

## 6. Comunicaciones: Estadística y sus Aplicaciones

### ESTIMACIÓN ROBUSTA EN MODELOS DE INDICE SIMPLE CON ERRORES ASIMÉTRICOS

Expositor: Ana M. Bianco (Universidad de Buenos Aires y CONICET, abianco@dm.uba.ar)  
Autor/es: Ana M. Bianco (Universidad de Buenos Aires y CONICET, abianco@dm.uba.ar);  
Claudio Agostinelli (Dipartimento di Matematica, Università degli Studi di Trento, Trento,  
Italia, claudio.agostinelli@unitn.it); Graciela Boente (Universidad de Buenos Aires y CONICET,  
gboente@dm.uba.ar)

Los Modelos de Índice Simple (SIM) han recibido una creciente atención en los últimos años en la literatura estadística. Una de las razones es que constituyen una interesante estrategia para burlar la conocida *la maldición de la dimensionalidad* al combinar el enfoque paramétrico con el no paramétrico.

En estos modelos se asume que la variable de respuesta  $y$  está relacionada con el vector de covariables  $x$  mediante la ecuación

$$y = \eta_0(\beta_0'x) + \epsilon, \quad (1)$$

donde  $\beta_0, x \in R^q$  and  $\eta_0 : R \rightarrow R$  es una función suave. Tanto la componente paramétrica del modelo,  $\beta_0$ , como la no paramétrica son desconocidas. Para preservar la identificabilidad del modelo, asumiremos, sin pérdida de generalidad, que  $\|\beta_0\| = 1$  y que su última componente  $\beta_q$  es positiva, donde  $\|\cdot\|$  denota la norma euclídea.

Estos modelos reducen la dimensionalidad de las covariables en tanto la función de enlace  $\eta_0$  puede ser estimada no paramétricamente por los índices simples  $\beta_0'x$ , que son univariados.

La mayor parte de la literatura asume que los errores tienen media 0 y varianza finita. Sin embargo, en la literatura estadística se reemplaza este supuesto por el de simetría de los errores a fin de obtener estimadores Fisher-consistentes. En muchas situaciones es necesario modelar errores asimétricos, por ejemplo con distribución en una familia exponencial, tal es el caso de la distribución log-gamma. Nos enfocaremos en el problema de estimar robustamente las componentes paramétrica y no paramétrica del modelo (1) cuando la densidad del error  $\epsilon$  es de la forma

$$g(\epsilon, \alpha) = Q(\alpha) \exp^{at(\epsilon)},$$

con  $\alpha > 0$  un parámetro nuisance parameter y  $t : R \rightarrow R$  es una función continua unimodal. Esta situación incluye la distribución Gamma con función de enlace log.

Se propondrán estimadores de  $\eta_0$  and  $\beta_0$  basados en un procedimiento de tres pasos en el que se combinan técnicas robustas de estimación paramétrica con técnicas de suavizado. Se consideran las propiedades asintóticas de los estimadores propuestos. Se presentarán los resultados de un estudio de simulación preliminar para evaluar el comportamiento de los estimadores propuestos bajo el modelo central y diferentes escenarios de contaminación.

---

### TEST ROBUSTOS PARA SUPERIORIDAD ENTRE DOS CURVAS DE REGRESIÓN

Expositor: Graciela Boente (Universidad de Buenos Aires y CONICET, gboente@dm.uba.ar)  
Autor/es: Graciela Boente (Universidad de Buenos Aires y CONICET, gboente@dm.uba.ar);  
Juan Carlos Pardo Fernandez (Universidad de Vigo, juanpc@uvigo.es)

El análisis de covarianza es una herramienta muy usada para comparar dos poblaciones cuando existen covariables y la relación respuesta-covariable es lineal. Para enfrentar el problema en el que dicha relación no tiene una forma paramétrica específica, se han desarrollado varios

procedimientos de test usando la distancia  $L_2$  entre estimadores de las funciones de regresión o bien contruyendo funciones de distribución empíricas a partir de los residuos del ajuste. Recientemente, para testear una hipótesis unilateral Neumeyer y Pardo-Fernández (2009) introdujeron un test con tasa  $n^{1/2}$  basado en la comparación de promedios de los residuos.

Como es bien sabido, datos atípicos (extremos o de otro tipo) presentan un desafío mayor en modelos de regresión noparamétrica que en modelos paramétricos. Como se menciona en Härdle (1990) “From a data-analytic viewpoint, a nonrobust behavior of the smoother is sometimes undesirable. . . . Any erratic behavior of the nonparametric pilot estimate will cause biased parametric formulations”. Más allá de la importancia de desarrollar estimadores robustos para proveer estimaciones más confiables, el problema de obtener procedimientos robustos de test de hipótesis necesita ser abordado. En el contexto de regresión noparamétrica, la literatura sobre test robustos en bastante mala. Para el caso de una población, Dette y Marchlewski (2010), Bianco *et al.* (2006) y Boente *et al.* (2013) consideraron test para homocedasticidad o para ajustes de la función de regresión, respectivamente. Para comparar dos funciones de regresión, Feng *et al.* (2015) consideran un test que permite detectar alternativas locales pero con tasa  $\sqrt{nh}$  y suponiendo segundo momento para los errores.

En esta charla presentamos una propuesta de test robustos para comparar dos funciones de regresión versus alternativas ordenadas. Nuestra propuesta no supone momentos para los errores del modelo de regresión y el test es capaz de detectar alternativas a tasa  $n^{1/2}$ . Obtenemos la distribución del estadístico del test bajo la hipótesis nula y bajo alternativas contiguas. A través de un estudio de simulación, se estudia la aproximación para muestras finitas del procedimiento basado en los valores críticos asintóticos y se analiza el efecto de introducir errores sin momentos o datos contaminados para mostrar la estabilidad del nuevo procedimiento.

## Bibliografía

- Bianco, A., Boente, G. and Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scandinavian Journal of Statistics*, **2**, 435-450.
- Boente, G., Cao, R., González-Manteiga, W. and Rodríguez, D. (2013). Testing in generalized partially linear models: A robust approach. *Statistics and Probability Letters*, **83**, 203-212.
- Dette, H. and Marchlewski, M. (2010). A robust test for homoscedasticity in nonparametric regression. *Journal of Nonparametric Statistics*, **22**, 723-736
- Feng, L., Zou, C., Wang, Z. and Zhu, L. (2015). Robust comparison of regression curves. *TEST*, **24**, 185-204.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
- Neumeyer, N and Pardo-Fernández, J. C. (2009). A simple test for comparing regression curves versus one-sided alternatives. *Journal of Statistical Planning and Inference*, **139**, 4006-4016.
- Stone, C. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

---

## ESTIMADORES ROBUSTOS BAJO EL MODELO PARCIALMENTE LINEAL ADITIVO

Expositor: Alejandra Mercedes Martinez (Universidad de Buenos Aires y Universidad Nacional de Luján, ale\_m\_martinez@hotmail.com)

Autor/es: Alejandra Mercedes Martinez (Universidad de Buenos Aires y Universidad Nacional de Luján, ale\_m\_martinez@hotmail.com); Graciela Boente (Universidad de Buenos Aires y CONICET, gboente@dm.uba.ar)

Los modelos parcialmente lineales aditivos suponen que se tienen observaciones independientes  $(Y_i, \mathbf{Z}_i^T, \mathbf{X}_i^T)^T$  con  $Y_i \in \mathbb{R}$ ,  $\mathbf{Z}_i \in \mathbb{R}^q$  y  $\mathbf{X}_i \in \mathbb{R}^d$  tales que

$$Y_i = \beta^T \mathbf{Z}_i + g(\mathbf{X}_i) + \epsilon_i$$

donde  $g(\mathbf{x}) = \mu + \sum_{j=1}^d g_j(x_j)$ .

Propuestas de estimación para este modelo pueden encontrarse en Härdle *et al.* (2004).

En este trabajo, presentaremos dos familias de estimadores robustos. La primera familia basada en un procedimiento que consta de tres etapas y que utiliza estimadores basados en integración marginal para la estimación de las componentes aditivas y estimadores robustos de regresión para la estimación de  $\beta$ . La segunda familia de estimadores robustos propuesta combina  $B$ -splines con estimadores robustos de regresión.

Para comparar las propuestas robustas entre ellas y con sus respectivas versiones clásicas se realizó un estudio de simulación bajo diferentes modelos parcialmente lineales aditivos y bajo distintos esquemas de contaminación.

Härdle, W., Müller, M., Sperlich, S. y Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer.

#### MODELOS PARA EL PROPENSITY SCORE QUE CONTEMPLAN LA POSITIVIDAD

Expositor: Julieta Molina (UBA, molina\_julieta@yahoo.com.ar)

Autor/es: Julieta Molina (UBA, molina\_julieta@yahoo.com.ar)

#### Modelos para el propensity score que contemplan la positividad

Supongamos que se dispone de una muestra aleatoria en la que cierto vector  $X$  es observado en cada unidad de la muestra, mientras que  $Y$ , la variable respuesta de interés, sólo es observada en algunos individuos. Denotamos por  $A$  a la variable binaria que indica cuando la respuesta es observada, es decir:  $A = 1$  si  $Y$  es observado y  $A = 0$  caso contrario. A partir de esta información, se desea estimar  $\beta_0 = E(Y)$ . Para poder identificar el parámetro de interés en función de la distribución de los datos observados, se asume que  $Y$  es independiente de  $A$  dado  $X$ . Este supuesto se conoce como missing at random y garantiza que  $E(Y) = E(YA/\pi(X))$ , siendo  $\pi(x) \equiv P(A = 1|X = x)$  el propensity score. Esta identidad sugiere estimar  $\beta_0$  con  $\hat{\beta}_n = \mathbb{P}_n YA / \hat{\pi}(X)$ , donde  $\hat{\pi}(x)$  estima el propensity score y  $\mathbb{P}_n$  representa la media empírica, definida por  $\mathbb{P}_n O = n^{-1} \sum_{i=1}^n O_i$ . El comportamiento asintótico de  $\hat{\beta}_n$  ha sido estudiado considerando estimadores paramétricos y no paramétricos de  $\pi(x)$ . Por lo general, se asume la llamada hipótesis de positividad, la cual establece que  $\pi(x) > b > 0$ . Sin embargo, los modelos de mayor uso para  $\pi$  raramente verifican dicha hipótesis. En este trabajo proponemos una familia de modelos para el propensity score que verifican la hipótesis de positividad pues cada elemento de esta familia toma valores en el intervalo  $(\epsilon, 1)$  siendo  $\epsilon > 0$  uno de los parámetros del modelo. Más precisamente asumimos que

$$\pi(x) = (1 - \epsilon)F(\beta^T x) + \epsilon, \tag{24}$$

donde  $F$  es una función link conocida,  $\epsilon \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^p$  y estimamos  $(\epsilon, \beta)$  mediante el método de máxima verosimilitud. Hacemos un estudio de Monte Carlo en el cual generamos un vector de covariables  $X$ , un vector  $A$  de forma tal que se verifique el modelo (1) para  $F(t) = \frac{e^t}{1+e^t}$  y una

variable respuesta  $Y$  faltante de acuerdo a este mecanismo de pérdida y calculamos el estimador ipw para la esperanza de  $Y$ .

---

## ESTIMACIÓN DE MÁXIMA $L_q$ -VEROSIMILITUD EN MODELOS CON ERRORES EN LAS VARIABLES FUNCIONALES

Expositor: Lucas Gastón Guarracino (Universidad Nacional de Mar del Plata, lucasguarra@hotmail.com)

Autor/es: Lucas Gastón Guarracino (Universidad Nacional de Mar del Plata, lucasguarra@hotmail.com);  
Patricia Cristina Giménez (Universidad Nacional de Mar del Plata, pcgimene@mdp.edu.ar)

En inferencia paramétrica, los métodos basados en la minimización de una medida de divergencia entre la densidad de los datos y la densidad del modelo, proveen estimadores con buenas propiedades de robustez respecto de los estimadores de máxima verosimilitud. Ha sido mostrado que el enfoque basado en la familia de divergencias potencia o  $q$ -divergencias, resulta equivalente a la minimización de una versión empírica de la entropía de Tsallis-Havdra-Charvat o a la maximización de la  $L_q$ -verosimilitud. El método resulta totalmente paramétrico, evitando los problemas que acarrea la estimación no paramétrica de densidades y preserva Fisher consistencia a partir de una transformación para reescalar el parámetro. La constante  $q$  controla el balance entre eficiencia y robustez. El estimador de máxima verosimilitud es obtenido como caso particular para  $q = 1$ .

En este trabajo estudiamos estimación de máxima  $L_q$ -verosimilitud en presencia de parámetros incidentales. El problema de estimar una relación funcional lineal, también conocido como estimación en modelos con errores en las variables funcionales es un importante ejemplo de este tipo de problema de estimación y será abordado especialmente. Resultados asintóticos y propiedades de robustez de los estimadores de los parámetros estructurales del modelo serán presentados. El comportamiento de los estimadores también será analizado mediante estudios de simulación.

---

## UN NUEVO MÉTODO DE REGISTRACIÓN PARA DATOS FUNCIONALES.

Expositor: Manuel Benjamin (IMAS UBA CONICET, manuelbenj@gmail.com)

Autor/es: Manuel Benjamin (IMAS UBA CONICET, manuelbenj@gmail.com)

La media muestral y los componentes principales funcionales son herramientas utilizadas habitualmente en el análisis de una muestra de curvas aleatorias. Sin embargo la calidad de la información que revelan puede verse seriamente afectada por la presencia de variación de fase. Las técnicas de registración apuntan a descomponer la variabilidad de la muestra en dos tipos: de fase (horizontal) y de amplitud (vertical). La primera es capturada por una función monótona creciente (“warping”) que deforma el dominio de las curvas de manera que al transformarlas solo presentan variación de amplitud. En la literatura existen numerosos métodos de registración que buscan alinear aspectos visibles o de forma de las curvas (máximos y mínimos o ceros de las derivadas). Sin embargo, estos procedimientos pueden conducir a análisis estadísticos poco satisfactorios.

Se presenta una propuesta para un nuevo procedimiento de registración que intenta superar las dificultades de los métodos descriptos. La misma se basa en los fundamentos del trabajo de *Kneip y Ramsay* donde las curvas registradas viven en un espacio de baja dimensión. Asimismo,

se exhiben las diferencias con la metodología de *Wagner y Kneip* también motivada por el mismo trabajo.

Referencias:

1. Kneip y Ramsay. *Combining registration and fitting for functional models*, (2008). Journal of the American Statistical Association Vol. 103, No. 483.
2. Wagner y Kneip. *Nonparametric Registration to Low-Dimensional Function Spaces*. (A ser publicado)

---

UNA MEDIDA DE PROFUNDIDAD LOCAL DEFINIDA EN ESPACIOS DE BANACH BASADA EN PROYECCIONES AL AZAR

Expositor: Lucas Raul Fernandez Piana (IMAS-CONICET, lucasfernandezpiana@gmail.com)  
Autor/es: Lucas Raul Fernandez Piana (IMAS-CONICET, lucasfernandezpiana@gmail.com);  
Marcela Svarc (Universidad de San Andrés, marcelasvarc@gmail.com)

Recientemente, con el objetivo de estudiar distribuciones multimodales a través de profundidades se introdujeron en la literatura la definiciones de *profundidades locales*. Las mismas buscan otorgar mayor profundidad a aquellas regiones más centrales y de este modo poder detectar regiones de centralidad local o caracterizar fluctuaciones locales. Las primeras definiciones de profundidades locales fueron dadas, para datos multivariados, Agostinelli y Romanazzi (2011) extienden los conceptos de profundidades simplicial y de semi-espacios y Paindavaine y Van Bever (2013) proponen una definición basada en entornos locales. Por otra parte, Agostinelli (2016) da una propuesta para el caso de datos funcionales, extendiendo el concepto de profundidades de semiespacios. Nuestro objetivo es dar una definición de profundidad local general, definida para elementos aleatorios en espacios de Banach. Como antecedente, Cuevas y Fraiman (2009), dieron una definición de profundidades para espacios generales basadas en proyecciones al azar. La idea se basa en combinar proyecciones unidimensionales y la noción de profundidad local unidimensional. Mostramos que la medida de profundidad local propuesta, bajo condiciones generales, satisface las propiedades básicas de las profundidades. Por otra parte, mostramos que el procedimiento es fuertemente consistente. Una de las principales ventajas de la técnica propuesta es su bajo costo computacional, lo que permite trabajar con datos en dimensiones altas. Dentro del marco de las aplicaciones, proponemos un criterio de cluster basado en profundidades locales, ejemplificamos su aplicación con ejemplos para datos multivariados y funcionales.

### Referencias

- Agostinelli, C. (2016). "Local half-region depth for functional data." *Preprint*
- Agostinelli, C. y Romanazzi, M. (2011). "Local Depth." *Journal of Statistical Planning and Inference*, **141**, 817-830.
- Cuevas, A. y Fraiman, R. (2009). "On depth measures and dual statistics. A methodology for dealing with general data." *Journal of Multivariate Analysis*, **100**(4), 753-766.
- Paindavaine y Van Bever (2013). "From depth to local depth: A focus in centrality." *Journal of the American Statistical Association*, **108**(503), 1105-1119.
-

Expositor: Maria Antonella Gieco (CONICET - Facultad de Ingenieria Quimica, UNL, antogiego@gmail.com)

Autor/es: Liliana Forzani (CONICET - Facultad de Ingenieria Quimica, UNL, liliana.forzani@gmail.com);

Maria Antonella Gieco (CONICET - Facultad de Ingenieria Quimica, UNL, antogiego@gmail.com);

Carlos Tolmasky (Institute for Mathematics and Its Applications and MCFAM, University of Minnesota, tolmasky@ima.umn.edu)

En aplicaciones que involucran datos en alta dimensión, frecuentemente unas pocas variables contienen la mayor parte de la información relevante. Resulta así de gran importancia identificar cuál es la dimensión adecuada para considerar en el análisis. Con el objetivo de comprender este tipo de situaciones, Johnstone [1] introdujo los modelos de covarianza “spiked”, en el cual todos los autovalores poblacionales, con excepción de los primeros (mayores, fijos) son iguales a uno. El comportamiento de los autovalores muestrales correspondientes a estos modelos ha sido ampliamente estudiado en la década pasada [2,3,4]. El problema de estimar cuántas componentes deben ser consideradas ha sido estudiado y resuelto en el caso tradicional de  $p$  (dimensión de los datos) fijo y  $n$  (tamaño de la muestra) creciendo a infinito [5]. En escenarios de alta dimensión, sin embargo, los estadísticos usuales ya no pueden aplicarse y la teoría estadística tradicional no resulta adecuada ni tampoco fácil de adaptar. Una de las contribuciones de este trabajo es estudiar la distribución asintótica del cociente de verosimilitud logarítmico bajo la hipótesis nula para el caso  $p, n \rightarrow \infty, p/n \rightarrow y \in (0, 1)$ . También se introduce una nueva definición de este estadístico adaptada al caso de ultra-alta dimensión; es decir,  $p, n \rightarrow \infty, p/n \rightarrow y \in [1, \infty)$ . Se estudia la distribución de este nuevo estadístico bajo la hipótesis nula y se extiende la definición y consistencia del criterio MLD (Minimum Description Length) a este caso. Se estudia la potencia de los tests propuestos y a partir de este análisis se introduce también una corrección que permite testear la dimensión del subespacio spiked incluso para valores de  $p/n$  cercanos a 1, situación en la que otros enfoques resultan deficientes [6].

## Referencias

- [1] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29, 295–327.
  - [2] Baik, J. and Silverstein, J. W. (2006). Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models. *Journal of Multivariate Analysis*, 97, 1382–1408.
  - [3] Debashis, P. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17, 1617–1642.
  - [4] Baik, J. and Ben Arous, G. (2005). Phase transition for the largest eigenvalue of large sample covariance matrices. *Annals of Probability*, 33, 1643–1697.
  - [5] Muirhead, R. J. (2005). Aspects of Multivariate Statistical Theory. *Wiley Series in Probability and Statistics*.
  - [6] Kritchman, S. and Nadler, B. (2009). Non-Parametric Detection of the Number of Signals: Hypothesis Testing and Random Matrix Theory. *IEEE Transactions on Signal Processing*, 57, 3930–3941.
-

## ESTIMACIÓN ROBUSTA BASADA EN INTERDISTANCIAS PARA CORRELACIÓN CANÓNICA FUNCIONAL.

Expositor: Nadia L. Kudraszow (IMAS-UBA-CONICET/UNLP, nkudraszow@mate.unlp.edu.ar)

Autor/es: Graciela Boente (IMAS-UBA-CONICET, gboente@dm.uba.ar); Nadia L. Kudraszow (IMAS-UBA-CONICET/UNLP, nkudraszow@mate.unlp.edu.ar)

El método de correlación canónica se utiliza para relacionar las variables en dos grupos. En general, esta metodología se usa cuando un conjunto de variables puede dividirse en dos grupos homogéneos y se desea identificar y cuantificar la relación entre ambos conjuntos de variables. Clásicamente el análisis canónico se realiza obteniendo las combinaciones lineales de cada subconjunto de variables que maximizan su correlación restringido a que las varianzas de dichas combinaciones sean iguales a uno. La extensión del método de correlaciones canónicas en el contexto de datos funcionales, donde los datos son curvas o funciones, no es directa. El enfoque basado en la extensión natural falla y es necesario utilizar alguna técnica que involucre suavizado. Una manera directa de introducir el suavizado es modificar la restricción sobre las varianzas de las combinaciones lineales agregando un término de penalización mediante la utilización de un operador de diferenciación. Este procedimiento es llamado análisis de correlación canónica suavizado y utiliza versiones muestrales de los operadores varianza y covarianza para estimar las variables y correlaciones canónicas, los cuales son altamente sensibles a observaciones atípicas, lo que lleva a la necesidad de desarrollar una alternativa robusta. Para obtener estimadores robustos proponemos reemplazar los operadores de covarianza muestrales por operadores de covarianza robustos, como los propuestos en Gervini (2010), los cuales utilizan como herramienta para detectar datos atípicos a las distancias entre las observaciones. Mediante un estudio de Montecarlo se comparará el desempeño de los estimadores propuestos con los ya existentes en la literatura en presencia y ausencia de contaminación.

## Referencias

- [1] Gervini, D. (2010). Outlier detection and trimmed estimation for general functional data. *Statistica Sinica*, **22**, 1639-1660.

---

## TESTS ROBUSTOS EN MODELOS DE REGRESIÓN PARCIALMENTE LINEALES SEMI-FUNCIONALES

Expositor: Alejandra Vahnovan (Facultad de Ciencias Exactas UNLP, avahnovan@mate.unlp.edu.ar)

Autor/es: Alejandra Vahnovan (Facultad de Ciencias Exactas UNLP, avahnovan@mate.unlp.edu.ar); Graciela Boente (CONICET, Facultad de Ciencias Exactas y Naturales UBA, gboente@dm.uba.ar)

El Modelo Parcialmente Lineal Semi-Funcional puede definirse como

$$Y = \beta' \mathbf{Z} + g(X) + \epsilon \quad (25)$$

donde  $\mathbf{Z} \in \mathbb{R}^p$ ,  $X$  es una variable aleatoria funcional, es decir toma valores en un espacio infinito-dimensional  $\mathcal{H}$ ,  $\epsilon$  es independiente de  $(\mathbf{Z}'; X)'$ ,  $\beta$  es un parámetro desconocido,  $g : \mathcal{H} \rightarrow \mathbb{R}$  es un operador suave desconocido que no se supone lineal y  $(H, d)$  es un espacio semimétrico.

Suponemos tener  $n$  observaciones  $(Y_i, \mathbf{Z}_i, X_i)$ ,  $i = 1, \dots, n$ , independientes e idénticamente distribuídas (i.i.d.), que satisfacen (1). Los estimadores clásicos de mínimos cuadrados en este modelo pueden verse seriamente afectados por la presencia de datos atípicos. En [1], Boente y Vahnovan introdujeron un procedimiento de estimación robusto de 3 pasos para este modelo y estudiaron el comportamiento asintótico del estimador.

Además de la importancia de desarrollar estimadores robustos, también merece atención el problema de test de hipótesis. En particular, deseamos hacer inferencias sobre el parámetro  $\beta$  o sobre alguna de sus componentes. Sean  $\beta = (\beta'_1, \beta'_2)'$ , donde  $\beta_1 \in \mathbb{R}^q$ . En este trabajo, nos concentramos en el problema de testear las siguientes hipótesis

$$i) H_{0\beta} : \beta = \beta_0 \qquad ii) H_{0\beta_1} : \beta_1 = \beta_{10}, \beta_2 \text{ no especificado}$$

y proponemos una clase de tests basados en el procedimiento de estimación de 3 pasos mencionado previamente. Se obtiene el comportamiento asintótico del estadístico del test bajo la hipótesis nula y bajo la alternativa y, a través de un estudio Monte Carlo se ilustra la performance de los tests propuestos.

## Referencias

[1] Boente, G. y Vahnovan, A.(2016), Robust estimators in semi-functional partial linear regression models, preprint.

## PROPUESTAS ROBUSTAS PARA EL ANÁLISIS DE CORRELACIONES CANÓNICAS

Expositor: Stella Maris Donato (Universidad de Buenos Aires, stelladonato@yahoo.com.ar)

Autor/es: Stella Maris Donato (Universidad de Buenos Aires, stelladonato@yahoo.com.ar); Jorge Gabriel Adrover (FAMAF, Universidad Nacional de Córdoba. CONICET, adrover@famaf.unc.edu.ar)

El análisis de correlación canónico (CCA) es una técnica de reducción de la dimensión en la cual dos vectores aleatorios de dimensión alta son reducidos a un nuevo par de vectores de dimensión más baja tras aplicar transformaciones lineales a cada uno de ellos, reteniendo tanta información como sea posible. Las componentes de los vectores transformados son llamados variables canónicas y proveen vectores cuya distancia cuadrada esperada ha sido minimizada. Como, clásicamente, la cercanía entre los dos vectores transformados es evaluada a través de una medida altamente sensible a observaciones atípicas, como es la pérdida cuadrática media, las transformaciones lineales que se buscan son también afectadas. Para superar esta situación, en esta presentación se proponen S-estimadores robustos para vectores y correlaciones canónicas por minimización de una M-escala basada en la distancia de transformaciones lineales de los vectores. Se analiza la consistencia bajo el modelo paramétrico y se deriva la función de Influencia. Por otra parte, la propuesta anterior resulta robusta cuando sólo una proporción esperada de los  $n$  vectores de la muestra resulta atípico, en este sentido se propone una nueva metodología robusta para tratar el caso donde la contaminación afecta a las variables de los vectores en forma independiente, por lo cual no se puede asegurar que menos de la mitad de los vectores de la muestra resultan contaminados y se analiza la performance de esta nueva propuesta por simulación.



## MODELOS DIFUSOS Y ANALISIS DE COMPONENTES PRINCIPALES PARA REDUCIR LA DIMENSION DE LA VARIABLE RESPUESTA EN MODELOS GEOESTADISTICOS.

Expositor: Mauro A.e. Chaparro Chaparro (Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN, CONICET-UNCPBA); Centro Marplatense de Investigaciones Matemáticas, Facultad de Cs. Exactas y Naturales, Univ. Nac. de Mar del Plata (CEMIM), chaparromauro76@gmail.com)

Autor/es: Mauro A.e. Chaparro Chaparro (Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN, CONICET-UNCPBA); Centro Marplatense de Investigaciones Matemáticas, Facultad de Cs. Exactas y Naturales, Univ. Nac. de Mar del Plata (CEMIM), chaparromauro76@gmail.com); Marcela Natal (Centro Marplatense de Investigaciones Matemáticas, Facultad de Cs. Exactas y Naturales, Univ. Nac. de Mar del Plata (CEMIM), mnatal@mdp.edu.ar); Marcos A.e. Chaparro (Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN, CONICET-UNCPBA), chapator@exa.unicen.edu.ar); Debora C. Marié (Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN, CONICET-UNCPBA), deboracmarie@gmail.com)

Los modelos geoestadísticos constituyen una importante herramienta que incorpora el efecto de la dependencia espacial en el estudio de la contaminación atmosférica en ambientes urbanos. En las últimas décadas, las técnicas del magnetismo ambiental han sido ampliamente estudiadas para su uso en el monitoreo de contaminantes en distintos ambientes dado que las propiedades magnéticas de los polvos atmosféricos (depositados/incorporados en vegetales) pueden dar información de las fuentes de contaminación. La evaluación del estado de los ambientes que se realizan a través de estas técnicas comprende una descripción sobre la concentración, la mineralogía y el tamaño de grano magnético. Es decir, desde un punto de vista estadístico, la variable respuesta que describe estado del ambiente es de naturaleza multivariada. En esta contribución se propone comparar dos estrategias para la reducción de la dimensión del problema para el ajuste de modelos geoestadísticos. En un caso se considera como variable respuesta a la primera componente principal obtenida a partir de un análisis de componentes principales sobre las variables magnéticas. En otro caso se utilizara como variable respuesta a un sistema de inferencia difusa (FIS) del tipo Mamdani construido por Chaparro y col (2015). El FIS utiliza como variables de entrada a 4 parámetros que describen la concentración (X y MRIS), mineralogía (Hcr) y tamaño de grano de los minerales magnéticos (KMRA/K). Los modelos se ajustaron a partir de un conjunto de datos (n=180) de la ciudad de Tandil (Argentina). De las mediciones realizadas a cada muestra se seleccionaron parámetros magnéticos que representan distintas propiedades magnéticas. Además se consideraron dos fábricas metalúrgicas ubicadas dentro de la ciudad como unas de las posibles fuentes de contaminación. La distancia a la fuente más próxima al punto de muestreo se definió como una covariable de los modelos geoestadístico. Los modelos ajustados se compararon utilizando el criterio de Akaike, el error cuadrático medio y el índice de la media cuadrática. En ambos casos se ajustaron los variogramas correspondientes, se obtuvieron mapas de predicción a partir de los modelos ajustados y de grillas definidas. Para los cálculos y gráficos se utilizó el software libre R.

---

## ANALISIS DE ASOCIACIONES ENTRE VARIABLES CATEGORIZADAS EN UN ESTUDIO SOBRE GRADUADOS UNIVERSITARIOS

Expositor: Jorgelina Del Valle Carrizo (Departamento de Matemática-Universidad Nacional de San Juan, jorgelinav\_carrizo@yahoo.com.ar)

Autor/es: Jorgelina Del Valle Carrizo (Departamento de Matemática-Universidad Nacional de San Juan, jorgelinav\_carrizo@yahoo.com.ar); Lilian Adriana Mallea (Departamento de Matemática-Universidad Nacional de San Juan, lamallea@gmail.com); Ana María Ruiz (Departamento de Matemática-Universidad Nacional de San Juan, anamaruiz5@hotmail.com)

El análisis con modelos Log-lineales y el Análisis de Correspondencias Múltiples son dos métodos estadísticos muy utilizados en el análisis de variables categorizadas con información presente en tablas de contingencias multidimensionales, y frecuentemente usadas en el ámbito de las Ciencias Sociales y de la Educación. Mientras que el Análisis de Correspondencias Múltiples se usa como técnica de análisis exploratorio, óptimo para la visualización gráfica de asociaciones entre las variables, para el análisis confirmatorio de tales asociaciones se usan los modelos Log-lineales. Se presenta en este trabajo un caso de aplicación de esta técnica estadística para el análisis de la estructura de dependencia entre variables categorizadas relevadas en un estudio sobre graduados universitarios de la Facultad de Filosofía Humanidades y Artes de la Universidad Nacional de San Juan. La interpretación de los parámetros del modelo permite describir relaciones entre variables de interés que describen el tránsito del graduado por la universidad.

Palabras Claves: datos categóricos; modelización log-lineal; graduado universitario.

---

#### OBJETO SIMBÓLICO: UNA FORMA DE TOMAR EN CUENTA LA VARIABILIDAD DE LOS DATOS

Expositor: Lilian Adriana Mallea (Departamento de Matemática-Universidad Nacional de San Juan, lamallea@gmail.com)

Autor/es: Lilian Adriana Mallea (Departamento de Matemática-Universidad Nacional de San Juan, lamallea@gmail.com); Myriam Beatriz Herrera (Instituto de Informática-Universidad Nacional de San Juan, mherrerarr@gmail.com)

En el Análisis de Datos clásico, los datos proceden de observaciones únicas, de determinadas variables sobre individuos únicos. Pero cuando los elementos de interés son clases o grupos de algún tipo, como los ciudadanos que viven en una ciudad determinada por ejemplo, hay variabilidad inherente en los datos. Reducir esta variabilidad mediante medidas de tendencia central lleva obviamente a una pérdida de información importante. Los datos simbólicos, introducidos por E. Diday en los ochenta, proporcionan un marco que permite representar datos con variabilidad mediante nuevos tipos de variables. Estos datos se pueden representar usando los arreglos usuales en forma de matrices, pero en los cuales los elementos de cada celda no son valores numéricos reales individuales, como es usual, sino conjuntos finitos de valores, intervalos o, de forma más general, distribuciones.

En los últimos años surgió el término Big Data, refiriéndose a conjuntos de datos tan grandes y complejos que se vuelven difíciles de procesar en un tiempo razonable con aplicaciones tradicionales de análisis de datos. El análisis de objetos simbólicos, al ofrecer la posibilidad de agregación de datos al nivel de granularidad elegido por el usuario, mientras se mantiene la información sobre la variabilidad intrínseca, desempeña un papel importante en este contexto. Los datos simbólicos pueden provenir de individuos considerando clases o grupos de los mismos y, en tal caso, representan las propiedades o descripciones de un elemento genérico de la clase que describen. Desde otro punto de vista, los datos simbólicos pueden establecerse por el conocimiento del experto sin necesidad de datos individuales o venir dados con incertidumbre. Los objetos simbólicos se describen por variables y datos simbólicos y proporcionan un mecanismo de vuelta a bases de datos o conjuntos de individuos en el sentido de conocer aquéllos

que se adecuan o relacionan con las descripciones simbólicas representadas por los objetos (las intenciones), según determinadas relaciones que también forman parte de las intenciones.

En este trabajo se introduce la noción de objeto simbólico, desde un punto de vista formal, y se muestra la obtención de este tipo de objetos a partir de una base de datos relacional, proveniente de encuestas a recién graduados universitarios.

Palabras Clave: Variable simbólica-Objeto simbólico-graduado universitario

---

#### CONVERGENCIA DE MODELOS EXPONENCIALES CON DISPERSIÓN BIVARIADOS

Expositor: Gabriela Boggio (UNR, FCEyE, gboggio@fcecon.unr.edu.ar)

Autor/es: Gabriela Boggio (UNR, FCEyE, gboggio@fcecon.unr.edu.ar); Lila Ricci (UNMDP, FCEyN, lricci@mdp.edu.ar); José Raúl Martínez (UNC, FaMAF, jmartine@famaf.unc.edu.ar)

Los modelos exponenciales con dispersión (ED) fueron definidos por Jørgensen (1987) y luego ampliamente utilizados en aplicaciones. Su difusión se debe, en gran parte, a que pueden ser tomados como distribuciones del error en modelos lineales generalizados; nos ocuparemos en particular de los modelos Tweedie que son ejemplos de ED y constituyen una herramienta sumamente enriquecedora, al ser una familia con dominio paramétrico  $\mathbb{R} \setminus (0, 1)$ . En cuanto a propiedades teóricas, se han probado importantes teoremas de convergencia que muestran a los modelos Tweedie como distribuciones límite (Jørgensen, Martínez y Tsao, 1994). En un artículo reciente Jørgensen y Martínez (2012) avanzaron en la construcción de modelos exponenciales con dispersión Multivariados (MED), aplicando el método de convolución. Obtuvieron así modelos con una estructura de correlaciones flexible y con distribuciones marginales en la misma familia.

El objetivo principal de este trabajo es obtener teoremas de convergencia que extiendan los teoremas univariados mencionados más arriba a modelos bivariados, probando así una propiedad sumamente relevante de los mismos: que son distribuciones límite. En el caso univariado, una hipótesis del teorema es que los ED sean generados por una medida de variación regular. Así es que en la demostración del teorema se usaron los teoremas Tauberianos. Otra herramienta fundamental fue el Teorema de Representación de Karamata (de Haan, 1975).

Se detallarán los pasos seguidos en la prueba de la extensión del teorema de convergencia, para MED con dominio en  $\mathbb{R}^2$ . Dado que los modelos exponenciales con dispersión tienen generalmente funciones de densidad que no pueden expresarse analíticamente en forma cerrada pero sus funciones generadoras de momentos son simples, la prueba se basa en mostrar la convergencia de estas últimas. En la misma se utilizó la definición de variación regular extendida a  $\mathbb{R}^n$  dada por Omeijer y Willekens (1989) y la extensión del teorema de representación de Karamata al caso multivariado hecha por de Haan y Resnick (1987).

---

#### ESTIMADORES ROBUSTOS PARA REGRESIÓN BINOMIAL NEGATIVA

Expositor: Marina Silvia Valdora (Universidad de Buenos Aires, mvaldora@gmail.com)

Autor/es: Marina Silvia Valdora (Universidad de Buenos Aires, mvaldora@gmail.com); Víctor Jaime Yohai (Universidad de Buenos Aires y Conicet, victoryohai@gmail.com); Michael Amiguet (Université de Lausanne, Michael.Amiguet@chuv.ch); Alfio Marazzi (Université de Lausanne, alfio.marazzi@chuv.ch)

El modelo de regresión binomial negativa se utiliza para modelar datos de conteo en los que se encuentra sobredispersión. El estimador de máxima verosimilitud se usa habitualmente para

estimar los parámetros del modelo pero tiene la desventaja de que es altamente sensible a la presencia de datos atípicos. El modelo de regresión binomial negativa asume que se observan  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  con  $y_i \in \mathbb{R}$  y  $\mathbf{x}_i \in \mathbb{R}^p$  donde  $y_i | \mathbf{x}_i$  tiene distribución  $NB(g(\boldsymbol{\beta}_0^T \mathbf{x}_i), \alpha_0)$  con  $\boldsymbol{\beta}_0$  y  $\alpha_0$  desconocidos. Asumimos que  $\mathbf{x}^T = (1, \mathbf{x}^{*T})$  y  $\boldsymbol{\beta}_0^T = (\beta_{01}, \boldsymbol{\beta}_0^{*T})$ . Proponemos estimadores robustos y altamente eficientes definidos tres pasos. En el primer paso, utilizando una propuesta de Han (1987) para otros modelos, estimamos  $\boldsymbol{\beta}_0^* / \|\boldsymbol{\beta}_0^*\|$ . En el segundo paso estimamos  $\|\boldsymbol{\beta}_0^*\|$ ,  $\beta_{01}$  y  $\alpha_0$  usando una combinación de M-estimadores. En el tercer paso, usamos los estimadores iniciales calculados en los pasos anteriores para detectar outliers y eliminarlos de la muestra usando un método similar al propuesto por Gervini y Yohai (2002) y Marazzi y Yohai (2004). Finalmente calculamos el estimador de máxima verosimilitud con las restantes observaciones. Probamos que el estimador final tiene eficiencia asintótica máxima y mostramos, mediante un estudio de simulación, que es altamente robusto.

### Referencias

- Marazzi A., Yohai V.J. (2004). Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference*, 122 (1-2), 271-291.
- Gervini, D. y Yohai, V.J. (2002) A class of robust and fully efficient regression estimators”, *Annals of Statistics*, **30**, 583-616.
- Han A.K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35(23), 303-316.

### UNA DEFINICIÓN CONSISTENTE DE CLASE MEDIA MULTIDIMENSIONAL

Expositor: Marcela Svarc (Universidad de San Andrés, msvarc@udesa.edu.ar)

Autor/es: Marcela Svarc (Universidad de San Andrés, msvarc@udesa.edu.ar); María Edo (Universidad de San Andrés, mariaedo@gmail.com); Walter Sosa Escudero (Universidad de San Andrés, wsosa@udesa.edu.ar)

En los últimos años los estudios sobre la clase media tomaron importancia en la literatura económica. El principal motivo radica en el rol que juega la clase media, no sólo en términos de consumo, sino además en la disminución de la polarización y en los potenciales conflictos sociales. Los principios teóricos a partir de los cuales se dan las definiciones de clase media son diversos, pero aún en los casos en donde la única variable representada es el ingreso, siempre es necesario definir dos cotas, una inferior que separa a los pobres de los no pobres y otra superior que separa a los ricos de los no ricos. Edo y Sosa Escudero (2014) hacen una detallada reseña de las más relevantes definiciones de clase media unidimensional existentes. Por otro lado, desde el año 1984, Sen introdujo la idea de que el bienestar es un concepto complejo y que era deseable medirlo en modo multidimensional. A partir de esas ideas muchos autores se abocaron a dar definiciones multidimensionales de bienestar y también a dar estudiar la de pobreza desde esta perspectiva. En este marco, nuestra propuesta es construir una noción multidimensional de clase media. El único antecedente de este análisis fue propuesto por Gigliarano y Mosler (2009), donde la clase media es conformada por una proporción fija de individuos, definida mediante el elipsoide de mínimo volumen. En este trabajo definimos a la clase media como el conjunto de individuos cuyo bienestar se encuentra entre dos percentiles multivariados, indicando una cota superior y una inferior. Para resolver este problema es necesario definir de percentiles multivariados que contemplen órdenes parciales, considerando que típicamente las definiciones de

cuantiles multivariados son equivariantes por transformaciones afines, es decir que no contemplan nuestros supuestos básicos. Nuestra propuesta se basa en suponer que se dispone de una dirección de crecimiento, definiendo posteriormente cuantiles univariados al proyectar los datos en esa dirección, obteniendo posteriormente cuantiles multidimensionales a partir de ellos. Se muestran resultados de consistencia fuerte. Por otra parte, proponemos un criterio de selección de variables para este tipo de problemas, extendiendo la ideas de Fraiman et al (2008). A modo de ejemplo, estudiamos el ejemplo de la Argentina, basándonos en datos provenientes de la Encuesta Permanente de Hogares para el período 2003-2014. La base de datos contiene información relativa al ingreso, a la vivienda, a la educación, a la ocupación y al consumo, entre otras.

#### Referencias

- Edo, M. y Sosa-Escudero, ó. (2014) “Assessing the measurement of the middle class: Argentina 1991-2012.” Preprint.
- Fraiman, R. Justel, A. y Svarc, M. (2008) “Selection of variables for cluster analysis and classification rules.” *Journal of the American Statistical Association*, 103 (483), 1294-1303.
- Cigliarano, C. y Mosler, K. (2009) “Measuring the Middle-Class Decline in one and many attributes.” *Cuaderno di Ricerca n 333*, Departamento di Economia, Universita Politecnica delle Marche.
- Sen, A. (1984) “The living standard.” *Oxford Economic Papers*, 73-90.

---

## INTRODUCCIÓN A LA ESTADÍSTICA BAYESIANA CON APLICACIONES A REGRESIÓN Y SUPERVIVENCIA

Expositor: Enrique E. Álvarez (Universidad Nacional de La Plata - Instituto de Cálculo (UBA-CONICET), [enriqueealvarez@fibertel.com.ar](mailto:enriqueealvarez@fibertel.com.ar))

Autor/es: Enrique E. Álvarez (Universidad Nacional de La Plata - Instituto de Cálculo (UBA-CONICET), [enriqueealvarez@fibertel.com.ar](mailto:enriqueealvarez@fibertel.com.ar))

En la Estadística Clásica (frecuentista) el parámetro  $\theta$  de un modelo estadístico se considera un elemento matemático desconocido pero fijo, del cual sólo se sabe su conjunto de valores posibles  $\theta \in \Theta$ , donde  $\Theta$  es el llamado Espacio Paramétrico. En los casos más usuales  $\Theta \subset \mathbb{R}^q$ , con  $q \in \mathbb{N}$ , es decir que  $\theta$  es un vector de constantes. Así, en la Estadística Clásica, el problema de Inferencia comienza con una muestra aleatoria  $\mathcal{X} := X_1, \dots, X_n$  con densidad conjunta  $f_{\mathcal{X}}(x; \theta) =: L_n(\theta)$ , la cual constituye la función de Verosimilitud una vez que se conoce la muestra y  $\theta$  varía a través de  $\Theta$ . A partir de allí la Estadística Clásica dispone de diversos métodos para estimar  $\theta$  o funciones del mismo ya sea en forma puntual o por intervalos.

En la Estadística Bayesiana, en cambio, se considera que el parámetro  $\theta$  de un modelo estadístico es un elemento del cual se posee información o presunciones previas, más allá de sus valores posibles. Esta información se resume en una llamada distribución *a priori* del parámetro. Convenientemente, esta información puede resumirse en una densidad  $f_{\Theta}(\theta)$ , sin que ello implique asignar a  $\theta$  el rol de una variable aleatoria. Lo que sí exige, en cambio, es una sustitución del concepto de probabilidad frecuentista por la probabilidad subjetiva. Los desarrollos matemáticos a partir de cualquiera de esas dos nociones son idénticos, ya que ambas cumplen los axiomas de Kolmogorov. Sólo se requiere tener presente su diferente interpretación epistemológica.

La inferencia Bayesiana se basa en obtener la distribución *aposteriori* del parámetro  $f_{\Theta|X}(\theta; x)$ , la cual es vista como una actualización o corrección de la información apriori luego de incorporar la evidencia muestral. La inferencia Bayesiana combina así dos fuentes de información: (i) la muestra, y (ii) conocimientos previos, a diferencia de la Estadística Clásica que solo utiliza información muestral.

En esta comunicación presentaremos los elementos fundamentales de la Estadística Bayesiana, comentando las diversas maneras de proponer distribuciones apriori y técnicas para la obtención las posterioris. De manera de ilustrar la utilidad y aplicabilidad de la Estadística Bayesiana, presentaremos ejemplos concretos relacionados con estimación del desgaste de piezas de un reactor nuclear mediante técnicas de regresión, o con un ejemplo de estimación de supervivencia en pacientes con leucemia.

---

## REDUCCIÓN SUFICIENTES DE DIMENSIONES EN REGRESIÓN CON PREDICTORES MIXTOS

Expositor: Pamela Llop (FIQ-UNL, pllop@santafe-conicet.gov.ar)

Autor/es: Pamela Llop (FIQ-UNL, pllop@santafe-conicet.gov.ar); Liliana Forzani (FIQ-UNL, liliana.forzani@gmail.com); Rodrigo García Arancibia (IECAL-FCE-UNL, rgarcia@fce.unl.edu.ar); Diego Tomassi (FIQ-UNL, diegot@santafe-conicet.gov.ar)

En ciertas áreas del conocimiento es muy frecuente contar con predictores que son de diferente naturaleza, como ser categóricos y continuos. Por ejemplo, en economía y ciencias sociales, los modelos que buscan explicar algún comportamiento o fenómeno social comúnmente incorporan variables explicativas de naturaleza mixta; esto es variables continuas (e.g. ingreso, edad), categóricas ordinales (e.g. escolaridad, materiales de la vivienda) y dicotómicas (e.g. género o la posesión de activos, como ser TV, radio, auto, etc.). A su vez, muchas veces se desea reducir la cantidad de predictores o combinarlos en un simple indicador a los efectos de simplificar el análisis procurando no perder información sobre el fenómeno en cuestión. El enfoque de Reducción Suficiente de Dimensiones (RSD) consiste en reducir la dimensión del espacio  $p$ -dimensional de las variables predictoras  $\mathbf{X}$  combinándolas en nuevas variables que vivan en un espacio de menor dimensión sin perder información sobre la variable respuesta  $Y$ . En su mayoría, los métodos de RSD suponen predictores continuos [2,3,5]. Sin embargo, recientemente se han desarrollado extensiones para predictores cuya distribución se encuentra dentro de una familia exponencial [1] o para predictores que son todos ordinales y para los que se asume la existencia de una latente continua [4]. En base a estas contribuciones, en el presente trabajo proponemos una extensión para el problema de RSD en regresiones con predictores mixtos (continuos, ordinales y dicotómicos). Más precisamente, en este trabajo obtenemos la reducción suficiente, la estimamos mediante el método de máxima verosimilitud, mediante ciertos estudios de simulación mostramos el buen comportamiento de dichas estimaciones, y finalmente ilustramos nuestro método aplicándolo a un conjunto de datos reales.

## Referencias

- [1] E. Bura, S. Duarte, and L. Forzani. *Sufficient reductions in regressions with exponential family inverse predictors*. To appear in Journal of the American Statistical Association.
- [2] R.D. Cook and L. Forzani. *Principal fitted components for dimension reduction in regression*. Statistical Science, 23:485–501, 2008.

- [3] R.D. Cook and L. Forzani. *Likelihood-Based sufficient dimension reduction*. Journal of the American Statistical Association, 104(485):197–208, 2009.
- [4] R. García Arancibia, P. Llop, L. Forzani, and D. Tomassi *Sufficient dimension reduction for ordinal predictors*. Submitted paper.
- [5] K.C. Li. *Sliced inverse regression for dimension reduction* (with discussion). Journal of the American Statistical Association, 86:316–342, 1991.

MODELIZACIÓN ESTRUCTURAL DE SERIES TEMPORALES DE RADIACIÓN SOLAR EN LA CIUDAD DE SALTA, MEDIANTE UN NUEVO FILTRO TIPO KALMAN ASOCIADO A UN MODELO SV(3) Y RUIDO BLANCO NO GAUSSIANO

Expositor: Orlando José Avila Blas (Depto. de Matemática-Facultad de Ciencias Exactas-UNSa, ojblas2012@gmail.com)

Autor/es: Orlando José Avila Blas (Depto. de Matemática-Facultad de Ciencias Exactas-UNSa, ojblas2012@gmail.com); Juan Carlos Abril (INIE-CONICET-Facultad de Ciencias Económicas-UNT, jabril@herrera.unt.edu.ar); Juan Carlos Rosales (Departamento de Matemática-Facultad de Ciencias Exactas-UNSa; IMECC-Univ. Est. de Campinas, Brasil, jcrmodeling@gmail.com.ar)

El ozono ( $O_3$ ) troposférico es uno de los principales componentes de la niebla tóxica ó smog. Se forma por la reacción fotoquímica entre contaminantes como los óxidos de nitrógeno ( $NOx$ ) y los compuestos orgánicos volátiles emitidos por los vehículos y por la industria, por lo que los eventos de mayor concentración de ozono se registran durante los períodos de tiempo soleado. El exceso de ozono en el aire puede producir efectos adversos de consideración en la salud humana, principalmente de tipo respiratorio y cardíacos; lo cual es de principal preocupación para la OMS. En este trabajo se presenta una modelización estadístico-estructural de series de tiempo de la variable concentración de ozono en algunos puntos de la ciudad de Salta, Argentina, con alta polución troposférica. Las series presentan una alta volatilidad estocástica que no puede modelarse con los modelos SV tradicionales planteados por Harvey [H] o por Avila Blas y otros, [AB]. Hemos solucionado el problema del modelado con la teoría dada en el siguiente **Teorema** : Definimos modelo de volatilidad estocástica SV(3) al dado por las ecuaciones estructurales:

$$Y_t = \sigma \varepsilon_t \exp\left(\frac{h_t}{2}\right)$$

$$h_t = \omega_1 h_{t-1} + \omega_2 h_{t-2} + \omega_3 h_{t-3} + \eta_t$$

Se consideran a los ruidos blancos  $\varepsilon_t$  y  $\eta_t$  como variables aleatorias i.i.d. con densidad no gaussiana, pero con media nula y varianza unitaria, y mutuamente no correlacionadas. Entonces:

- el modelo estructural es convergente si y sólo si el vector  $(\omega_1, \omega_2, \omega_3) \in \mathbf{B}(\mathbf{0}, \mathbf{1}) \subset \mathbf{R}^3$  a fin de garantizar la no existencia de raíces unitarias.
- la estimación de los hiperparámetros del modelo se realiza maximizando el logaritmo de la verosimilitud truncada del nuevo filtro asociado, con núcleo  $\Upsilon$  restringido a la bola unitaria en  $\mathbf{R}^3$ , mediante la aproximación

$$\left(\frac{\partial \log p(Y_n \Phi)}{\partial \Phi}\right)_{\Phi=\tilde{\Phi}} \approx \left(\frac{\partial E[\log p(\omega, Y_n \Phi)]}{\partial \Phi}\right)_{\Phi=\tilde{\Phi}}$$

donde  $E$  es la esperanza con respecto a  $p(\omega/Y_n, \tilde{\Phi})$  y  $\tilde{\Phi}$  es un valor experimental tomado convenientemente como semilla para el inicio del proceso de filtrado.

**Referencias** 1) [H] Harvey Andrew C., (1996), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press. 2) [AB] Avila Blas O. J., Abril J. C. y Lesino, G. (2000) *Radiación y temperaturas diarias: Un modelo de correlación estructural*. AVERMA (Avances en Energías Renovables y Medio Ambiente) 4, 2, 11.31-11.36

---