

V Jornada
de
Matemática
y
Soluciones Aplicadas

10 Años de la Maestría en Modelos Aleatorios

Memorias
18 de Noviembre de 2010
Universidad Central de Venezuela
Facultad de Ciencias
Caracas, 2010



IVIC
INSTITUTO VENEZOLANO
DE INVESTIGACIONES
CIENTÍFICAS



Postgrado en
**MODELOS
ALEATORIOS**
UCV-IVIC

Editores:

Dra. Glaysar Castro (UCV)

Dra. Mairene Colina (UCV)

Dra. Carenne Ludeña (IVIC – UCV)

Edición y Montaje:

Dr. Manuel Maia (UCV)

Editorial

La Maestría en Modelos Aleatorios (MMA), programa conjunto de la Universidad Central de Venezuela y del IVIC, se enorgullece al haber cumplido sus primeros 10 años (y casi 11) de funcionamiento. Para celebrar este acontecimiento decidimos dedicar la V Jornada de Matemática y Soluciones Aplicadas, un evento bianual ideado para compartir temas de interés y retos futuros, a esta primera década de la MMA, invitando a viejos y nuevos compañeros de ruta.

La MMA se creó con el propósito de formar profesionales altamente capacitados en el área de modelización probabilística y aplicaciones estadísticas, con amplia experiencia en la resolución de problemas. Dos grandes ejes impulsaron esta iniciativa. El primero, propiciar la construcción de puentes con los diversos interlocutores dentro y fuera del ámbito universitario, que permitiesen la incorporación de nuevas tecnologías, saberes y destrezas evidenciando el enorme potencial de desarrollo en esta área. Este objetivo se ha ido cumpliendo de manera decidida, aunque más lentamente de lo que a veces quisiéramos. En estos años por nuestras aulas han pasado más de 30 economistas, biólogos, físicos, ingenieros además de matemáticos, todos ellos dando fe de gran solvencia y capacidad. En estas notas, que se suman a varias ediciones anteriores que les invitamos a consultar, presentamos 9 resúmenes de trabajos realizados por nuestros estudiantes y que constituyen una muestra cabal de la calidad y alcance de nuestros egresados, que ocupan hoy por hoy puestos de trabajo en prestigiosas universidades y cada vez más en el sector productivo, para nuestro gran regocijo.

El segundo eje lo constituyó la idea de lograr establecer un núcleo de I+D+i integrando modelización aleatoria, análisis estadístico de datos y manejo de grandes volúmenes de información con aplicaciones en riesgo, finanzas, petróleo y energía, telecomunicaciones,

ambiente, medicina y salud, gestión, minería de datos, control y automatización de procesos y agroindustria. Para ello ha sido necesario pensar en esquemas de organización novedosos con el sector productivo que vinculen proyectos, asesorías, docencia e investigación fomentando el desarrollo de productos de alto valor tecnológico agregado y nuevas maneras de trabajar que potenciaran las capacidades de todos los actores. Esto ha implicado inventar nuevos conceptos de inter-relación con el sector productivo y en este contexto nace Modalea, una propuesta de Empresa de Base Tecnológica Universitaria que permitirá el desarrollo y comercialización de productos y proyectos de asesoría, formación especializada y colaboración en el marco de la Locti. La experiencia lograda en estos años, desde nuestra práctica profesional, la de nuestros estudiantes y la convicción institucional de que la universidad debe seguir otros rumbos diferentes a los tradicionales, nos hace ser optimistas con respecto al futuro de esta propuesta.

No podemos finalizar este balance sin rendir homenaje a aquellos que por su tesón hicieron posible este proyecto, a los docentes, directores y estudiantes, a las autoridades y personal de apoyo, a los que desde sus empresas o instituciones nos han apoyado con tiempo y recursos.

Muy especialmente, queremos dedicar esta V Jornada a uno de sus más conspicuos alentadores, el Dr. José Rafael León, es decir, Chichi. A él nuestro más caluroso agradecimiento.

Carenne Ludeña
Noviembre de 2010

Índice general

• Estudio del Comportamiento de Citas en el Instituto de Previsión del Profesorado de la Universidad Central de Venezuela. (Lic. María Verónica Carvajal Rivero, Dr. Ricardo Ríos).	1
• Un Modelo de Probabilidad de Quiebra Aplicado a las Instituciones que Conforman el Sistema Bancario Venezolano. 1999-2009. (David Ceccato Grau, Ricardo Ríos).	7
• Valoración de Activos Financieros en Mercados Incompletos Mediante Métodos de Dualidad. (Henryk Gzyl, Luis Paredes).	11
• Aplicación de la Modelización ARMAX a la Dinámica de la Malaria. (Francisco J. Laguna, José Rafael León, María Eugenia Grillet).	15
• Aplicación de métodos de ecuaciones diferenciales estocásticas para el desarrollo de un modelo de difusión de contaminantes en la superficie del Lago de Valencia. (Jocelyn León, José León).	19
• Una cooperación del Análisis de Componentes Principales con el Análisis Envolvente de Datos. (Lic. Tomás León, Dr. Ricardo Ríos).	27
• Programa que Aproxima la Función Voltaje, Corriente y Resistencia Efectiva de Redes Eléctricas en Grafos Conectados, por medio de Simulaciones de Caminatas Aleatorias sobre los Grafos. (Miguel A. Padrino).	31
• Reconocimiento Litológico a partir de Registros de Pozo. (Eddre Peña, Carenne Ludeña).	41
• Modelo estocástico de equilibrio general dinámico (MEEGD) para la construcción de densidades a priori de var Bayesianos: una aplicación a la economía venezolana. (Mariela Perdomo, Daniel Barráez).	45
• Caracterización de deudores de Tarjetas de Crédito a través de técnicas de Aprendizaje Estadístico. (Carlos David Salazar Paredes, Mairene Colina, Carenne Ludeña).	49

ESTUDIO DEL COMPORTAMIENTO DE CITAS EN EL INSTITUTO DE PREVISIÓN DEL PROFESORADO DE LA UNIVERSIDAD CENTRAL DE VENEZUELA.

Lic. María Verónica Carvajal Rivero¹, Dr. Ricardo Ríos²

¹*Universidad Central de Venezuela*
mavecari@gmail.com

²*Universidad Central de Venezuela*
ricardo.rios@ciens.ucv.ve@gmail.com

RESUMEN

El Instituto de Previsión del Profesorado (IPP) de la Universidad Central de Venezuela (UCV) adscrito a la Asociación de Profesores de la Universidad Central de Venezuela (APUCV) es una institución de carácter no lucrativo y órgano operativo, creada por la UCV y la APUCV para favorecer la protección social de sus profesores en tres grandes programas: Servicio Médico-Odontológico (SEMO), Servicio de Atención Médico y Hospitalario Integral (SAMHOI) y Créditos. Sus usuarios se clasifican en afiliados y beneficiarios de los afiliados (generalmente familiares), lo cual constituye un universo de aproximadamente 27.000 personas que pueden generar la necesidad de recibir atención médica-odontológica consuetudinaria o de atender algún siniestro hospitalario.

La información de cobertura asistencial del SEMO es activa y variable, por lo cual se consideró como series temporales y con herramientas clásicas de análisis de series temporales se estudiaron. Los datos estudiados representan el número de consultas solicitadas, realizadas y de emergencias, en las diversas especialidades del *Servicio Médico* (Cardiología,

Dermatología, Endocrinología, Fisiatría, Gastroenterología, Ginecología, Medicina Interna, Oftalmología, Otorrinolaringología, Pediatría, Psiquiatría, Traumatología, Urología) y *Servicio Odontológico* (Cirugía Bucal, Endodoncia, Odontología, Odontología Infantil, Ortodoncia, Periodoncia, Prótesis Dentales) que se ofrecen actualmente en el servicio, los cuales fueron extraídos del Sistema de Administración y Estadísticas Médicas para Clínicas y Hospitales (S.A.M.) del IPP. Se posee información desde Mayo 2004 hasta Marzo 2010.

METODOLOGÍA

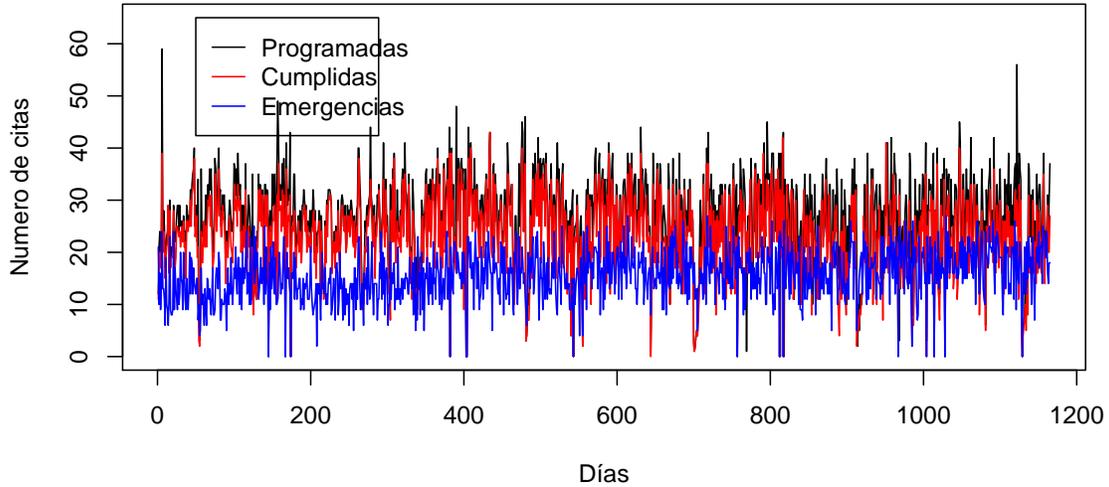
Se utilizó la metodología de *Box-Jenkins*, la cual consiste en la sistematización de modelos llamados Autorregresivos Integrado de Media Móvil (ARIMA, por sus siglas en inglés) o Autorregresivos Estacional Integrado de Media Móvil (SARIMA, por sus siglas en inglés) que utilizan variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Los modelos ARIMA determinan cuántos valores del pasado se deben utilizar para predecir la siguiente proyección, así como también valores de las series e identifican los coeficientes y número de regresiones que se utilizarán. Estos modelos son muy sensible en la precisión para determinar sus coeficientes.

También se realizaron pruebas no paramétricas para medir la bondad de ajuste de las distribuciones de probabilidad del comportamiento de citas en los diversos días de la semana: Test de Kolmogorov-Smirnov, Test de Wilcoxon y Test T-Student.

En la siguiente gráfica (página siguiente) se muestra la relación de citas solicitadas, cumplidas y emergencias en Medicina Interna, una de las especialidades con mayor demanda en el instituto. Se observa un incremento considerable en las emergencias en el transcurso del

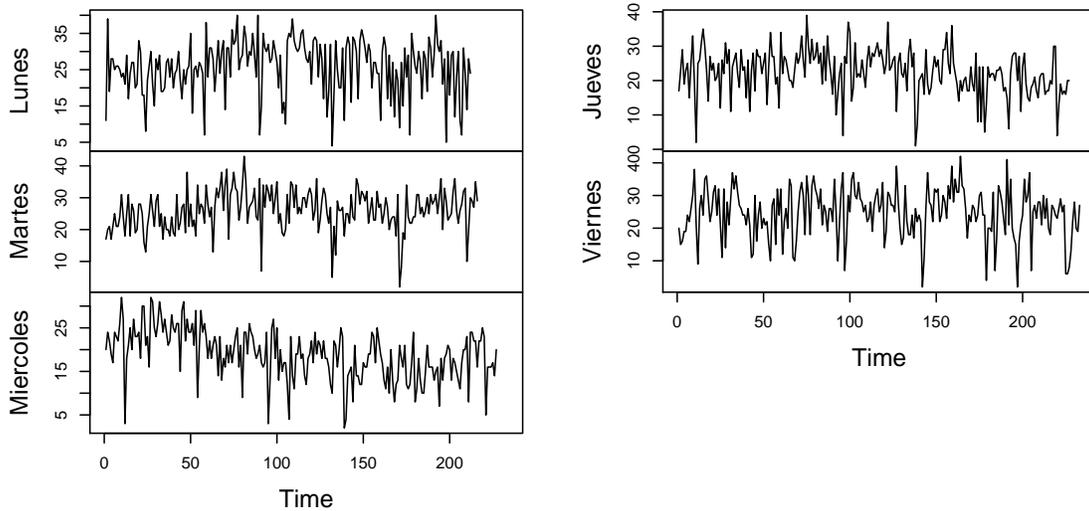
tiempo. Se observaron patrones estacionales, por lo cual se descompuso cada serie en cinco

Series temporales de las citas de Medicina Interna



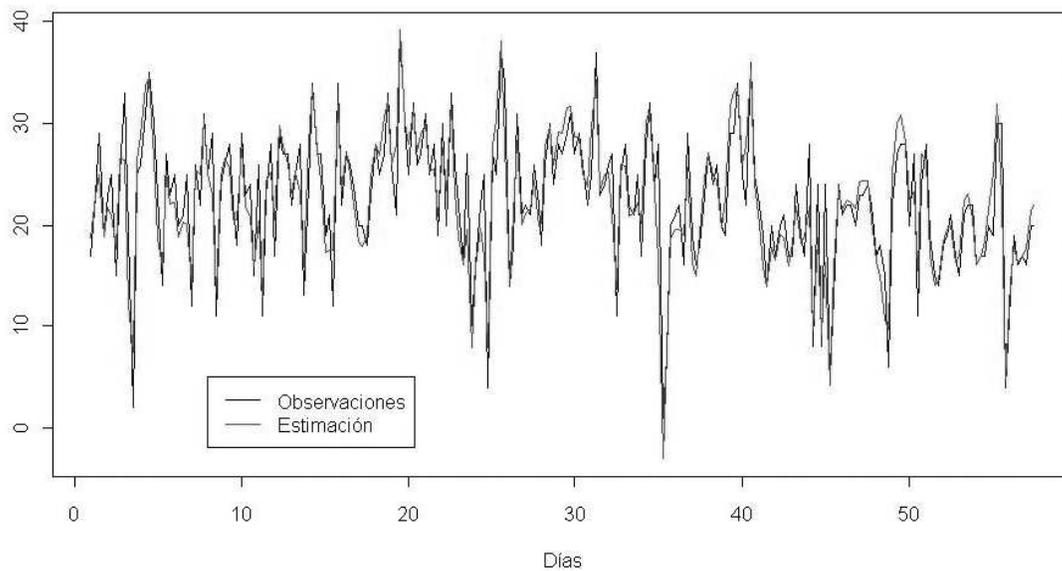
series, las cuales representan los día de la semana (un ejemplo visual se puede apreciar en la siguiente imagen) y luego, se realizaron ajustes para sus respectivos modelos.

Cumplidas en Medicina Interna



En el caso de las citas cumplidas en Medicina Interna se percibe comportamiento similar en los diversos días de la semana, sin embargo, luego de medir la bondad de ajuste se rechazó el hecho que los miércoles y jueves tienen igual distribución al resto de los días.

En la siguiente grafica muestra el ajuste ARIMA(2,0,3) de las citas cumplidas los jueves en la especialidad de Medicina Interna.



RESULTADOS Y CONCLUSIONES

Las series generadas por especialidad que representan la cantidad de citas tanto solicitadas como realizadas presentan un período semanal, lo cual caracteriza el comportamiento de citas en SEMO. Se observó, en la mayoría de especialidades, que a menor número de citas solicitadas o realizadas, mayor es el número de emergencias. Y a mayor número de citas, menor número de emergencias. Las especialidades que presentan mayor cantidad de emergencias son las especialidades con menor cantidad de profesionales disponibles actual-

mente. Los días en los cuales se presenta mayor número de citas solicitados y realizadas son los lunes, martes y jueves, viernes. Así como también a principio, final de un período vacacional o asuertos laborales.

Al realizar las predicciones de citas tanto solicitadas, cumplidas como emergencias para el período Mayo-Diciembre de 2010, se comparó con la capacidad de citas en el Instituto y se reveló la necesidad de realizar planes de contingencia puesto la demanda estimada fue superior a la capacidad actual. Por ello, se recomienda contratar profesionales en las especialidades de cardiología, psiquiatría, otorrinolaringología, urología o contratar servicio de extramuros para consultas de dichas especialidades.

REFERENCIAS

- [1] Brockwell, P.J., AND Davis, R.A. "Introduction to Time Series and Forecasting", Springer-Verlag, New York (1996).
- [2] Dalgaard, P. "Statistic and Computing", Springer, New York (2002).
- [3] Correa, J. and González, N. "Gráficos Estadísticos con R", Medellín (2002).
- [4] Kirchgässner, G. and Wolters, J. "Introduction to Modern Time Series Analysis", Springer, New York (2007).
- [5] Hamilton, J. "Time Series Analysis", Princeton University Press, New Jersey (1994).
- [6] Zivot, E. and Wang, J. "Modeling Financial Time Series with S-Plus", Springer (2006).

UN MODELO DE PROBABILIDAD DE QUIEBRA APLICADO A LAS INSTITUCIONES QUE CONFORMAN EL SISTEMA BANCARIO VENEZOLANO. 1999-2009.

David Ceccato Grau¹, Ricardo Ríos²

¹*Universidad Central de Venezuela*
davceccato@gmail.com

²*Universidad Central de Venezuela*
ricardo.rios@ciens.ucv.ve

RESUMEN

El objetivo de este trabajo es estudiar el proceso asociado a la crisis financiera del año 2009, poniendo particular interés en los elementos que la desencadenaron, de manera de formular recomendaciones de políticas específicas al sector. Para tal fin se utiliza el modelo propuesto por Peresetsky, Karminskyy y Golovan (2004), el cual se basa Este trabajo utiliza la regresión logística para estimar la probabilidad individual de quiebra de las instituciones bancarias, basándose en las observaciones de un conjunto de factores tanto individuales a cada banco, como macroeconómicos. La principal conclusión de este trabajo es que las regulaciones impuestas al sector bancario han incrementado el efecto de las condiciones macroeconómicas sobre la actividad de intermediación, por lo que la probabilidad de quiebra de las instituciones financieras venezolana aumenta de forma significativa en el momento en que el ciclo económico se encuentra en su fase recesiva.

El tema propuesto en el presente trabajo es el análisis de los factores determinantes de la probabilidad de quiebra de las instituciones que conforman el sistema financiero venezolano, para el período comprendido entre los años 1999 y 2009, cuando ocurrió una crisis

bancaria, que culminó con la intervención de diez instituciones financieras por parte de la autoridad regulatoria.

Dado que las instituciones financieras venezolanas poseen diferentes tamaños, manejan políticas de riesgo diferenciadas, tienen distintos perfiles de negocios y la crisis bancaria del año 2009 no implicó una quiebra de la totalidad del sistema, se realizó un análisis de conglomerados previó.

Dentro de los procedimientos a empleados para tal fin se utilizaron los algoritmos K -medias –con distintas variantes– y K -medias recortadas.

Posteriormente, se utilizó una regresión logística para estimar la probabilidad individual de quiebra de las instituciones bancarias, basándose en las observaciones de un conjunto de factores tanto individuales a cada banco –tales como: la configuración, el tamaño y calidad de su portafolio de activos, las fuentes de financiamiento a las que posee acceso, entre otros–, como macroeconómicos –entre los que se consideraron: el nivel y la estabilidad de la actividad económica existente, las regulaciones legales a las cuales se encuentra sometido y el nivel de concentración de mercado–.

La principal conclusión de este trabajo es que las regulaciones impuestas al sector bancario han incrementado el efecto de las condiciones macroeconómicas sobre la actividad de intermediación. La aplicación de medidas tales como la implementación de tasas de interés preferenciales, establecimiento de requisitos de reserva inadecuados y la instauración de cuotas la cartera de crédito hacia sectores específicos de la economía han sido identificados como factores que puede impulsar o apresurar el inicio de una crisis de liquidez o solvencia.

La vuelta de los controles sobre las tasas de interés ha imposibilitado que el sector financiero pueda trasladar las variaciones tanto a nivel de riesgo como en el costo de los fondos a los usuarios finales, por lo que ante un deterioro en las calidad de los activos – asociados a cambios del entorno macroeconómico– afectaría en mayor medida la rentabilidad

de las instituciones, e incluso podría poner en entredicho su viabilidad futura.

La mayor dependencia del sector bancario a la evolución del ciclo económico aumenta su vulnerabilidad de verse afectado por una crisis cuando éste entre en su fase recesiva, por un incremento de la morosidad crediticia, cuando se agudizan los problemas de solvencia de los agentes económicos.

Estas medidas, además de afectar la rentabilidad de las instituciones bancarias, conllevan a una reducción de los índices de capitalización, aumentando la vulnerabilidad del sistema financiero a una crisis, por lo que su conveniencia en el tiempo podría estar en entredicho.

REFERENCIAS

- [1] Baltagi, B. (1995). *Econometric Analysis of Panel Data*. John Wiley and Sons.
- [2] Bikker, J. y H. Hu (2002). "Cyclical patterns in profits, provisioning and lending of banks". De Nederlandsche Bank NV. DNB Staff Reports N.-86.
- [3] Carrasco, R. (2001). "Modelos de elección discreta para datos de panel y modelos de duración: una revisión de la literatura". Universidad Carlos III de Madrid.
- [4] Demirgüç-Kunt, A., L. Laeven y R. Levine (2003). "The impact of bank regulations, concentration, and institutions on bank margins". World Bank Policy Research. Working Paper Series N.-3030.
- [5] Freixas, Xavier y J.C Rochet (2001). "Economía Bancaria". Antoni Bosch Editor. Barcelona, España.
- [6] Hanweck, G, y L. Ryu (2005). "The sensitivity of bank net interest margins and profitability to credit, interest rate, and term structure shocks across bank product specializations". FDIC Working Paper 2005-02.
- [7] Hedeker, D (2007). "Multilevel Models for Ordinal and Nominal Variables". *Handbook of Multilevel Analysis*.
- [8] Hsiao: C. (1993). *Analysis of Panel Data*. Cambridge University Press.
- [9] Luo, R. y H. Wang (2008). "A composite logistic regression approach for ordinal panel data regression". *Data Analysis Techniques and Strategies*, Vol. 1, No. 1.

- [10] 10. Peresetsky, Karminskyy y Golovan (2004). "Probability of default models of Russian banks". Bank of Finland. Discussion Papers. 2004. No. 21
- [11] 11. Wooldridge, J (2006). "Cluster-sample methods in applied econometrics: extended analysis". Department of Economics. Michigan State University.

VALORACIÓN DE ACTIVOS FINANCIEROS EN MERCADOS INCOMPLETOS MEDIANTE MÉTODOS DE DUALIDAD.

Henryk Gzyl¹, Luis Paredes²

¹*Instituto de Estudios Superiores de Administración*

²*Escuela de Matemática, Universidad Central de Venezuela*

`luis.paredes@ciens.ucv.ve`

RESUMEN

El creciente desarrollo de las matemáticas, que en particular es aplicado a las finanzas, fue causado por la relación encontrada en los diferentes campos tales como el análisis funcional, el análisis numérico, los procesos estocásticos, entre otros.

Estas áreas son también, hoy en día, de gran interés en los estudios financieros, ellas han permitido solucionar muchos problemas que no son posibles de resolver de forma cualitativa o mediante un planteamiento de leyes que rijan los aspectos más importantes de las finanzas.

Usando tópicos de probabilidades, análisis y ecuaciones diferenciales, se resuelve un problema financiero, el cual plantea que si un inversionista se encuentra en un mercado financiero y posee un conjunto de activos el espera no incurrir en pérdidas por las variaciones de los precios. Black y Scholes[7], desarrollaron la idea de elaborar una garantía, la cual es obtenida a partir de los activos poseídos, que posteriormente fue llamada “derivado”[1].

Por el crecimiento del volumen de las operaciones financieras de estos instrumentos, surge la necesidad de estudiar a fondo la valoración de los mismos. Esto llevó, a la consideración de los costos de transacción y a la clasificación de los tipos de mercados. El objetivo

es obtener un precio justo para todas las partes involucradas, es decir, el emisor y comprador deben coincidir en la valoración del derivado. Para ello, se introducirá el precio de indiferencia que plantea una política de inversión óptima que resulta de la posesión de un conjunto de activos. Este valor será estudiado con diferentes funciones de utilidad y es obtenido bajo las hipótesis de incompletitud. El enunciado básico para el inversionista es comprar un derivado a un precio v , que a le produzca la misma utilidad no comprar el derivado o pagar la cantidad v hoy para recibir el valor del derivado en el futuro.

Los orígenes del concepto de precio de indiferencia está en el trabajo de Hodges y Neuberger [5], sin embargo ha sido extendido por Elliott y Van der Hoek en [3]. Esas ideas fueron aplicadas en el contexto de opciones reales por Smith y McCradle en [8].

METODOLOGÍA

Se introducirán todos los conceptos teóricos básicos en el contexto de los modelos de mercado a un lapso de tiempo, con la noción de arbitraje y su relación con el concepto de medida neutra al riesgo, esto último permitirá definir los derivados que son de gran importancia en la clasificación de los modelos de mercados. Además, es presentado el modelo de Arrow-Debreu [2] que permite replantear el modelo de mercado.

Seguidamente se considero una clasificación de las funciones de utilidad introducida en [4], que con la ayuda de la teoría dual planteada en [9] es posible obtener funciones duales, esto permite desarrollar los portafolios óptimos en mercados incompletos bajo costos de transacción, los cuales son comparados con su valor dual.

Por último presentamos la definición de precio de indiferencia conjuntamente con sus propiedades, para luego desarrollar un algoritmo que permite obtener el valor del derivado con la implementación de métodos numéricos.

RESULTADOS Y CONCLUSIONES

Desarrollado todo el marco teórico del precio de indiferencia con sus propiedades, se presenta un algoritmo que permite obtener el valor del derivado con la implementación de métodos numéricos. Resultando que la valoración del derivado no depende de la función de utilidad, si no del reordenamiento del mercado con el modelo de Arrow- Debreu.

Para obtener el resultado anterior se considero dos modelos de mercados, uno con dos tipos de acciones y dos posibles fluctuaciones en el valor de los precios y el otro con dos tipos de acciones y cinco tipos de cambios en los precios de los activos.

REFERENCIAS

- [1] M. ARRIOJAS Y H. GZYL, Derivados financieros, Curso básico, *Libros El Nacional*, Caracas,(2000).
- [2] K. ARROW., Essays in the Theory of Risk Bearing, *North-Holland, Amsterdam*, (1970).
- [3] R. ELLIOTT Y J. VAN DER HOEK, Pricing Claims on Non-tradable Assets, *Contemporary Mathematics*. Vol. 351, (2004), 103-114.
- [4] V. HENDERSON Y D. HOBSON, Utility Indifference Pricing- An Overview, *Volume on Indifference Pricing*, (ed. R. Carmona), *Princeton University Press*, (2004).
- [5] S. HODGES Y A. NEUBERGER, Optimal replication of contingent claims under transaction costs, *Rev. Futures Markets*, Vol. 8, (1999), 222-239.
- [6] W. MANFRED , Theory of Duality in Mathematical Programmig, *SPRINGER-VERLAG*, Vol.(1989).
- [7] S. PLISKA , Introduction to Mathematical Finance, Discret time models, *BLACKWELL*, (1997).
- [8] J. SMITH AND K. MCCARDLE , Valuing oil properties: integrating option pricing and decision analysis approaches, *Operations Research*, Vol. 46, (1998), 198-217.
- [9] K. WILFRED , Maxima and Minima with Application, *JOHN WILEY & SONS, INC*,(1999).

APLICACIÓN DE LA MODELIZACIÓN ARMAX A LA DINÁMICA DE LA MALARIA.

Francisco J. Laguna¹, José Rafael León², María Eugenia Grillet³

¹*Modelos Aleatorios, U.C.V.*
gargola72@gmail.com

²*Matemática, U.C.V.*
jose.leon@ciens.ucv.ve

³*Laboratorio de Biología de Vectores, I.Z.T., U.C.V.*
mgrillet@strix.ciens.ucv.ve

RESUMEN

Uno de los problemas que presenta la dinámica de la malaria, es que a menudo las series de incidencia parasitaria son series no estacionarias. Mediante la aplicación de la modelización ARMAX se desea mejorar la bondad de ajuste de las serie de incidencia parasitaria. La modelización ARMAX consiste en estimar un modelo de regresión a partir de una variable exógena (X_t) y restar el predictor (pY_t) a la serie original (Y_t). A la serie ya ajustada (W_t) se le estima un modelo ARMA. Se trabajo con el registro de Paria, Sucre (variables regresoras: precipitación, humedad y temperatura), el de Sifontes, Bolivar (variables regresoras: precipitación, humedad y temperatura), y el de Atures, Amazonas (variables regresoras: precipitación y nivel de río).

METODOLOGÍA

La metodología usada se puede resumir en los siguientes pasos:

1) se estima el predictor pY_t mediante una regresión lineal múltiple. El predictor $p(Y_t|X)$ esta definido como

$$p(Y_t|X)Y_t = \beta^* + \sum_{j=0}^r \beta_j x_{tj-1},$$

donde β^* y β_j son los coeficientes de la regresión múltiple y r es un entero arbitrario.

2) Se calcula la serie ajustada W_t , restando a la serie Y_t el predictor $p(Y_t|X)$.

3) Se estima el mejor modelo ARMA(p, q)(P, Q) para la serie ajustada (W_t) según el criterio BIC.

Para validar los resultados se compara el p-valor del test de Ljung-Box del ARMA ajustado a la serie original (Y_t) con el de la serie ajustada (W_t), de tal modo que si se incrementa el p-valor del test de Ljung-Box del ARMA ajustado a W_t se considera idónea a la variable climática como variable regresora.

RESULTADOS

Paria (2003-2009)

	Regresor	σ^2	p-valor de LB test	BIC	p	P	q	Q	k
Y	no	0,006	9,3E-05	-678,60	2	2	2	0	7
W_t	Temperatura	0,006	7,2E-04	-682,51	2	2	0	0	5
W_t	Humedad	0,012	1,9E-01	-457,86	1	2	2	0	6
W_t	Precipitacion	0,009	1,4E-01	-557,92	2	2	1	0	6

Se puede ver que la aplicación de la modelización ARMAX es satisfactoria en la medida que el p-valor del test de Ljung-Box incrementa (aunque ese incremento sea bastante

pequeño con la temperatura como regresor externo), lo cual hace que el $ARMA(p, q)(P, Q)$ estimado sea mucho mas fiable en sus pronósticos. Por el contrario el BIC favorece la temperatura como regresor externo debido sobre todo a que con el número de parámetros del $ARMA(k)$ se reduce y no aumenta el valor de σ^2 .

Atures (1999-2009)

	Regresor	σ^2	p-valor de LB test	BIC	p	P	q	Q	k
Y_t	no	0,064	2,7E-01	81,19	1	1	1	0	4
W_t	Precipitación	0,054	2,9E-01	-13,92	1	1	2	0	5
W_t	Nivel de río	0,058	7,5E-02	18,74	2	1	0	0	4

Si bien el residuo del $ARMA(p, q)(P, Q)$ ajustado a la serie original se distribuye como ruido blanco y la modelización ARMAX no aporta nada en lo que respecta la test de Ljung-Box, la varianza del residuo (σ^2) se reduce, incrementando así la bondad de ajuste tal como lo indica el valor del BIC.

Sifontes

	regresor	σ^2	p-valor de LB test	BIC	p	P	q	Q	k
Sifontes (1999-2009)	Y_t no	1,506	3,3E-02	1716,59	1	0	2	0	4
	W_t Temperatura	1,193	3,0E-07	1594,56	1	0	2	0	4
	W_t Humedad	1,101	4,0E-08	1552,94	1	0	2	0	4
	W_t Precipitación	1,433	9,7E-12	1690,60	1	0	1	1	4
Sifontes (2004-2009)	Y_t	1,271	3,3E-02	851,03	5	2	0	0	8
	W_t Temperatura	1,738	5,6E-07	904,44	1	2	0	0	4
	W_t Humedad	1,626	7,6E-04	899,36	1	2	0	0	4
	W_t Precipitación	2,807	5,1E-09	1033,19	1	2	0	0	4

En Sifontes (1999-2008), si bien el BIC se reduce con cualquiera de las variables usadas

como regresor externo ya que para todos los casos σ^2 disminuye, mientras el número de parámetros se mantiene (k), los resultados no son satisfactorios ya que en todos los casos p-valor del test de Ljung-Box disminuye sensiblemente, por lo que no puede decirse que el residuo se distribuya como ruido blanco lo cual quita validez al ARMA estimado.

Suponiendo que tal vez el incremento tan evidente en la incidencia de casos de malaria que se observaba el registro del IPS de Sifontes a partir de 2004 podría haber afectado las estimaciones, sólo se usó el registro de IPS desde 2004 hasta 2008. En este caso no sólo se disminuye el valor del p-valor del test de Ljung-Box sino que además el BIC aumentó, a pesar de que se redujo a la mitad el número de parámetros (k). De modo que en Sifontes es posible que la tendencia no esté supeditada a factores climáticos, sino a factores de otro tipo (la más posible que sean de carácter antrópicas, como la actividad minera en el área).

REFERENCIAS

- [1] Brockwell P, David R, Introduction to Time Series and Forecasting, Springer-Verlag, 2002.
- [2] Chowell G et al, The spatial and temporal patterns of falciparum and vivax malaria in Perú: 1994-2006, Malaria Journal 2009, 8:142.
- [3] Grillet M.E, Barrera R, Disentangling the Effect of Local and Global Spatial Variation on a Mosquito-Borne Infection in a Neotropical Heterogeneous Environment. Am. J. Trop. Med. Hyg., 82(2), 2010, pp. 194-201.
- [4] SAEZ-SAEZ, Vidal, MARTINEZ, Jean, RUBIO-PALIS, Yasmín et al. Evaluación semanal de la relación malaria, precipitación y temperatura del aire en la Península de Paria, estado Sucre, Venezuela. Bol Mal Salud Amb, dic. 2007, vol. 47, no.2, p.177-189. ISSN 1690-4648.
- [5] Time Series Modeling: Conditional Mean and Variance Models [publicación en línea]. <http://www.mathworks.com/help/toolbox/econ/f8-82329.html#f8-81155>

APLICACIÓN DE MÉTODOS DE ECUACIONES DIFERENCIALES ESTOCÁSTICAS PARA EL DESARROLLO DE UN MODELO DE DIFUSIÓN DE CONTAMINANTES EN LA SUPERFICIE DEL LAGO DE VALENCIA.

Jocelyn León¹, José León²

*¹Universidad Central de Venezuela
jocelynleon@gmail.com*

*²Universidad Central de Venezuela
jleon@euler.ciens.ucv.ve*

RESUMEN

El área circundante al Lago de Valencia está rodeada de complejos industriales, zonas residenciales y ríos tributarios que son descarga de grandes cantidades de sustancias químicas, entre ellas, los detergentes sintéticos de origen industrial y doméstico. Con esta situación se ha venido presentando un grave desequilibrio ecológico que ha estado provocando una serie de problemas ambientales.

De allí la importancia de realizar estudios que permitan, considerando las características de la región, tomar las medidas necesarias para la recuperación del Lago de Valencia. Uno de los aspectos de gran importancia, relacionado con diversas áreas de investigación, es la simulación de la trayectoria de los contaminantes, de tal modo que sea posible una toma de decisión adecuada para atacar el problema ambiental.

Estudiamos el comportamiento y trayectoria en dos dimensiones del contaminante mediante el proceso de difusión de sus partículas, cuya suma constituye hipotéticamente el

derrame del contaminante. El desplazamiento de cada partícula está determinado por ecuaciones diferenciales estocásticas. Utilizamos como método de aproximación numérica a las ecuaciones diferenciales estocásticas propuestas el *método de aproximación de Milstein*. Se considera como punto de inicio del proceso de difusión los afluentes del Lago, por ser estos los de mayor incidencia en el índice de contaminación presente en el Lago de Valencia. En cuanto a la representación de condiciones y procesos que se dan a lugar en el Lago, se cuenta con una data que contiene las velocidades presentes en el medio en diversos puntos del Lago, además, se tomarán en consideración los procesos de evaporación y de hundimiento de partículas, así como también el comportamiento del derrame al aproximarse al borde del Lago. Complementamos el estudio incorporando una imagen satelital del Lago de Valencia, que nos permite ilustrar el esparcimiento de contaminante, bajo determinadas condiciones, con el uso de técnicas de procesamiento digital de imágenes.

METODOLOGÍA

La descripción del movimiento de los fluidos se puede realizar desde dos enfoques, el enfoque Euleriano y el enfoque Lagrangiano. Esta investigación está sustentada bajo el enfoque Lagrangiano. Se verá el fluido de contaminante como un gran número de partículas cuya trayectoria está determinada por las corrientes debidas a la acción del viento, del oleaje y del flujo del agua vertida por los ríos que caen en el Lago. Cada partícula es gobernada por una fuerza que la podemos dividir en dos categorías, Una fuerza determinista (fuerza viscosa continua) y una fuerza estocástica (proveniente de la discontinuidad de la materia), (ver referencias [2] y [5]). Es aquí donde entra en juego una de las ramas ampliamente desarrollada de la probabilidad, con aplicaciones importantes dentro y fuera de la matemática, estamos hablando de la teoría de los procesos estocásticos.

Será base de nuestra investigación el proceso estocástico ampliamente estudiado denominado movimiento Browniano o proceso de Wiener. Otro concepto clave lo constituye la integral estocástica. La integral estocástica surge de la necesidad de resolver determinados problemas, en los que aparecen implicados ciertos procesos estocásticos, como el Movimiento Browniano, por ejemplo las integrales del tipo

$$\int_0^t f(s, w) dW_s(w).$$

El hecho de que las trayectorias de los Movimientos Brownianos no sean diferenciables ni de variación acotada (ver referencias [6], [5] y [1]) impide integrar respecto a éste movimiento en el sentido de Riemann- Stieltjes. Surge por tanto la necesidad de crear una nueva integral, que en casos de regularidad del integrando sí coincidirá con la integral de Riemann-Stieltjes.

Esta integral es la denominada integral de Itô. Además, Itô nos dota de una herramienta de cálculo enunciada en el lema de Itô, en su versión más simple (para una función dos veces continuamente diferenciable), que es la llamada fórmula de Itô.

Existen tres extensiones para el lema de Itô,

- La extensión I, para procesos que dependen tanto del tiempo como de la trayectoria Browniana, entonces se tienen funciones del tipo $f(t, x)$.
- La extensión II, es para procesos que dependen del tiempo y de procesos de Itô, los procesos de Itô son de la forma:

$$X_t = X_0 + \int_0^t A_s^{(1)} ds + \int_0^t A_s^{(2)} dW_s,$$

$A^{(1)}$ y $A^{(2)}$ son adaptados al movimiento Browniano.

- La extensión III, para procesos estocásticos que dependan del tiempo y de dos procesos de Itô.

Ahora bien, como lo hemos mencionado, cada partícula estará influenciada por dos fuerzas, una aleatoria y otra determinista, por lo que requerimos de un modelo que describa una trayectoria que presente perturbación, pero guarde cierta estructura determinista. El modelo que está propuesto, dentro del enfoque Langragiano, en este trabajo de investigación es una ecuación diferencial estocástica, que claramente cumple con las condiciones antes expuestas, (ver referencias [6] y [5]).

Hemos hallado un modelo matemático que nos permitirá simular la trayectoria de las partículas de contaminante. Sin embargo, no basta tener la definición del modelo, pues pueden surgir preguntas como: ¿este tipo de ecuación tiene solución?, y en caso de tenerla, ¿es única?. El teorema de existencia y unicidad de las ecuaciones diferenciales estocásticas nos responde afirmativamente las interrogantes planteadas. No obstante, las ecuaciones diferenciales estocásticas que admiten una solución explícita son una excepción a la regla, por ende necesitamos hallar una técnica que nos permita aproximar esta solución explícita, a dicha aproximación se le denomina solución numérica (ver referencia [3] y [5]).

Una manera sencilla de hallar la aproximación numérica es considerar los diferenciales como diferencias, es así como se obtiene el esquema de aproximación de Euler. Por otro lado tenemos el esquema de aproximación de Milstein, en el cual se aplicará a los integrandos el lema de Itô (extensión II), hallándose un término de corrección adicional que contiene los incrementos cuadrados del movimiento Browniano, (ver referencias [6] y [5]).

En este caso, el modelo específico a utilizar está representado por la siguiente ecuación diferencial estocástica,

$$dX_t = \begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} + \begin{pmatrix} b_1(X_t^1, X_t^2) \\ b_2(X_t^1, X_t^2) \end{pmatrix} dt$$

donde $(b_1(X_t^1, X_t^2), b_2(X_t^1, X_t^2))$ está dado por el flujo de las corrientes superficiales del Lago

de Valencia, (dW_t^1, dW_t^2) es un movimiento Browniano estándar bidimensional, a_{11} y a_{22} constantes. Esta ecuación será aproximada por el método de aproximación de Milstein.

Una vez que se tiene la aproximación numérica de la ecuación diferencial estocástica, se hará uso de técnicas básicas de procesamiento digital de imágenes (ver referencia [4]) para el desarrollo de una interfaz gráfica que nos permita apreciar la solución bajo el entorno aportado por una imagen satelital del Lago de Valencia.

Finalmente, utilizando técnicas predicción espacial, Kriging (ver referencia [8]), se estimará la difusión de una sustancia contaminante presente en el Lago, y posteriormente se comparará con la solución del modelo planteado con respecto a dicha sustancia.

RESULTADOS Y CONCLUSIONES

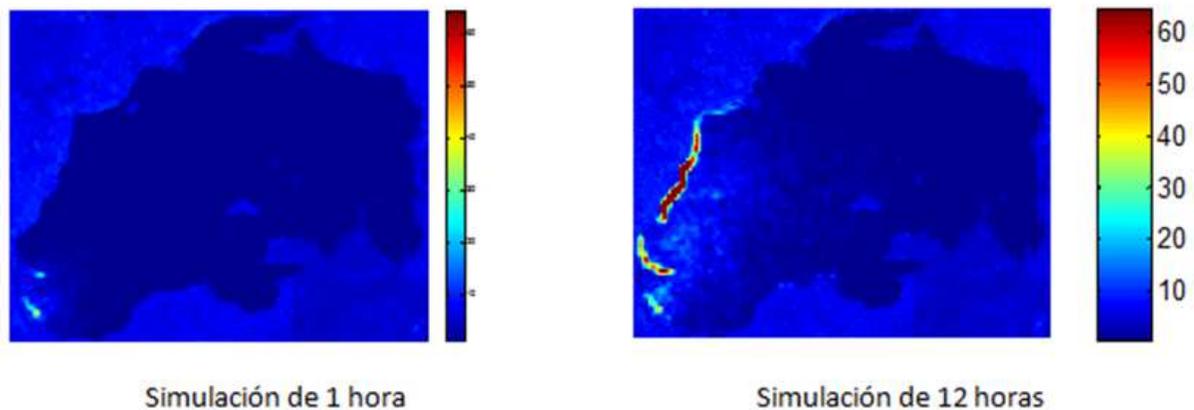


Figura 1: Solución del Modelo de Dispersión.

En la Figura 1, se muestra la solución obtenida en una simulación de 1 hora y en una de 12 horas. Se observa que el desplazamiento se orienta hacia la parte norte del Lago, y esto tiene sentido ya que en las estimaciones de las Velocidades del Lago obtenidas la componente Y de

la velocidad toma valores superiores a los de la componente X en esa región. Para verificar que efectivamente los valores obtenidos tienen sentido, se tomó la información referente a la velocidad del viento en el Lago, donde se tiene que esta varía de 4 a 13 Km/h . Al multiplicar el intervalo por 0.03 tenemos una aproximación en Km/h de la variación de las corrientes superficiales inducidas por el viento, como nuestra simulación es la capa superficial del Lago, estas cotas en m/s deben coincidir, a menos en el orden, con las cotas de las velocidades de las corrientes estimadas por Kriging. Al satisfacerse este razonamiento, se verificó que el desplazamiento promedio de las partículas estuviese en concordancia con el desplazamiento lineal que deberían tener con respecto a la velocidad con la que se mueven, por ejemplo, en el caso de la simulación de 1 hora, si las partículas se mueven entre $[0,0203, 0,1040]$ m/s equivalente a $[73,08, 374,4]$ m/h , entonces el promedio de desplazamiento de una partícula en una hora debe estar dentro del intervalo $[73,08, 374,4]$ m , en este caso se obtuvo 133,5 m . Para la simulación de 12 horas, se tiene un intervalo de $[876,96, 4492,8]$ m , y se obtuvo que en promedio se desplazó 1157,77 m . El tiempo de ejecución, para 100 partículas, fue en la simulación de 1 hora de 37.333821 segundos y en el caso de 12 horas fue de 411.637397 segundos. Por lo tanto, podemos afirmar que se observa el efecto del campo de velocidades, y además los desplazamientos obtenidos guardan concordancia con los planteamientos teóricos en los que nos hemos basado, con lo que el comportamiento de la dispersión reproduce en cierta manera lo que sucede en la realidad.

REFERENCIAS

- [1] Feller, W. An Introduction to Probability Theory and Its Applications. (1957) 437-474.
- [2] Flores T., Henry "Estudio Numérico Experimental del Esparcimiento de Hidrocarburos en la Superficie del Agua", Universidad Central de Venezuela, 17-38 (1996).
- [3] Gihman, I.I. and Skorohod, A.V. Stochastic Differential Equations. Springer-Verlag, 39-43 (1972).

- [4] Gonzalez, Rafael C. Digital Image Processing. Second Edition by Prentice-Hall,(2002).
- [5] León P., Jocelyn "Aplicación del cálculo estocástico en el modelaje del esparcimiento de contaminantes en la superficie del agua", Universidad Central de Venezuela, 5-52 (2008).
- [6] Mikosch, T. Elementary Stochastic Calculus. World Scientific.(1998).
- [7] Gihman, I.I. AND Skorohod, A.V. Stochastic Differential Equations. Springer-Verlag (1972) 14-163.
- [8] Schabenberger, Oliver and Gotway, Carol A. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC(2004).

UNA COOPERACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES CON EL ANÁLISIS ENVOLVENTE DE DATOS.

Lic. Tomás León¹, Dr. Ricardo Ríos²

¹*Universidad Central de Venezuela*
tomasleon2@gmail.com

²*Universidad Central de Venezuela*
ricardo.rios@ciens.ucv.ve

RESUMEN

Actualmente la toma de decisiones en muchos ámbitos se hace de manera subjetiva. Junto con esto, están técnicas o métodos estadísticos que se usan para tratar de realizar dicha tarea de manera más objetiva y así equilibrar el proceso subjetivo intrínseco en el proceso.

Para realizar dicha tarea muchas veces se habla de la palabra *eficiencia* donde se toman en cuenta las entradas (INPUT) y las salidas (OUTPUT) del sistema, unidad u objeto, siendo entonces una unidad más eficiente si produce una mayor cantidad de salidas usando una pequeña cantidad de entradas. Por lo tanto, lo que se busca es medir cuáles unidades son eficientes y cuáles no, en el caso de las ineficientes se busca como mejorarlas, ya sea aumentando las salidas una cierta cantidad y dejando las entradas constante o reduciendo las entradas una cierta cantidad y dejando las salidas constante. Dado esto, la toma de decisiones debe de orientarse de cierta manera hacia la búsqueda de esas cantidades, para así lograr la eficiencia de las unidades, orientando entonces el proceso de toma de decisiones hacia dicha tarea.

En el presente trabajo se tomaran en cuenta dos procesos orientados hacia la búsqueda de la eficiencia, por un lado el Análisis Envlovente de Datos (DEA por sus siglas en inglés)

y por otro lado el Análisis de Componentes Principales (ACP), con la finalidad de describir ambos métodos en cuanto a la toma de decisiones, para luego buscar la manera de integrar ambos sin pérdida de generalidad, tratando entonces de incorporar las bondades de ambos métodos, y viendo como opera la toma de decisiones subjetivas en el marco del Análisis Envolverte de Datos.

METODOLOGÍA

Para probar la eficiencia de la Unidades se uso la metodología DEA, y la metodología ACP, con la finalidad de ver el comportamiento de ambos en términos de medir a las unidades, esto dado que ambos métodos presentan una serie de ventajas y desventajas; para el trabajo se usaron datos reales y sintéticos, a los cuales se les aplico la técnica bootstrap para obtener una mayor variabilidad dentro de los datos y un mayor número de muestras para así ver el comportamiento de las metodologías antes mencionadas. También se usó una integración de ambos métodos, con la finalidad de ver si era posible involucrar las ventajas de ambas metodologías, y así tratar que el proceso de toma de decisiones sea más ventajoso y con menos posibilidad de error.

RESULTADOS Y CONCLUSIONES

Mediante la técnica Bootstrap se obtuvieron diferentes muestras a las cuales se les aplicaron las metodologías antes planteadas, se pudo observar que el DEA es muy sensible a pequeños cambios, a la introducción o extracción de Unidades, mientras que el ACP es un método más robusto ante variaciones, inclusión o exclusión de más Unidades; también fue posible aplicar el método integrado de ACP con DEA y de igual manera como se esperaba los resultados fueron mejores que con solo la aplicación del DEA, lo cual brinda entonces

muchas ventajas, dado que el ACP solo sirve para obtener una posición de las Unidades brindando así solo conocimiento de cuales serian las eficientes y cuales ineficientes, mientras que el DEA brinda mucho conocimiento sobre como tratar a las unidades ineficientes para convertirlas en eficientes y de igual manera un posicionamiento de las unidades dependiendo de su eficiencia. Pudiendo concluir entonces que para la toma de decisiones es mejor aplicar el método integrado dado que se involucran las ventajas de ambas metodologías y presenta más robustes ante variaciones que pueden ser intrínsecas a la realidad.

REFERENCIAS

- [1] Arieu A. **Eficiencia técnica comparada en elevadores de granos de Argentina, bajo una aplicación de análisis de envolvente de datos. La situación del puerto de Bahía Blanca.** Universidad Tecnológica Nacional. Consorcio de gestión del Puerto de Bahía Blanca.
- [2] Casimiro A, Díaz L, Martínez M. **Análisis de la eficiencia e innovación mediante DEA: Aplicación a la industria forestal española.** Departamento de Economía y Gestión forestal. ETS. Madrid.
- [3] Charnes A, Cooper W, Rhodes E. (1978). **Measuring the efficiency of decision making units.** European Journal of Operational Research. Volumen 2 Pag: 429-444.
- [4] Cooper W, Seiford L, Tone, K. (2007). **DATA ENVELOPMENT ANALYSIS, A Comprehensive Text with Models, Applications, References and DEA-Solver Software.** Second Edition, Springer, New York.
- [5] Peña D. (2002). **Análisis de datos multivariados.** McGraw Hill, Madrid, España
- [6] Shanmugam R, Johnson, C. (2007). **At a crossroad of data envelopment and principal component analyses.** OMEGA. *the International Journal of Management Science.* Volumen (35). Pag: 351-364.
- [7] Serrano V, Blasco O. (2006). **Evaluación de la eficiencia mediante el análisis envolvente de datos. Introducción a los modelos básicos.** Universidad de Valencia.
- [8] Zhu, J. (1998). **Data envelopment analyses vs. principal component analysis: An illustrative study of economic performance of chinese cities.** European Journal of Operational Research. Volumen 111 Pag: 50-61.

- [9] Zuleyka Diaz-Martinez y Jose Fernandez-Menendez. (2008). **DEA: Data Envelopment Analysis**. R package version 0.1-2.
- [10] R Development Core Team (2009). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

PROGRAMA QUE APROXIMA LA FUNCIÓN VOLTAJE, CORRIENTE Y RESISTENCIA EFECTIVA DE REDES ELÉCTRICAS EN GRAFOS CONECTADOS, POR MEDIO DE SIMULACIONES DE CAMINATAS ALEATORIAS SOBRE LOS GRAFOS.

Miguel A. Padrino

Universidad de Oriente
padrinomiguel@gmail.com

RESUMEN

Se diseña un programa que da aproximaciones del análisis, de redes eléctricas resistivas sobre grafos conectados, que además da una aproximación de la resistencia efectiva de la red, para los casos donde la batería conectada es de r voltios y la diferencia de potencial aplicada es tal que $v(1) = 0$ y $v(N) = 0$ o a la inversa $v(1) = r$ y $v(N) = r$, donde $v(k)$ es la función voltaje.

METODOLOGÍA

Dado un conjunto de puntos $S = \{1, 2, \dots, N\}$, llamados nodos y una relación c de S en S , llamada relación de conexión, se tiene el grafo $G = (S, c)$, si el grafo no es muy complicado este puede ser representado gráficamente uniendo los nodos por líneas o lazos, según la relación de conexión. Un grafo será llamado conectado si dado dos nodos cualesquiera existe un camino de nodos conectados, según la relación de conexión, que los une. Si a cada par de nodos k y j conectados se le asocia un número real, $\infty > R_{k,j} > 0$, llamado

resistencia, entonces se tiene una red eléctrica en un grafo conectado. En algunos cálculos eléctricos se emplea el inverso de la resistencia, que se denomina conductancia, en el caso planteado se denota por $C_{k,j}$ a la conductancia entre los nodos k y j , si estos están conectados. Para indicar que los nodos k y j están conectados según la relación de conexión se escribe $k \vee j$. Se denota por C_k a la conductancia en el nodo k , la cual es definida como la suma de las conductancias $C_{k,j}$ tales que $j \vee k$, es decir, $C_k = \sum_{j:j \vee k} C_{k,j}$.

A la red eléctrica sobre el grafo conectado $G = (S, c)$ se le conecta una batería de un voltio en los nodos 1 y N , aplicando una diferencia de potencial tal que $v(k) = r$ y $v(N) = 0$, donde $v(k)$ para cada $k \in S$, es la función voltaje. El conjunto de los nodos donde se conecta la batería es llamado la frontera de la red y denotado por $F = \{1, N\}$, el resto de los nodos $S - F$, es llamado el conjunto de los nodos interiores, el cual es denotado por I .

Según la diferencia de potencial aplicada, una corriente $i_1 = \sum_{j:j \vee 1} i_{1,j}$, fluirá dentro de la red desde la batería, donde $i_{1,j}$ representa la corriente que fluye del nodo 1 al nodo j si $1 \vee j$. La cantidad de corriente que fluye depende de la resistencia total de la red.

Se puede considerar esta red de frontera $F = \{1, N\}$, como una resistencia simple entre los nodos 1 y N , llamada resistencia efectiva de la red y denotada por R_{ef} .

A esta red eléctrica en el grafo conectado $G = (S, c)$, se le asocia una Cadena de Markov sobre el grafo, considerando el conjunto de los nodos $S = \{1, 2, \dots, N\}$, como el espacio de estados, las probabilidades de transición están asociadas a las conductancias de la red, si dos nodos k y j no están conectados según la relación c , entonces la probabilidad de transición de k a j es 0. Para formalizar estas ideas, se define la sucesión de variables aleatorias $(X_n)_{n \geq 0}$, sobre el espacio de estados S , con matriz de transición Q como:

X_n ; indica la posición aleatoria sobre S en el instante n .

Los elementos $q_{kj} = P(X_{n+1} = j | X_n = k)$ de la matriz de transición $Q = (q_{kj})_{k,j \in S}$ que representan, la probabilidad de saltar al estado j dado que el proceso esta en el estado k , se

toman de la siguiente forma:

$$\begin{cases} \frac{C_{k,j}}{C_k} & \text{si } k \in S \text{ y los estados } k \text{ y } j \text{ estan conectados} \\ 0 & \text{si } k \in S \text{ y los estados } k \text{ y } j \text{ no están conectados} \end{cases}$$

Como las resistencias son finitas, las conductancias deben ser no nulas y como el grafo es conectado, siempre es posible pasar de un estado a otro en un número finito de pasos. Este tipo de cadenas de Markov asociadas a redes eléctricas, reciben el nombre de caminatas aleatorias sobre el grafo conectado $G = (S, c)$, o simplemente caminata aleatoria asociada a la red eléctrica.

Ahora si se impone la condición de que la caminata aleatoria asociada sea absorbente en la frontera de la red $F = 1, N$, entonces la matriz de transición estará dada por $P = (p_{kj})_{k,j \in S}$, donde p_{kj} , la probabilidad de saltar al estado j dado que el proceso esta en el estado k , se toma de la siguiente forma:

$$\begin{cases} \frac{C_{k,j}}{C_k} & \text{si } k \in I \text{ y los estados } k \text{ y } j \text{ estan conectados} \\ 0 & \text{si } k \in I \text{ y los estados } k \text{ y } j \text{ no están conectados} \\ 1 & \text{si } k \in F \text{ y } j = k \\ 0 & \text{si } k \in F \text{ y } j \neq k \end{cases}$$

Por las definiciones de $Q = (q_{kj})_{k,j \in S}$ y $P = (p_{kj})_{k,j \in S}$, es claro que las matrices de transición de las caminatas aleatorias asociadas, la no absorbente y la absorbente en la frontera de la red, coinciden en todas sus filas interiores.

Dada la caminata aleatoria asociada, absorbente en la frontera de S con matriz de transición $P = (p_{kj})_{k,j \in S}$, se define la función, $h(k)$, para cada $k \in S$, como la probabilidad de que la caminata comenzando en el estado k , llegue a la frontera 1 antes que a N .

Según la situación planteada, en Doyle and Snell (2000), se afirma que $v(k) = h(k)$, para cada $k \in S$, esto basado en propiedades de la función voltaje, de la función de probabilidad $h(k)$ y de funciones armónicas en S respecto a la matriz de transición P . Esto último permite en el caso planteado interpretar probabilísticamente la función voltaje en el nodo k , como la probabilidad de que la caminata comenzando en el estado k , llegue a la frontera 1 antes que a N .

En Doyle and Snell (2000), se da la interpretación probabilística de una corriente unitaria fluyendo en la red, tomando una visión ingenua del proceso de conducción eléctrica. Se imagina que partículas positivamente cargadas entran en la red al nodo 1 y deambulan de nodo en nodo hasta que ellas finalmente llegan al nodo N , donde ellas salen de la red. Para determinar la corriente $i_{k,j}$ a lo largo de la rama de k a j , se considera que en el transcurso de esta peregrinación las partículas pueden pasar a lo largo del tiempo una o varias veces la rama de k a j , y en la dirección opuesta de j a k .

Se demuestra que la corriente $i_{k,j}$ es el número neto esperado de movimientos a lo largo del lazo de k a j , donde los movimientos al revés de j a k son contados como negativos. Para ello se considera la caminata aleatoria asociada no absorbente en la frontera de la red, con matriz de transición $Q = (q_{kj})_{k,j \in S}$ y se plantea la siguiente situación, la caminata comienza en el estado 1 y es detenida al llegar al estado N . Si la caminata retorna al estado 1 antes de llegar a N , la caminata continua ya que el estado 1 no es absorbente. Se define la función,

$u(k)$; número esperado de visitas al estado k antes de llegar a N , para cada $k \in S$.

Utilizando esta función se demuestra que si en la frontera de la red eléctrica se impone una diferencia de potencial tal que $v(1) = \frac{u(1)}{C_1}$ y $v(N) = 0$, entonces la corriente que fluye desde k a j , $i_{k,j}$ cumple que,

$$i_{k,j} = u(k)q_{kj} - u(j)q_{jk}$$

Donde $u(k)q_{kj}$, es el número esperado de tiempo que la caminata tarda desde k a j y $u(j)q_{jk}$ es el número esperado de tiempo que tarda de j a k , de lo que se concluye que la corriente $i_{k,j}$ es el valor esperado para el número neto de tiempo que la caminata pasa a lo largo del lazo de k a j .

Una propiedad característica de la corriente $i_{k,j}$, determinada al imponer la diferencia de potencial, $v(1) = \frac{u(1)}{C_1}$ y $v(N) = 0$, es que el total de flujo de corriente entrando en la red en el nodo 1 (y saliendo de N) es 1. En símbolos esto es, $i_1 = \sum_{j:j \neq 1} i_{1,j} = 1$.

Esta corriente unitaria que fluye desde 1 hasta N , no es igual a la obtenida en el problema eléctrico, cuando a la red se le conecta una batería de un voltio, aplicando un diferencia de potencial de forma tal que $v(1) = 1$ y $v(N) = 0$.

Para obtener la interpretación probabilística de la resistencia efectiva de la red eléctrica en cuya frontera se aplica una diferencia de potencial de forma tal que $v(1) = 1$ y $v(N) = 0$, se vuelve a considerar la caminata aleatoria asociada, no absorbente en la frontera de la red, con matriz de transición $Q = (q_{kj})_{k,j \in S}$ y en este caso también se considera la función $h(k)$, para cada $k \in S$, como la probabilidad de que la caminata comenzando en el estado k , llegue a la frontera 1 antes que a N . A partir de la definición de $h(k)$ se tiene que, $1 - h(1)$, es la probabilidad de comenzando en 1, llegar a N antes que retornar a 1, lo cual es llamado la probabilidad de escape de 1 y se denota por $P_1(\text{escape})$. En Ferrari (1987) se afirma y se dan indicaciones para demostrar que:

$$R_{ef} = \frac{1}{C_1 P_1(\text{escape})}.$$

Esto último permite interpretar probabilísticamente la resistencia efectiva en el caso plateado, como el inverso del producto entre la conductancia en el nodo 1 y la probabilidad de escape de 1.

Tanto Doyle and Snell (2000) como Ferrari (1987), plantean formulas que permiten en términos de múltiples simulaciones de caminatas aleatorias sobre el grafo conectado, dar aproximaciones de, la probabilidad de que la caminata comenzando en el estado k , llegue a la frontera 1 antes que a N , número neto esperado de movimientos a lo largo del lazo de k a j y la probabilidad de escape de 1. Estas formulas permiten dar aproximaciones de la función voltaje cuando la diferencia de potencial aplicada es tal que $v(1) = 1$ y $v(N) = 0$, de una corriente unitaria fluyendo en la red y de la resistencia efectiva cuando la diferencia de potencial aplicada también es tal que $v(1) = 1$ y $v(N) = 0$. Además se dan indicaciones para determinar mediante el Teorema Central del Límite y la Ley fuerte de los grandes números el error estadístico cometido en las aproximaciones con un número determinado de simulaciones.

Para diseñar el programa que da aproximaciones del análisis, de redes eléctricas resistivas sobre grafos conectados, que además da una aproximación de la resistencia efectiva de la red, para los casos donde la batería conectada es de r voltios y la diferencia de potencial aplicada es tal que $v(1) = r$ y $v(N) = 0$ o a la inversa $v(1) = 0$ y $v(N) = r$. Se desarrollaran con detalles cada una de las indicaciones de los autores que tratan el tema para lograr demostrar y entender cada uno de los planteamientos antes señalados y utilizando argumentos similares se demostraron las siguientes relaciones, que extienden los resultados a los casos en que la magnitud de la batería es distinta de un voltio.

Si a la red eléctrica sobre el grafo $G = (S, c)$, se le conecta una batería de r voltios en la frontera $F = \{1, N\}$ aplicando una diferencia de potencial tal que $v(1) = r$ y $v(N) = 0$, al considerar la caminata aleatoria sobre el grafo conectado $G = (S, c)$ asociada a la red eléctrica, pero absorbente en la frontera $F = \{1, N\}$, es decir con matriz de transición $P = (p_{kj})_{k,j \in S}$ y denotar por $h(k)$ a la probabilidad de que la caminata comenzando en k finalice en 1 antes que en N , para cada $k \in S$, entonces $v(k) = rh(k)$, para todo $k \in S$. Ahora si la

diferencia de potencial aplicada en la frontera de la red es tal que $v(1) = 0$ y $v(N) = r$, en este caso se tendrá que $v(k) = r(1 - h(k))$, para todo $k \in S$.

Por la definición de resistencia efectiva de una red eléctrica sobre un grafo conectado y como se demostró que el flujo de corriente entrando en la red y la función voltaje, al conectar en la frontera de la red una batería de r voltios, son proporcionales al flujo de corriente y la función voltaje, respectivamente, cuando se conecta en la frontera de la red una batería de un voltio, entonces es claro que el valor de la resistencia efectiva es independiente de la magnitud de la batería que se conecte en la frontera de la red, lo cual permite usar en cualquier caso, la expresión de la resistencia efectiva en términos de la conductancia en el nodo 1 y la probabilidad de escape de 1, planteada en Ferrari (1987).

Ya se sabe que si en la frontera de la red se impone una diferencia de potencial de forma tal que $v(1) = \frac{u(1)}{C_1}$ y $v(N) = 0$, entonces la corriente que fluye desde k a j , $i_{k,j}$ cumple que,

$$i_{k,j} = u(k)q_{kj} - u(j)q_{j,k}.$$

Pero si en la frontera de la red, se aplica una diferencia de potencial tal que $v(1) = r$ y $v(N) = 0$, y se denota por $\bar{i}_{k,j}$ a la corriente que fluye desde k a j , en este caso se demostró que,

$$\bar{i}_{k,j} = r \frac{1}{R_{ef}} \{u(k)q_{kj} - u(j)q_{j,k}\}$$

Esta extensión de los resultados a los casos en que la batería conectada es distinta de un voltio y considerando los casos en que la diferencia de potencial es invertida en la frontera, permite reescribir las formulas que dan las aproximaciones de la función voltaje, la corriente fluyendo en la red y la resistencia efectiva, en términos de múltiples simulaciones de caminatas aleatorias sobre los grafos, para este caso general. Las formulas que dan las aproximaciones se especifican para un número de 10.000 simulaciones de caminatas aleatorias sobre

los grafos y se realizan los cálculos que determinan que con esa cantidad de simulaciones el error cometido es menor a 0,01 con una probabilidad no menor a 0,95.

Para lograr simular las caminatas aleatorias asociadas, se propone un método de simulación de cadenas de Markov, sobre un espacio finito de estados, con matriz de transición dada, este método es una adaptación del método de la transformada inversa pero aplicado las filas de la matriz de transición de la cadena de Markov sobre un conjunto finito de estados, para generar valores de la cadena. Dicho método se escribe en forma de algoritmo, basado en ese algoritmo se desarrolla un segundo algoritmo que dada una caminata aleatoria sobre un grafo conectado, permita iniciar una caminata aleatoria en un estado k y en un número finito de pasos finalizar en un estado r .

Con todos los resultados previos se diseña un programa que aproxima la función voltaje, la corriente y la resistencia efectiva de redes eléctricas en grafos conectados, su codificación es realizada en el paquete de cómputo Matlab y su nombre asignado RedesE.

El programa Redes E, le permite a un usuario configurar una red eléctrica en un grafo conectado, las entradas del programa son, cantidad de nodos de la red, valores de las conductancia entre los nodos, magnitud de la batería a conectar en la frontera y diferencia de potencial aplicada, las salidas del programa son las aproximaciones de la función voltaje, magnitudes y direcciones de la corriente fluyendo en la red y la resistencia efectiva.

Para verificar los resultados que se obtienen con el programa, se diseñan cuatro ejemplos de redes eléctricas sobre grafos conectados, a las cuales se les aplica una diferencia de potencial específica en la frontera. Con el programa RedesE se obtienen las aproximaciones correspondientes a cada red eléctrica y estos resultados son comparados con los valores exactos, los cuales son calculados con los procedimientos clásicos de física eléctrica y se observa que el error cometido de las aproximaciones en cada caso, es consistente con el error de aproximación teórico para este tipo de aproximaciones.

RESULTADOS Y CONCLUSIONES

El flujo de corriente entrando en la red y la función voltaje, al conectar en la frontera de la red una batería de r voltios, son proporcionales al flujo de corriente y la función voltaje, respectivamente, cuando se conecta en la frontera de la red una batería de un voltio, entonces es claro que el valor de la resistencia efectiva es independiente de la magnitud de la batería que se conecte en la frontera de la red.

El programa RedesE, dada cualquier red eléctrica en un grafo conectado, a la cual se le aplica una diferencia de potencial en la frontera $F = \{1, N\}$, especificada por el usuario, tiene como salidas aproximaciones de la función voltaje, de la corriente fluyendo en la red y de la resistencia efectiva.

Se plantean cuatro ejemplos de redes eléctricas, cuyos datos fueron introducidos en el programa RedesE obteniendo las aproximaciones del análisis de las redes eléctricas, los cuales al compararlos con los resultados exactos, se puede observar que no difieren en gran medida, como era de esperarse.

REFERENCIAS

- [1] Doyle y Snell, Random walks and electric networks, The Mathematical Association of America, 1984, Version 3.02, 5 January 2000.
- [2] Ferrari, Pablo, Passeios aleatorios e redes eléctricas, 16^o Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 1987.
- [3] Guyon, Xavier, Métodos numéricos por cadenas de Markov, Decimosegunda escuela venezolana de Matemáticas, IVIC, Caracas 1999.
- [4] Norris, J. R, Markov Chains. Cambridge University Press, New York, 2006.

RECONOCIMIENTO LITOLÓGICO A PARTIR DE REGISTROS DE POZO.

M.SC. Eddre Peña¹, Carenne Ludeña²

¹Universidad Central de Venezuela

eddrepen@gmail.com

²Instituto Venezolano de Investigaciones Científicas (IVIC)

cludena@ivic.ve

RESUMEN

Se quiere desarrollar un sistema inteligente que segmente y clasifique a partir de registros geofísicos de pozo, que sea capaz de reconocer la definición (principio y fin) de las capas del subsuelo de un pozo así como también el tipo de litología presente.

El sistema estima las capas litológicas mediante un proceso de segmentación, que minimiza una función de energía definida a partir de los registros, usando una variante del Recocido Simulado, y las clasifica implementando una combinación de Maquinas de Soporte Vectorial bajo entrenamiento supervisado.

METODOLOGÍA

- Preproceso de los datos: detección de datos atípicos y remuestreo de los datos.
- Entrenamiento de los modelos: construcción de dos modelos de clasificación SVMs con núcleo de Laplace, uno basado en maquinas 1-versus 1 y el otro en maquinas 1-versus-resto.

- Segmentación: luego de tener los modelos entrenados debemos construir un vector de rupturas con los datos que queremos clasificar.
- Clasificación de las capas: con el vector de rupturas podemos evaluar los datos en los modelos de clasificación, uno de los modelos nos da la probabilidad de pertenencia a cada clase y el otro nos da la clase a la que pertenece.

RESULTADOS Y CONCLUSIONES

El sistema inteligente de reconocimiento litológico (S.I.R.LI) en promedio presenta un 82 % de buena clasificación, un nivel de certeza superior al 90 % cuando clasifica con una probabilidad superior a 0.7 lo que indica que el sistema es muy confiable si tomamos en cuenta que solo utiliza información de registros.

El usuario de S.I.R.LI esta en la capacidad de conocer la distribución de las facies litológicas con un alto nivel de acierto y detectar al mismo tiempo presencia de litologías complejas al mirar las capas con problemas de clasificación.

Conocer la distribución de las facies litológicas del yacimiento es el primer paso para la caracterización mecánica del mismo, ya que se puede asociar ciertas propiedades a dichas capas. Propiedades como las constantes elásticas estáticas que se derivan de las pruebas de las muestras de rocas en el laboratorio, tales como la medición de las deformaciones para un esfuerzo aplicado, pueden asociarse al tipo litológico de la muestra y así tratar de obtener una relación entre deformación y litología.

REFERENCIAS

- [1] CRAMMER K., SINGER Y. "On the learnability and design of output codes for multi-class problems". *Computational Learning Theory*, 35-46, 2000.
- [2] ESCALONA, J. "Modelos Estadísticos en Selección y Clasificación de Registros de Pozos". *I.V.I.C.*, 2003.
- [3] FRIEDMAN, J. HASTIE, T. TIBSHIRANI, R. "The Elements of Statistical Learning". *Springer*, 2001.
- [4] GONZÁLEZ, L. "Modelos de Clasificación basados en Maquinas de Vectores Soporte". *Universidad de Sevilla*, 2003.
- [5] CHIH-CHUNG C, CHIH-JEN L. "LIBSVM: a library for Support Vector Machines". URL:<http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [6] MARTÍNEZ, Z. "Simulación, Segmentación Y Clasificación Litológica a Partir de Registros de Pozo". *U.C.V.*, 2001.
- [7] PEÑA, E. "Reconocimiento Litológico a Partir de Registros de Pozo ". *U.C.V.*, 2010.
- [8] PLATT, J. "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods". URL:<http://citeseer.nj.nec.com/platt99probabilistic.html>.
- [9] R DOCUMENTATION. "Support vector machines. Package: Kernlab". *R help on Ksvm*.
- [10] RÍOS, R. RODRIGUEZ, L. "Estimación uniparamétrica en procesos autoregresivos con régimen de Markov". Aceptado para su publicación en *Divulgaciones Matemáticas*.
- [11] Schlumberger. "Schlumberger, Interpretación de Perfiles. Volumen I."
- [12] WESTON J., WATKINS C. "Multi-class support vector machines". In *M. Verleysen, Proceedings of ESANN99 Brussels*, 1999.

MODELO ESTOCÁSTICO DE EQUILIBRIO GENERAL DINÁMICO (MEEGD) PARA LA CONSTRUCCIÓN DE DENSIDADES A PRIORI DE VAR BAYESIANOS: UNA APLICACIÓN A LA ECONOMÍA VENEZOLANA.

Mariela Perdomo¹, Daniel Barráez²

¹Banco Central de Venezuela

mariperd@bcv.org.ve

²Universidad Central de Venezuela

dbarraez@bcv.org.ve, daniel.barraez@ciens.ucv.ve

RESUMEN

En este trabajo se estudió y se implementó en el computador el método de Marco Del Negro y Frank Schorfheide para construir densidades a priori de un VAR. Las densidades a priori definidas, permitieron estimar la densidad a posterior usando conjugados naturales. Se implementó esta técnica con el fin de efectuar predicciones para la producción, inflación y tasas de interés. Se realizaron las predicciones y se compararon con un VAR frecuentista y un BVAR de Litterman. Las estimaciones y predicciones se efectuaron para las economías de EEUU y Venezuela.

METODOLOGÍA

La metodología consistió en generar simulaciones de las series macroeconómicas a partir de un MEEGD estimado (Schorfheide [1]), para construir un prior de los parámetros de un VAR. Combinando el prior con la verosimilitud de la serie observada se obtiene una estimación de la densidad a posteriori de los parámetros.

Entre las principales ventajas de la metodología, cabe destacar en primer lugar, que permite construir priors fundamentados económicamente, a diferencia de los priors comúnmente empleados cuya motivación es de naturaleza estadística. Además, permite estimar de manera conjunta los parámetros del BVAR y del MEEGD. Es computacionalmente rápido, puesto que el cálculo de la verosimilitud del BVAR y del MEEGD no requiere del Algoritmo de Kalman, que es la herramienta con la que usualmente se calcula la verosimilitud de los MEEGD.

RESULTADOS Y CONCLUSIONES

La combinación del VAR y del MEEGD presentó un buen desempeño predictivo en el corto plazo para los dos países considerados. En el caso de Venezuela, las predicciones se deterioran al ampliar el horizonte temporal, en comparación con los otros dos VAR. Probablemente, este deterioro se deba al comportamiento irregular de las series macroeconómicas venezolanas, que dificulta su modelaje por un modelo estilizado como el que aquí consideramos. Sería de interés considerar modelos que incorporen especificidades propias de la economía venezolana.

REFERENCIAS

- [1] M. Del Negro and F. Schorfheide. Priors from General Equilibrium Models for VARS. *International Economic Review*, 45(2):643-673, 2004.
- [2] J. Geweke. Using simulation methods for bayesian econometric models: inference, development, and communication. Technical report, 1998.
- [3] R. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1):19-46, January 1976.
- [4] M. Perdomo. *Modelo estocástico de equilibrio general (MEEG) para la construcción de densidades a priori de VAR bayesianos: una aplicación a la economía venezolana*. Trabajo de gra-

do de Maestría, Postgrado en Modelos Aleatorios, Universidad Central de Venezuela, Caracas, 2008.

- [5] F. Schorfheide. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics*, pages 645-670, 2000.
- [6] C. Sims. Solving Linear Rational Expectations Models. *Computational Economics*, 20(1):1-20, 2002.
- [7] F. Smets and R. Wouters. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123-1175, 09 2003.
- [8] M. Woodford. *Interest and prices: foundations of a theory of monetary policy*. Princenton University Press, 2003.
- [9] A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. New York, John Wiley y Sons, 1971.

CARACTERIZACIÓN DE DEUDORES DE TARJETAS DE CRÉDITO A TRAVÉS DE TÉCNICAS DE APRENDIZAJE ESTADÍSTICO.

Carlos David Salazar Paredes¹, Mairene Colina², Carene Ludeña³

¹*Modelos Aleatorios UCV*
salazar.cd@gmail.com

²*Universidad Central de Venezuela*
mairene.colina@ciens.ucv.ve

³*Instituto Venezolano de Investigaciones Científicas (IVIC)*
cludena@ivic.ve

RESUMEN

Haciendo uso de técnicas de aprendizaje estadístico se desea determinar las características de los deudores de tarjetas de crédito de una institución financiera, con el fin de determinar cuáles son las posibles variables asociadas al proceso de caracterización, el cual se lleva a cabo en 2 etapas: la primera es un proceso de clasificación a través de la técnica two steps cluster, el cual permite hacer clasificación para grandes volúmenes de información, donde se permiten variables categóricas y continuas de manera conjunta, y la segunda es estimar el número de veces que el deudor estará en gestión una vez que ingresa al sistema de gestión de cobro, dado el cluster en el que fue clasificado previamente y otras variables, usando árboles de clasificación y regresión logística Multinomial.

METODOLOGÍA

Two steps cluster: el algoritmo de Two steps Cluster es un algoritmo de análisis cluster de manera escalonada, el cual está diseñado para manejar de manera conjunta variables

categorías y variables continuas así como grandes volúmenes de información, y se lleva a cabo en 2 etapas a saber: **Etap 1 (Pre Cluster):** se produce a través de una agrupación previa de los datos en pequeños cluster a través de una aproximación secuencial en el cual se evalúa registro a registro y se determina si este puede ser agregado a alguno de los pequeños cluster formados previamente o bien puede formar algún nuevo sub cluster basado en algún criterio de distancia. **Etap 2 (Cluster):** es esta etapa se toman los subcluster resultantes de la primera etapa y luego se agrupan dentro del número de cluster que se deseen tener como resultados finales, usando el método de agrupamiento jerárquico aglomerativo.

Árboles de Clasificación:

Los árboles de clasificación (también llamados árboles de decisión), es uno de los métodos de aprendizaje estadístico supervisado no paramétrico bastante utilizado, los árboles de clasificación destacan por su sencillez y pueden utilizarse en diversas áreas.

Modo de Construcción del Árbol: Digamos de una manera sencilla que el algoritmo parte de un conjunto de datos que son los denominados datos de entrenamiento los cuales son el resultado de dividir el conjunto inicial de datos en 2 partes estos primeros de entrenamiento y unos segundos llamados de prueba, en el cual los datos están etiquetados con un conjunto de pertenencia, se comienza entonces con un nodo inicial y surge la pregunta de cómo dividir este conjunto de datos en 2 o más partes homogéneas haciendo uso de alguna de las variables disponibles, la cual es escogida de manera tal que la partición resulte de la mejor manera posible, si por ejemplo seleccionamos la variable X_i la división sería de forma tal que en el lado izquierdo del nodo inicial estén los datos donde $X_i \leq c$ y del lado derecho estarán entonces aquellos datos donde $X_i > c$, y luego se repite de modo este proceso el cual termina cuando todas las observaciones hayan sido clasificadas.

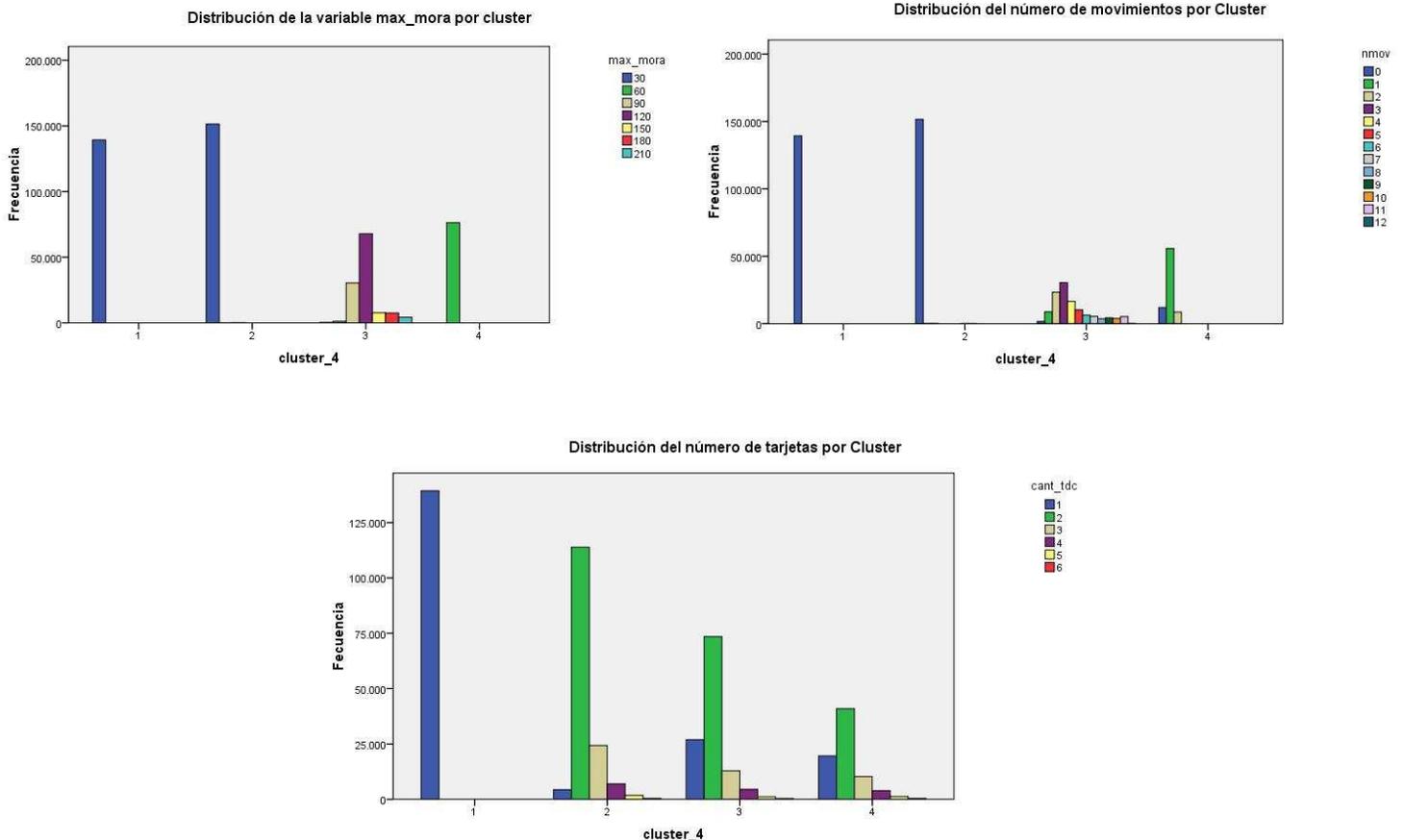
RESULTADOS Y CONCLUSIONES

Descripción de los datos.

Para el caso de estudio de segmentación de los deudores de Tarjetas, se consideraron los clientes que estuvieron al menos 1 vez en el sistema de gestión de cobro desde octubre de 2008 hasta diciembre de 2009 (15 meses) donde a cada uno se le registro a lo largo de este periodo variables cuantitativas como: Monto Vencido (Pago mínimo), Limite de la TDC, Monto Pagado, Luego estos se promediaron, de igual manera, se determino el nivel máximo de mora que había alcanzado, así como una variable que cuantifica el número de veces que el cliente sube de altura de mora entre un mes y otro (nmov), durante los 15 meses del periodo en estudio donde en total se obtuvieron cerca de 496.000 clientes.

Etapa 1: Clasificación por medio de two steps cluster.

Distribución de las variables categóricas por Cluster:



Predicción del número de veces en gestión usando Regresión Logística Multinomial

Correcta clasificación del numero de veces en Gestión a través de Regresion Logistica Multinomial

Observed	Predicted					Percent Correct
	1	2	3	4	5	
1	97645	40	1	0	8	99,9%
2	5629	70355	172	25	48	92,3%
3	89	5631	50284	367	715	88,1%
4	16	696	7692	37748	6249	72,0%
5	7	106	1458	10193	191412	94,2%
Overall Percentage	21,2%	15,8%	12,3%	9,9%	40,8%	92,0%

Predicción del número de veces en gestión (ngest2) usando árboles de clasificación

Correcta clasificación del numero de veces en Gestión a través de Arboles

Sample	Observed	Predicted					Percent Correct
		1	2	3	4	5	
Training	1	78138	240	56	102	44	99,4%
	2	4654	56294	86	143	333	91,5%
	3	122	4404	40033	305	1083	87,1%
	4	26	591	5977	31359	4289	74,2%
	5	56	108	1288	8959	152168	93,6%
	Overall Percentage	21,2%	15,8%	12,1%	10,5%	40,4%	91,6%
Test	1	19510	73	8	18	9	99,4%
	2	1093	13875	19	35	96	91,8%
	3	25	1056	10002	75	298	87,3%
	4	8	140	1508	7971	1141	74,0%
	5	18	41	308	2227	38111	93,6%
	Overall Percentage	21,1%	15,5%	12,1%	10,6%	40,6%	91,6%

Una vez llevada a cabo la caracterización se puede llegar a las siguientes conclusiones.

- Al momento de realizar el proceso de clasificación de los deudores a través de la técnica de two steps cluster las variables categóricas tuvieron predominaron sobre las variables continuas.

- Al momento de predecir el número de veces en gestión que estará un cliente la variable cluster pertenecía obtenida en la primera etapa resultó de ayuda en la construcción de los árboles de clasificación.
- La predicción de manera correcta usando árboles de clasificación para el número de veces que estará el cliente en gestión haciendo uso de los árboles de clasificación superó el 90 % de correcta clasificación.
- La Regresión Logística Multinomial sirvió como técnica que permitió comprobar la eficiencia de los árboles de clasificación en técnicas de aprendizaje estadístico supervisado ya que se obtuvieron resultados similares a los árboles de clasificación.

REFERENCIAS

- [1] Agresti, A. (2007). *Categorical Data Analysis*, 2nd edn. New York, Wiley.
- [2] Friedman J, Hastie T, Tibshiriani R (2008), *the elements of statistical learning*. 2nd Edn. California.
- [3] Zhang, T., Ramakrishnon, R., and Livny M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. *Proceedings of the ACM SIGMOD Conference on Management of Data*, p. 103-114, Montreal, Canada.
- [4] ZHEXUE HUANG, *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*, *Data Mining and Knowledge Discovery* 2, 283-304 (1998).
- [5] *TwoStep Cluster Analysis*, SPSS.

