 or 50000 variables where the algorithms of Maronna were not able to calculate solutions anymore. Our new algorithms were also able to identify outliers in real datasets. Our contribution focused on the estimation of methods in [4] but similar ideas can be applied to other robust estimators such as proposed in [1] or [5].

Keywords: robust subspace, algorithm, computational time, outliers, high-dimensional data.

References

[1] Candès, E. J., Li, X., Ma, Y. and Wright, J. (2011). "Robust Principal Component Analysis?", *J. ACM, 58*, p. 11:1–11:37.

[I] Wednesday/Miércoles, 11:50a.m., Room/aula 1, Session: LACSC–1, LACSC–Statistical Computing 4

[II] Escuela Superior Politécnica del Litoral (ESPOL), Matemáticas, Ecuador, Guayaquil, holgceva@espol.edu.ec

[III] KU Leuven, Leuven, Belgium, stefan.vanaelst@kuleuven.be

[IV] The University of British Columbia, Vancouver, Canada, matias@stat.ubc.ca

[2] Croux, C. and Haesbroeck, G. (2000). "Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies", *Biometrika, 87*, p. 603–618.

[3] Hubert, M., Rousseeuw, P. J. and Verdonck, T. (2012). "A deterministic algorithm for robust location and scatter", *Journal of Computational and Graphical Statistics, 21*, p. 618–637.

[4] Maronna, R. (2005). "Principal Components and Orthogonal Regression Based on Robust Scales", *Technometrics, 47*, p. 264–273.

[5] She, Y., Li, S., and Wu, D. (2016). "Robust Orthogonal Complement Principal Component Analysis", *Journal of the American Statistical Association, 111*, p. 763–771.

Copula Functions[I]

Short course / Curso corto

CHERUBINI, UMBERTO[II] Mulinacci, Sabrina [III]

Italia

Copula functions represent the most general tool to describe the dependence structure among random variables. Almost all applications of copula functions have been devoted to modeling the dependence in a static setting. Only more recently copula functions have been used to model dynamic dependence, that is the dependence across time in a stochastic process. Having in mind financial applications, relevant restrictions on the classes of stochastic processes to be considered must be imposed: the most typical ones are the Markov property and the martingale requirement. This short course is based on Cherubini et al. (2012). The first part will be devoted to introducing the basic notion and the main properties of copula functions: we will analyze main copula functions families and main statistical properties. After that, in the second part, we will analyze the seminal result due to Darsow et al. (1992): it allows to express the Markov property in terms of a specific relation among the copula functions representing the dependence structure among the random variables of the stochastic process at different times. Therefore, one can implement a technique to build stochastic processes modeling the increments and the dependence structure between levels and increments in order to disentangle processes with independent and dependent increments.

Keywords: Copula Functions, Markov processes.

References

[1] Cherubini U., Gobbi F., Mulinacci S., Romagnoli S. (2012): Dynamic copula methods in Finance. John Wiley & Sons.

[2] Darsow W.F., Nguyen B., Olsen E.T. (1992): Copulas and Markov Processes. Illinois Journal of Mathematics, 36, 600–642.

[I] Thursday/Jueves, 8:00a.m., Room/aula 1, Session: SC–LACSC, Tutorial LACSC

[II] University of Bologna, umberto.cherubini@unibo.it, Bologna, Italia, umberto.cherubini@unibo.it

[III] University of Bologna, sabrina.mulinacci@unibo.it, , Italia, sabrina.mulinacci@unibo.it

Optimality of the mean absolute deviation about the mean and its generalization to higher-way arrays[I]

Communication / Ponencia

CHOULAKIAN, VARTAN OHANES[II]

Canada

Optimisation has two faces: minimisation of a loss function or maximisation of a gain function. In this talk we will show that: 1) Mean absolute deviation about the mean maximises a gain function based on the power set of the individuals. 2) For 2-way arrays which are 2-way centered, the same maximisation criterion corresponds to a well known matrix norm used in Grothendieck's inequality and used in multivariate analysis. 3) For 3-way arrays or tensors which are 3-way centered, the same maximisation criterion corresponds to 3-way tensor norm. Examples from multivariate statistical data analysis will be provided for the three cases.

Keywords: Taxicab correspondence analysis, taxicab matrix norm, 3-way anova.

References

[1] Choulakian, Vartan (2016). Matrix factorizations based on induced norms. Statistics, Optimization and Information Computing, 4, 1-14.

[I] Wednesday/Miércoles, 10:30a.m., Room/aula 1, Session: LACSC–1, LACSC–Statistical Computing 4

[II] Université de Moncton, Math/Statistics, Canada, Moncton, vartan.choulakian@umoncton.ca

On optimal direction gibbs sampling[I]

Invited Session / Sesion invitada

CHRISTEN, J ANDRÉS[II] Montesinos, Cricelio[III] Santana-Cibrian, Mario [IV]

México

The MCMC generalized Gibbs samplers are those that may simulate from any direction not necessarily bounded to each axis along the parameters of the objective function. In this paper we study how to optimally choose such directions in a directional, random scan, Gibbs sampler setting. We start by considering that the optimal directions will be those that minimize the mutual information (Kullback-Leibler divergence) of two steps of the MCMC. Then we propose two optimal directions distributions for the particular case of the multivariate Normal objective function. Moreover, this result is generalized to be used when a local Multivariate Normal approximation is available for the objective function. To test the resulting algorithms we use highly skewed non-normal objective functions as our target distribution. Finally, the performance of the algorithms that we present is compared with the performance of two cases of a random walk Metropolis.

Keywords: MCMC Methods, Gibbs Sampler; Optimal Gibbs Direction; Metropolis Random Walk.

[I] Wednesday/Miércoles, 4:40p.m., Room/aula 1, Session: L4–2, LACSC–Forecasting using Simulation

[II] Centro de Investigación en Matemáticas (CIMAT-CONACYT), Jalisco S/N, Guanajuato, GT, 36023, México, jac@cimat.mx

[III] (CIMAT, México), Guanajuato, México. jose.montesinos@cimat.mx

[IV] (Universidad de Querétaro, México), Colin Fox (Otago University, New Zealand).

Scatter and Joint Dependence Ranking of Multivariate Populations with Applications to Brain Cytoarchitecture Complexity[I]

Communication / Ponencia

CORAIN, LIVIO[II] Ceccato, Riccardo [III] Salmaso, Luigi [IV] Peruffo, Antonella [V]

Italy

Without assuming any kind of random distribution, i.e. within a fully nonparametric framework, we propose a combination and permutation-based method for separately testing and ranking on both scatter and joint dependence aspects of two multivariate distributions. We theoretically motivate the conditions under which the proposed methodology is a valid solution and numerically prove its effectiveness throughout an extensive simulation study. In particular, we show that our method is distribution-free and acceptable for even quite small sample sizes. Finally we apply the proposed procedure to the analysis of brain cytoarchitecture complexity of some populations defined on the basis of factors such as sex, age and specie.

Keywords: generalized Beherens-Fisher problem, permutation tests, non-parametric combination.

Mathematics Subject Classification (2010): 62G09 Nonparametric inference - Resampling methods, 62G10 Nonparametric inference - Hypothesis testing, 62H15 Multivariate Analysis - Hypothesis testing.

References

[1] Corain, L., Arboretti, R., and Bonnini, S. (2016), *Ranking of Multivariate Populations - A Permutation Approach with Applications*, Boca Raton, Florida, Chapman and Hall/CRC.

[2] Cozzi, B., De Giorgio, A., Peruffo, A., Montelli, S., Panin, M., Bombardi, C., Grandis, A., Pirone, A., Zambenedetti, P., Corain, L. and Granato, A. (2017). "The Laminar Organization of the Motor Cortex in Monodactylous Mammals: a Comparative Assessment Based on Horse, Chimpanzee and Macaque". *Brain Structure and Function*, DOI:10.1007/s00429-017-1369-3.

[3] Corain, L. and Salmaso, L. (2015). "Improving Power of Multivariate Combination-based Permutation Tests", *Statistics and Computing*, DOI: 10.1007/s11222-013-9426-0, 25 (2), pp. 203–214.

[I] Thursday/Jueves, 2:55p.m., Room/aula 1, Session: LACSC–4, LACSC–Statistical Computing 7

[II] University of Padova, Management and Engineering, Italy, Vicenza, livio.corain@unipd.it

[III] Department of Management and Engineering, University of Padova, Vicenza, Italy, ceccato@gest.unipd.it

[IV] Department of Management and Engineering, University of Padova, Vicenza, Italy, luigi.salmaso@unipd.it)

[V] Department of Comparative Biomedicine and Food, University of Padova, Legnaro (PD), Italy, antonella.peruffo@unipd.it

A new log-linear bimodal Birnbaum-Saunders regression model[I]

Plenary Talk / Conferencia Plenaria

CRIBARI NETO, FRANCISCO[II]

Brazil

The log-linear Birnbaum-Saunders model has been widely used in empirical applications. We introduce an extension of this model based on a recently proposed version of the Birnbaum-Saunders distribution which is more flexible than the standard Birnbaum-Saunders law since its density may assume both unimodal and bimodal shapes. We show how to perform point estimation, interval estimation and hypothesis testing inferences on the parameters that index the regression model we propose. We also present a number of diagnostic tools, such as residual analysis, local influence, generalized leverage, generalized Cook's distance and model misspecification tests. We investigate the usefulness of model selection criteria and the accuracy of prediction intervals for the proposed model. Results of Monte Carlo simulations are presented. Finally, we also present and discuss an empirical application.

Keywords: Birnbaum-Saunders distribution, diagnostic analysis, model selection criteria, prediction interval, misspecification test.

[I] Friday/Viernes, 2:45p.m., Auditorium, Session: C–LACSC, Conference LACSC 3

[II] Universidad Federal de Pernambuco, Departamento de Estatística, Pernambuco, Brazil, cribari@de.ufpe.br

Prediction of Complex Human Traits using Big Data from Biobank[I]

Invited Session / Sesion invitada

DE LOS CAMPOS, GUSTAVO[II]

United States

Despite the important discoveries reported by Genome-Wide Association studies, for most human traits and diseases the prediction R-squared (R-sq.) achieved with genetic scores remains considerably lower than the trait heritability. Many factors contribute to the missing heritability problem; among them, sample size is one of the most important. Modern biobanks (e.g., UK-Biobank, Million Veteran Program) will provide unprecedentedly large biomedical data sets: Will the advent of Big Data close the gap between the trait heritability and the proportion of variance that can be explained by a genomic predictor? I will address this question from a theoretical and an empirical perspective. After a brief discussion of the factors that affect the prediction accuracy of a genomic predictor I will present results obtained with data from the UK-Biobank that demonstrate that the use of Big Data (N 400,000) in combination with regularized regression methods (I will discuss both applications based on Bayesian models and Lasso) can eliminate a sizable fraction of the gap between trait heritability and prediction R–squared.

Keywords:
Complex Traits, Prediction Accuracy, High Dimensional Regressions, LASSO, Bayesian.

[I]Tuesday/Martes, 1:30p.m., Room/aula 1, Session: L3–1, LACSC–New developments in Genomic Selection and Prediction studies

[II]Michigan State University, Department of Epidemiology and Biostatistics, Michigan, United States, gdeloscampos@epi.msu.edu

Análisis clásico y Bayesiano de series temporales para la predicción del precio del café[I]

Poster / Cartel

DUARTE RIVERA, ROBERTO CARLOS[II]

Honduras

A comparison of two theories of time series applied to the prediction of the price of coffee.

Keywords: Time series, Bayesian, DLM

References

[1] Analysis of Financial Time Series, Ruey S. Tsay ; Bayesian Forecasting and Dynamic Models, West Harrison

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Nacional Autónoma de Honduras, Estadística Matemática, Honduras, Tegucigalpa, roberto.duarte@unah.edu.hn

Bayesian Estimation of the Shrinkage Parameter in Ridge Regression[I]

Invited Session / Sesion invitada

FIRINGUETTI LIMONE, LUIS[II] Pereira-Barahona, Manuel[III]

Chile

A common problem encountered in regression modeling is that of multicollinear regressors. It is also the case that in Data Science high dimensional data matrices are commonplace, with explanatory variables which are highly collinear or even linearly dependent because the number of variables are often larger than the number of observations. This condition makes least squares methods infeasible for parameter estimation and special regression techniques are required to deal with estimation in the presence of multicollinearity.

Ridge Regression is an estimation procedure known to be useful when multicollinearity is present. This estimator however depends on a shrinkage parameter which, being dependent on unknown coefficients, needs to be estimated.

This paper presents a Bayesian view of the shrinkage parameter. The resulting Ridge Regression estimator with Bayesian shrinkage is compared through simulations to other well known RR estimators.

Keywords: Bayesian estimation, linear regression, ridge regression, simulations.

References

[1] Hoerl, A. E., and Kennard, R. W. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics. 1970; 12, 69–82.

[2] Hoerl, A. E., and Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 1970; 12, 55–67.

[3] Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. Ridge Regression: Some Simulation. Communications in Statistics: Theory and Methods.1975; 4, 105–123.
Kibria, B. M. G. Performance of Some New Ridge Regression Estimators. Communications in Statistics-Simulation and Computation. 2003; 32, 419–435.

[4] Kibria, B. M. G. Performance of the Shrinkage Preliminary Test Ridge Regression Estimators Based on the Conflicting of W, LR and LM Tests. Journal of Statistical Computation and Simulation. 2004; 74, 793–810.

[5] Lawless, J. and Wang, P. A Simulation Study of Ridge Regression and Other Regression Estimators. Communications in Statistics-Theory and Methods, 1976; 5, 1615–1624.

[6] van Wieringen, W. N. (2015). Lecture notes on ridge regression. arXiv preprint arXiv:1509.09169.

[I] Tuesday/Martes, 4:20p.m., Room/aula 1, Session: L4–1, LACSC–Statistical Computing for Data Science

[II] Departamento de Estadística, Universidad del Bío-Bío, Concepción, Chile, lfiringu@ubiobio.cl

[III] Departamento de Estadística, Universidad, del Bío-Bío, Concepción, Chile, mpereira@ubiobio.cl

Efficient Monte Carlo evaluation of resampling-based hypothesis tests[I]

Plenary Talk / Conferencia Plenaria

FUNG, WING KAM[II]

China

Monte Carlo evaluation of resampling-based tests is often conducted in statistical analysis. However, this procedure is generally computationally intensive. The pooling resampling-based method has been developed to reduce the computational burden but the validity of the method has not been studied before. In this talk, we first investigate the asymptotic properties of the pooling resampling-based method, and then propose a novel Monte Carlo evaluation procedure namely the n-times pooling resampling-based method. Theorems as well as simulations show that the proposed method can give smaller or comparable root mean squared errors and bias with much less computing time, thus can be strongly recommended especially for evaluating highly computationally intensive hypothesis testing procedures.

Keywords: Bootstrap test, Monte Carlo sample, power estimate, permutation test, resampling, size estimate.

[I] Tuesday/Martes, 11:20a.m., Auditorium, Session: C–LACSC, Conference LACSC 1

[II] The University of Hong Kong, Hong Kong, Department of Statistics and Actuarial Science, China, wingfung@hku.hk

Foreign Exchange Dependence through Diferent Copula Models[I]

Invited Session / Sesion invitada

GARCÍA, JESÚS[II]

Brazil

We postulate di erent scenarios based on copula models speci ed by C(u; v) = uv + f(u)g(v) for suitable functions f and g, see Rodriguez-Lallena and Ubeda-Flores (2004). We used these copula families to model the dependence between two currencies quoted relative to the U.S. Dollar, the Canadian Dollar and the South Korean Won. We adopt a Bayesian approach to estimate the corresponding parameters, and as a result we estimate the impact of losses on the South Korean Won on the Canadian Dollar.

References

[1] Rodriguez-Lallena, J. and Ubeda-Flores, M. (2004). A new class of bivariate copulas. Statistics and Probability Letters 66, 315-325.

[I] Thursday/Jueves, 1:50p.m., Room/aula 1, Session: LACSC–3, LACSC–Copulas

[II] University of Campinas, IMECC, Brazil,jegime@gmail.com

Bayesian Sensitivity Analysis for Asymmetric Copulas with Cubic Sections[I]

Invited Session / Sesion invitada

GONZALEZ LOPEZ, VERONICA A.[II]

Brazil

The use of copulas to model the dependence between indicators leads us to observe the di erent methods of estimation and its applicability, given practical circumstances, such as having small databases. For this reason, Bayesian methods under the scope of copulas come to show immense utility. In this paper, we investigate how the responses of the Asymmetric Cubic Sections copula model are a ected when we vary the prior distributions display over the model parameters. We use as a reference setting a non-informative prior distribution, and we observe the e ect of the other prior distributions in relation to it. We used this diversity of scenarios to map the possible degrees of dependence, between two educational scores obtained by students of the undergraduate course of statistics at the University of Campinas in 2014.

Keywords: Copulas, Dependence Analysis, Measures of Dependence, Applications.

[I]Thursday/Jueves, 2:10p.m., Room/aula 1, Session: LACSC–3, LACSC–Copulas

[II]University of Campinas, IMECC, Brazil, veronica.g.l@gmail.com

The risk of amputation of the lower extremities of patients with diabetes mellitus: an application using the e-extended regression family of models[I]

Invited Session / Sesion invitada

GRANZOTTO, DANIELE C T[II]

Brazil

An important issue in parametric statistical methods is the identification of a distribution that fits the data well. Focusing in the quality of the adjustment of the model, in this paper we present a new way to extend lifetime distributions in the presence of covariates, regression exponential extended (or e-extended) model. Including a new parameter that is set in an unrestricted parameter space, the new model retains the characteristics of a good model: it is simple, more flexible and continues being interpretable. In order to show the applicability of the e-extended models, we present the e-extended Weibull regression showing analytically expressions for the cumulative distribution and survival functions as well as related expression for the MLE's. A simulation study was carried out, by using Monte Carlo and bootstrap methods, in order to show the finite samples properties of the corresponding estimates. The usefulness of this new family of models was showed in the study of the risk of amputation in patients with injury of the lower extremities due to the type 2 diabetes mellitus. This is a work co-authored by K.A. Anaya-Izquierdo and F. Louzada.

Keywords: e–extended, regression, survival, risk.

References

[1] Louzada, F., Granzotto, D.C.T.: The transmuted log-logistic regression model: a new model for time up to first calving of cows. Statistical Papers (2015)

[2] Marshall, A., Olkin, I.: Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families. Springer Series in Statistics. Springer New York (2007)

[3] Tollenaar, N., Mooijart, A.: Type I errors and power of the parametric goodness of fit test: Full and limited information. Journal of Mathematical and Statistical Psychology 56, 271–288 (2003)

[4] Verdinelli, I., Wasserman, L.: Bayesian goodness of fit testing using infinite–dimensional exponential families. The Annals of Statistics 26(4), 1215–1241 (1998)

[I] Wednesday/Miércoles, 2:55p.m., Room/aula 1, Session: L3–4, LACSC–Statistics and Applications

[II] Universidade Estadual de Maringá, Department of Statistics, Maringá, Brazil, danigranzotto@outlook.com

A bayesian hierarchical model for extreme concentration of carbon monoxide pollution in Mexico City[I]

Communication / Ponencia

JIMÉNEZ HERNÁNDEZ, JOSÉ DEL CARMEN[II] Pérez Rodríguez, Paulino[III]
López Cerino, Marisol[IV]

México

In recent years the statistical modeling of extreme events has been a large number of applications in many areas, for example, in hydrology, environment, to mention a few, in which the theory of extreme values has been widely developed, Resnick (2013) present the mathematical sustenance while Beirlant et al. (2006) present their statistical basis. Within this, modeling with spatial extremes presents, recently a great interest, this has given rise to two different forms of modeling. The first is based on max-stable processes, which are an infinite dimensional generalization of the multivariate extreme value theory (Huser and Davison, 2014). The second uses Bayesian hierarchical models with a latent spatial structure, in which the spatial variation of the parameters in the marginal distribution can be described. In this work, we propose a spatial Bayesian hierarchical model to analyze pollution data by carbon monoxide in Mexico City. Spatial trends are proposed in the parameters of the extreme values distribution, we produce predictive maps, in these we can observe the spatial behavior of the parameters in a marginal form which allows to capture the local variation, which is not possible only with the deterministic trend surfaces.

The model.

To analyze our data, we propose a model that includes spatial variation in the parameters of the extreme value distribution, this can be described by a latent spatial process. Specifically, we suppose that the response variables $\{Y(x)\}$ are independent conditionally on an unobserved latent process $\{S(x)\}, x \in \mathcal{X}$ and the parameters of the generalized extreme-value (GEV) distribution $\{\mu(x), \sigma(x), \xi(x)\}$ vary smoothly for $x \in \mathcal{X}$ according to $\{S(x)\}$, in addition, we assume that the $\{S(x)\}$ for each GEV parameter are mutually independent, though this assumption can be relaxed. For instance, we model to $\mu(x)$ as,

$$\mu(x) = f_\mu(x; \beta_\mu) + \mathbf{S}_\mu(\mathbf{x}; \alpha_\mu, \lambda_\mu), \tag{10}$$

where f_μ is a deterministic function depending on regression parameters β_μ, and S_μ is a zero mean, stationary Gaussian process with covariance function $\gamma_\mu(h) = \alpha_\mu \exp(- \parallel h \parallel / \lambda_\mu)$ and unknown sill and range parameters α_μ and λ_μ. We use similar formulations for $\sigma(x)$ and $\xi(x)$. Then conditional on the values of the three Gaussian processes at the sites $(x_1, \ldots, x_D)$, the maxima are assumed to be independent with,

$$Y_i(x_d) \mid \{\mu(x_d), \sigma(x_d), \xi(x_d)\} \sim \text{GEV}(\{\mu(x_d), \sigma(x_d), \xi(x_d)\}), \tag{11}$$

[I] Wednesday/Miércoles, 1:50p.m., Room/aula 1, Session: L3–3, LACSC–Application of classification techniques in the study of student's success in propauedeutic college math courses

[II] Universidad Tecnológica de la Mixteca, Instituto de Física y Matemáticas, México, Huajuapan de León, Oaxaca, jcjim@mixteco.utm.mx

[III] Departamento de Estadística, Colegio de Postgraduados, Texcoco, Estado de México, perpdgo@gmail.com

[IV] Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca, Huajuapan de León, Oaxaca, México, marisol@mixteco.utm.mx

$i = 1, \ldots, n$, $d = 1, \ldots, D$. And from the Bayesian approach, a joint prior density must be defined for the parameters $\alpha_\mu, \alpha_\sigma, \alpha_\xi, \lambda_\mu, \lambda_\sigma, \lambda_\xi, \beta_\mu, \beta_\sigma, \beta_\xi$. We use conjugate priors whenever possible, taking independent inverse Gamma and multivariate normal distributions for α_μ and β_μ, respectively. No conjugate prior exists for λ_μ, for which we take a relatively uninformative Gamma distribution. The prior distributions for the two remaining GEV parameters are defined similarly. For samples of the joint posterior distribution we use an efficient implementation of the Gibb sampler algorithm included in the package `SpatialExtremes` de `R`.

Results.

We obtained the estimates of the parameters of the posterior distribution after 10000 iterations of the Markov chian, thinned by a factor 10, preceded by a burn-in of 5000 iterations. A sensitivity analysis of the results with respect to the choice of priori was performed and, although the posterior distributions for α and λ were different, the predictive pointwise return level maps were similar. We produce a predictive maps for each parameter of the GEV distributions, these maps were produced by first generating one conditional simulation of three independent Gaussian processes for each state of the Markov chain given its then current values of μ, σ y ξ.

Conclusions.

We found that there is no variation in the β_1 and β_2 parameters to $\mu(x)$, the variations seems reasonable, with the decrease as latitude increases and longitude decreases corresponding to a general reduction in altitude away from the Mexico City. The pattern of variation for the scale parameter is similar. We found that it was not possible to learn from the data simultaneously about the parameters α and λ, for which there is an identifiability problem. The predictive maps shows the main strength of the latent variable approach: the use of stochastic processes to model the spatial behavior of the marginal parameters enables us to capture complex local variation in the return levels that deterministic trend surfaces cannot reproduce.

Keywords: Extreme value distribution, Stochastic process, Bayesian inference.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2006). *Statistics of Extremes: Theory and Applications.* John Wiley & Sons.

[2] Cooley, D. (2010). "Spatial hierarchical modeling of precipitation extremes from a regional climate model", *Journal of Agricultural, Biological, and Environmental Statistics, n^o 15*, p. 381–402.

[3] Davison, A. C., Padoan, S. A. and Ribatet, M. (2012). "Statistical modeling of spatial extremes". *Statistical Science n^o 27*, p. 161–186.

[4] Huser, R. and Davison, A. C. (2014). "Space–time modelling of extreme events", *J. R. Statist. Soc. B, n^o 76*, p. 439–461.

[5] Resnick, S. I. (2013). *Extreme Values, Regular Variation and Point Processes*, New York: Springer.

Bivariate Teissier's Copula and Applications[I]

Invited Session / Sesion invitada

KOLEV, NIKOLAI[II]

Brazil

The baseline model under consideration has been first introduced by the French biologist Teissier (1934) considering a mortality of several domestic animal species protected from accidents and disease, i.e., dying as a result of a pure aging. Specifically, for a non-negative random variable X, the model is defined by the survival function

$$P(X > x) = \exp\{x + 1 - e^x\} \ \ para \ \ x > 0.$$

It is direct to check that the corresponding mean residual life function $E[X - x|X > x] = e^{-x}$. Teissier´s distribution is motivated by the empirical fact that many vital functions are decaying exponentially. We first give historical remarks about the forgotten univariate Teissier´s model. We introduce a bivarite version of the Teissier distribution and outline its basic properties. The corresponding copula is obtained and applications are discussed.

Observation: Master level in Statistics is a must to understand the exposition.

Keywords: Copulas, Dependence measures, Extreme events, Inference, Sibuya's dependence function.

References

[1] Teissier, G. (1934). Recherches sur le vieillissement et sur les lois de mortalite. Annales de Physiologie et de Physicochimie Biologique 10, 237–284.

[I] Thursday/Jueves, 1:30p.m., Room/aula 1, Session: LACSC–3, LACSC–Copulas

[II] Institute of Mathematics and Statistics, University of Sao Paulo, Statistics, Brazil, Sao Paulo, kolev.ime@gmail.com

Exact queue length distribution in an equilibrium M/G/1 queue with Weibull Service[I]

Poster / Cartel

LACAYO, RAMON A.[II]

Colombia

We derive the exact, closed form distribution of the queue length in an equilibrium M/G/1 queue with Weibull service and FCFS discipline. The distribution is obtained by straightforward differentiation at the origin of the probability generating function as given by the Pollaczek-Khinchin formula. As a result of using Faà di Bruno's formula for the differentiation of a composite function, the relevant probabilities are given as finite sums of partial Bell polynomials. An important feature of these computations is that, contrary to a widespread belief, they require no version of the Weibull Laplace transform other than that given by the definition.

Keywords: Weibull transform, Bell polynomial, queue length, equilibrium, Weibull distribution.

References

[1] Asmussen S., Klüppelberg C., Sigman K., Sampling at subexponential times, with queueing applications. Stochastic Processes and their Applications. Vol.79, Issue 2, 1999, Pages 265–286

[2] Mehdi, J. Stochastic models in queueing theory. Academic Press, 2002

[3] Ramsay C., Exact waiting time and queue size distributions for equilibrium M/G/1 queues with Pareto service. Queueing Syst (2007) 57: 147–155

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad del Magdalena, Economics, Colombia, Santa Marta, lacayo.ramon@gmail.com

A model to estimate the probability of wining the USA presidential elections[I]

Invited Session / Sesion invitada

LÓPEZ, EDGAR D.[II] Montiel, Luis V.[III]

México

This work proposes a method to optimize the distribution of the electoral polls along the 50 states of The United States of America (USA). The method is rooted on a stochastic model that estimates the probability of winning The USA Presidential Elections. In previous elections, Trump Vs Clinton, and Bush Vs Gore, it has been clear that the Electoral College is key on the final outcome, and such outcome is not necessarily representative of the popular vote. Hence, this work proposes a better approximation to the mechanics impose by the Electoral College, and can be used to optimize the recollection of information provided by the polls and, at the same time, provides the probability distribution of the number of Electoral College votes a candidate might have. Previous models assume independence among states, and provide ad-hoc data points to argue that this assumption is valid. However, independence among some states might exist at different levels of correlation. Hence, our model relaxes the assumption of independence providing a way to include different levels of pairwise correlation. This opens the possibility to consider that two states "A´´ and "B´´ that share similar economic and social characteristics, might have some level of correlation such that by knowing "A´´ we might have some level of information about "B´´. We use copulas mix with a vector simulation method called JDSIM to implement the model and present results.

Keywords: USA presidential election, votes forecasting, JDSIM, Copulas.

[I] Wednesday/Miércoles, 5:00p.m., Room/aula 1, Session: L4–2, LACSC–Forecasting using Simulation

[II] Instituto Tecnológico Autónomo de México, Department of Industrial and Operations Engineering, C. México, México, ed_gar_da_vid134@hotmail.com

[III] Instituto Tecnológico Autónomo de México, México luis.montiel@itam.mx

Fraud detection in bank loans via zero-inflated non-default rate regression models[I]

Invited Session / Sesion invitada

LOUZADA, FRANCISCO[II] Oliveira, M[III] Moreira, F.F.[IV]

Brazil

In this paper, we introduce a methodology based on zero-inflated non-default rate regression model in order to deal with fraud rate estimation in bank loan portfolios. Our approach enables us to accommodate three different types of loan borrowers, i.e., fraudsters, those who are susceptible to default and finally, those who are not susceptible to default. Regarding to the survival analysis framework, an advantage of our approach is to accommodate zero-inflated times, which is not possible in the usual cure rate models. The parameter estimation is reached by maximum likelihood estimation procedure. The proposed method is illustrated in a real Brazilian dataset of loan survival times. This paper is authored by Mauro Ribeiro de Oliveira Júnior (Caixa Econômica Federal, Brazil) and Fernando F. Moreira (Credit Research Centre, University of Edinburgh Business School, Scotland, UK).

Keywords: Survival Analysis, new lifetime distributions, MLE.

[I] Wednesday/Miércoles, 2:35p.m., Room/aula 1, Session: L3–4, LACSC–Statistics and Applications

[II] Institute of Mathematical Sciences and Computing, Universidade of São Paulo, São Carlos, Brazil, louzada@icmc.usp.br

[III]

[IV]

On Some Theoretical Aspects of SABSCOR, SSQCOR, SUMCOR and MAXVAR Criteria[I]

Invited Session / Sesion invitada

MARTINEZ RUIZ, ALBA[II]

Chile

Several criteria have been proposed to relate a set of blocks of variables measured in the same individuals, such as SABSCOR, SSQCOR, SUMCOR and MAXVAR. For these criteria, different approaches may be used to find solutions. In this work, I compare the theoretical aspects of some of the procedures for each criterion and some considerations are made for the analysis of large data sets.

Keywords: SABSCOR, SSQCOR, SUMCOR, MAXVAR.

References

[1] Kettenring, J. R. (1971). "Canonical analysis of several sets of variables", *Biometrika, n^o 58* , p. 433–451.

[2] Mathes, H. (1993). "Global optimization criteria of the PLS–algorithm in recursive path models with latent variables", In *Haagen, K., Bartholomew, D.J., and Deistler, M. (Eds.), Statistical modeling and latent variables* (pp. 229–248). Elservier Sciece Publishers.

[3] Hanafi, M. (2007). "PLS path modeling: Computation of latent variables with the estimation mode B", *Computation Stat, n^o 22* , p. 275–292.

[I] Tuesday/Martes, 4:00p.m., Room/aula 1, Session: L4–1, LACSC–Statistical Computing for Data Science

[II] Universidad Católica de la Santísima Concepción, Departamento de Ingenieria Industrial, Chile, Concepción, alba.martinez.ruiz@gmail.com

Robust clustering approaches based on the joint application of trimming and constrains[I]

Invited Session / Sesion invitada

MAYO ISCAR, AGUSTÍN[II]

España

We introduce TCLUST methodology which corresponds to robust procedures for model based clustering, which robustness is based in the joint application of trimming and constraints. The application of trimming allows to avoid the influence of contaminating observations in the sample and the application of constraints prevent to choose singulatities or spurious local maximizers as the estimation. TCLUST approaches have been successfully applied in several clustering settings, such as, for multivariate observations or for identifying groups around linear models. For them, there are available versions using crisp and fuzzy assigments. An important issue when applying TCLUST procedures is related with choosing the input parameter values. Currently, there are available methodology for providing to the users sensible strategies for selectingg them.

Keywords: model-based clustering, fuzzy clustering, trimming, scatter constraints.

[I]Friday/Viernes, 11:10a.m., Room/aula 1, Session: L4–3, LACSC–Recent advances in robust statistics and distribution theory

[II]Universidad de Valladolid, Departamento de Estadística e Investigación Operativa and IMUVA, España, agusmayo@med.uva.es

Formal implementation of the semi-supervised K-means algorithm with hard constraints[I]

Communication / Ponencia

MELNYKOV, IGOR[II] Melnykov, Volodymyr, [III]

United States

K-means algorithm is a popular clustering procedure due to its ease of implementation and fast performance. In an unsupervised setting, where no restrictions on class memberships of observations are imposed, the algorithm has a well-motivated objective function. At the same time, for a semi-supervised case, when certain constraints on observations exist, no rigorous formal methods that would account for these constraints have been suggested. To date, the practitioners have used ad-hoc procedures that add the checks of additional conditions as intermediate steps in the classic algorithm [1, 2]. Consequently, the restrictions that exist in the semi-supervised setting do not become a part of the fabric of the algorithm and its objective function remains unchanged.

We consider the semi-supervised clustering framework in the presence of hard positive and negative constraints. Positive constraints require certain data points to be in the same class, whereas negative restrictions prevent them from being placed in the same class in the solution. Using recent advances in semi-supervised clustering [3], we introduce a method making hard positive and negative constraints an integral part of the K-means algorithm and show the modifications that occur to its objective function when the constraints are accommodated. In addition to making hard constraints an organic part of the K-means algorithm, the proposed method has the flexibility to be applied both in the cases when training data are labelled according to their class assignments and in the situations when constraints only define blocks of observations in the data set, but class determinations have not been made. The developed approach shows good practical results in a variety of situations with respect to cluster overlap and number of dimensions.

Keywords: K-means, semi-supervised clustering, hard constraints.

References

[1] Wagstaff, K., Cardie, C., Rogers, S., and Schroedl, S. (2001). “Constrained K-means clustering with background knowledge", *Proceedings of The Eighteenth International Conference on Machine Learning (ICML-2001)*, p. 577–584.

[2] Basu, S., Bilenko, M. and Mooney, R. (2004). “A probabilistic rramework for semi-supervised clustering", *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 59–68.

[3] Melnykov, V., Melnykov, I. and Michael, S. (2016). “Semi-supervised model-based clustering with positive and negative constraints", *Advances in Data Analysis and Classification*, v. 10, p. 327–349.

[I] Wednesday/Miércoles, 2:10p.m., Room/aula 1, Session: L3–3, LACSC–Application of classification techniques in the study of student’s success in propauedeutic college math courses

[II] Colorado State University - Pueblo, Department of Mathematics and Physics, United States, Pueblo, igor.melnykov@csupueblo.edu

[III] Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL 35487, United States, vmelnykov@cba.ua.edu

Robust inference for the ROC curve[I]

Invited Session / Sesion invitada

MILHEIRO LOURENÇO, VANDA[II]

Portugal

The receiver operating characteristic (ROC) curve is a mainstay on the statistical evaluation of medical tests. When covariate information is available, ROC regression is usually employed as a way to evaluate if the discrimination accuracy of a medical diagnostic test, may change as a function of such covariate. In this paper we propose robust inference methods for ROC regression which can be used to safeguard against the presence of outlying biomarker values. Simulation results suggest that the methods perform well recovering the true conditional ROC curve and corresponding area under the curve, on a variety of data contamination scenarios. We illustrate our methods in a case study on prostate cancer diagnosis.

Keywords: Robust regression, ROC curve, diagnostic tests.

References

[1] Farcomeni, A., Ventura, L. (2012). An overview of robust methods in medical research. Statistical Methods in Medical Research, 21(2):111–133.

[2] Lourenço, V.M., Pires, A.M., Kirst, M. (2011). Robust linear regression methods in association studies. Bioinformatics, 27(6):815–821.

[3] Pepe, M.S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. Biometrics, 54(1):124–135.

[I] Friday/Viernes, 10:30a.m., Room/aula 1, Session: L4–3, LACSC–Recent advances in robust statistics and distribution theory

[II] NOVA University of Lisbon, Portugal, Department of Mathematics, Faculty of Sciences and Technology, Portugal, vmml@fct.unl.pt

Comparison of changes of scale for the simultaneous analysis of variables, qualitative and quantitative, using multivariate statistical methods.[I]

Poster / Cartel

MORALES GARCIA, EMMANUEL[II] Del callejo-Canal, Diana[III]
Canal-Martínez Margarita [IV]

México

The management of mixed databases, in other words, a mixture of qualitative and quantitative variables, regularly presents the difficulties for their analysis. One of the alternatives to solve this problem is to homogenize the scale, that is change the scale of measurement of the variables so that all are equal, this can be achieved, according to the literature consulted, through two ways: 1) Categorization and 2) Quantification (Correa, 2008) and (Gower et al.2016).As a result of a consultancy with a mixed database from the area of biology, the research questions that arose with the alternative of homogenizing the scale were: Which of these two routes has better results? What advantages and disadvantages does one have over another? What packages are available for these procedures? This paper is the route of a statistical research process that involves the management of mixed multivariate systems, which is a problem that frequently occurs in various areas of knowledge and is regularly working inadequately. The biology database was used to make a comparison between categorization and quantification as ways to homogenize the scale and try to answer the research questions posed above. We identified and exemplified the multivariate statistical methods relevant to both cases and thus know the advantages and disadvantages of each process. The categorization process was carried out using the quartiles method and later a multiple correspondence analysis (MCA) was used, while in the quantification the HOMALS technique was used to later use a principal component analysis (PCA), the analyzes for both processes were realized with software R Project 3.4.1. In general, both processes worked well for this application, although the categorization process is a bit slower than the quantification process since the assignment of the categories is done by the user. The categorized data had a better representation in the descriptive univariate analysis, while the quantified data had a better representation in the multivariate analysis, which for this particular exercise was paramount. For this reason, it is that for this example with real data the quantification turns out to be more efficient than the categorization.

Keywords: mixed multivariate systems, change of scale, categorization, quantification.

Resumen

El manejo de bases de datos mixtas, en otras palabras, mezcla de variables cualitativas y cuantitativas presenta con regularidad dificultades para su análisis. Una de las alternativas para solucionar este problema es

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Veracruzana, Estadística, México, Xalapa,Veracruz, emmanuelgarciam94@gmail.com

[III] Instituto de Investigaciones de Estudios Superiores, Económicos y Sociales. Universidad Veracruzana, Xalapa, Veracruz, México, ddelcallejo@uv.mx

[IV] Instituto de Investigaciones de Estudios Superiores, Económicos y Sociales. Universidad Veracruzana, Xalapa, Veracruz, México, mcanal@uv.mx

homogeneizar la escala, es decir cambiar la escala de medición de las variables para que todas sean de un mismo tipo, esto se puede lograr, según la bibliografía consultada, mediante dos vías: 1) Categorización y 2) Cuantificación (Correa,2008) & (Gower et. al.2016). Como resultado de una consultoría con una base de datos mixta proveniente del área de biología, las preguntas de investigación que surgieron ante la alternativa de homogenizar la escala fueron ¿Cuál de estas dos vías presenta mejores resultados? ¿Qué ventajas y desventajas presenta una sobre otra? ¿Qué paqueterías hay disponibles para estos procedimientos? Este escrito es el recorrido de un proceso de investigación estadística que involucra el manejo de sistemas multivariantes mixtos, el cual que es un problema que se presenta frecuentemente en diversas áreas del conocimiento y con regularidad se trabaja inadecuadamente. Se utilizó la base de datos de biología para realizar una comparación entre la categorización y la cuantificación como vías para homogeneizar la escala y tratar de responder a las preguntas de investigación antes planteadas. Se identificaron y ejemplificaron los métodos estadísticos multivariantes pertinentes para ambos casos y así conocer las ventajas y desventajas de cada proceso. El proceso de categorización se llevó a cabo mediante el método de cuartiles y posteriormente se utilizó un análisis de correspondencias múltiple (MCA por sus siglas en inglés), mientras que en la cuantificación se ocupó la técnica HOMALS para posteriormente utilizar un análisis de componente principales (PCA por sus siglas en inglés), los análisis para ambos procesos se realizaron con el software R Project 3.4.1. De manera general ambos procesos funcionaron adecuadamente para esta aplicación, aunque el proceso de categorización resulta un poco más tardado que el de cuantificación ya que la asignación de las categorías las realiza el usuario. Los datos categorizados tuvieron una mejor representación en el análisis univariante descriptivo, mientras que, los datos cuantificados tuvieron una mejor representación en el análisis multivariante, que para este ejercicio en particular era primordial. Por esta razón es que para este ejemplo con datos reales la cuantificación resulta ser más eficiente que la categorización.

Palabras clave: sistemas multivariantes mixtos, cambio de escala, categorización, cuantificación.

References

[1] Correa,G.(2008) ; Gower et. al. (2016); Lê, S., Josse, J. & Husson, F. (2008),R Development Core Team (2008).

Dominio afectivo de la estadística en estudiantes universitarios[I]

Poster / Cartel

MUJICA BETANCOURT, RUBEN DARIO[II] García Bolívar, Judith Josefina[III]

Ecuador

The research was based on the interpretation of the affective domain of Statistics of university students from an ethnographic perspective. For this, qualitative data were analyzed, coming from the same students and the professors who dictate the statistical subject. In-depth interviews were used as a technique and as methods of analysis, categorization and triangulation. After the qualitative analysis of the data, it was found that there are two transversal axes in the students' domain, which are: the personal motivations and the causal attributions.

Keywords: Affective domain, Statistics, Students.

Resumen

La investigacion se basó en la interpretación del dominio afectivo de la Estadística de los estudiantes universitarios desde una perspectiva etnográfica. Para ello, se analizaron datos de forma cualitativa, provenientes de los mismos estudiantes y de los profesores que dictan la asignatura estadística. Se utilizaron entrevistas a profundidad como técnica y como métodos de análisis, la categorización y la triangulación. Luego del análisis cualitativo de los datos se obtuvo que son dos ejes transversales en el dominio de los estudiantes, los cuales son: las motivaciones personales y las atribuciones causales.

Palabras clave: Dominio afectivo, Estadística, Estudiantes.

References

[1] Batanero, C. (2001). Didáctica de la Estadística. Editado por Grupo de Educación Estadística Universidad de Granada. Departamento de Didáctica de la Matemática. Universidad de Granada. España.

[2] Coello, Y., Villasmil, J. y Mujica, R. (2014). Motivos sociales presentes en Estudiantes y Docentes de Postgrado en la Universidad Rafael María Baralt, Núcleo Coro. Memorias del Evento Pre humanísticas 2014. Universidad Bolivariana de Venezuela. Punto Fijo. Venezuela.

[3] Goetz, J. y LeCompte, M. (1988). Etnografía y diseño cualitativo en investigación cualitativa. Madrid.

[4] Gómez-Chacón, I. (2000). Matemática Emocional. Edit. Narcea. Madrid. España.,

[5] Haedo, A. (2001). Una revisión a la enseñanza de la Estadística en escuelas y Universidades de Argentina. (Libro 2, pp. 165-167).

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad Politécnica Estatal de Carchi, Logística y Transporte, Ecuador,Tulcan, ruben.mujica@upec.edu.ec

[III] Universidad Politécnica Estatal del Carchi, Tulcán, Ecuador, judithj.garcia@upec.edu.ec

Estimation of expectations in two-level nested simulation experiments[I]

Invited Session / Sesion invitada

MUÑOZ, DAVID F.[II]

México

Two–level nested simulation methods have been recently applied for the analysis of simulation experiments under parameter uncertainty. On the outer level of the nested run, we generate (n) observations of the parameters, while on the inner level; we fix the parameter on its corresponding value and generate (m) observations using a simulation model. In this paper, we focus on the output analysis of two-level stochastic simulation experiments for the case where the observations of the inner level are independent, showing how the variance of the simulated observations can be decomposed in the sum of parametric and stochastic components. Furthermore, we derive central limit theorems that allow us to compute asymptotic confidence intervals to assess the accuracy of the simulation–based estimators for the point forecast and the variance components.

Theoretical results are validated through experiments using a forecasting model for sporadic demand, where we have obtained analytical expressions for the point forecast and the variance components.

Keywords:

Keywords: Guaranteed–Service Level, Safety Stock, Bayesian Forecasting, Inventory Simulation.

[I] Wednesday/Miércoles, 4:00p.m., Room/aula 1, Session: L4–2, LACSC–Forecasting using Simulation

[II] Instituto Tecnológico Autónomo de México, Department of Industrial and Operations Engineering, C. México, México, davidm@itam.mx

Development and Execution of a technique to generate models of Credit Scoring[I]

Communication / Ponencia

MUÑOZ TRIGUEROS, SAMUEL[II]

Costa Rica

Nowadays diversity of risks affect the financial entities, being one of them the credit risk. This credit refers to the risk that a lender assumes when awarding a credit to a borrower, being a possibility a non-payment of the obligations in the future. The risk of non-payment may be common to the entire client portfolio of the entity. If an appropriate measurement of the credit risk is not performed, then a bankruptcy or financial insolvency can occurred.

In the statistics field an important group of tools and techniques are being developed for an optimal and adequate credit risk measurement, best known as Credit Scoring models. A very important reason of these models´ apogee is the variety of national and international regulations, like the International Financial Reporting Standard (IFRS9).

The main purpose of this academic investigation is to develop and execute a technique to generate Credit Scoring Models based on nonparametric information statistics and validation methods. Below is detailed a brief summary of the main results from the investigation: the elaborated methodology, the phases and algorithms of the technique in the language R, which are: variables transformation with a grouping method, using the weight of evidence (WoE) and the value information (IV). Model estimation using the philosophy of the generalized additive models (GAM), where the aim is to maximize the divergence using optimization algorithms. Measurement of model performance using the KS and ROC. Use of three validation methods (training/test, cross-validation and bootstrap 0.632). An easy interpretation for the end user by giving a scale to the model.

Finally, the execution of this technique is performed with data from a financial entity and the performance is compared with other techniques commonly used in the industry, like: logistic regression, decision trees and neural networks.

Keywords: Credit Scoring, GAM, Risk, Nonparametric Statistics, Validation Methods.

Resumen

Las entidades financieras se ven constantemente afectadas por diversidad de riesgos, siendo uno de ellos el riesgo crediticio. En términos generales este es el riesgo que una entidad financiera asume cuando otorga un crédito a un cliente, dándose la posibilidad de que en un futuro éste incumpla con sus obligaciones de pago. Ahora bien, si este riesgo de incumplimiento de pago es alto para la cartera de clientes de la entidad y el mismo no se mide oportunamente puede llevar eventualmente a la quiebra financiera.

En las últimas décadas se han desarrollado una variedad de herramientas y técnicas para la medición del riesgo creditico, que se conocen como modelos de Credit Scoring. El desarrollo de estos modelos es un campo que está en auge, siendo potenciado por las diversas regulaciones nacionales e internacionales, como es el caso del "International Financial Reporting Standard 9" (IFRS9 por sus siglas en ingles).

[I] Thursday/Jueves, 3:15p.m., Room/aula 1, Session: LACSC–4, LACSC–Statistical Computing 7

[II] Universidad de Costa Rica, Escuela de Estadistica, Costa Rica, San José, samutriz@gmail.com

El objetivo de la investigación es: elaborar y ejecutar una técnica para generar modelos de Credit Scoring partiendo de estadísticos de información no paramétricos y de métodos de validación.

El resultado de la misma es la metodología elaborada; las fases y los algoritmos en el lenguaje R de la técnica, que son: transformación de la variables por medio de un método de agrupación usando el peso de la evidencia (WoE) y el valor de la información (IV), estimación del modelo utilizando la filosofía de los modelos aditivos generalizados (GAM) donde se busca maximizar la divergencia por medio de algoritmos de optimización, la medición de desempeño del modelo utilizando el KS y ROC, la utilización de tres métodos de validación (desarrollo/prueba, cruzada y bootstrap 0.632) y la forma de darle una escala al modelo, para facilitar la interpretación al usuario final.

Finalmente, se realiza la ejecución de esta técnica con datos de una entidad y se compara el desempeño con otras técnicas comúnmente utilizadas en la industria (regresión logística, árboles de decisión y redes neuronales).

Palabras clave: Credit Scoring, No paramétricos, Riesgo Crediticio, GAM, Métodos de Validación.

References

[1] Anderson, R. (2007). The Credit Scoring Toolkit. Great Britain: OXFORD UNIVERSITY PRESS.

[2] Efron, B. & Tibshirani, R. (1997). Improvements on Cross-Validation: The .632+ Bootstrap Method. Journal of the American Statistical Association , 92 (438), 548 - 560.

[3] Hastie, T. & Tibshirani, R. (1990). Generalized Additive Models. United States of America: Chapman & Hall/CRC.

[4] Ince, H. & Aktan, B. (2009). A Comparison of Data Mining Techniques for Credit Scoring in Banking: A Managerial Perspective. Journal of Business Economics and Management , 10 (3), 233 – 240.

[5] Kullback, S. (1968). Information Theory and Statistics. New York: DOVER PUBLICATIONS, INC.

Statistics Without Borders: Overpassing between conscious and brain response[I]

Invited Session / Sesion invitada

NASCIMENTO, DIEGO[II] Toostani, I.G. Vaziri, Z.S. Colacique, M. Louzada, F. Delbem, A.C.B. Barros, C.G.C. Oliveira, A.A. Hyppolito, M.A. Leite, J.P.

Brazil

This work aimed to analyze the tinnitus loudness perception among positive emotional induction. It was conducted a randomized cross-over study designed in three sessions. Thus was combined with three different interventions such as visual stimulation, visual stimulation with high definition tDCS (HD- tDCS), and visual stimulation with Sham.

Forty-five patients were instructed to sit on a fixed chair with 85 cm distance from the 40-inch installed LCD. Each session contained two blocks of neutral pictures presentation, pre and post interventions. Results from the patient self-assessment, shown from a Generalized Linear Mixed model, presented meaningful difference between visual stimulation with HD-tDCS treatment and others (One-Level, $P<0.001$). In this manner, clinical analysis was proceeded to confront conscious cognitive-intervention vs. physiological (brain) response. EEG session least around 45 minutes each, with a 1000 sample per second, considering 256 channels per subject. Therefore, this EEG dataset summarized in 2.5gb data. Thus, concept of regularity of a series is expressed as Fourier frequencies, in terms of periodic variations, driven by sines and cosines, shows properties of the time-invariant linear filter, among the frequency domain, discriminating the spatial-temporal dependence of the brain. Furthermore, gradual decrement in tinnitus loudness scaling over the time (from the beginning of stimulation) was indicated. The outcomes are in line with our hypothesis that positive emotional induction can reduce the loudness perception in tinnitus patients. Further investigation is needed to find out its effect on bothersome and the ability to ignore tinnitus.

This is a work co-authored by I.G. Toostani, Z.S. Vaziri, M. Colacique, F. Louzada, A.C.B. Delbem, C.G.C. Barros, A.A. Oliveira , M.A. Hyppolito, J.P. Leite.

Keywords: Multivariate Time Series, biosignals, Neuromodulation, Tinnitus.

[I] Wednesday/Miércoles, 3:15p.m., Room/aula 1, Session: L3–4, LACSC–Statistics and Applications

[II] Institute of Mathematical Sciences and Computing, Universidade of São Paulo, São Carlos, Brasil, dcn@usp.br

Generalized Gamma distribution: Different methods of estimations [I]

Poster / Cartel

NASCIMIENTO, DIEGO[II] Luiz Ramos, Pedro[III] Louzada, Francisco[IV]

In this study, we present different estimation procedures for the parameters of the generalized gamma distribution such as the maximum likelihood, penalized maximum likelihood, method of moments, modified moments, ordinary and weighted least-squares, percentile, maximum product of spacings, Cramer-von Mises and the Anderson-Darling maximum goodness-of-fit estimators and compare them using extensive numerical simulations. We showed that the penalized maximum likelihood estimator is the most efficient for estimating the parameters of the proposed distribution. Finally, our proposed methodology was also illustrated in two real data sets related to the lifetime of the components of harvest sugarcane machines.

Keywords: Generalized gamma distribution, Penalized maximum likelihood estimator, Simulated annealing.

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Institute of Mathematical Sciences and Computing, Universidade of São Paulo, São Carlos, Brazil, dcn@usp.br

[III] Institute of Mathematical Science and Computing, University of Sao Paulo, São Carlos - Brazil, ramospedroluiz@gmail.com

[IV] Institute of Mathematical Sciences and Computing, Universidade of São Paulo, São Carlos, Brazil, louzada@icmc.usp.br

Genomic Selection: Is it worth the sequence?[I]

Invited Session / Sesion invitada

PEREZ ENCISO, MIGUEL[II]

España

Genomic Selection (GS) is the procedure by which DNA polymorphisms (predominantly biallelic markers or SNPs) is used to predict breeding values, that is, the future performance of offspring from the selected individuals. Traditionally, GS has been implemented using genotyping arrays but the use of sequence data in genomic prediction models is a topic of high interest, given the decreasing prices of current sequencing technologies (NGS). Two main reasons are given to advocate the use of sequence in GS: (i) the causal mutations are in the data and (ii) therefore, disequilibrium of phenotype with genotype does not decay, avoiding the need to reevaluate the model. Further, SNP arrays only interrogate a minor fraction of total variability, since array SNPs have been obtained in individuals that are not selection candidates. NGS data have also drawbacks: they are expensive (at least 10x larger than arrays) and noisy. Here I investigate whether sequence can overcome GS performance when using arrays. I show, using extensive simulations and theory, that the potential benefit is very small (0-4%) unless at least part of the SNPs in strong disequilibrium with the causal mutations can be identified. These results closely match the few experimental data that are available so far. The main reasons for this conclusion are: (i) most SNPs found with sequence are rare and therefore of no utility for prediction, (ii) most sequence information is redundant due to disequilibrium, and (iii) the practical impossibility of faithfully identifying all causal mutations due to high noise/signal ratio.

Keywords: Genomic selection, Genominc Prediction, SNP data.

References

[1] Pérez-Enciso M, JC Rincón, A Legarra. 2015. Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised Genet Sel Evol 47:43, doi: 10.1186/s12711-015-0117-5

[2] Pérez-Enciso M, Forneris N., De Los Campos G, Legarra A. 2017. Evaluating sequence-based genomic prediction with an efficient new simulator. Genetics 205:939-953.

[I] Tuesday/Martes, 1:50p.m., Room/aula 1, Session: L3–1, LACSC–New developments in Genomic Selection and Prediction studies

[II] Centre for Research in Agricultural Genomics, Consortium, Barcelona, España, miguel.perez@uab.es

An R implementation for the aggregation of factor levels in the binomial logit model[I]

Communication / Ponencia

PONSOT BALAGUER, ERNESTO[II]

Ecuador

The binomial logit model (Agresti, 2015) is one of the most widely used statistical techniques for the analysis of categorical data. When after a logit adjustment, it is decided to group levels of the explanatory factors and repeat the analysis, Ponsot et al (2009) and Ponsot et al (2012) show that a distributional error is incurred. The obtained distribution is binomial if and only if the probabilities of success are the same. The authors also suggest an alternative course of action that decreases the variance of the estimators, improves the accuracy of the estimates and increases the power of hypothesis tests on the parameters. In the practical implementation of the results obtained by the authors, it is required to automate the aggregation of levels in some explanatory factor of interest. In the unsaturated multifactorial case this task is of considerable complexity. The R programs (R Core Team, 2017) implemented to give computational support to the methodology are presented. Specifically. The following functions are described: AgrFac(): Which receives the original data set, the number of factors present in the model, the factor some of whose levels will be aggregated and the list of levels to be aggregates. Using the database facilities provided by the routine sqldf (Grothendieck, 2017), select the grouped levels by adding both the observed and total responses. A Data Frame object is returned with the transformation of the original data. ModP(): Whose objective is to adjust a model in this situation, using the developed theory. From the probability estimates of the original logit model and the desired aggregation indications, the function obtain the predicted probability vector and its variance and covariance matrix, as well as the number of elements of the new vector and the number of levels that must be added. Then a column is added as an indicator of whether the row in question is involved (1) or not (0) in the aggregation requested, and the new levels that will be used after the aggregation as those of the factor in question. By convention, what is involved is assigning the lowest of those levels of the factor in each set of aggregation, to all the remaining levels of the set, for each of the sets and without altering the levels of the rest of the factors, nor those of the same factor that do not participate in the aggregation. It follows the extraction of the new indexing of the probabilities vector (and, therefore, of its variance and covariance matrix), necessary after applying the aggregation of levels. Finally, the routine compute new estimates of the parameters, their standard errors, p-values and confidence intervals for the model with grouped data.

Keywords: logit model, aggregation of factor levels, R Statistical Software.

References

[1] Agresti, A. (2015) Foundations of Linear and Generalized Linear Models. John Wiley & Sons. Wiley series in probability and statistics, p. 165–201.

[I] Tuesday/Martes, 2:55p.m., Room/aula 1, Session: L3–2, LACSC–Recent advances in Statistical Computing 1

[II] Yachay Tech, School of Mathematical Sciences and Information Technology, Ecuador, Urcuqui, eponsot@yachaytech.edu.ec

[2] Ponsot-Balaguer, E.; Sinha, S. and Goitia, A. (2009) On the grouping of levels of the explanatory factor in the binary logit model. Colombian Journal of Statistics, v 32, no 2, p. 157–187.

[3] Ponsot-Balaguer, E.; Sinha, S. and Goitia, A (2012) Aggregation of explanatory factor levels in a binomial logit model: Generalization to the multifactorial unsaturated case. Colombian Journal of Statistics, v 35, no 1, p. 139–166.

[4] R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, https://www.R-project.org/

[5] Grothendieck, G. (2017) sqldf: Manipulate R Data Frames Using SQL. R package version 0.4-11, https://cran.r-project.org/package=sqldf

Outlier detection on microarrays applying mixtures of Gaussian and heavy-tailed distributions[I]

Invited Session / Sesion invitada

POSEKANY, ALEXANDRA[II]

Portugal

Normal distributions aare a basic assumption on inear models, however many data in fields like biology or economics do not apply to the standard even though the observation sizes are large. We propose a robustification of Bayesian inference employing mixture models which simultaneously allow density estimation and outlier detection. For this, we suggest mixing standard Gaussian components with Student´s t distributed ones to identify the over-dispersed part of data a part of which is extremely noisy. As an application, we present several microarray data, as they are well–known for their complicated, over-dispersed noise behaviour. Microarrays have found their way from research into clinical practice which makes detecting problems in the molecular and bio-informatical analyses even more important. Our goal in a addition to a better inference of differential expression is to identify noisy genes and in addition to recognise whether single arrays are responsible for this behaviour allowing for a means of quality control.

[I]Tuesday/Martes, 2:10p.m., Room/aula 1, Session: L3–1, LACSC–New developments in Genomic Selection and Prediction studies

[II]Danube University Krems, Department for Clinical Neurosciences and Preventive Medicine, Austria, Krems, alexadra.posekany@gmail.com

Estudio fractal de la serie de tiempo de la variable CO2 en el volcan de San Salvador-año 2005 [I]

Poster / Cartel

RIVAS MORALES, MILTON ARNOLDO[II]

El Salvador

We work with a database of measurements of different physical factors, such as C02, air temperature, atmospheric pressure, among other factors, taken at the San Salvador volcano in the period from January 1 to December 31 of year 2005. An estimate is made of the fractal dimension of the time series for the variable carbon dioxide ($C0_2$) and predictions are made for three months.

Keywords: fractal dimension, time series, predictions.

References

Donald, L., (1997), Fractals and Chaos in Geology and Geophysics, Cornell Univesity, Cambridge University Press Kennetj, F,. (2003), Fractal Geometry, Mathematical Foundations and Aplications, University of St Andrews, Wiley, UK Cryer, J. & Chan, K,. (2010), Time Series Analysis: With Applications in R, Springer Texts in Statistics

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Universidad de El Salvador, Escuela de Matemática-Facultad de Ciencias Naturales y Matemática, El Salvador, San Salvador, miltonrivas11@gmail.com

Copula Density Estimation by Chebyshev Orthogonal Polynomials at the Padua Points[I]

Communication / Ponencia

QU, LEMING[II]

United States

A multivariate distribution with continuous margins can be uniquely decomposed via a copula and its marginal distributions. Copula density estimates that are expressible as the product of a base copula density and a linear combination of Chebyshev orthogonal polynomials at the Padua Points are considered here. The base copula is a mixture of parametric copulas that are capable of capturing symmetrical, lower, and upper tail dependence, respectively [1]. The so-called "Padua points"[2] are the first known example of a unisolvent point set over bivariate polynomials which have a provably minimal order of growth in Lebesgue constant (log square of the degree). Fast algorithms exist for bivariate Lagrange interpolation at the Padua points in a square [3]. The coefficients of the Chebyshev polynomials are estimated by method of moments. The symmetry constraints for a copula density are enforced by symmetry of the polynomial coefficient matrix. The uniform margin constraints for a copula density are enforced by weighted line sum scaling on the coefficient matrix. Several criteria are proposed for determining the number of terms to be included in the polynomial adjustment component.

Keywords: Copula, Density Estimation, Padua Points.

References

[1] Hu, L. (2006) "Dependence patterns across financial markets: a mixed copula approach´´. *Applied Financial Economics*, 16(10):717–729

[2] Caliari, M., De Marchi, S., & Vianello, M. (2005) "Bivariate polynomial interpolation on the square at new nodal sets´´,*Applied Mathematics and Computation* 165: 261–274.

[3] Caliari, M., De Marchi, S., Sommariva, A. & Vianello, M. (2011) "Padua2DM: fast interpolation and cubature at the Padua points in Matlab/Octave´´. *Numerical Algorithms*, 56: 45–60.

[I] Wednesday/Miércoles, 10:50a.m., Room/aula 1, Session: LACSC–1, LACSC–Statistical Computing 4

[II] Department of Mathematics, Boise State University, Boise, Idaho, United States, LQU@boisestate.edu

Bayesian spatial inflated beta regression model for assessment of reading level in school districts of Peru[I]

Invited Session / Sesion invitada

QUIROZ CORNEJO, ZAIDA J.[II] Sal y Rosas, Víctor G. [III] Bayes, Cristian [IV]
Valdivieso, Luis [V]

Perú

The Peruvian Ministry of Education organizes from time to time Student Census Evaluations in order to track the national assessment of educational progress and allocate efforts strategically. One relevant variable in this analysis is the reading level and, particularly, the proportion of high schools that achieve the highest (third) reading level (3RL) at each district. If we aim to determine which factors affect to this proportion, we could perform a regression analysis. However, such a proportion turns out to be occasionally zero, especially in rural areas commonly associated with low-level reading skills. In addition, the 3RL district proportions may be spatially referenced, that is, influenced by its neighbors. To overcome these problems we propose the use a beta inflated mean regression analysis, which allows us to not only manage the zero values (and introduce a spatial effect) but also to incorporate covariates directly to the expected 3RL district proportions. Parameter estimation in the proposed model is performed using the Hamiltonian Monte Carlo algorithm implemented in the Stan software

Keywords: Beta Inflated Regression, Bayesian Inference, MCMC Methods, Hamiltonian Monte Carlo, Education level assessment.

[I] Wednesday/Miércoles, 4:20p.m., Room/aula 1, Session: L4–2, LACSC–Forecasting using Simulation

[II] Pontificia Universidad Católica del Perú, Departamento de Ciencias e Ingeniería, Lima, Perú, zquiroz@pucp.edu.pe

[III] Pontificia Universidad Católica del Perú,

[IV] Pontificia Universidad Católica del Perú,

[V] Pontificia Universidad Católica del Perú,

Grouped p-generalized spherical distributions[I]

Communication / Ponencia

RICHTER, WOLF-DIETER[II]

Germany

The Gauss-exponential distribution plays an important role as a high risk limit law, see the lectures on high risk scenarios and extremes in [1]. The Gauss-exponential distribution can be considered as a particular asymmetric derivation of the Gauss-Laplace law. In particular, Gauss-exponential probabilities of arbitrary events can be dealt with by considering the corresponding Gauss-Laplace probabilities. Density level sets of the standard Gauss-Laplace distribution are topological boundaries of star bodies centered at the origin. The Minkowski functionals of the corresponding star bodies, however, are not homogeneous functions of order one as it is often assumed in the literature on star-shaped distributions. Instead, the bodies corresponding to different density levels reflect different geometric properties and are typically directed to different directions. The aim of the talk is to first recall (p, q)-spherical generalizations of the two-dimensional Gauss-Laplace distribution from [2] and to secondly extend this consideration to the grouped $(p_1, ..., p_k)$-spherical case. We prove geometric and stochastic representations which can be considered as standard tools for dealing with the present distributions in a way being similar to that one has already successfully been dealing for a long time with spherical and, since more recently, even with more general homogeneous star-shaped distributions. This will be demonstrated by generalizing the Box-Muller simulation method following [3] and [4] and discussing further statistical properties following [5].

Keywords:

Gauss-exponential distribution, Gauss-Laplace distribution, grouped $(p_1, ..., p_k)$-spherical uniform distribution, stochastic vector representation, dynamic geometric disintegration.

References

[1] Balkema, G., Embrechts, P. (2007). High risk scenarios and extremes. In: Zurich Lectures in advanced mathematics. Zurich: European Mathematical Society.

[2] Richter, W.-D. (2017). The class of (p,q)-spherical distributions with an extension of the sector and circle number functions. Risks 2017, 3. Doi: 103390/risks5030040

[3] Kalke, S., Richter, W.-D. (2013). Simulation of the p-generalized Gaussian distribution. J. Stat.Comput. Simul. 83 (4), 641-667. Doi: 10.1080/00949655.2011.631187

[4] Richter, W.-D., Schicker, K. (2017). Simulation of polyhedral convex contoured distributions. Journal of Statistical Distributions and Applications 4:1. Doi: 10.1186/s40488-017-0055-6

[5] Richter, W.-D. (2017). Statistical reasoning in dependent p-generalized elliptically contoured distributions and beyond. Journal of Statistical Distributions and Applications 4:21. Doi: 10.1186/s40488-017-0074-3

[I] Friday/Viernes, 11:30a.m., Room/aula 1, Session: L4–3, LACSC–Recent advances in robust statistics and distribution theory

[II] University of Rostock, Institute of Mathematics, Germany, Germany, wolf-dieter.richter@uni-rostock.de

A robust DF-REML framework for genetic association studies[I]

Invited Session / Sesion invitada

RODRIGUES, PAULO[II]

Brazil & Finland

In genetic association studies, linear mixed models (LMMs) are used to test for associations between phenotypes and candidate single nucleotide polymorphisms (SNPs). These same models are also used to estimate heritability, which is central not only to evolutionary biology but also to the prediction of the response to selection in plant and animal breeding, as well as the prediction of disease risk in humans. However, when one or more of the underlying assumptions are violated, the estimation of variance componentsmay be compromised and therefore somay the estimates of heritability and any other functions of these. Considering that datasets obtained from real life experiments are prone to several sources of contamination, which usually induce the violation of the assumption of the normality of the errors, a robust derivative-free restricted-maximum-likelihood framework (DF-REML) together with a robust coefficient of determination are proposed for the LMM in the context of genetic studies of continuous traits.

Keywords: Derivative-free REML, Linear mixed model, Heritability estimation.

References

[1] Lourenço, V. M., Pires, A. M., Kirst, M. (2011). Robust linear regression methods in association studies. Bioinformatics 27(6), 815–821.

[2] Meyer K. (1989). Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative free algorithm. Genetics Selection Evolution 21, 317–340.

[I] Friday/Viernes, 10:50a.m., Room/aula 1, Session: L4–3, LACSC–Recent advances in robust statistics and distribution theory

[II] Center for Applied Statistics and Data Analytics, University of Tampere, Finland and Department of Statistics, Federal University of Bahia, Brazil, paulocanas@gmail.com

Bootstrap technique for hydrocarbon exploration and production[I]

Poster / Cartel

SUNJAY, SUNJAY[II]

India

Geophysical signals are nonstationary and multiscale in character.statistical signal processing is the area of signal processing where mathematical statistics is used to solve signal processing problems.In most statistical signal processing applications where a certain parameter is of interest there is a need to provide a rigorous statistical performance analysis for parameter estimators.Another important problem in signal processing is to make certain probability statements with respect to a true but unknown parameter. Bootstrapping is a method for statistical inference requiring intensive resampling of input data to estimate statistical properties characterizing the data such as mean, variance, and confidence intervals. The main advantage of the bootstrap over Monte Carlo simulations is that the bootstrap does not require the experiment to be repeated. The basic idea in bootstrapping is that the actual data are resampled to produce a large number of data sets. The individual estimates of a given parameter may then be used to compute its standard errors and/or confidence intervals.From a data manipulation point of view, the main idea encapsulated by the bootstrap is to simulate as much of the real world probability mechanism as possible, substituting any unknowns with estimates from the observed data. Bootstrapping in the seismic context as a random reordering of seismic traces so that the traces initially arranged in increased offset order will be reorganized in an unpredictable manner using random number generation algorithms. A bootstrap procedure for high-resolution velocity analysis : extended information criterion based on the bootstrap for the estimation of the number of harmonics actually present in geophysical data.The bootstrap procedure may be used to assign errors to seismic velocities and to improve the velocity spectrum computed from high-resolution coherence measures. Bootstrap method to construct an average coherence measure and a kernel density estimator of the velocity that maximizes the coherence. The average coherence exhibits an important attenuation of spurious events while retaining enough resolution to model reflections properly with similar moveout curves. Lower frequencies provide better penetration into the deep section and reduce side lobes of the wavelet for easier structural interpretation, velocity analysis and inversion; higher frequencies sharpen the central peak of the wavelet for thin bed differentiation. Bootstrap deghosting reduces the side lobes of the wavelet and makes the geologic structure easier to interpret.Bootstrap is applied for Radan transform- slant stacking seismic denoising, AVO amplitude variation with offset interpretation. Bootstrap deghosting also has its own limitations. Since it is data driven, strong noise contamination can reduce the reliability of the deghosting filter. The bootstrapped differential semblance (BDS) estimator allows for efficient event detection and more accurate parameter tracking. The bootstrapping detects slight wavelets shifts and allows the discrimination between the parameters giving a good fit to the reflection moveout. Application of the bootstrap method for DCA(Decline curve analysis) of reservoir adopted ordinary bootstrap to resample the original production data. Bootstrap technique is applied to detect outliers in Dynamic Data Analysis(DDA) of Permanent Downhole Guage Data (PDG). the application of detectors based on bootstrap techniques used in a GPR–Ground Penetrating Radar/BHR- Bore Hole Radar

[I] Thursday/Jueves, 3:35p.m., pasillo, Session: Cart, Poster/Carteles

[II] Bihar, India, sunjay_sunjay@hotmail.com

system.Bootstrap procedure for estimating the sampling variability of principal component analysis (PCA) , Fast bootstrap PCA – resampling is a low dimensional transformation– interpretation of principal components depends on the coordinate vectors on which the sample is measured.The key to improving computational efficiency of PCA in bootstrap samples is to realize that all resampled observations are contained in the same low dimensional subspace as the original sample. . PCA is often invoked as a method of reducing the dimensionality of subsequent analysis such as independent component analysis or clustering Independent component analysis (ICA) produces spatially independent components (component images). Bootstrap methods have been widely applied in image processing, pattern recognition and computer vision.

Keywords: Statistical signal processing, Bootstrap technique, PCA, seismic signal.

Minimum quadratic distance estimation based on the cumulative hazard function[I]

Communication / Ponencia

TITA GRANZOTTO, DANIELE CRISTINA[II] Louzada, Francisco[III]
Anaya-Izquierdo, Karim[IV]

Brazil

In this research, we introduce a new method of parameter estimation based on minimizing a quadratic distance from the cumulative hazard of the model to the Nelson-Aalen estimator. Given that the cumulative hazard is function restricted to be non-decreasing and therefore is easier to estimate than the hazard rate function, is natural to think about a parametric estimation method that compares the cumulative hazard of the model with that of the data in order the minimize a measure of distance between them. This is specially convenient when the analytical form of the cumulative hazard of a parametric model is simple and the natural candidate to represent the data in this context is the Nelson-Aalen non-parametric estimator of the cumulative hazard. Also, such minimum distance estimators are known to have robustness (to model misspecification) properties when compared to maximum likelihood. We illustrate the use of the new estimation method performing a simulation study to illustrate the finite sample properties of the derived estimators and showing some of the robustness properties compared to maximum likelihood.

Keywords: Cumulative hazard function, Nelson-Aalen estimator, Minimum quadratic distance estimation.

References

[1] Hjort, N. (1990). Goodness of Fit Tests in Models for Life History Data Based on Cumulative Hazard Rates. Ann. Statist., 10(3), 1221–1258.

[2] Basu, A., Shioya, H., and Park, C. (2011). Statistical Inference. The minimum distance approach. CRC Press.

[I] Wednesday/Miércoles, 11:10a.m., Room/aula 1, Session: LACSC–1, LACSC–Statistical Computing 4

[II] State University of Maringa, Department of Statistics, Brazil, Maringa, danigranzotto@outlook.com
[III] University of São Paulo, São Carlos-SP, Brazil, louzada@icmc.usp.br
[IV] University of Bath, Bath, England, K.Anaya-Izquierdo@bath.ac.uk

Índice de palabras clave

Index of keywords

Índice de autores

Índice de contribuciones por país