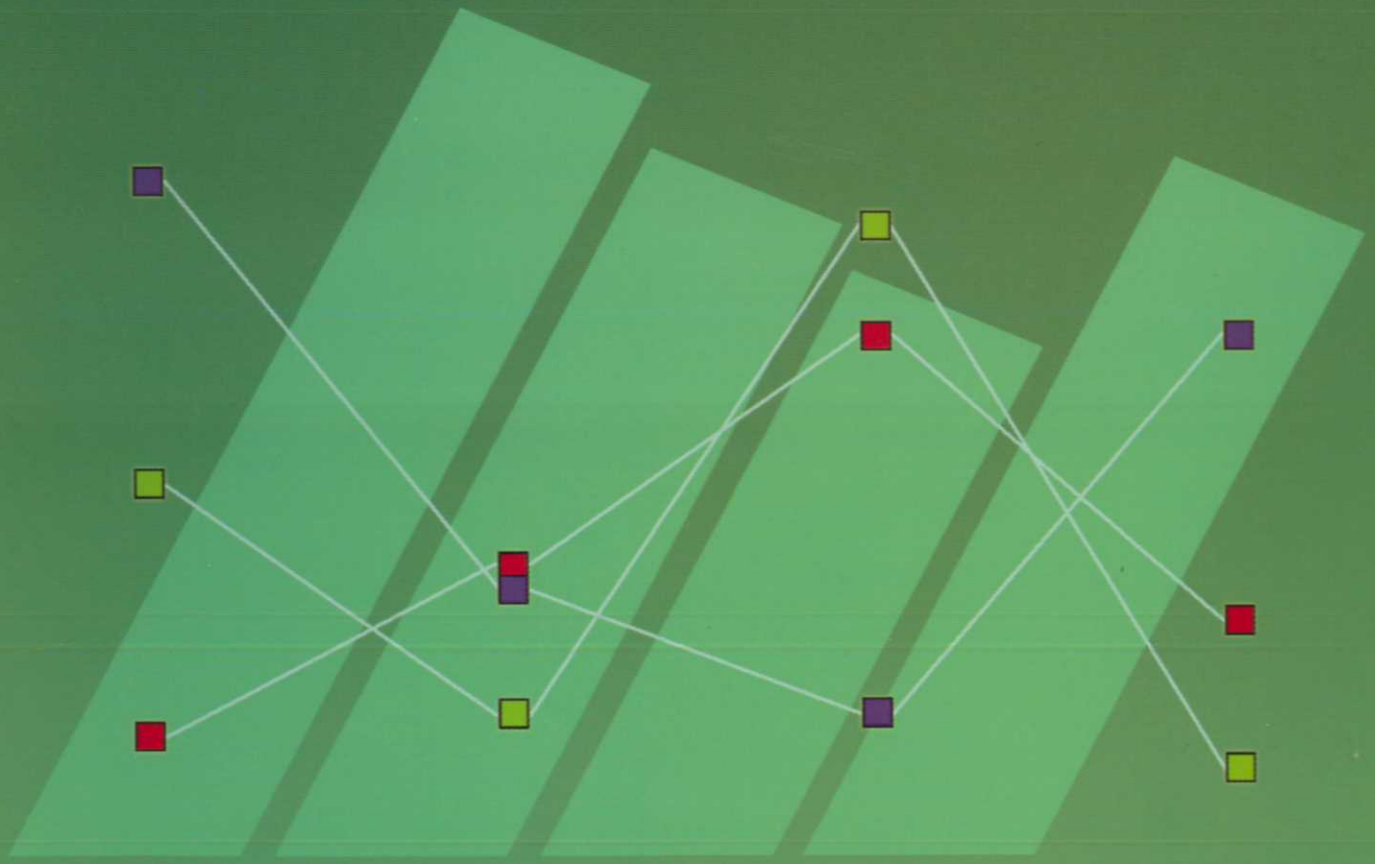
Memorias del XVII Foro Nacional de Estadística



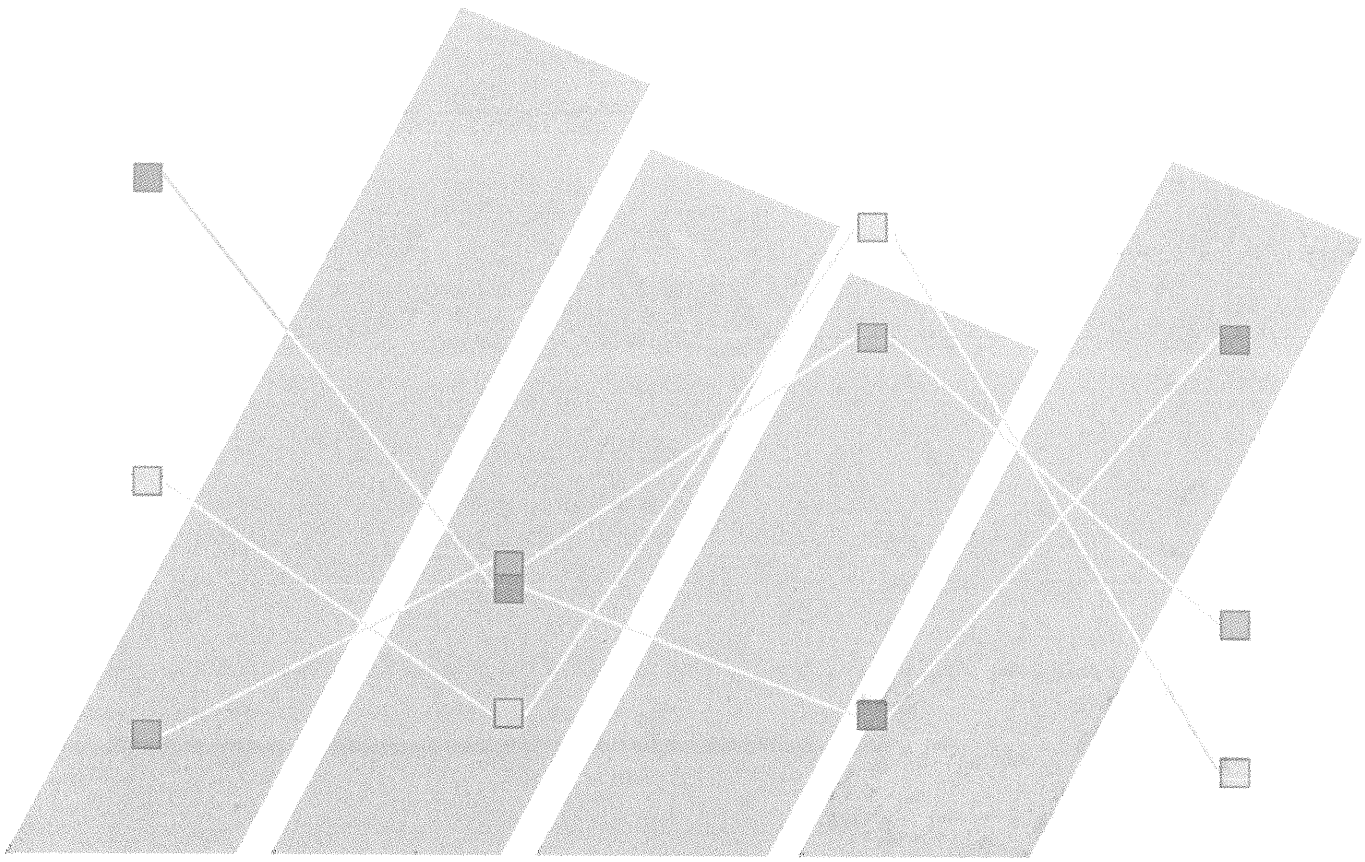
INEGI

www.inegi.gob.mx

Memorias del

XVII

Foro Nacional de Estadística



INEGI

www.inegi.gob.mx

Memorias del XVII Foro Nacional de Estadística

Impreso en México
ISBN 970-13-4339-5

Presentación

El XVII Foro Nacional de Estadística se llevó a cabo de 9 al 13 de septiembre de 2002 en la Universidad de la Américas en la ciudad de Puebla, organizado por el departamento de Actuaría de dicha Universidad.

En estas memorias se presentan algunos de los resúmenes de las contribuciones presentadas en este foro. Todos los resúmenes recibidos fueron incluidos sin un proceso de arbitraje, aunque con una detallada revisión.

La Asociación Mexicana de Estadística agradece a la Universidad de las Américas Puebla su apoyo para la realización de este foro y al Instituto Nacional de Estadística, Geografía e Informática el apoyo para la edición de estas memorias.

El Comité Editorial:

Antonio González Fragoso

Rafael Perera Salazar

Karim Anaya Izquierdo

Contenido

Uso de Excel en la Enseñanza de la Probabilidad y la Estadística

Víctor Aguirre Torres
Departamento de Estadística, ITAM

1. Introducción

En el presente trabajo se presenta una idea para comunicar conceptos importantes relacionados con las materias de Probabilidad y Estadística que hace uso de Excel. El propósito es la visualización del concepto empleando hojas de cálculo de una manera interactiva. Excel tiene bastantes funciones matemáticas y estadísticas interconstruidas, esto facilita la construcción de los ejemplos.

Las hojas de cálculo pueden usarse ya sea para presentarse en clase o bien como ejercicios de tarea, en esta última modalidad, el hecho de que Excel no sea un paquete estadístico requiere que los alumnos manejen los conceptos a un nivel mas allá de lo superficial para realizar los ejercicios que se les soliciten.

Otra ventaja de usar Excel y no otras aplicaciones como Mathematica, Maple o Mathlab, es su disponibilidad casi universal así como su facilidad de uso. Obviamente la manera de presentar un mismo concepto variará de profesor a profesor por lo que el propósito de este trabajo es solo el de presentar la idea y un ejemplo que la ilustre de manera parcial. Toca a cada persona interesada en usar este enfoque en adecuarlo a sus propias vivencias, formación, etc. Las posibilidades del enfoque son muy amplias, por ejemplo para el tema de Probabilidad, el autor ha diseñado hojas para los siguientes temas: Cálculos para la Densidad Normal, Densidad t de Student, Densidad Trinomial, Distribuciones Discretas de Probabilidad, Ley Débil de los Grandes Números, Momios de Póker, Propiedades del Modelo Normal, Transformación de Variables Aleatorias, Teorema de Bayes, Variable Aleatoria y Vector Aleatorio Bivariado. Mientras que para el área de Estadística: Análisis de Series de Tiempo, Asociación de Variables Cualitativas, Estadísticas de Orden, Estimación por Máxima Verosimilitud, Intervalos de Confianza para una Media, Modelos AR(1), Regresión Lineal Múltiple, Regresión Lineal Simple y Variación Muestral.

En Gudiño (1999) se presenta un marco metodológico, y más aplicaciones de este enfoque a la enseñanza de la teoría de vectores aleatorios. A continuación se presenta un ejemplo en la enseñanza de la Distribución Trinomial.

2. Ejemplo: Distribución Trinomial

Cuando se presenta el vector aleatorio trinomial, típicamente el alumno antes ya estuvo expuesto a la variable aleatoria binomial. Por esta razón se esperaría que el paso fuese fácil, ya que en lugar de que cada ensayo tenga solo dos resultados, ahora cada ensayo tiene tres resultados, o más como es el caso multinomial. La realidad es que esto no es así de fácil. En primer lugar la memoria del alumno no es tan precisa como para recordar los detalles de la distribución binomial tanto como el profesor quisiera. En segundo lugar el paso de variable aleatoria a vector aleatorio da lugar a nuevas fuentes de confusión. Por ejemplo, se puede confundir el número de ensayos independientes con el número de posibles resultados en cada ensayo, etc. La notación que seguiremos es la siguiente, (X_1, X_2) es un vector aleatorio trinomial con parámetros $(4, p_1, p_2)$. Para tratar de que el alumno visualice, identifique, pueda separar y concatenar mas fácilmente los conceptos y notación involucrados en el tema, se construyó la hoja de cálculo que se muestra en la figura 1. Los conceptos clave que se desean ilustrar son los siguientes:

1. Los 4 ensayos independientes.
2. El hecho de que cada ensayo tiene tres posibles resultados.
3. La probabilidad de ocurrencia de cada resultado permanece constante en los 4 ensayos.
4. (X_1, X_2) cuenta conjuntamente el número de veces que se obtuvieron los resultados 1 y 2 en los 4 ensayos.
5. El soporte del vector (X_1, X_2), el cual determina el soporte de la función de densidad.

Una de las cuestiones de fundamental importancia en la enseñanza de distribuciones de probabilidad es que el alumno alcance a diferenciar la variable o vector aleatorio de su

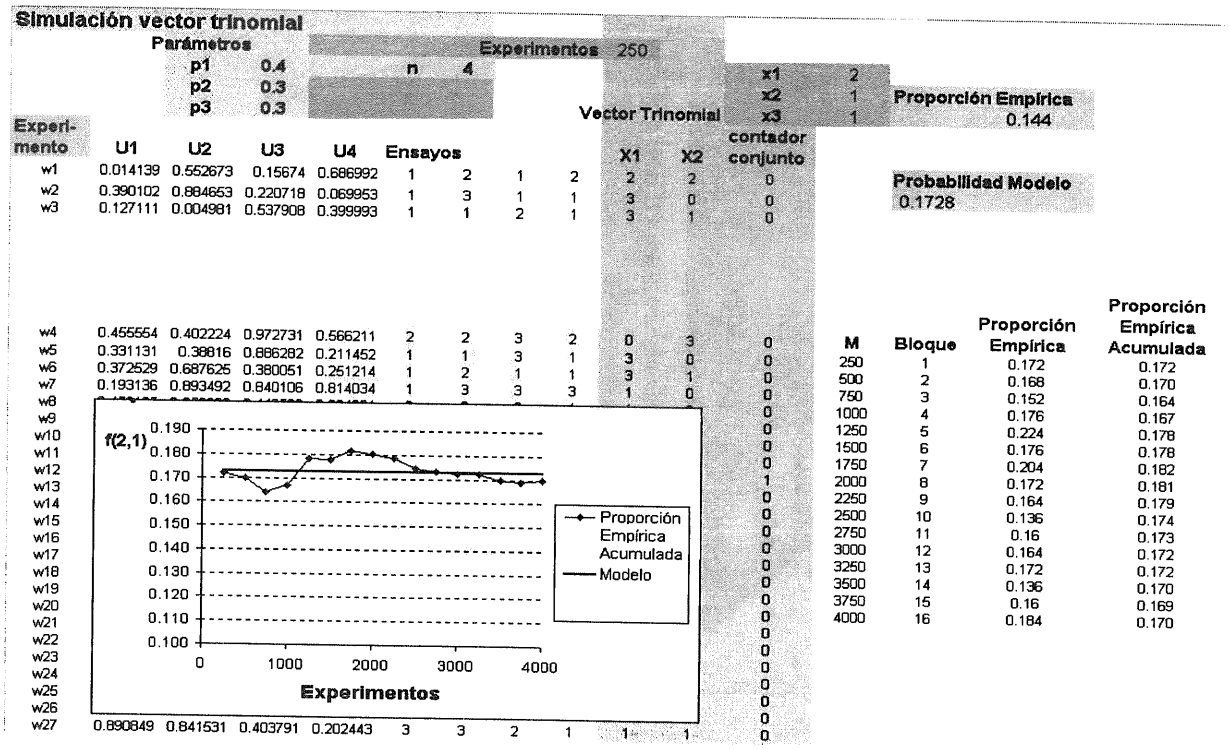


Figura 1. Demostración Distribución Trinomial (4, 0.4, 0.3)

modelo probabilístico. Por ejemplo, alguien que durante los inicios de su educación sobre Probabilidad únicamente haya visto enunciados como el siguiente: "sea X variable aleatoria con función de densidad $f(x)$..." y $f(x)$ siempre aparece explícitamente, pero X nunca o rara vez, un efecto probable es que tenga una confusión por decir lo menos. Tal diferenciación es una de las cosas que se hacen evidentes en esta hoja de cálculo. Otra idea que se ilustra es el concepto de probabilidad frecuentista, seguramente hemos escuchado la frase "si repetimos una gran cantidad de veces un experimento aleatorio y calculamos la proporción de veces que ocurre un cierto resultado...", bueno, en la hoja de cálculo esto se visibiliza fácilmente acumulando la proporción de veces que aparece una cierta combinación, por ejemplo que (X_1, X_2) tome el valor $(2, 1)$. La hoja consta de los siguientes elementos:

1. Los parámetros de la distribución. Los parámetros p_1 y p_2 se pueden cambiar libremente siempre y cuando se cumplan las restricciones de que $p_1 > 0$, $p_2 > 0$ y que $p_1 + p_2 < 1$. Cambiando estos valores ayuda a ilustrar su significado en el experimento. Por ejemplo, un valor de $p_2 = 0{,}9$ hará que en los ensayos haya muchos doses. El parametro n es

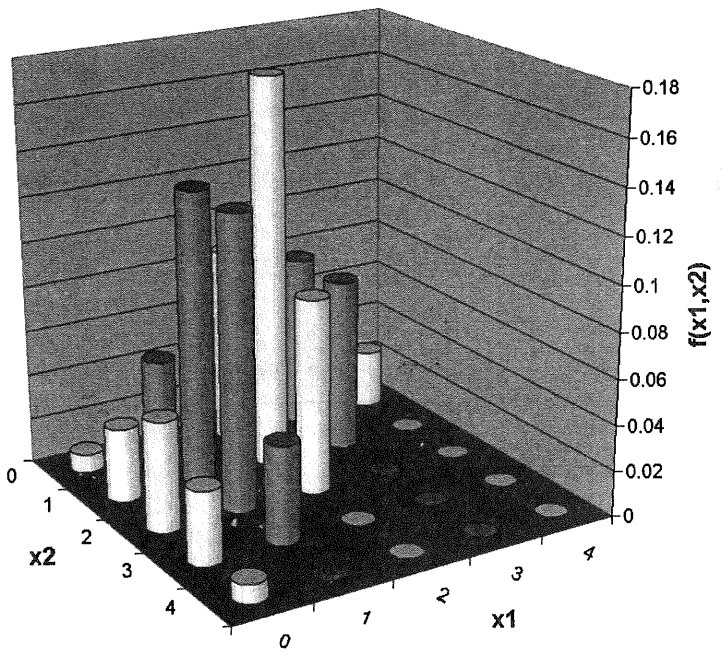


Figura . rico Distribución Trinomial (4, 0.4, 0.3)

fijo. El hecho de que haya 3 $p's$ ilustra la idea de que cada ensayo tiene 3 resultados posibles.

2. Cuatro columnas de numeros aleatorios uniformemente distribuidos, marcadas como U1 hasta U4, los cuales cambian cada vez que se da un "Intro" en la tabla. Una columna marcada como "Experimento" y que identifica cada una de las 250 veces que se repite el experimento.

3. Cuatro columnas marcadas como "Ensayos", las cuales registran los resultados de cada uno de los ensayos. En esta parte se visualiza la diferencia entre número de ensayos y número de resultados en cada ensayo.

4. Dos columnas etiquetadas como X1 y X2 que son el valor observado del vector trinomial para cada experimento. Aquí se hace visible el concepto de vector aleatorio discreto. Esta es la parte medular de la hoja de cálculo.

5. Una columna etiquetada como "contador conjunto" que ilusta el concepto de distribución conjunta del vector. En las celdas etiquetadas como x1, x2 se puede introducir

cualquier combinación de valores, la celda x3 se calcula como 4-x1-x2. El contador conjunto vale 1 si el vector aleatorio toma esa combinación de valores y cero de otra forma. Completa este elemento de la hoja una celda denominada "Proporción Empírica" que calcula la proporción de unos en los 250 experimentos. Esta celda ilustra el concepto de probabilidad frecuentista.

6. Una celda denominada "Probabilidad Modelo" que da el valor de la probabilidad de la combinación seleccionada de acuerdo al modelo trinomial. Con esto se ilustra el papel del modelo probabilístico para describir el comportamiento de un vector aleatorio. La "Proporción Empírica" y la "Probabilidad Modelo" deberían ser parecidas. Hay que explicar que difieren precisamente por la aleatoridad de la simulación.

7. Para menguar el efecto de la aleatoridad sobre la proporción empírica se inserta una tabla donde se acumula la proporción empírica para 16 bloques de 250 experimentos cada uno. La columna Proporción Empírica de la tabla se obtiene fácilmente copiando y pegando el valor de la celda "Proporción Empírica" cuantas veces se desee. En el caso de la figura 1 se hizo un total de 16 veces.

8. El último elemento de esta tabla es la gráfica de la proporción empírica acumulada junto con la probabilidad dada por el modelo. Aquí se ilustra el concepto de probabilidad frecuentista, y el papel del modelo en la descripción del comportamiento del vector aleatorio.

Otro concepto interesante que se puede ilustrar con la hoja de cálculo es el del soporte de la función de densidad. Hay combinaciones que el vector nunca toma. Para visualizarlo en términos del modelo se puede mostrar la figura 2, la cual también se puede obtener fácilmente con Excel.

3. Consideraciones

El uso de este tipo de hojas de cálculo puede ser útil en cualquier fase del proceso de enseñanza aprendizaje. Por ejemplo, al comienzo de la exposición puede servir para que el alumno comprenda y arme mejor todos los conceptos involucrados en el tema, esto seguramente le permitirá adentrarse mas fácilmente en los aspectos mas teóricos del tema. Este

tipo de ejercicios también se pueden usar al final del tema como parte de las experiencias de aprendizaje de los alumnos. Otra consideración importante es el hecho de que una hoja que aparentemente sirve para presentar un tema también puede servir para reforzar otros temas vistos con anterioridad. Se ha cuestionado en la literatura la exactitud de los cálculos con Excel, en la experiencia del autor no ha habido problema alguno a este respecto, probablemente debido a la finalidad didáctica del enfoque el cual no requiere demasiada exactitud. Finalmente, es conveniente recalcar que por su flexibilidad y lo portátil que son, estas hojas de cálculo pueden tener una amplia gama de usos. Comentarios: aguirre@itam.mx.

Referencias

Gudiño-Antillón, J. (1999). *Propuesta para la Enseñanza de la Teoría de Vectores Aleatorios.* Tesis de Licenciatura en Actuaría, Instituto Tecnológico Autónomo de México, 264 páginas.

Técnicas de registro de datos funcionales y sus consecuencias en su análisis estadístico

Hugo Eduardo Calva Díaz
Eduardo Castaño Tostado
Universidad Autónoma de Querétaro
Juan de Dios Figueroa Cárdenas
Reyna Araceli Mauricio Sánchez
CINVESTAV- unidad Querétaro

1. Introducción

En la actualidad hay procesos en los cuales se miden características de unidades experimentales o de muestreo de manera casi continua; en muchos casos esto motiva a que estas mediciones se pueden pensar como manifestaciones discretas de una función subyacente, un dato funcional. Esto sugiere el analizar este tipo de procesos considerando a los datos agrupados como funciones.

Partiendo de datos $y_1, y_2, ..., y_n$ obtenidos de medir una unidad bajo estudio en tiempos $t_1, t_2, ..., t_n$, podemos suponer un modelo de la siguiente forma

$$y_j = x(t_j) + \epsilon_j, \ \epsilon_j \sim (0, \sigma^2).$$

Un primer paso en análisis funcional de datos es, después de plantear el modelo anterior, estimar eficientemente a x. El enfoque fue por mínimos cuadrados a partir del uso funciones base del tipo B - spline de orden 4 penalizando la rugosidad del ajuste(ver, por ejemplo, Green y Silverman,1994 o Ramsay y Silverman, 1997). El objetivo de este trabajo se centra en dos pasos posteriores a la estimación de x, descripción de técnicas de registro y la aplicación de componentes principales funcionales.

2. Registro de datos funcionales

Teniendo N unidades experimentales, las $x_i(t), i = 1, .., N$ estimadas, éstas pueden diferir debido a dos tipos de variación: variación de amplitud ("altura"), o variación de fase (tiempos diferentes de un mismo estado). Pensemos que cada x_i está definida en un intervalo $[0, T_i]$. Denotemos por a un intervalo estándar por $[0, T_0]$. Sea $h_i(t)$ una transformación del tiempo t tal que

$$h_i(t) : [0, T_0] \rightarrow [0, T_i]$$

con h_i estrictamente creciente, satisfaciendo las condiciones de frontera $h_i(0) = 0$ y $h_i(T_0) = T_i$. Algunas veces se podría necesitar que $h_i(t)$ sea una función suave, en el sentido de que sea diferenciable un cierto número de veces.

Sea $x_0(t)$ una función objetivo fija definida sobre $[0, T_0]$, en el sentido de que las características de las curvas x_i serán alineadas en algún sentido con las características de x_0. Entonces podemos proponer el modelo

$$x_i[h_i(t)] = x_0(t) + \varepsilon_i(t) \qquad ó \qquad x_i \circ h_i = x_0 + \varepsilon_i, \tag{1}$$

donde ε es pequeño relativo a x_i y centrado en cero. Si a x_0 la definimos a través de valores discretos x_{j0}, $j = 1, ..., n$, entonces nuestro modelo(1)se convierte en:

$$x_i[h_i(t_j)] = x_{j0} + \varepsilon_{ij}. \tag{2}$$

Debido al supuesto de que ε es pequeño relativo a x_i, este modelo postula que diferencias mayores en forma entre la función objetivo y una función específica son debido únicamente a la variación de fase.

Un modelo más complejo para variación inter - curvas, que combina variación de fase y amplitud podría ser

$$x_i[h_i(t)] = A_i(t)x_0(t) + \varepsilon_i(t), \tag{3}$$

donde $A_i(t)$ es una función que modula la amplitud; así la versión discreta de (3) será

$$x_i[h_i(t_j)] = A_i(t_j)x_{j0} + \varepsilon_{ij}.$$

Supongamos, ahora, que se ha podido identificar estas N funciones $h_i(t)$, adjetivadas "warping". Podemos calcular las funciones registradas $x_i^*(t)$ de la siguiente forma:

1. Calcular valores de la función inversa $h_i^{-1}(t)$ para una fina malla de valores de t, $i = 1, ..., N$.
2. Calcular los valores de $x_i(t)$ para la misma malla de valores de t.

La tarea de registro, entonces, es estimar las funciones de tiempo "warping" h_i.

Un método para alinear curvas es mediante la identificación del tiempo de ocurrencia de características comunes sobresalientes (landmarks) en las curvas (ver por ejemplo Ramsay, 1998), suponiendo que la variación por corrimiento no es importante. Estas características son algunas veces picos o valles. Los que es esencial en este tipo de alineación es que el l-ésimo landmark, de L de interés, para la curva $x_i(t)$ sea claramente localizable en un único tiempo t_{0l}. Podemos designar el comienzo y el fin de las curvas como características con tiempos 0 y T_i, respectivamente, e indicar la sucesión entera de los tiempos landmark t_{il}, $l = 0, ..., L+1$ del intervalo $[0, T_i]$. El registro de una curva $x_i(t)$ respecto a una función de referencia x_0 se reduce a encontrar una función "warping" $h_i(t)$, tal que

$$x_i[h_i(t_{il})] = x_0(t_{0l}), \qquad l = 0, ..., L+1.$$

Usando esta estrategia, las curvas son alineadas transformando el tiempo, de tal manera, que los eventos landmark de las curvas ocurran aproximadamente en el mismo tiempo. Utilizando la siguiente notación para el operador de integración:

$$D^{-1}h(t) = \int_0^t h(s)ds,$$

la atención se enfoca a encontrar cada h_i con atributos de ser monótona, que $\ln(Dh)$ sea diferenciable y que $w = D\{\ln(Dh)\} = \frac{D^2h}{Dh}$ sea Lebesgue cuadrado integrable. Estas condiciones asegurarán que $h(t)$ sea estrictamente monótona ($Dh > 0$) y que su primera derivada sea suave y acotada casi dondequiera. El criterio de ajuste generalmente utilizado es

$$F_\lambda(y|w) = N^{-1}\sum_i \{y_i - \beta_0 - \beta_1 m(t_i)\}^2 + \lambda \int_0^T w^2(t)dt \tag{4}$$

donde

$$m(t) = \left\{D^{-1}\exp(D^{-1}w)\right\}(t)\,. \tag{5}$$

Ramsay (1998) desarrolló un algoritmo de 2 etapas para la minimización de la expresión (4.8).

Otra técnica de registro es la que se conoce como registro continuo (Ramsay, 2000). Se pueden registrar dos curvas, x_0 y x_1, optimizando alguna medida global de similitud entre las curvas completas. Consideremos la siguiente matriz de covarianzas entre x_0 y x_1.

$$\begin{bmatrix} \int x_0^2(t)dt & \int x_0(t)x_1(t)dt \\ \int x_0(t)x_1(t)dt & \int x_1^2(t)dt \end{bmatrix}. \tag{6}$$

Supongamos ahora que con (6) se realiza un análisis de componentes principales funcionales (para detalles de componentes principales funcionales ver Ramsay y Silverman, 1997). Si los valores de las curvas son proporcionales, entonces tal análisis produciría únicamente un sólo componente principal funcional, esto es, únicamente uno de los 2 eigenvalores de la matriz (6)será distinto de cero. Es por esta razón que se propuso que se puede escoger una función "warping" $h_i(t)$ de tal manera que

$$\min_h F(h_i) = \min_h \log \mu_2 \begin{bmatrix} \int x_0^2(t)dt & \int x_0(t)x_i[h_i(t)]dt \\ \int x_0(t)x_i[h_i(t)]dt & \int x_i^2[h_i(t)]dt \end{bmatrix}.$$

La búsqueda de las funciones "warping" h se realiza en la misma familia utilizada para el caso del registro landmark vista antes.

3. AFD en perfiles térmicos de maíces mexicanos

Es de importancia caracterizar al maíz, ya que en México es uno de los alimentos de mayor consumo y de importancia como insumo industrial. Una forma recientemente propuesta de caracterizarlo es por medio del estudio de la gelatinización del almidón componente (Mauricio Sánchez, 2001); dependiendo de las características de la gelatinización, el uso más adecuado del grano en procesos industriales. Para ello se genera un perfil térmico de un maíz a partir de hacer pasar un voltaje fijo a través de una muestra de maíz, pero variando la temperatura en el medio y midiendo la corriente (amperaje) conducida.

La gelatinización comienza, matemáticamente, cuando la primera derivada de la corriente con respecto a la temperatura es igual a cero; posterior al comienzo de la gelatinización, el pico de gelatinización se caracteriza matemáticamente cuando se alcanza un mínimo en el perfil térmico.

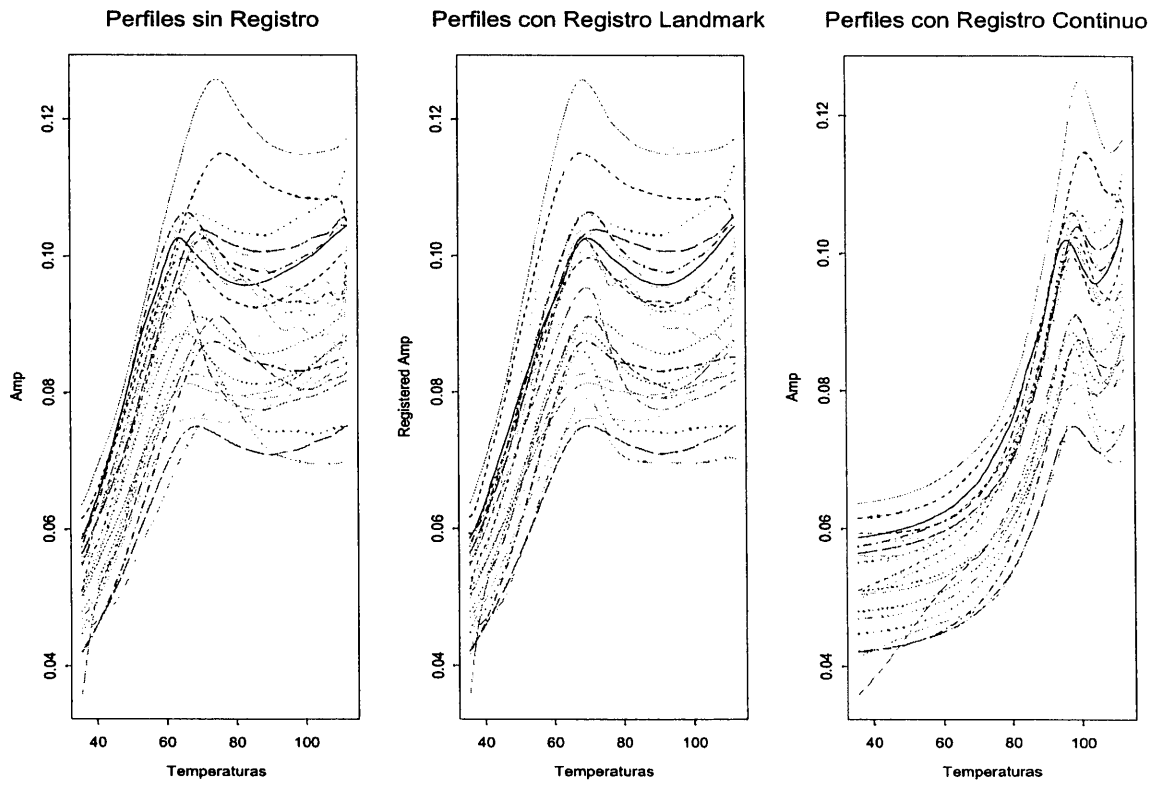


Figura 1: Perfiles suavizados sin registro y registrados

Los datos analizados en este trabajo provienen de 28 tipos de maíces mexicanos cada uno con 116 lecturas en un intervalo de temperaturas que va de 40 °C a 120°C. Nuestro objetivo en esta investigación, fue describir las fuentes de variación más importantes en el proceso de gelatinización entre diferentes razas de maíces, vía registro y por medio de análisis de componentes principales funcionales.

En el primer panel de la Figura 1 se muestran los 28 perfiles ya suavizados sin registro, para lo que se utilizó una representación por 30 funciones base tipo B - spline, splines de orden cuatro y 18 nodos. Se puede apreciar la variación tanto en amplitud como en desfasamiento en la ocurrencia del inicio y del pico de la gelatinización.

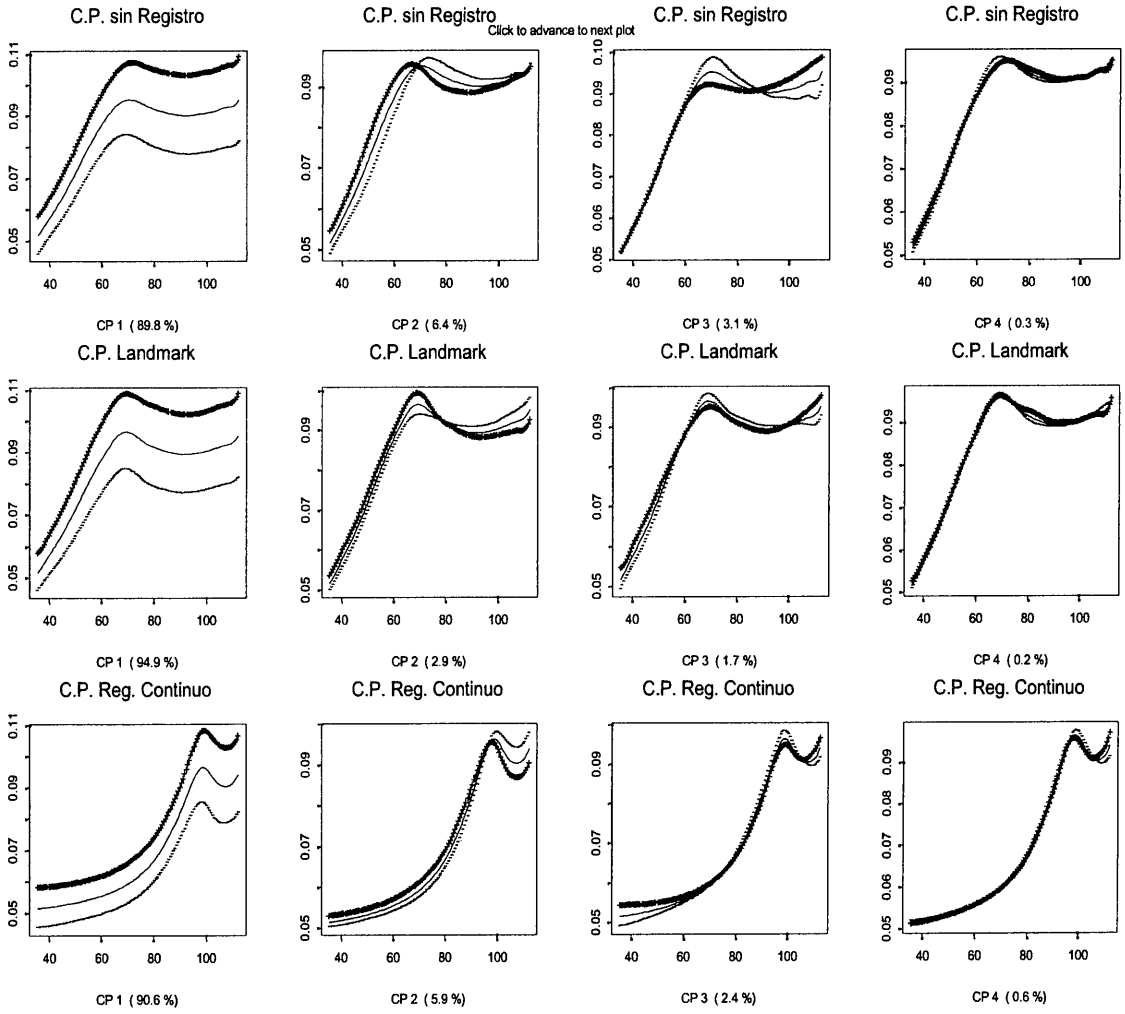


Figura 2: Componentes principales funcionales de perfiles de maíz

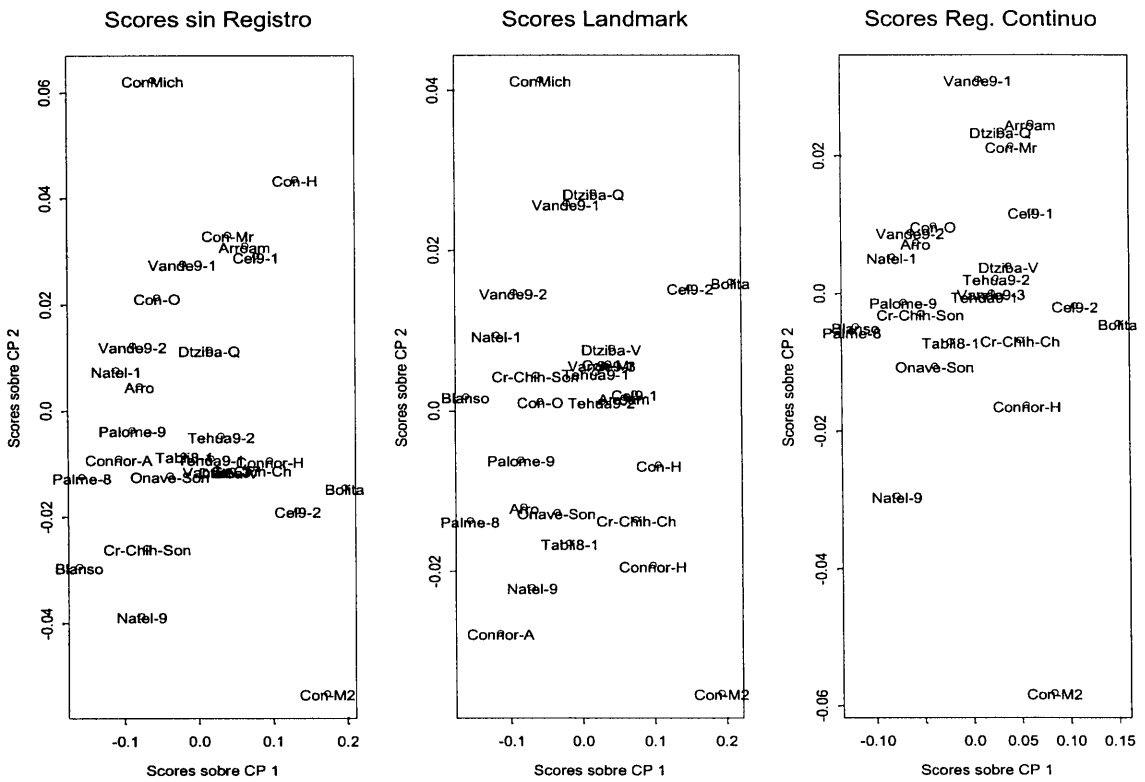


Figura 3: Scores de los perfiles sin registrar, con registro landmark y con registro continuo respectivamente

En el segundo panel de la Figura 1 se muestran los perfiles ya registrados por landmarks; éstos fueron identificados como el inicio y el pico de la gelatinización de cada perfil, mediante la inspección de los perfiles de derivadas respectivos. El alineamiento logrado en este conjunto de datos por registro landmark es eficiente en el sentido de generar perfiles que sólo muestran variación por amplitud.

El panel extremo izquierdo de la Figura 1 se presentan los perfiles registrados con el criterio continuo; los perfiles registrados muestran una gran deformación; con respecto al inicio y al pico de gelatinización el registro global muestra resultados insatisfactorios de alineamiento; esto es natural dado que de entrada no se señala a landmarks de interés porque la técnica responde a características globales y no locales de los perfiles. Así si el interés radica en comparar landmarks específicos debe usarse el registro landmark; el registro global es más natural de aplicarse cuando se desea registrar perfiles en los que no hay características comunes específicas de interés por parte del investigador.

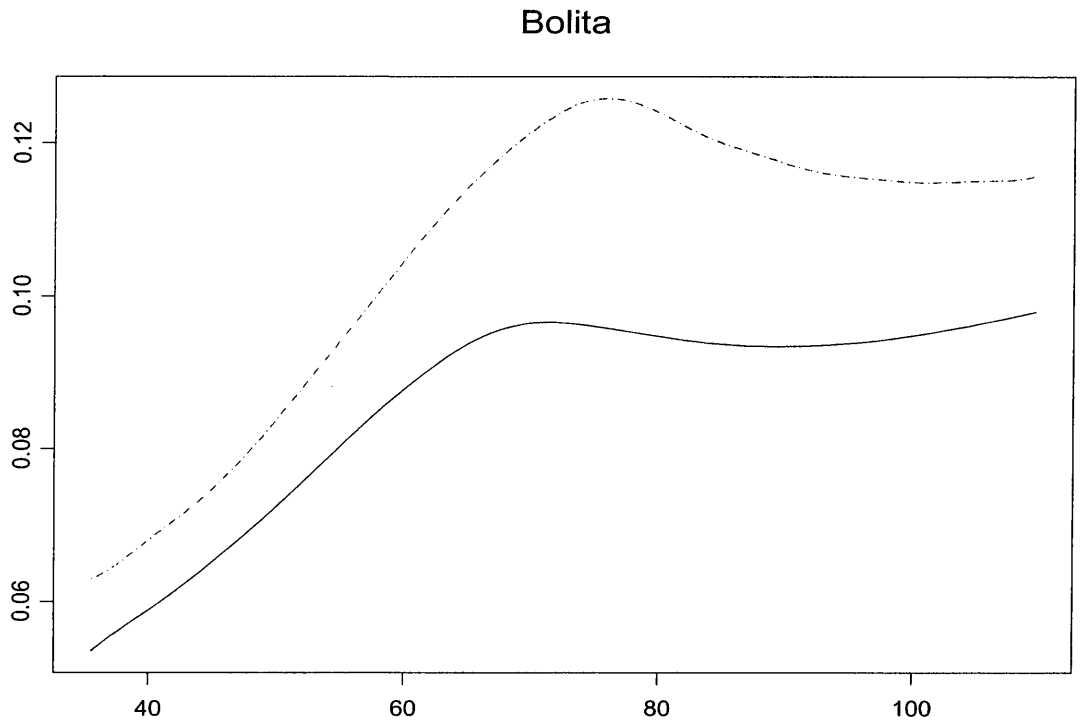


Figura 4: Perfil térmico de maíz Bolita (punteado) versus el perfil promedio

En la Figura 2 se muestran los componentes principales funcionales como múltiplos del perfil promedio. Respecto a perfiles sin registro previo, en el primer renglón de la Figura 2, se puede apreciar que el primer componente principal representa el 90 % de la variabilidad total, siendo ésta debida a perfiles que de manera uniforme se alejan del promedio ya sea con amperajes inferiores o ya sea con amperajes inferiores de manera preponderante a partir de los 70^oC. El segundo componente principal, a pesar de que reporta un porcentaje bastante menor, representa variabilidad aportada por perfiles que muestran un comportamiento mixto respecto al promedio: perfiles que hasta antes del inicio de la gelatinización están uniformemente por arriba o por abajo del promedio, pero que toda la gelatinización se da en amperajes menores o mayores, respectivamente, a los del promedio. Observando el primer panel la Figura 3, en donde se muestran los scores de los perfiles de maíces respecto a los dos primeros componentes principales sin registro, Bolita es el maíz con el mayor peso positivo en el primer componente principal funcional. En la Figura 4 se muestra el perfil respectivo comparándolo con el perfil promedio.

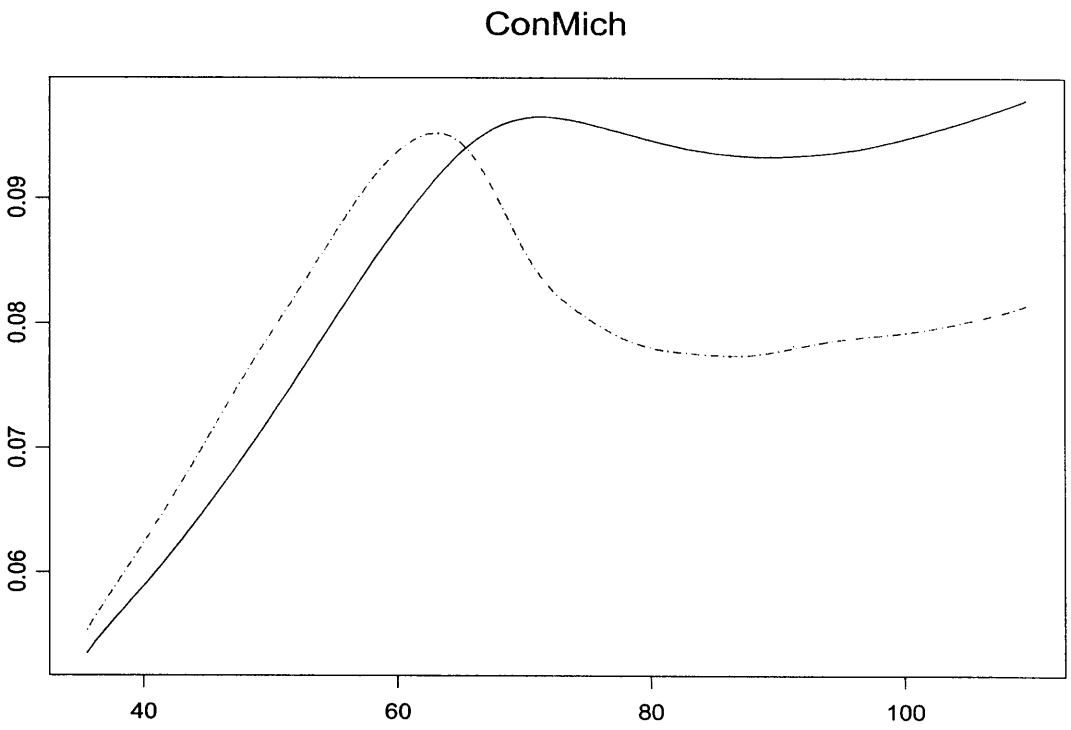


Figura 5: Perfil térmico de maíz ConMich (punteado) versus el perfil promedio

Se puede apreciar que tiene un amperaje mucho mayor que el promedio a lo largo de todo el rango de temperaturas. ConMich es un maíz cuyo perfil tiene un peso muy importante en el segundo componente funcional; en la Figura 5 se muestra su perfil en relación al perfil promedio, en donde se evidencia su comportamiento mixto respecto al promedio. ConM2 es un perfil con pesos importantes en ambos componentes principales. En la Figura 6 se muestra su perfil que evidencia las características combinadas representadas por los dos primeros componentes principales funcionales pero sin alejarse en global mucho del promedio.

Respecto a los componentes principales de perfiles registrados con landmarks, se aprecia que el primer componente principal reporta un porcentaje mayor (95 %) que en el caso sin registro, pero cualitativamente representando el mismo tipo de variación del primer componente de los perfiles sin registro. El segundo componente principal representa variación por perfiles que uniformemente están por arriba o por abajo del promedio pero así hasta el inicio de la gelatinización, para después alcanzar el pico por abajo o por arriba del promedio.

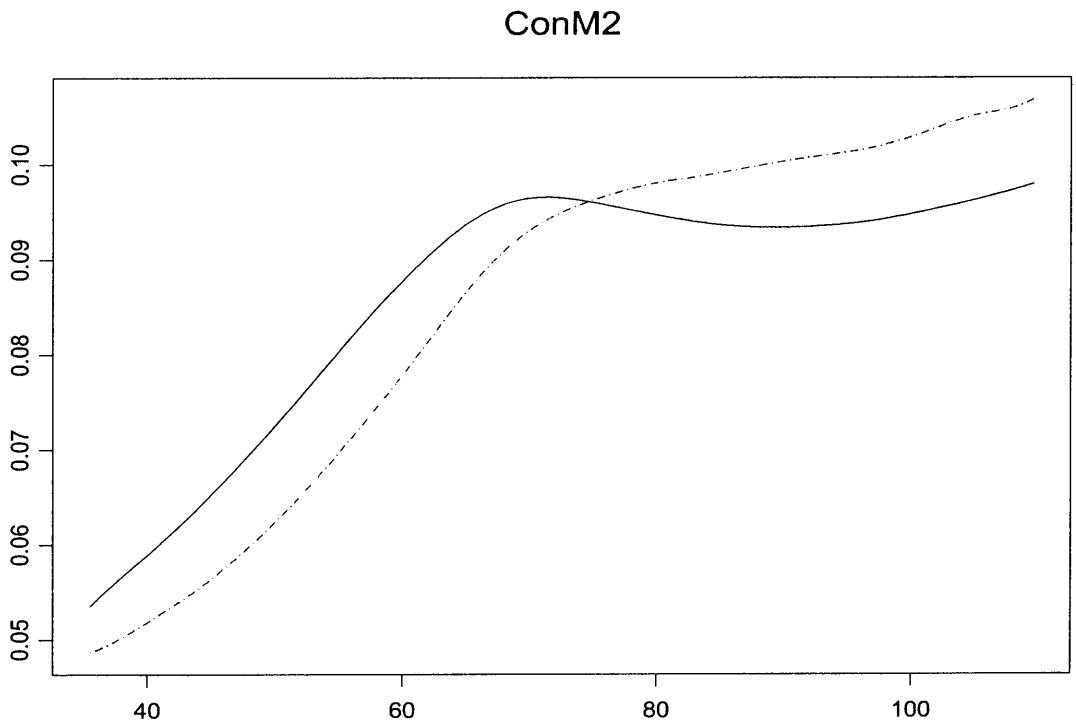


Figura 6: Perfil térmico de maíz ConM2 (punteado) versus el perfil promedio

Respecto a los componentes principales a partir de datos registrados globalmente, al ser muy parecidos a los perfiles sin registro, salvo la deformación mencionada, arrojan interpretaciones cuantitativa y cualitativamente similares.

4. Agradecimientos

El trabajo de segundo autor fue financiado parcialmente a través del proyecto CONACYT 36616E.

Referencias

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models, A roghness penalty approach.*. Chapman & Hall /CRC.

Mauricio Sánchez, R.A. (2001). Caracterización Fisicoquímica, Térmica y Eléctrica de Razas Mexicanas de Maíz y Evaluación de sus Posibles Usos en la Industria Alimentaria. *Tesis de Maestría en Ciencia y Tecnología de los Alimentos,* Universidad Autónoma de Querétaro, Facultad de Química. Programa de Posgrado en Alimentos del Centro de la República.

Ramsay, J.O. y Silverman, B.W. (1997). *Functional Data Analysis.* Springer-Verlag, New York.

Ramsay, J.O. (1998). Curve Registration. *Journal of the Royal Statistical Society*, B, 60, 365-375.

Ramsay, J.O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society*, B, 60, part 2, 365-375.

Ramsay, J.O. (2000). *A Guide to Curve Registration.* FTP ego.psych.mcgill.ca/pub/ ramsay.

Inference for Mixtures Of Distributions for Censored Data with Partial Identification

Alberto Contreras-Cristán
Eduardo Gutiérrez-Peña
Federico O'Reilly
IIMAS-UNAM

1. Introduction

In this supplementary report we discuss an implementation of the EM algorithm for the statistical analysis of mixtures of distributions in the context of partial identification. Our implementation is related to the type 4 case in Redner and Walker (1984), but we account for censoring in the likelihood function given by our specific context, described in the manuscript and in Mendenhall and Hader (1958). Although the EM algorithm approach is interesting and feasible to implement in our particular case, we conclude that it is computationally more involved and expensive than the methods proposed in our paper referred to above.

2. An Implementation Of The Em Algorithm

We refer the reader to Redner and Walker (1984) and Tanner (1996) for a description and examples of the EM algorithm. Here we only give a sketch of our implementation.

Let $\Xi \equiv (\boldsymbol{\theta}', \boldsymbol{\pi}')' = ((\theta_1', \ldots \theta_k'), (\pi_1, \ldots, \pi_k))'$ denote the vector of all the parameters, where $\theta_j' = (\mu_j, \varphi_j)$, with $\varphi_j = \log \sigma_j$. Recall from equation (1) in the manuscript, that the likelihood $L(\Xi|\boldsymbol{x})$ for the observed data is proportional to

$$\left\{ \prod_{j=1}^{k} \prod_{i=1}^{r_j} \frac{\widetilde{f}_j(y_{ji}; \theta_j)}{\widetilde{F}_j(C; \theta_j)} \right\} \prod_{j=1}^{k} \{\pi_j \widetilde{F}_j(C; \theta_j)\}^{r_j} \times \{1 - \widetilde{G}(C; \boldsymbol{\theta}, \boldsymbol{\pi})\}^{N-r}. \tag{1}$$

We will assume that the original failure times follow a Weibull distribution. Thus after

standardizing the data we have

$$f_j(x_{ji};\mu_j,\varphi_j) = e^{-\varphi_j} e^{(x_{ji}-\mu_j)e^{-\varphi_j}} \exp\left\{-e^{(x_{ji}-\mu_j)e^{-\varphi_j}}\right\}$$

and

$$F_j(x_{ji};\mu_j,\varphi_j) = 1 - \exp\left\{-e^{(x_{ji}-\mu_j)e^{-\varphi_j}}\right\}.$$

Now, consider as *augmented data* the pairs $z_1 = (x_{01}, i_1), \ldots, z_{N-r} = (x_{0\,N-r}, i_{N-r})$, where the first component in each pair is an unobserved (censored) datum and the second component is an integer which indicates the population corresponding to the first component. Thus, the augmented likelihood $L(\boldsymbol{x}, \boldsymbol{z}|\Xi)$ is proportional to

$$\prod_{j=1}^{k}\prod_{i=1}^{r_j}\{\pi_j f_j(x_{ji};\theta_j)\} \times \prod_{l=1}^{N-r}\{\pi_{i_l} f_{i_l}(x_{0l};\theta_{i_l})\}. \tag{2}$$

From equations (1) and (2) it follows that the predictive distribution for the augmented data is

$$P(z_1,\ldots,z_{N-r}|\boldsymbol{x};\Xi) = \prod_{l=1}^{N-r}\left\{\frac{\pi_{i_l} f(x_{0l};\theta_{i_l})}{1-G(0;\Xi)}\right\}, \tag{3}$$

for $z_l = (x_{0l}, i_l)$, $i_l = 1,\ldots,k$, $l = 1,2,\ldots,N-r$ and $x_{0l} \geq 0$. It can be shown that

$$\int_0^\infty \cdots \int_0^\infty \sum_{i_1=1}^{k} \cdots \sum_{i_{N-r}=1}^{k} \prod_{l=1}^{N-r}\left\{\frac{\pi_{i_l} f(x_{0l};\theta_{i_l})}{1-G(0;\Xi)}\right\} dx_{01}\ldots dx_{0\,N-r}$$

$$= \int_0^\infty \cdots \int_0^\infty \prod_{l=1}^{N-r}\left\{\frac{\sum_{i_l=1}^{k}\pi_{i_l} f(x_{0l};\theta_{i_l})}{1-G(0;\Xi)}\right\} dx_{01}\ldots dx_{0\,N-r} = 1$$

We assume a uniform prior for Ξ, so the augmented posterior density is given by

$$p(\Xi|\boldsymbol{z};\boldsymbol{x}) \propto \prod_{j=1}^{k}\left\{\prod_{i=1}^{r_j}\pi_j e^{-\varphi_j} e^{(x_{ji}-\mu_j)e^{-\varphi_j}} \exp\{-e^{(x_{ji}-\mu_j)e^{-\varphi_j}}\}\right\}$$

$$\times \prod_{l=1}^{N-r}\left\{\pi_{i_l} e^{-\varphi_{i_l}} e^{(x_{0l}-\mu_{i_l})e^{-\varphi_{i_l}}} \exp\{-e^{(x_{0l}-\mu_{i_l})e^{-\varphi_{i_l}}}\}\right\}.$$

Thus, the expectation of $\log\{p(\Xi|\boldsymbol{x},\boldsymbol{z})\}$ with respect to the augmented predictive density (3) evaluated on a current value Ξ^u of Ξ, is given by

$$Q(\Xi,\Xi^u) \equiv E\left\{\log\{p(\Xi|\boldsymbol{x},\boldsymbol{z})\}|\Xi^u;\boldsymbol{x}\right\} =$$

$$\int_0^\infty \cdots \int_0^\infty \sum_{i_1=1}^{k} \cdots \sum_{i_{N-r}=1}^{k} \log\{p(\Xi|\boldsymbol{x},\boldsymbol{z})\} \prod_{l=1}^{N-r} \frac{\pi_{i_l}^u f(x_{0l};\theta_{i_l}^u)}{1-G(0;\Xi^u)} dx_{01}\ldots dx_{0\,N-r}.$$

In the following computations, $\stackrel{\triangle}{=}$ denotes equality up to an additive constant (with respect to Ξ)

$$\begin{aligned}
Q(\Xi,\Xi^u) &\stackrel{\triangle}{=} \sum_{j=1}^{k}\sum_{i=1}^{r_j}\left\{\log\{\pi_j\} - \varphi_j + (x_{ji}-\mu_j)e^{-\varphi_j} - e^{(x_{ji}-\mu_j)e^{-\varphi_j}}\right\} \\
&+ \sum_{l=1}^{N-r}\sum_{i_l=1}^{k}\log\{\pi_{i_l}\}\left\{\frac{\pi_{i_l}^u\{1-F(0;\theta_{i_l}^u)\}}{1-G(0;\Xi^u)}\right\} \\
&- \sum_{l=1}^{N-r}\sum_{i_l=1}^{k}\varphi_{i_l}\left\{\frac{\pi_{i_l}^u\{1-F(0;\theta_{i_l}^u)\}}{1-G(0;\Xi^u)}\right\} \\
&+ \sum_{l=1}^{N-r}\sum_{i_l=1}^{k}\int_0^\infty (x_{0l}-\mu_{i_l})e^{-\varphi_{i_l}}\left\{\frac{\pi_{i_l}^u f(x_{0l};\theta_{i_l}^u)}{1-G(0;\Xi^u)}\right\}dx_{0l} \\
&- \sum_{l=1}^{N-r}\sum_{i_l=1}^{k}\int_0^\infty e^{(x_{0l}-\mu_{i_l})e^{-\varphi_{i_l}}}\left\{\frac{\pi_{i_l}^u f(x_{0l};\theta_{i_l}^u)}{1-G(0;\Xi^u)}\right\}dx_{0l} \\
&= \sum_{j=1}^{k}\left\{r_j + (N-r)\left\{\frac{\pi_j^u\{1-F(0;\theta_j^u)\}}{1-G(0;\Xi^u)}\right\}\right\}\log\pi_j \\
&- \sum_{j=1}^{k}\left\{r_j + (N-r)\left\{\frac{\pi_j^u\{1-F(0;\theta_j^u)\}}{1-G(0;\Xi^u)}\right\}\right\}\varphi_j \\
&+ \sum_{j=1}^{k}\sum_{i=1}^{r_j}(x_{ji}-\mu_j)e^{-\varphi_j} - \sum_{j=1}^{k}\sum_{i=1}^{r_j}e^{(x_{ji}-\mu_j)e^{-\varphi_j}} \\
&+ (N-r)\sum_{j=1}^{k}\left\{\frac{\pi_j^u\{m(\theta_j^u)-\mu_j(1-F(0;\theta_j^u))\}}{1-G(0;\Xi^u)}e^{-\varphi_j}\right\} \\
&- (N-r)\sum_{j=1}^{k}\left\{\frac{\pi_j^u e^{-\mu_j e^{-\varphi_j}}}{1-G(0;\Xi^u)}\nu(\theta_j^u;\varphi_j)\right\}, \qquad (4)
\end{aligned}$$

where

$$m(\theta_j^u) = e^{\varphi_j^u} \int_{l(\theta_j^u)}^{\infty} \log\{v\} e^{-v} dv + \mu_j^u \int_{l(\theta_j^u)}^{\infty} e^{-v} dv,$$

$$\nu(\theta_j^u; \varphi_j) = e^{\mu_j^u e^{-\varphi_j}} \int_{l(\theta_j^u)}^{\infty} v^{e^{\varphi_j^u - \varphi_j}} e^{-v} dv$$

and $l(\theta_j^u) = e^{-\mu_j^u e^{-\varphi_j^u}}$.

Note that we can write expression (4) as

$$Q(\Xi, \Xi^u) \triangleq h(\pi_1, \ldots, \pi_k) + g(\varphi_1, \ldots, \varphi_k; \mu_1, \ldots, \mu_k),$$

where

$$h(\pi_1, \ldots, \pi_k) = \sum_{j=1}^{k} \left\{ r_j + (N-r) \left\{ \frac{\pi_j^u \{1 - F(0; \theta_j^u)\}}{1 - G(0; \Xi^u)} \right\} \right\} \log \pi_j$$

and

$$\begin{aligned} g(\varphi_1, \ldots, \varphi_k; \mu_1, \ldots, \mu_k) &= -\sum_{j=1}^{k} \left\{ r_j + (N-r) \left\{ \frac{\pi_j^u \{1 - F(0; \theta_j^u)\}}{1 - G(0; \Xi^u)} \right\} \right\} \varphi_j \\ &+ \sum_{j=1}^{k} \sum_{i=1}^{r_j} (x_{ji} - \mu_j) e^{-\varphi_j} - \sum_{j=1}^{k} \sum_{i=1}^{r_j} e^{(x_{ji} - \mu_j) e^{-\varphi_j}} \\ &+ (N-r) \sum_{j=1}^{k} \left\{ \frac{\pi_j^u \{m(\theta_j^u) - \mu_j (1 - F(0; \theta_j^u))\}}{1 - G(0; \Xi^u)} e^{-\varphi_j} \right\} \\ &- (N-r) \sum_{j=1}^{k} \left\{ \frac{\pi_j^u e^{-\mu_j e^{-\varphi_j}}}{1 - G(0; \Xi^u)} \nu(\theta_j^u; \varphi_j) \right\}. \end{aligned}$$

In order to maximize Q with respect to Ξ, we maximize in turn each of the terms h and g. Let us start this procedure with h, by using Lagrange multipliers we get that h attains a maximum at the point

$$\pi_j = \frac{r_j + (N-r)\{\frac{\pi_j^u (1 - F(0; \theta_j^u))}{1 - G(0; \Xi^u)}\}}{N}, \qquad j = 1, 2, \ldots k. \tag{5}$$

Correspondingly, by differentiating g with respect to μ_j and φ_j, $j = 1, 2, \ldots, k$, we get that g attains a maximum at the solution $(\mu_1, \ldots, \mu_k, \varphi_1, \ldots, \varphi_k)$ to the equations

$$\nabla_{\boldsymbol{\theta}}\, g = \mathbf{0}$$

or in a component-wise fashion

$$\begin{aligned} -r_j \;+\; & \sum_{i=1}^{r_j} e^{(x_{ji}-\mu_j)e^{-\varphi_j}} - (N-r)\frac{\pi_j^u\{1-F(0;\theta_j^u)\}}{1-G(0;\boldsymbol{\Xi}^u)} \\ +\; & (N-r)\frac{\pi_j^u\{\nu(\theta_j^u;\varphi_j)\}}{1-G(0;\boldsymbol{\Xi}^u)} e^{-\mu_j e^{-\varphi_j}} = 0, \quad j = 1, \ldots, k \end{aligned} \tag{6}$$

and

$$\begin{aligned} -\; & \left\{ r_j + (N-r)\frac{\pi_j^u\{1-F(0;\theta_j^u)\}}{1-G(0;\boldsymbol{\Xi}^u)} \right\} e^{\varphi_j} - \sum_{i=1}^{r_j}(x_{ji}-\mu_j) \\ +\; & \sum_{i=1}^{r_j}(x_{ji}-\mu_j)e^{(x_{ji}-\mu_j)e^{-\varphi_j}} - (N-r)\frac{\pi_j^u\{m(\theta_j^u)-\mu_j(1-F(0;\theta_j^u))\}}{1-G(0;\boldsymbol{\Xi}^u)} \\ -\; & (N-r)\frac{\pi_j^u e^{-\mu_j e^{-\varphi_j}}}{1-G(0;\boldsymbol{\Xi}^u)} \left\{ \nu(\theta_j^u;\varphi_j)\mu_j - \omega(\theta_j^u;\varphi_j) \right\} \\ =\; & 0, \qquad j = 1, 2, \ldots, k, \end{aligned} \tag{7}$$

where

$$\omega(\theta_j^u;\varphi_j) = \mu_j^u \nu(\theta_j^u;\varphi_j) + e^{\mu_j^u e^{-\varphi_j}} e^{\varphi_j^u} \int_{l(\theta_j^u)}^{\infty} v^{e^{\varphi_j^u - \varphi_j}} \log\{v\} e^{-v} dv.$$

The numerical solution of these equations can be carried out with the aid of *Mathematica* (code available from the first author). Iteration of the *expectation* (compute equation (4)) and *maximization* (solve equations (5), (6) and (7)) steps of the EM algorithm will provide us with estimates $\boldsymbol{\Xi}^\eta$ for $\boldsymbol{\Xi}$.

3. Using The Em Iterates To Estimate The Asymptotic Variance-Covariance Matrix

Meng and Rubin (1991) and Tanner (1996) explain how to estimate the asymptotic variance-covariance matrix for the EM estimates. The idea is to conceptualize the EM algorithm as a mapping $M : \Omega \longrightarrow \Omega$ from the parameter space onto itself such that $M(\Xi^u) = \Xi^{u+1}$. Provided that after a number of iterations M reaches a fixed point Ξ^η (the EM estimate) $M(\Xi^\eta) = \Xi^\eta$, then we can use the sequence $\Xi^0, \Xi^1, \ldots, \Xi^m$, $m < \eta$, to numerically differentiate $M(\Xi)$. Let us denote by $\frac{\partial M(\Xi)}{\partial \Xi}$ the resulting matrix and by $\frac{\partial^2 Q(\Xi,\Xi^\eta)}{\partial \Xi^2}$ the Hessian matrix for Q. For the case $k = 2$ and considering $\Xi' = (\mu_1, \varphi_1, \mu_2, \varphi_2, \pi_1, \pi_2)$ we have

$$-\frac{\partial^2 Q(\Xi, \Xi^\eta)}{\partial \Xi^2} = \begin{bmatrix} Q_{1,1} & 0 & Q_{1,3} & 0 & 0 & 0 \\ 0 & Q_{2,2} & 0 & Q_{2,4} & 0 & 0 \\ Q_{1,3} & 0 & Q_{3,3} & 0 & 0 & 0 \\ 0 & Q_{2,4} & 0 & Q_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & Q_{5,5} & 0 \\ 0 & 0 & 0 & 0 & 0 & Q_{6,6} \end{bmatrix}, \tag{8}$$

since the expressions for those $Q_{i,j} = Q_{i,j}(\Xi, \Xi^\eta)$ which are different from zero are quite lengthy, we omit to include them here. This matrix must be inverted in order to compute (9) below. However it is a sparse matrix, a condition which usually makes inversion feasible even for moderate dimensions.

Meng and Rubin (1991) state that the variance-covariance matrix

$$V(\Xi^\eta) = \left\{ -\frac{\partial^2 \log\{p(\Xi|\boldsymbol{x})\}}{\partial \Xi^2}\bigg|_{\Xi=\Xi^\eta} \right\}^{-1},$$

can be computed as

$$\begin{aligned} V(\Xi^\eta) &= \left\{ -\frac{\partial^2 Q(\Xi, \Xi^\eta)}{\partial \Xi^2}\bigg|_{\Xi=\Xi^\eta} \right\}^{-1} \left\{ I - \frac{\partial M(\Xi)}{\partial \Xi}\bigg|_{\Xi=\Xi^\eta} \right\}^{-1} \left\{ \frac{\partial M(\Xi)}{\partial \Xi}\bigg|_{\Xi=\Xi^\eta} \right\} \\ &+ \left\{ -\frac{\partial^2 Q(\Xi, \Xi^\eta)}{\partial \Xi^2}\bigg|_{\Xi=\Xi^\eta} \right\}^{-1}, \end{aligned} \tag{9}$$

where I stands for the $3k$-dimensional identity matrix. Meng and Rubin's results are used

in the next section in order to compute standard errors and confidence intervals for the EM estimates.

4. A Study Of Coverage For Em Confidence Intervals

The next simulation exercise for $k = 2$ populations was carried out in order to study the coverage of the confidence intervals produced with the method described in the previous sections. *Mathematica* code is available from the first author of this report upon request.

For each of the three parameter values in Table 1, the next steps were repeated 100 times

- A sample of the censored mixture described in Section 1 of our manuscript was simulated. We used $N = 369$ as sample size and $C = 630$ as censoring threshold.
- The EM estimates Ξ^{η} are computed iterating the *expectation* and *maximization* steps.
- Standard errors for these estimates were computed using the asymptotic variance-covariance matrix described in Section 3. Then considering an approximation to the normal distribution, a confidence interval is computed for each component of Ξ.

The coverage (percentage of intervals containing the true value of the parameter) is reported in Table 2 for each of the three parameter values.

Table 1: True parameter values ($\pi_2 = 1 - \pi_1$)

	Ξ_1	Ξ_2	Ξ_3
μ_1	-1.0451	-0.9836	-1.1374
μ_2	-0.5781	-0.5163	-0.6353
φ_1	-0.2362	-0.1776	-0.3273
φ_2	-0.1180	-0.0607	-0.1714
π_1	0.2954	0.26	0.33

Table 2: Coverage for 95 % intervals

	Ξ_1	Ξ_2	Ξ_3
μ_1	0.91	0.88	0.94
μ_2	0.96	0.88	0.96
φ_1	0.96	0.89	0.98
φ_2	0.98	0.91	0.97

References

Mendenhall, W. and Hader, R.J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, **45**, 504-520.

Meng, X.L. and Rubin, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association*, **86** (416) 899-909.

Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**(2) 195-239.

Tanner, M.A. (1996). *Tools for Statistical Inference* (3rd. ed.), Springer Verlag: New York.

Medición de la Desigualdad: Breve Revisión

Patricia Covarrubias Aguirre

Colegio de Postgraduados,Facultad Latinoamericana de Ciencias Sociales, Sede México

1. Introducción

Desde principios de siglo se han propuesto diferentes métodos y formas de medir la desigualdad, surgidas de áreas como la Teoría de la Información, la Economía y la Estadística. El desarrollo de estos métodos ha dado origen a la definición de diferentes tipos de órdenes estocásticos. En la actualidad el tema de la desigualdad ha cobrado gran importancia, debido a la relación que tiene con el problema social de la pobreza y la distribución inequitativa del ingreso. En este trabajo se presenta una reseña de algunos trabajos relevantes sobre medición de la desigualdad y se describen algunos de los diferentes métodos y medidas que se han utilizado para medir la desigualdad.

2. Definición general de una medida relativa de desigualdad

Es una relación funcional I entre un conjunto D de estados sociales y un conjunto R de puntos de comparación ordenados por una relación binaria $\gtrsim$. Esta medida extrae de un estado social dado $d \in D$, aspectos que son relevantes a la desigualdad.

$$\begin{aligned} I : D &\rightarrow R \\ d &\longmapsto I(d) \end{aligned}$$

La relación binaria satisface

1. $I(d) \gtrsim I(d^*) \Rightarrow d$ tiene un nivel de desigualdad por lo menos tan alto como el de d^*.

2. $I(d) \gtrsim I(d^*)$, $I(d) \lesssim I(d^*) \Rightarrow d$ y d^* tienen el mismo nivel de desigualdad, denotado como $I(d) \sim I(d^*)$.

3. $I(d) > I(d^*) \Rightarrow I(d) \gtrsim I(d^*) \wedge \neg\, (I(d) \backsim I(d^*))$, d tiene más desigualdad que d^*.

Como estructura adicional, se pide que $\gtrsim$ sea:

reflexiva, $\qquad (r \gtrsim r \qquad \forall\, r \in R)$,

transitiva, $\qquad (r \gtrsim r^*,\ r^* \gtrsim\ r^{**} \Rightarrow r \gtrsim\ r^{**})$,

antisimétrica, $\qquad (r \gtrsim\ r^*,\ \ r^* \gtrsim r\ \Rightarrow r \backsim\ r^*)$.

Las medidas numéricas como el coeficiente de Gini y la medida de entropía de Theil toman R como los reales y la relación " $\gtrsim$ " como " $\geqslant$ " (mayor o igual). Estas satisfacen las tres propiedades anteriores y además " $\geqslant$ " es completa en R, ya que todo $x \in R$ puede ser comparado con cualquier otro elemento $y \in R$, usando " $\geqslant$ " (incluso consigo mismo).

El dominio D está formado por hogares o personas y el elemento típico en D es un vector $\underline{X} = (X_1, X_2, ..., X_n)$, donde X_i es el ingreso de la i-ésima persona y n es el número de individuos. $D = U_{n=1}^{\infty} D_n$, donde $D_n = \{\underline{X} \in R^{\,n} : \sum_{i=1}^{n} X_i > 0$ y $X_i \geqslant 0 \ \ \forall i\}$, es decir, hay un ingreso positivo a distribuirse y nadie percibe un ingreso negativo; en algunos casos se usa un dominio $D_+ = \{\underline{X} \in D : X_i > 0 \ \ \forall i\}$ más pequeño.

3. Propiedades de las medidas relativas de desigualdad

Existen problemas al usar diferentes medidas de desigualdad para comparar diferentes poblaciones. Uno de ellos es que el ordenamiento que se obtiene puede depender de la medida que se utilice. Para tratar de resolver este problema se han utilizado dos enfoques: el de ranqueo parcial y el axiomático.

Ranqueo parcial

Cualquier clase (aceptable) de medidas puede no coincidir en algunas comparaciones, pero hay algunas otras en las que todas las medidas de esa clase deben coincidir. Entonces cabe preguntarse por la relación que debe haber entre dos distribuciones para asegurar que todas las medidas de desigualdad en una clase lleven a las mismas conclusiones. En otras palabras, ¿cuál es el orden inducido por una clase de medidas?, o suponiendo que ya se tiene un ordenamiento, ¿cuál es la clase de medidas consistentes con él?

Enfoque axiomático

El hecho de que dos medidas no coincidan puede ser irrelevante si se tienen buenas razones para preferir una sobre la otra. La pregunta es entonces, ¿cómo escoger una? Resulta apropiado en este momento preguntarse cuáles serían propiedades deseables en una medida de desigualdad y evaluar a las medidas de acuerdo a esas propiedades.

Se piden 4 propiedades básicas que las medidas relativas de desigualdad deben satisfacer:

1. Principio de transferencia (PD).- Algunas transferencias (las de ricos a pobres) disminuyen la desigualdad. De la misma manera, se requiere también que la desigualdad crezca siempre que un ingreso sea transferido de una persona pobre a otra menos pobre (más rica). A este tipo de transferencia se le llama transferencia regresiva. Por ejemplo, se obtiene $\underline{X} \in D$ a partir de $\underline{Y} \in D$ mediante una transferencia regresiva si dadas i, j tenemos que

$$\begin{aligned} i) &\qquad Y_i < Y_j \ , \\ ii) &\qquad X_j - Y_j = Y_i - X_i > 0 \ , \\ iii) &\qquad X_k = Y_k \qquad \forall k \neq i, j \ , \end{aligned}$$

   i.e., una transferencia regresiva implica un incremento en la desigualdad.

2. Simetría (S).- La medida es "ciega" al hecho de quién tiene qué. $\underline{X} \in D$ se obtiene de $\underline{Y} \in D$ mediante una permutación $\Rightarrow I(\underline{x}) \backsim I(\underline{y})$, es decir, la desigualdad no cambia si los individuos intercambian lugares.

3. Homogeneidad (H).- Doblando el ingreso de todos, la medida no debe cambiar. El cambio proporcional en los ingresos no modifica la desigualdad, es decir, si $\underline{X} \in D$ se obtiene de $\underline{Y} \in D$ por un cambio proporcional en el nivel de ingresos $\Rightarrow I(\underline{x}) \backsim I(\underline{y})$.

   Las 3 propiedades antes descritas se satisfacen para poblaciones del mismo tamaño. Para permitir poblaciones de diferentes tamaños se usa la noción de "replicación". $\underline{Y}$ es una replicación de $\underline{X}$ si $\underline{X} \in D$ y para alguna $m \geqslant 2$, se tiene $\underline{Y} = (Y^{(1)}, Y^{(2)}, ..., Y^{(m)})$, donde cada $Y^{(i)} = x$.

4. Principio de Población (PP): Si $\underline{Y}$ es una replicación de $\underline{X} \Rightarrow I(\underline{x}) \backsim I(\underline{y})$.

   El nivel de desigualdad no se debe afectar si el número de individuos con cada ingreso se duplica.

De estas 4 propiedades, no todas parecen tener una aceptación generalizada, por ejemplo, Arnold (1987) argumenta que únicamente el principio de Tranferencia (Pigou–Dalton), al que él llama el axioma de Robin Hood, es el único que cuenta con aceptación casi universal.

4. Algunas medidas comunes de desigualdad

Criterio de Percentiles

Una manera tradicional de ilustrar la desigualdad en la distribución del ingreso es dividir a la población en clases de igual tamaño o "percentiles" ordenando a los individuos de pobres a ricos y ver entonces el porcentaje total de los ingresos que le corresponde a cada clase. Estos métodos tabulares permiten observar cambios generales en la distribución, sin embargo tienen la desventaja de que ignoran la distribución dentro de las clases, es decir, de algún modo suponen uniformidad intra-clase.

(El Cuadrado del) Coeficiente de Variación

Si se considera la distribución del ingreso como una distribución de probabilidad, entonces $P[X = x]$ es la proporción de la población cuyo ingreso es x, de modo que pueden definirse de

la manera usual $E(X)$, $Var(X)$, $Desv.est.(X)$ y el coeficiente de variación $CV(X) = S/\overline{X}$. Este coeficiente funciona igual que la varianza para poblaciones con la misma media, pero cuando la media difiere, toma en cuenta efectos potenciales de escala y no depende de las unidades de medición. Resulta ventajosa su familiaridad, pero hereda de la varianza una "neutralidad en la transferencia" que no es aceptable en una medida de desigualdad. Una redistribución del ingreso en la cola izquierda de la distribución tiene el mismo efecto que una redistribución equivalente del lado derecho.

Algunos autores han sugerido que deben pesar más las transferencias en la cola inferior de la distribución. El coeficiente de variación puede tomar valores en $(0, \infty)$, pero generalmente se transforma para obtener una medida en el intervalo $[0, 1]$.

La Varianza Muestral

La varianza muestral $V(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $x \in D_n$, tiene una descomposición simple en dos términos, entre-grupos e intra-grupos, sin embargo no es una medida relativa de desigualdad. Si los ingresos en una distribución dada se duplican, el factor de incremento de la varianza es 4 y no 2. Una manera de convertir a V en una medida homogénea de desigualdad, es aplicarla al logaritmo de los ingresos en vez de a los ingresos directamente, i.e.,

$$Var\left[\ln\left(ingreso\right)\right] = V_L(x) = \frac{1}{n}\sum_{i=1}^{n}(\ln(x_i) - \ln(\widetilde{x}))^2, \qquad x \in D_n^+,$$

donde $\widetilde{x} = (\Pi_{i=1}^{n} x_i)^{1/n}$ es la media geométrica muestral. Con esto, se obtiene una medida homogénea que, sin embargo, no satisface el principio PD. Quizá la manera más natural de convertir a la varianza en una medida relativa de desigualdad es aplicar V a $\left(\frac{x}{\overline{x}}\right)$ en vez de a x, con lo cual se obtiene el cuadrado del coeficiente de variación

$$C^2(x) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i}{\overline{x}} - 1\right]^2, \qquad x \in D_n.$$

La Medida de Entropía de Theil

Surge de la Teoría de la Información y en un contexto de eventos aleatorios mutuamente excluyentes puede interpretarse como el grado de incertidumbre inherente o como la información esperada que se obtiene al realizarse un evento aleatorio. Alcanza su máximo cuando todos los eventos son igualmente probables y su mínimo cuando alguno de ellos tiene probabilidad 1 de ocurrir. Aplicando la medida de Shannon a la distribución de la participación de los ingresos se obtiene una medida de igualdad, que restada de su valor máximo nos dá la medida de (desigualdad) entropía de Theil. Esta última hereda una propiedad de aditividad que permite partir la desigualdad en "entre-grupos" e "intra-grupos", lo que facilita identificar las fuentes de desigualdad.

En 1967 H. Theil introdujo dos medidas de desigualdad con propiedades aditivas de descomposición. La primera de ellas, $T_1 : D \to R$ puede estimarse utilizando

$$T_1(x) = \sum_{i=1}^{n} \frac{x_i}{|x|} \ln \left[n \frac{x_i}{|x|} \right], \qquad x \in D_n,$$

donde se toma la convención $0 \ln(0) = 0$. Esta medida se conoce como la medida de entropía de Theil pues $T_1(x) = \ln(n) - S\left[\frac{x}{|x|}\right]$, donde $S \equiv -\sum_{i=1}^{n} p_i \log p_i$ (con $x \log x \equiv 0$, cuando $x = 0$), es la medida de Shannon, proveniente de la Teoría de la Información. Con la convención mencionada anteriormente, T_1 es continua en D_n y la propiedad más útil de esta medida es la descomposición que permite partir la desigualdad en dos piezas:

1. Desigualdad "intra-grupos", dada por la suma de la desigualdad del grupo ponderada por la participación en el ingreso total correspondiente al grupo.
2. Desigualdad "Entre-grupos", obtenida aplicando la medida a una distribución del ingreso "suavizada" donde la media del grupo reemplaza a los ingresos de los miembros de cada grupo.

La segunda medida de Theil T_2, usa las proporciones de la población en cada grupo como ponderadores en vez de la participación del grupo en el ingreso total. Se define en un dominio menor, $T_2 : D_n^+ \to R$ y se estima como $T_2(x) = \sum_{i=1}^{n} \frac{1}{n} \ln\left(\frac{|x|}{nx_i}\right)$, con $x \in D_n^+ = D_+ \cap D_n$.

Otras medidas

Función de bienestar social

Una medida muy aceptada por los economistas, debido a que se basa en conceptos puramente económicos, es la llamada Función de bienestar social, que asigna a cada distribución de ingresos un número que indica el nivel de bienestar social de una población dada. Esta función es creciente en el ingreso y "prefiere la igualdad", en el sentido de que crece siempre que dos ingresos se acercan. Dada una distribución específica, la cantidad más pequeña de ingresos que, otorgada a cada uno de los miembros de la población, resultará en el mismo nivel de bienestar social se conoce como Ingreso equivalente igualmente distribuido. Este es el concepto en el que se basa la construcción de esta medida.

Las medidas de Theil y el cuadrado del coeficiente de variación son todas medidas de una clase general de medidas de desigualdad que pueden descomponerse. Shorrocks y Cowell introdujeron en 1980 una clase (de un parámetro) generalizada de medidas de entropía:

$$I_c(x) = \begin{cases} \frac{1}{n}\frac{1}{c(c-1)}\sum_{i=1}^{n}\left[(\frac{x_i}{\overline{x}})^c - 1\right], & c \neq 0, 1 \\ \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{\overline{x}}\ln(\frac{x_i}{\overline{x}}), & c = 1 \\ \frac{1}{n}\sum_{i=1}^{n}\ln(\frac{x_i}{\overline{x}}), & c = 0 \end{cases} \qquad x \in D_n\ .$$

Las medidas de Atkinson pertenecen a una clase de medidas agregadas que, en general, no pueden descomponerse aditivamente. Sea $W : D_+ \rightarrow R$ una función de bienestar social de la forma:

$$W(x) = \begin{cases} [\sum_{i=1}^{n}\frac{i}{n}(x_i)^{1-\varepsilon_j}]^{\frac{1}{1-\varepsilon}}, & \varepsilon > 0,\ \varepsilon \neq 1, \\ \Pi_{i=1}^{n}(x_i)^{1/n}, & \varepsilon = 1, \end{cases} \qquad x \in D_n.$$

$W(x)$ es el nivel de ingresos que si fuera igualmente distribuído llevaría al mismo nivel de bienestar social que x. La porción del ingreso que se "desperdicia" debido a la desigualdad estaría dada por

$$A_\varepsilon(x) = \begin{cases} 1 - [\sum_{i=1}^{n}\frac{i}{n}(\frac{x_i}{\overline{x}})^{1-\varepsilon_j}]^{\frac{1}{1-\varepsilon}}, & \varepsilon > 0, \varepsilon \neq 1, \\ 1 - \Pi_{i=1}^{n}(\frac{x_i}{\overline{x}})^{1/n}, & \varepsilon = 1, \end{cases} \qquad x \in D_n.$$

y define a la clase de Medidas de desigualdad de Atkinson. A_ε corresponde a I_c con $\varepsilon = 1 - c$ vía la transformación

$$A_\varepsilon = \begin{cases} 1 - [c(c-1)I_c + 1, & c < 1, \\ 1 - e^{-I_c}, & c = 0. \end{cases}$$

La Curva de Lorenz

Sea $L(p)$ la proporción acumulada del ingreso total que corresponde a la fracción acumulada p de la población, cuando los ingresos se ordenan de manera ascendente. Es decir $p =$ proporción de la población $(0 < p < 1)$ y $L(p) =$ % del total de ingresos recibidos por la proporción p de la población con más bajos ingresos.

Si todos los ingresos fueran iguales, los individuos que forman la proporción p reciben $(100 \times p)\,\%$ del ingreso total. En ese caso la curva de Lorenz es la diagonal $x = y$. A mayor desigualdad, la curva se desplaza más a la derecha por debajo de la diagonal. Se dice que la distribución 1 tiene menos desigualdad que la 2 si $L_1(p) \geqslant L_2(p)$, $\forall p$, $0 < p < 1$, con desigualdad estricta para alguna p. Si las curvas se cruzan, entonces no son comparables bajo este criterio. El uso de la curva de Lorenz como medida de desigualdad se ha extendido por varias razones:

1. Tiene sentido.
2. Es fácil de entender e interpretar por las personas que hacen política pública.
3. Al tratarse de una función de distribución (en el sentido estadístico) , es posible estudiar sus propiedades y ventajas con métodos estadísticos formales.

La construcción de esta curva, propuesta por Max Otto Lorenz en 1905 y formalizada por Gaswirth en 1971, descansa sobre la misma lógica que el criterio de percentiles, pero tiene la ventaja de que contempla a toda la distribución. Sin embargo, en algunas ocasiones, cuando se quieren comparar curvas que se cruzan, sólo ofrece un ordenamiento parcial, por lo que en algunos casos se ha propuesto usar de manera complementaria una medida numérica que asigne un nivel de desigualdad a cada distribución.

Para cualquier variable aleatoria no negativa X, se denota su función de distribución como F_X y su curva de Lorenz correspondiente

$$L(u) = \frac{\left[\int_0^u F_X^{-1}(y)\,dy\right]}{\left[\int_0^1 F_X^{-1}(y)\,dy\right]}\,, \qquad 0 \leq u \leq 1\,,$$

donde

$$F_X^{-1}(y) = \sup\{x : F_X(x) \leq y\}\,, \qquad 0 < y < 1.$$

Con esta notación *el orden parcial de Lorenz* $\leq_L$ se define como sigue: $X \leq_L Y$ (i. e. X no exhibe más desigualdad que Y, en el sentido de Lorenz) si $L_X(u) \geq L_Y(u), \forall u \in [0,1]$.

La Consistencia de Lorenz es una propiedad que una medida de desigualdad debe tener para satisfacer el criterio de Lorenz, cuando éste aplica. La clase de medidas que satisfacen la propiedad de consistencia de Lorenz es, precisamente, la clase definida por las 4 propiedades PD, S, H y PP mencionadas anteriormente.

Consistencia de Lorenz: Una medida de desigualdad $I : D \to R$ se llama consistente en el sentido de Lorenz si $\forall x, y \in D$

i) $L(x) > L(y) \Rightarrow I(x) > I(y)$, y

ii) $L(x) \backsim L(y) \Rightarrow I(x) \backsim I(y)$.

El Coeficiente de Gini

Es una de las medidas numéricas más usadas, debido a que ofrece una interpretación intuitiva en términos de la curva de Lorenz. Este coeficiente se define como

$$G(X) = \int_0^1 (u - L_X(u))du\,,$$

y corresponde al doble del área contenida entre la curva de Lorenz y la diagonal $x = y$, que se tendría si todos los ingresos fueran iguales, es decir si la distribución del ingreso fuera igualitaria. De modo que mientras mayor es la desigualdad mayor será el valor del coeficiente de Gini (G). En los casos extremos, si todos los ingresos fueran iguales $G = 0$ y si hubiera sólamente un ingreso mayor que cero, entonces $G = 1$.

Además, para cada $x \in D_n$, el Coeficiente de Gini puede escribirse como

$$G(x) = \frac{1}{2n^2\overline{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \ ,$$

y a partir de esto se obtiene que

$$G(x) = \frac{1}{n^2\overline{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{mín}(x_i, x_j) \ .$$

En la primera expresión puede observarse que $G(x)$ es el promedio de las n^2 diferencias $|x_i - x_j|$ sobre 2 veces la media muestral. Es decir, mientras más grande sea la distancia promedio entre los ingresos, dada una media constante, más desigualdad habrá. De la segunda expresión, vemos que G no es diferenciable, aunque sí continua en D_n.

En la literatura pueden encontrarse varias interpretaciones intuitivas del coeficiente de Gini. Como ya se mencionó, una de ellas consiste en interpretarlo como dos veces el área entre la Curva de Lorenz y la diagonal. La segunda interpretación supone que si X_1 y X_2 son dos variables aleatorias obtenidas (con reemplazo) de una población con función de distribución F_X, entonces

$$G(x) = \frac{E\,|X_1 - X_2|}{E\,|X_1\,| + E|\,X_2|} \ .$$

G es la diferencia (absoluta) esperada entre los dos ingresos seleccionados aleatoriamente, dividida entre la suma de sus esperanzas. Esta segunda interpretación lleva a definir un nuevo coeficiente que, a juicio de algunos autores, constituye una medida más natural de desigualdad.

Sean X_1 y X_2 la primera y segunda realizaciones de X tomadas de la población, pero esta vez sin reemplazo, de modo que no hay posibilidad de que el ingreso de una misma persona sea elegido dos veces. Esto cambia la probabilidad de cada pareja X_1 y X_2 de $1/n^2$ a $1/n(n-1)$ y nos lleva a un coeficiente de Gini alternativo, definido como sigue

$$\widetilde{G}(x) = \frac{1}{2n(n-1)\overline{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \, .$$

Como $\widetilde{G}(x) = \frac{n}{n-1} G(x)$, las dos medidas asignan diferentes rangos de acuerdo a los tamaños de la población. $\widetilde{G}$ no es una medida relativa de desigualdad pero tiende al valor de G si el tamaño de la población es grande.

Otras medidas menos conocidas son el Índice de Pietra y el de Kakwani. El primero corresponde a la máxima distancia vertical entre la curva de Lorenz y la línea igualitaria y coincide además con el doble del área del triángulo más grande que puede inscribirse entre ellas, su expresión es

$$IP(X) = \max_{u \in (0,1)} [u - L_X(u)],$$

cuando F_X es estrictamente creciente en su soporte el máximo se alcanza en $u = F_X(E(X))$ y el índice puede expresarse como $IP(X) = E\left|X - E(X)\right| / 2E(X)$.

La misma longitud de la curva de Lorenz puede pensarse como una medida de desigualdad, Kakwani en 1980 definió un índice basado en esta longitud como $K(X) = \frac{l_X - \sqrt{2}}{2-\sqrt{2}}$, donde $l_X = \frac{1}{E(X)} E(\sqrt{E(X)^2 + X^2})$ es la longitud de la Curva de Lorenz.

5 Desarrollos más recientes

En la literatura se presentan elgunos intentos de extensiones de la Curva de Lorenz al caso multivariado. Para el caso bivariado Taguchi en 1972 y Arnold en 1983 propusieron dos candidatos para superficies de Lorenz. La superficie de Lorenz sugerida por Taguchi es la función $L(s,t)$, definida implícitamente por

$$\begin{aligned} s &= \int_0^{x_1} \int_0^{x_2} f_{X_1,\, X_2}(\xi, \eta)\ d\xi d\eta, \\ t &= \int_0^{x_1} \int_0^{x_2} \xi\ f_{X_1,\, X_2}(\xi, \eta)\ d\xi d\eta / E(X_1), \\ L(s,t) &= \frac{\int_0^{x_1} \int_0^{x_2} \eta\ f_{X_1, X_2}(\xi, \eta)\, d\xi\ d\eta}{E(X_2)}. \end{aligned}$$

mientras que la de Arnold, un tanto más fácil de interpretar, es la superficie $L(s,t)$ y se define implícitamente como

$$
\begin{aligned}
s &= \int_0^{x_1} f_{X_1}(\xi)\ d\xi, \\
t &= \int_0^{x_2} f_{X_2}(\eta)\ d\eta, \\
L(s,t) &= \frac{\int_0^{x_1}\int_0^{x_2} \xi\eta\ f_{X_1,X_2}(\xi,\eta)\, d\xi\ d\eta}{E(X_1X_2)}
\end{aligned}
$$

Koshevoy y Mosler (1996) presentan una generalización al caso multivariado, conocida como Zonoide de Lorenz que, para una distribución $d-variada$, es un subconjunto convexo cerrado en el hipercubo unitario en R^{d+1}. Sobre esa misma línea, en trabajos posteriores Koshevoy ha definido diferentes tipos de Zonoides. Otros autores como Mosler y Muliere (1998) trabajan sobre transformaciones que preservan el orden, considerando algunos principios diferentes al de Pigou-Dalton, en los que no cualquier transferencia del estilo Robin Hood es válida; sino que se fija un umbral que divide clases y sólo se permiten transferencias cerca del umbral o de una clase a otra. En estos últimos trabajos, la elección de dicho umbral por sí misma, constituye un problema abierto.

Referencias

Arnold, B. C. (1987). *Majorization and the Lorenz Order: A Brief Introduction*, VI, Vol. 43, Lecture Notes in Statistics, Springer Verlag.

Cowell, F. A. (1997). Measurement of Inequality, London School of Economics.

Foster, J. E. (1985) . Inequality Measurement, Proceedings of Symposia in Applied Mathematics, Vol. 33.

Koshevoy, G. , Mosler, K. (1996).The Lorenz Zonoid of a Multivariate Distribution, *Journal of the American Statistical Association*, Vol. 91, No 434, Theory and Methods.

Mosler, K., Muliere, P. (1998). Welfare Means and Equalizing Transfers, *METRON*, Vol. LVI, n. 3 – 4.

Regresión No Lineal Mediante Optimización por Enjambre de Partículas

Sergio de-los-Cobos-Silva
John Goddard C.
Departamento de Ingeniería Eléctrica, Universidad Autónoma Metropolitana-Iztapalapa
Blanca R. Pérez S.
Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa
Javier Trejos Z.
CIMPA, Escuela de Matemática, Universidad de Costa Rica

1. Introducción

En regresión no lineal se encuentra el problema, tan frecuente en distintos métodos estadísticos, de la obtención de mínimos locales (Draper y Smith(1968),Tomassone *et al* (1992)). En efecto, el método clásico de regresión no lineal es el conocido como de Gauss-Newton, el cual está basado en una aproximación de la función de ajuste mediante el uso de un polinomio de Taylor de primer orden, y en un procedimiento de búsqueda de un óptimo local (Bates y Watts(1988), Draper y Smith(1968)). Por ello, es razonable pensar en la implementación de técnicas de optimización global. Entre estas técnicas, podemos citar el recocido simulado, la búsqueda tabú y los algoritmos genéticos, las cuales han sido ampliamente usadas en distintos problemas estadísticos, de investigación operacional o de ingeniería.

Todas estas implementaciones de las técnicas de optimización mencionadas han mejorado o al menos igualado los resultados conocidos,(los detalles pueden consultarse en de-los-Cobos *et al* (1996) y Trejos *et al* (1998)). En regresión no lineal ya se ha aplicado la técnica de recocido simulado y búsqueda tabú (ver Trejos y Villalobos(1999) y Trejos *et al*(2001)) con resultados satisfactorios.

En el presente trabajo se propone la utilización del método heurístico de optimización por enjambre de partículas (PSO) (del inglés Particle swarm optimization) para encontrar soluciones al problema de regresión no lineal. Conviene señalar que nos restringimos al caso de la

regresión con una variable explicativa, pudiendo generalizarse fácilmente el trabajo al caso de varias variables explicativas; además, no se abordan los problemas relativos a la estimación estadística, como sería la obtención de intervalos de confianza para los parámetros.

El trabajo está organizado de la siguiente manera: primero se describirá el problema de la regesión no lineal como un problema de optimización, y se presentarán algunos de los métodos más conocidos para llevarla a cabo. Posteriormente se hace una presentación de la optimización por enjambre de partículas. Enseguida se presenta la implementación efectuada de PSO para el problema de regresión no lineal. Finalmente, se dan algunas conclusiones.

2. Regresión no lineal

Dadas dos variables $\mathbf{x}$ y $\mathbf{y}$ observadas sobre n objetos, donde $\mathbf{x}$ es una variable explicativa y $\mathbf{y}$ es una variable a explicar que depende de $\mathbf{x}$, se quiere describir la relación de dependencia de $\mathbf{y}$ respecto a $\mathbf{x}$ mediante una función f; es decir, se quiere establecer la relación funcional $\mathbf{y} = f(\mathbf{x}) + \epsilon$, donde ϵ es un término de error (en este trabajo no supondremos que ϵ sigue alguna distribución de probabilidad en particular.). La función f depende generalemente de ciertos parámetros, cuyo vector denotaremos $\vec{\theta}$, por lo que escribiremos a la función de regresión $f_{\vec{\theta}}$. Se utiliza el criterio de mínimos cuadrados para medir la calidad de la aproximación funcional propuesta:

$$S(\vec{\theta}) = ||\mathbf{y} - f(\mathbf{x})||^2 = \sum_{i=1}^{n} [y_i - f_{\vec{\theta}}(x_i)]^2 \tag{1}$$

donde $\mathbf{x} = (x_1, \ldots, x_n)^t$ y $\mathbf{y} = (y_1, \ldots, y_n)^t$ son los vectores de las observaciones de las variables, y $||\ ||$ es la norma Euclídea usual.

Salvo cuando $f_{\vec{\theta}}$ es una función lineal, se conoce una solución general a este problema. Debe notarse que en algunos casos la función $f_{\vec{\theta}}$ no es en sí lineal pero el problema de regresión puede *linealizarse*. Por ejemplo, el modelo $\mathbf{y} = \theta_1 e^{\theta_2 \mathbf{x} + \epsilon}$ es linealizable aplicando logaritmo natural, obteniéndose $\mathbf{y}' = \theta_3 + \theta_2 \mathbf{x} + \epsilon$, donde $\mathbf{y}' = \ln y$ y $\theta_3 = \ln \theta_1$. Sin embargo, el modelo $\mathbf{y} = \theta_1 e^{\theta_2 \mathbf{x}} + \epsilon$ no es linealizable mediante la función logaritmo, ya que el error ϵ es aditivo. En este trabajo no nos ocuparemos de los modelos que son linealizables, ya que éstos se

resuelven fácilmente mediante la regresión lineal clásica.

Algunos ejemplos de modelos que no son linealizables, son los siguientes:

- crecimiento logístico: $\mathbf{y} = \frac{\theta_1}{1+\exp[-\theta_2(\mathbf{x}-\theta_3)]} + \epsilon$
- crecimiento con decaimiento: $\mathbf{y} = \theta_1 \exp[-\theta_2(\mathbf{x} - \theta_3)^2] + \epsilon$
- Draper y Smith(1968): $\mathbf{y} = \frac{\theta_1}{\theta_1 - \theta_2}[e^{-\theta_2 \mathbf{x}} - e^{-\theta_1 \mathbf{x}}] + \epsilon$

2.1. Métodos de solución

El método más usado en regresión no lineal es el de Gauss-Newton, que se basa en una aproximación lineal de la función $f_{\vec{\theta}}$ cuando ésta es derivable. La aproximación de $f_{\vec{\theta}}(\mathbf{x})$ se basa en el polinomio de Taylor de primer orden alrededor del punto $\vec{\theta}^0 \in \mathbf{R}$:

$$f_{\vec{\theta}}(x_i) = f_{\vec{\theta}^0}(x_i) + \sum_{j=1}^{p} \left[\frac{\partial f_{\vec{\theta}}(x_i)}{\partial \theta_j}\right]_{\vec{\theta}=\vec{\theta}^0} (\theta_j - \theta_j^0). \tag{2}$$

Poniendo $f_i^0 = f_{\vec{\theta}^0}(x_i)$, $\beta_j^0 = \theta_j - \theta_j^0$ y $z_j = \left[\frac{\partial f_{\vec{\theta}}(x_i)}{\partial \theta_j}\right]_{\vec{\theta}=\vec{\theta}^0}$, entonces se puede ver que el modelo no lineal se aproxima por uno lineal de la forma:

$$\mathbf{y} - f^0 = \sum_{j=1}^{p} \beta_j^0 z_j^0 + \epsilon,$$

por lo que se puede obtener una estimación $\mathbf{b}^0$ de $\vec{\beta}^0$ usando regresión lineal múltiple.

Se procede usando un valor inicial $\vec{\theta}^0$, y por iteraciones sucesivas se construye una sucesión de valores de los parámetros de la regresión $\vec{\theta}^0, \vec{\theta}^1, \vec{\theta}^2, \ldots$, reemplazando durante la $k+1$-ésima iteración $\vec{\theta}^k$ por $\vec{\theta}^{k+1}$ en la aproximación, hasta que se converge a un valor estable. Claramente esta convergencia (esta convergencia no está ni siquiera garantizada, ver Draper y Smith (1968)[pp.464–465]) puede ser hacia un óptimo local del criterio $S(\vec{\theta})$ en (1), ya que

no se garantiza la convergencia a un óptimo global siendo la búsqueda local en los contornos elipsoidales de los distintos puntos del proceso iterativo. Por ello se ha pensado en el uso de técnicas de optimización combinatoria que traten de evitar esos mínimos locales del criterio, entre las que están como se mencionó anteriormente el recocido simulado y la búsqueda tabú, u otro procedimiento para optimización global como PSO, el cuál se presenta en este trabajo.

Otra técnica ampliamente usada en optimización es la llamada de descenso del gradiente. Se trata de un método iterativo que busca la dirección de máximo descenso en cada punto de la iteración. En el caso de la regresión, se debe mover una estimación $\vec{\theta^k}$ de θ en la dirección del vector $\left(-\frac{\partial S(\vec{\theta})}{\partial \theta_1}, \ldots, -\frac{\partial S(\vec{\theta})}{\partial \theta_p}\right)$. Debe notarse que si bien teóricamente el método converge, esta convergencia puede ser muy lenta.

Marquardt(1963) propuso un método que lleva su nombre y que trata de mejorar los defectos de los métodos de Gauss-Newton y de descenso de gradiente. El método está basado en una interpolación entre las direcciones que escogen esos dos métodos en cada iteración.

3. Regresión No Lineal mediante PSO

3.1. Optimización por enjembre de Partículas

La optimización por enjambre de partículas (PSO), es una parte de lo que se conoce como inteligencia de enjambre, y tiene sus raices en la vida artificial, psicología social, ingeniería y ciencia de la computación. Dicho método fué descubierto a través de la simulación de modelos sociales simplificados, como son el vuelo de las parvadas de pájaros o el movimiento de las escuelas de peces y particularmente en la teoría de enjambres. Las partículas o agentes en PSO se piensan como pájaros a prueba de colisión y el intento original fué el de la simulación de la elegante pero impredecible coreografía de la parvada. Un concepto importante fué el de considerar que uno de los objetivos del vuelo que realizaba la parvada era el de localizar alimento. En este punto la idea de que la dinámica de la parvada posibilitaba a los miembros de la parvada el capitalizar el conocimiento que tenian uno del otro. Por un lado cada agente "conocia"su mejor posición, así como la mejor posición del miembro que la había localizado.

La PSO se basa en el uso de un conjunto de partículas o agentes que corresponden a estados de un problema de optimización, haciendo que cada partícula se mueva en el espacio de soluciones en busca de una posición óptima. Una característica de PSO es que los agentes se comunican entre sí, y entonces -como en un sistema social- un agente con una buena posición (medida de acuerdo a una función objetivo) *influye* en los demás atrayéndolos hacia él. Cuando la población se inicializa, adicionalmente a que a las variables se les asigna valores aleatorios también se les asigna una velocidad aleatoria. En cada iteración, la velocidad de cada partícula es aleatoriamente acelerada hacia su mejor posición (donde el valor de la función de aptitud u objetivo es mejor) y a través de las mejores posiciones de sus vecinos. En términos generales las principales características de PSO son: evaluación, comparación e imitación.

3.2. Detalles de la Implementación

Se tomó cada partícula como un valor específico del vector de parámetros $\vec{\Theta}$ las que fueron generadas de manera aleatoria. Se consideraron $i = 1, \ldots, n$ partículas y en cada iteración t se identificaba la mejor posición de cada partícula $\vec{\Theta}_i^*(t)$ y la mejor partícula de todas en la iteración $\vec{\Theta}^*(t)$, esta evaluación se realizó mediante el criterio de mínimos cuadrados. Posteriormente se realizó el movimiento: $\vec{\Theta}_i^*(t+1) = \vec{\Theta}_i^*(t) + V(\vec{\Theta}_i^*(t+1))$, donde $V(\vec{\Theta}_i^*(t+1)) = V(\vec{\Theta}_i^*(t)) + \gamma_1 * rand_1 * (\vec{\Theta}_i^*(t) - \vec{\Theta}_i(t)) + \gamma_2 * rand_2 * (\vec{\Theta}^*(t) - \vec{\Theta}_i(t))$, $rand_i$ son valores aleatorios con distribución uniforme [0,1] y $\gamma_1 + \gamma_2 \leq 4{,}2$.

4. Ejemplo didáctico

Considérese el siguiente ejemplo didáctico (Antoniadis, et al 1992), en el que se quiere ajustar el modelo $\mathbf{y} = \theta_1 e^{-\theta_2 \mathbf{x}}$ para el conjunto el de datos siguiente:

x	−2,5	−1	1	2
y	1	1,1	−1,1	0,2

La función presenta dos mínimos locales de $S(\vec{\theta})$: $\vec{\theta}^* = (0{,}669, 0{,}214)$ con $S(\vec{\theta})^* = 1{,}968$ y $\vec{\theta}^1 = (0 - {,}764, -0{,}0298)$ con $S(\vec{\theta})^1 = 3{,}436$.

4.1. Resultados y conclusiones

Se realizaron varias experimentos donde se obtuvieron resultados bastantes satisfactorios. Así en el ejemplo anteriormente citado se obtuvieron resultados como los siguientes: Promedio de eficiencia del 98.01 % con una desviación estándar de 1.92 % en tan sólo 40 corridas con 50 partículas y 100 iteraciones por corrida, donde la eficiencia se calculó como: $1 - \frac{SC(\vec{\Theta}) - 1{,}968}{1{,}968}$, donde $SC()$ es la suma de cuadrados de la desviación.

Una característica interesante de PSO es su facilidad de programación e implementación computacional. Por falta de espacio no se pueden presentar otros resultados comparativos exhaustivos, pero mencionaremos que e n muchos otros conjuntos de datos hemos obtenido resultados similares. A manera de conclusión diremos que PSO no necesita que la función del modelo sea derivable, y si lo es no se necesita conocer su derivada. Aún se deben realizar comparaciones más amplias, estudiando la sensibilidad de los parámetros de PSO, trabajo que se realizará en el futuro cercano.

Referencias

Antoniadis, A.; Berruyer, J.; Carmona, R. (1992) *Régression Non Linéaire et Applications.* Economica, Paris.

Bates, D.M.; Watts, D.G. (1988) *Nonlinear Regression Analysis and its Applications.* John Wiley & Sons, New York.

de-los-Cobos-Silva S.; Pérez-Salvador B.; Gutiérrez Andrade M. (1996) *Programación Estocástica en Optimización.* IMSIO-DECFI,Universidad Nacional Autónoma de México, México.

Draper, N.R.; Smith, H. (1968) *Applied Regression Analysis.* John Wiley & Sons, New York.

Kennedy, J.: Eberhart, R.C. (2000) *"Intelligent Swarm Systems"*. Academic Press, New York.

Kennedy, J.: Eberhart, R.C.; Shi Y. (2001) *"Swarm Intelligence"*. Morgan Kaufmann.

Marquardt, D.W. (1963) "An algorithm for least squares estimation of nonlinear parameters", *Journal of the Society for Industrial and Applied Mathematics* **2**: 431–441.

Ratkowsky, D.A. (1983) *Nonlinear Regression Modeling. A Unified Practical Approach.* Marcel Dekker, Inc., New York.

Reeves, C. (Ed.) (1995) *Modern Heuristic Techniques for Combinatorial Problems.* McGraw-Hill, London.

Tomassone, R.; Audrain, S.; Lesquoy, E.; Millier, C. (1992) *La Régression. Des Nouveaux Regards sur une Ancienne Méthode Statistique.* Masson, Paris.

Trejos, J.; Murillo, A.; Piza, E. (1998) Global stochastic optimization for partitioning, In: A. Rizzi et al. (Eds.) *Advances un Data Science and Classification*, Springer-Verlag, Berlin.

Trejos, J.; Villalobos, M. (1999) Optimización mediante recocido simulado en regresión no lineal. *Memorias del XII Foro Nacional de Estadística*, Monterrey: 183–190.

Trejos-Zelaya, J.; de-los-Cobos, S. S.; Villalobos, M. (2001) Aplicación de la Búsqueda Tabú en Regresión No Lineal. *VIII Congreso Latinoamericano de Probabilidades y Estadística Matemática*, 12-16 de noviembre, La Habana, Cuba.

Trejos, J.; Goddard, J.; Piza, E.; de-los-Cobos, S. S.(2002) Clasificación de Datos Numéricos Mediante Optimización por Enjambre de Partículas. *5th. International Conference of Operations Research*, March 4-8, Havana, Cuba.

Análisis Espacial de la Calidad del Agua Marina en el Norte de Quintana Roo

Carlos Díaz Ávalos
Ma. Esther Pérez Trejo
IIMAS, UNAM
Jorge Herrera Silveira
CINVESTAV IPN Unidad Mérida

1. Introducción

La presencia de correlación espacial en datos ecológicos es más la regla que la excepción. La rama de la estadística relacionada a este tipo de análisis es la llamada Estadística Espacial, la cual se ha desarrollado desde principios del siglo XX y que actualmente ha cobrado un gran auge debido a la disponibilidad de software más potente, los cuales permiten ajustar modelos espaciales complejos. La teoría de variables regionalizadas (Matheron, 1965) proporciona herramientas estadísticas adecuadas para el análisis espacial de dichos datos, haciendo posible la obtención de interpoladores óptimos en cuanto a los criterios de insesgamiento y minimización de la varianza de errores de interpolación. Este error de predicción puede incorporarse tanto a los mapas geoestadísticos como a los procesos de toma de decisiones (por ejemplo, la delimitación de áreas contaminadas). En este trabajo se presentan los resultados de un análisis espacio-temporal de la calidad del agua en tres regiones marinas del norte del estado de Quintana Roo. Las áreas consideradas fueron: Cancún, Punta Nizuc e Isla Mujeres, debido a que se encuentran dentro de la zona de influencia de un desarrollo turístico. El objetivo es evaluar el impacto que ha tenido la presencia de la zona turística en la calidad del agua y, por consiguiente, sobre el ecosistema de los lugares. Para esto, es necesario ubicar las áreas de influencia y la intensidad de las descargas de agua dulce provenientes del continente. Las variables bajo estudio fueron: fluorescencia, temperatura, salinidad y transparencia. Los datos fueron tomados utilizando un muestreador automático equipado con un geoposicionador durante las temporadas de nortes (enero 2001), secas (mayo 2001) y lluvias (septiembre 2001).

2. Geoestadística

La estadística tradicional se basa en el supuesto de que las observaciones son una realización de variables aleatorias independientes e idénticamente distribuidas. En este caso, z representa la variable de interés y x denota la ubicación en que dicho valor de z fue observado (de hecho x es un vector cuyas componentes son latitud y longitud). Los detalles y fundamentos teóricos de la teoría de variables regionalizadas pueden consultarse en textos especializados como Cressie (1993), Goovaerts (1997) y Chiles y Delfiner (1999). Nosotros solo describiremos aquellos conceptos que metodológicamente son importantes.

La base del análisis geoestadístico es la función variograma, definida por

$$\gamma(h) = \tfrac{1}{2}E\{[z(x) - z(x+h)]^2\}$$

donde h es la distancia que separa las observaciones realizadas en los puntos x y $x+h$. De la definición, es claro que $\gamma(h) \to 0$ a medida que $h \to 0$. Es decir, aquellas observaciones realizadas en puntos cercanos tienden, en promedio, a ser semejantes. En la práctica $\gamma(h)$ no se conoce debe estimarse a partir de las observaciones. El estimador de , o variograma empírico utilizado con más frecuencia es el de momentos (Matheron, 1965).

Una vez que se ha obtenido el variograma empírico, se le ajusta un modelo teórico con el propósito de modelar la asociación espacial entre las observaciones. Este modelo es posteriormente incorporado en el proceso de predicción espacial conocido como *kriging*. En general, los modelos ajustados a los variogramas empíricos contienen tres parámetros: la pepita, la meseta y el rango. El rango corresponde a la distancia aproximada para la cual dos observaciones se vuelven no-correlacionadas. La pepita y la meseta son proporcionales a la variabilidad de $z(x)$ a corta y gran escala, respectivamente. Armstrong (1998) explica que la tasa de crecimiento del variograma con respecto a la distancia indica que tan rápidamente la influencia de una muestra disminuye al aumentar la distancia. Una vez que el variograma ha alcanzado su valor límite (meseta), ya no existe correlación alguna entre muestras. Esta distancia crítica, conocida como rango, proporciona una definición de la noción de "zona de influencia".

Un concepto importante es el de *anisotropía*. De acuerdo a lo descrito por Armstrong (1998),

cuando el variograma se calcula para diferentes direcciones, en ocasiones se comporta de manera distinta para algunas de ellas (es decir, hay anisotropía). Si esto no ocurre, el variograma depende solamente de la magnitud de la distancia entre dos puntos y se dice que es *isotrópico*. Se distinguen dos casos de anisotropía: la geométrica y la de zona. En la anisotropía de zona los variogramas tienen la misma meseta en todas las direcciones, aunque sus rangos son distintos. En la anisotropía de zona, las mesetas de los variogramas no son las mismas en todas las direcciones.

3. Metodología

Las coordenadas geográficas latitud y longitud para cada muestreo fueron transformadas a distancias horizontales y verticales con respecto a la posición del radiofaro (VOR) del aeropuerto internacional de Cancún. Lo anterior se llevó a cabo con el propósito de poder interpretar el rango de los variogramas como distancias de dependencia espacial. La selección de este punto fue arbitraria y no tiene efecto sobre las conclusiones resultantes del análisis de datos. Los variogramas empíricos, así como los ajustes de modelos y las predicciones de kriging para las cuatro variables en los tres polígonos se realizaron empleando el paquete estadístico S-Plus (Mathsoft, Seattle, WA). Una vez obtenidas las predicciones de kriging y sus correspondientes desviaciones estándar del error se procedió a elaborar los mapas de predicciones y sus correspondientes varianzas de error. Cabe señalar que algunas variables presentan problemas estadísticos, tales como presencia de outliers, anisotropía de zona o geométrica, etc. Por consiguiente, la manera de llevar a cabo las predicciones diferirá de un caso a otro. La tabla 1 hace referencia a estos casos para cada una de las variables bajo estudio. Estas situaciones se describen en la siguiente sección.

4. Resultados

La tabla 1 muestra los modelos ajustados a los variogramas empíricos, así como los parámetros respectivos. La fluorescencia presentó los valores más bajos tanto de pepita como de meseta. Este hecho sugiere que esta variable tiene varianza global baja en todos los polígonos muestreados. La variabilidad espacio-temporal se ilustra con los mapas (Figura 1) de

valores krigados para tres variables en todas las épocas del año: temperatura y transparencia para Punta Nizuc, y fluorescencia para Isla Mujeres.

4.1. Referencia a casos en tabla 1

1) Estas variables se analizaron con la metodología estándar, ya que no presentaron ningún problema o anomalía (no hay observaciones atípicas o algún tipo de anisotropía). Por consiguiente, fue posible ajustar modelos teóricos a sus variogramas empíricos. Haciendo uso de estos modelos, se llevó a cabo la predicción mediante kriging, así como la realización de los mapas de zona correspondientes.

2) Haciendo uso de las gráficas de las variables con respecto a su índice, se encontró que estas variables presentan observaciones atípicas las cuales distorsionan el comportamiento de los variogramas empíricos. Las observaciones atípicas fueron eliminadas para el ajuste de los variogramas. Una vez ajustados los variogramas, los puntos atípicos fueron eliminados para realizar la predicción mediante kriging, debido a que existen elementos para pensar que se trata de errores de medición.

3) Misma situación que en 2), sin embargo en este caso los puntos atípicos sí fueron tomados en cuenta para realizar la predicción mediante kriging. 4) Esta variable tiene anisotropía geométrica. La predicción mediante kriging se realizó tomando en cuenta este hecho.

5) Estas variables presentaron anisotropía de zona. A fin de llevar a cabo una predicción que tome en cuenta dicha anisotropía, se ajustó un modelo gam con predictores longitud y latitud. Se exploraron los variogramas empíricos de los residuos para cada variable. Las predicciones zonales de cada variable y la elaboración de los mapas de zona están dados por la suma de las predicciones los residuos krigados (parte aleatoria de los modelos gam) más las predicciones de las partes deterministas de los modelos.

6) Esta variable presentó anisotropía de zona. A diferencia de las variables en 4), el variograma empírico de los residuos de su modelo gam no es isotrópico. Se llevaron a cabo predicciones separadas para cada zona diferenciada. La estimación por zona siguió la metodología estándar.

Variable	Polígono	Época	Caso	Modelo	Pepita	Meseta	Rango (km)
Fluorescencia	Isla Mujeres	Nortes	1	Esférico	0.0009	0.0006	0.3
		Lluvias	1	Esférico	0.00005	0.000039	0.94
		Secas	1	Gaussiano	0.000028	0.000019	0.8
Fluorescencia	Cancún	Nortes	1	Esférico	0.00032	0.00022	1
		Lluvias	1	Esférico	0.000009	0.000011	2
		Secas	1	Esférico	0.00004	0.000135	1
Fluorescencia	Punta Nizuc	Nortes	7	Pepita (*)			
		Lluvias	5	Esférico	0.000012	0.000031	0.55
		Secas	1	Esférico	0.00005	0.00004	0.7
Temperatura	Isla Mujeres	Nortes	5	Esférico	0.014	0.015	0.6
		Lluvias	1	Seno	1.6	2	0.125
		Secas	1	Esférico	0.013	0.04	1.45
Temperatura	Cancún	Nortes	1	Gaussiano	0.013	0.015	1
		Lluvias	1	Gaussiano	0.05	2.3	0.4
		Secas	3	Exponencial	0.006	0.013	0.075
Temperatura	Punta Nizuc	Nortes	7	Pepita (*)			
		Lluvias	1	Esférico	0.055	0.37	0.77
		Secas	7	Pepita (*)			
Salinidad	Isla Mujeres	Nortes	2	Seno	0.0029	0.0018	0.058
		Lluvias	1	Seno	0.56	1.4	0.1
		Secas	1	Gaussiano	0.007	0.022	0.8
Salinidad	Cancún	Nortes	4	Exponencial (†)	0.0023	0.0032	0.8
		Lluvias	1	Gaussiano	0.01	0.9	0.4
		Secas	3	Seno	0.009	0.016	0.09
Salinidad	Punta Nizuc	Nortes	3	Esférico	0.014	0.01	0.35
		Lluvias	1	Gaussiano	0.02	0.15	0.4
		Secas zona 1	6	Esférico	0.0023	0.0005	0.22
		Secas zona 2		Exponencial	0.0015	0.001	0.6
Transparencia	Isla Mujeres	Nortes	10		0.025	-	-
		Lluvias	1	Esférico	0.015	0.031	0.5
		Secas	1	Gaussiano	0.022	0.03	0.45
Transparencia	Cancún	Nortes	10		0.031	-	-
		Lluvias	10		0.00073	-	-
		Secas	8	Exponencial	0.002	0.0023	0.38
Transparencia	Punta Nizuc	Nortes	3	Gaussiano	0.026	0.02	1
		Lluvias	1	Esférico	0.005	0.0225	0.8
		Secas	1	Esférico	0.0075	0.0105	0.5

* Predicciones con GAM
† La proporción de anisotropía es de 0.111 y el ángulo es de 22.5° (NE)

Tabla 1: Modelos Ajustados

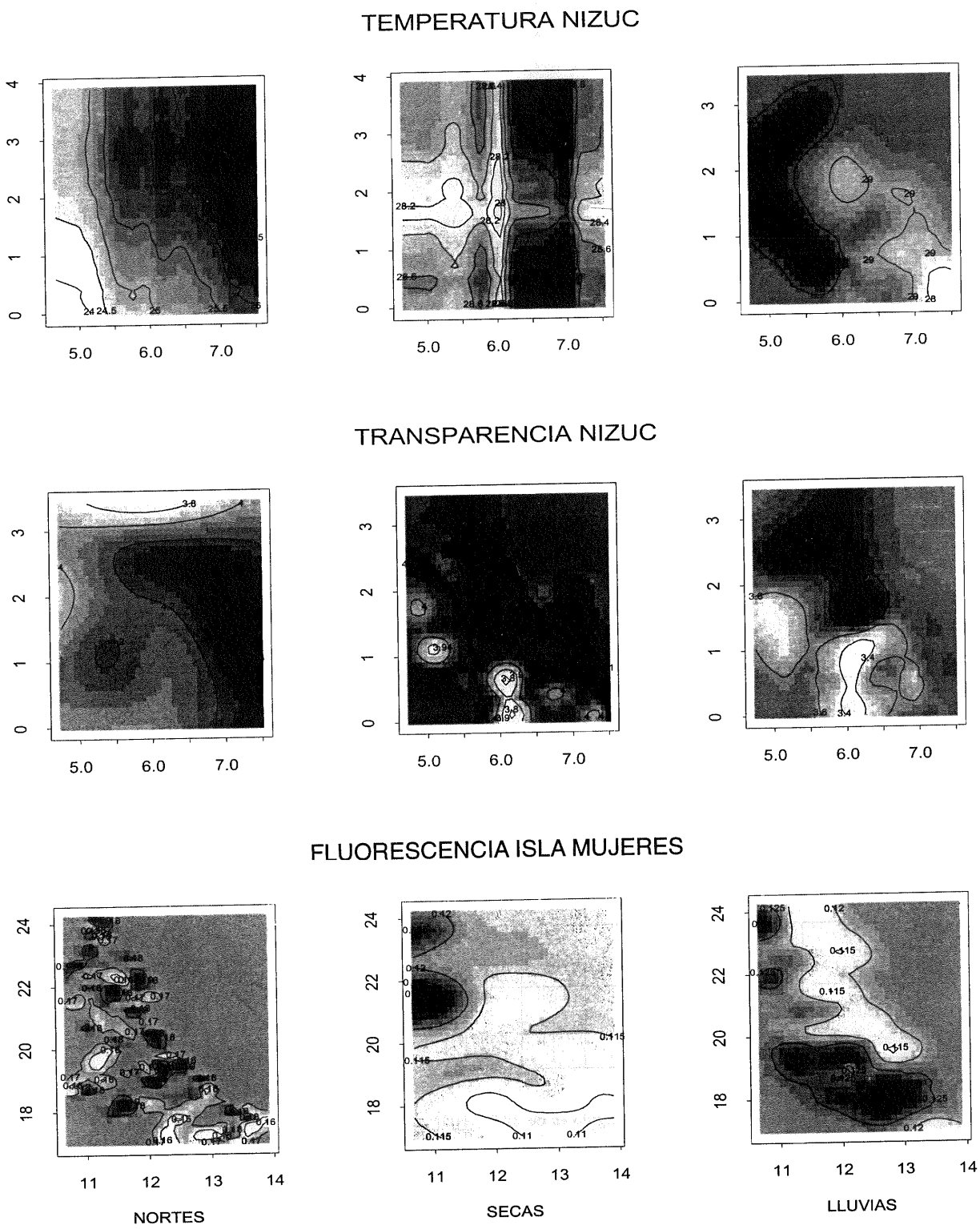


Figura 1: Mapas de valores Krigados

7) Estas variables muestran un comportamiento con tendencia. Se ajustaron modelos gam a las variables respuesta con predictores longitud y latitud. En estos casos, se encontró que los residuos del gam no presentan estructura espacial. Las predicciones y la elaboración de mapas de zona se realizaron utilizando únicamente las partes deterministas de los modelos. Este hecho está indicando que las variables se explican únicamente en términos de la posición geográfica.

8) Misma situación que en 7), sin embargo en este caso la predicción está dada por la suma de los residuos krigados (parte aleatoria del gam) mas la predicción de la parte determinista del modelo.

9) Estas variables no mostraron estructura espacial alguna. Sus variogramas empíricos, son sólo efecto pepita. Por lo tanto, no tiene sentido realizar ningún tipo de predicción.

10) Para estas variables, también se encontró que existen observaciones atípicas cuya presencia distorisionaba el comportamiento de sus variogramas empíricos. Por lo tanto, los outliers fueron eliminados para ajustar los variogramas. éstos revelaron que las variables no tienen estructura espacial alguna, ya que sólo presentan un efecto pepita. Por consiguiente, no tiene caso realizar predicción alguna.

5. Conclusiones

Los resultados indican fuertes diferencias en el comportamiento de las variables tanto espacial como temporalmente, sugiriendo que los modelos de administración y manejo de las descargas deben ser distintos para cada polígono. Este estudio puede ser la base para estudios más avanzados a realizar en un futuro cercano. Por ejemplo, sería posible llevar a cabo una predicción más precisa si se contara con más observaciones en el tiempo (por decir algo, observaciones mensuales durante un período de tres años). En este caso, se podría calcular y modelar un variograma empírico que fuese función no solamente de la distancia sino también del tiempo.

Referencias

Armstrong Margaret. (1998). *Basic Linear Geostatistics.* Springer.

Chilés Jean-Paul & Pierre Delfiner. (1999). *Geostatistics.* Modeling Spatial Uncertainty. Wiley & Sons, Inc.

Cressie N.A. (1993). *Statistics for Spatial Data.* Wiley NY.

Goovaerts Pierre. (1997). *Geostatistics for Natural Resources Evaluation.* Oxford University Press.

Hastie T.J. & Tibshirani R.J. (1990). *Generalized Additive Models.* Chapman & Hall.

Isaaks Edward H. & Srivastava Mohan R. (1989). *Applied Geostatistics.* Oxford University Press.

Matheron G. (1965). *Les Variables Régionalisees et leur Estimation. Una Application de la Theorié des Fonctions Aleatories aux Sciences de la Nature* . Massou, París.

Wackernagel Hans. (1998). *Multivariate Geostatistics.* Springer.

Diseños factoriales 2^k o 2^{k-p} en doble arreglo

Jorge Domínguez Domínguez
Centro de Investigación en Matemáticas

1. Introducción

Los diseños factoriales o factoriales fraccionados en un doble arreglo consideran tanto los factores de control como los de ruido. Los factores de ruido sólo se pueden controlar durante el desarrollo del experimento, pero prácticamente es imposible o costoso controlarlos durante el proceso de producción.

La idea de realizar un experimento a través de esta estrategia fue inicialmente presentada por Taguchi (1986) mediante el diseño robusto. Por la importancia que adquirió este tema en la industria, aproximadamente en los últimos 20 años se ha realizado una amplia investigación en esta dirección. Un excelente resumen de las ideas principales del diseño robusto las expone en un panel de discusión Nair (1992).

El objetivo de este trabajo es describir la estrategia experimental que se utiliza en la aplicación del doble arreglo factorial completo o fraccionado. Para ello se ilustran tres ejemplos en diferentes procesos industriales donde se presenta la necesidad de aplicar la estrategia del diseño robusto

2. Planteamiento del problema

Los ingenieros que desarrollan productos desean que estos funcionen de manera adecuada el mayor tiempo posible y que no les afecten condiciones externas. Esto dió lugar a lo que se conoce como desarrollo de productos robustos. Es frecuente que en la planeación experimental se proponga un diseño factorial completo o fraccionado en dos niveles (2^k o 2^{k-p}) para identificar que factores tienen un efecto en la respuesta. Si en los k factores se incluyen los de ruido o los que son difíciles de controlar en el proceso entonces se pueden incluir en la

relación definitoria, cómo se verá más adelante. Por lo general, con los factores de ruido se pueden caracterizar restricciones de aleatorización. Los diseños que se plantean con esas restricciones dan lugar a los diseños factoriales en bloques o en parcelas divididas.

La motivación por aboradar este tema surge de la consultoría y colabaración con algunos ingenieros en diferentes procesos de manufactura. A partir de este actividad se han generado diferentes estrategias experimentales las cuales tienen varias alternativas para realizar el experimento. Estas alternativas caen dentro de la conceptualización del diseño robusto y puede tratarse en los esquemas de los diseños factoriales en dos niveles. Se cuenta con tres casos, estos se describirán tal como se plantearon.

Situación 1. Un ingeniero estudia la igualación de tonos para aplicarlo al acabado final de un producto. Él plantea un esquema para determinar que factores afectan la brillantez. Se tienen cuatro factores de control y tres factores de ruido todos en dos niveles. Los factores de control A: posición de quemado, B: cantidad de óxido tipo r, C: cantidad de óxido tipo s, D: espesor de aplicación. Los factores de ruido son: P velocidad de la cadena, Q: temperatura de quemado, R carga del horno.

Situación 2. Se consideran seis factores que dan lugar a una formulación que se aplica en la manufactura de pieles. Ésta se pone a dos tipos diferentes de pieles. Se desea encontrar la fórmula que reduzca los costos de operación, tal que no se afecte a las caraterísticas de calidad del acabado en la piel. En total se tienen siete factores, seis corresponden a diferentes substancias y un al tipo de piel, cada factor tiene dos niveles.

Situación 3. Mediante el proceso de fermentación sólida del bagacillo de café, se produce la β mananaza. Ésta subtancia se utiliza en la industria para blanquear el papel. Los factores, en dos niveles que intervienen en el proceso son A: pH, B: flujo de aire, C: tiempo, D: glucosa, E: petona, F: humedad. Se realizan 16 pruebas experimentales, sin embargo por las condiciones del laboratorio es necesario dividir el experimento en cuatro bloques.

Estas tres situaciones ilustran el tipo de problemas que ocurren en la práctica cuando se desarrolla un producto, o cuando se planea la necesidad de realizar estudios de robustez. Abordar la discusión de diferentes estrategias experimentales en estas situaciones requiere de un mayor espacio. Aquí sólo bosquejaremos la solución a los tres casos citados.

Ante las situaciones planteadas existen otros diseños alternativos para realizar la experimentación, este procedimiento se ilustrará de manera breve con la problemática de la situación 1. A la vez, éste da un guía para obtener la estructura de confusión y con ello estudiar el efecto de los factores de control y de ruido así como sus posibles interacciones.

3. Alternativas de esquemas experimentales

En las tres situaciones, la estrategia experimental apropiada es el diseño factorial en dos niveles. En estos experimentos, el número de tratamientos por realizar es grande dado que se tienen varios factores. Por ello es necesario dividir el experimento y eso da lugar a un diseño factorial fraccionado, es decir, 2_R^{k-p} donde p son las veces que se divide el experimento y R la resolución. En este esquema se define un generador que es la guía para fraccionar, con este se obtiene una estructura de confusión. Esta última indica que los efectos no se pueden estimar independientemente uno de otro, uno está completamente o parcialmente correlacionado con el otro.

Planteado así, k representa tanto a los factores de control como a los de ruido, en las dos primeras situaciones descritas en el apartado anterior. En la tercera situación k representa a los factores de control y a los bloques. De esta manera, al escribir la estructura de confusión, se tiene que los factores de ruido están confundidos con el efecto de interacción triple de los factores de control. Esta información se describe en la Tabla 1, ésta también contiene los generadores para construir la estructura de confusión para otros diseños 2_R^{k-p} que servirán de referencia más adelante.

En la Tabla 1, las columnas que representan a $E = ABC$ y a $F = BCD$ indican los generadores para el diseño 2_{IV}^{6-2}. De manera análoga tenemos los generadores para 2_{IV}^{7-3} y 2_{III}^{8-4} con ellos obtenemos la estructura de confusión de cada una de estos esquemas. Es conveniente indicar que con esta estrategia de confusión y estos esquemas planteados en un doble arreglo se obtendrá la estrategia experimental para cada una de las tres situaciones planteadas.

Trat	2^4_V	A	B	C	D	ABC	BCD	ACD	ABD	$ABCD$	
1		-1	-1	-1	-1	-1	-1	-1	-1	1	y_1
2		1	-1	-1	-1	1	-1	1	1	-1	y_2
3		-1	1	-1	-1	1	1	-1	1	-1	y_3
4		1	1	-1	-1	-1	1	1	-1	1	y_4
5		-1	-1	1	-1	1	1	1	-1	-1	y_5
6		1	-1	1	-1	-1	1	-1	1	1	y_6
7		-1	1	1	-1	-1	-1	1	1	1	y_7
8		1	1	1	-1	1	-1	-1	-1	-1	y_8
9		-1	-1	-1	1	-1	1	1	1	-1	y_9
10		1	-1	-1	1	1	1	-1	-1	1	y_{10}
11		-1	1	-1	1	1	-1	1	-1	1	y_{11}
12		1	1	-1	1	-1	-1	-1	1	-1	y_{12}
13		-1	-1	1	1	1	-1	-1	1	1	y_{13}
14		1	-1	1	1	-1	-1	1	-1	-1	y_{14}
15		-1	1	1	1	-1	1	-1	-1	-1	y_{15}
16		1	1	1	1	1	1	1	1	1	y_{16}

2^{5-1}_V	A	B	C	D			F
2^{6-2}_{IV}	A	B	C	D	E	F	
2^{7-3}_{IV}	A	B	C	D	$P=ABC$	$Q=BCD$	$R=ACD$
2^{8-4}_{III}	A	B	C	D	$G=ABC$	$E=BCD$	$F=ACD$ $H=ABD$

Tabla 1 Diseño factorial 2^4 y diferentes fracciones a partir de este esquema.

Al efectuar el experimento 2^{k-p}_R es frecuente que exista dificultad en mover los niveles de los factores de ruido. Para suavizar esta, un alternativa es considerar el doble arreglo propuesto por Taguchi, para ello tendríamos que dividirlo como se muestra en la Tabla 2. Ahora, en esta aproximación no se contempla el posible efecto de los factores de ruido, ni su relación con los factores de control, información que es relevante para los ingenieros en los procesos cuando su interés es diseñar productos robustos.

	P Q R	$2^{l-r}_{R_2}$	
$A \quad B \quad C \quad D$			
	y_{11}	$\cdots$	y_{m1}
	$\cdot$	$\cdots$	$\cdot$
$2^{k-p}_{R_1}$	$\cdot$	$\cdots$	$\cdot$
	$\cdot$	$\cdots$	$\cdot$
	y_{1n}	$\cdots$	y_{mn}

Tabla 2. Doble arreglo ortogonal en $2^{k-p}_{R_1}$ y $2^{l-r}_{R_2}$

Debido a la restricción en la aleatorización en el diseño robusto, una segunda alternativa es plantear el doble arreglo como un diseño en parcelas divididas, donde toda la parcela (WP) corresponde al arreglo interno y la subparcela (SP) al arreglo externo. Un diseño factorial fraccionado se puede proponer en la parcela, es decir $2^{k-p}_{R_1}$ y otro en la subparcela, o sea $2^{l-q}_{R_2}$ finalmente este esquema concide con el de la Tabla 2. En este enfoque hay varios trabajos de investigación, por ejemplo Box-Jones (1991), Shoemeker, Tsui y Wu (1991), Bisgaard (2000) y Loeppky y Sitter (2002).

4. Discusión

Una vez que tenemos descrito el esquema de doble arreglo ortogonal, utilizamos éste para plantear la estrategia experimental en las tres situaciones propuetas. En las primeras dos situaciones las planteamos como parcelas divididas y la tercera como un factorial fraccionado en bloques.

En la primera proponemos como WP: $2^{4-1}_{R_1}$ los factores de control y SP: $2^{3-1}_{R_2}$ los factores de ruido. De aquí se originan varias preguntas ¿cómo identificar factores activos en WP y en SP? ¿Cómo identificar interacciones entre los factores en WP y SP? Para contestar a éstas, varias alternativas se presentan al proponer diferentes generadores (algunas de éstas se comentan de manera muy breve y sin mucho detalle):

1. Para el diseño $2^{4-1}_{R_1}$ se propone como generador $D = ABC$ y en el diseño $2^{3-1}_{R_2}$ se considera $R = PQ$. Así la relación definitoria es: $I = ABCD = PQR = ABCDPQR$ y se genera la estructura de confusión. De aquí se destaca que, los efectos principales de los factores de control estan confundidos con efectos de tercer orden, así $R_1 = IV$. Los efectos principales para los factores de ruido estan confundidos con efectos dobles, así $R_2 = III$.

2. Ahora, si como generador se propone $D = PQR$, el factor de control se confunde con el efecto de interacción triple de los de ruido los generadores son $I = ABCD = DPQR = ABCPQR$. Al presentar la estructura de confusión se generan entre si las relaciones de los factores de control y ruido.

3. Se puede intercambiar el papel de los factores de control y de ruido, es decir $P = ABC$. Lo que claramente presenta otros escenarios en la estructura de confusión.

En la situación 2, el esquema experimental puede ser 2^{7-3}_R. Sin embargo un alternativa adecuada es plantear como WP el diseño $2^{6-3}_{R_1}$ y como SP el factor: tipo de piel.

Finalmente para la situación tres, nos auxiliamos del esquema 2^{8-4}_R presentado en la Tabla 1. Ahí se tienen cuatro generadores, éstos se aplican de tal manera que den lugar a un diseño factorial fraccionado en bloques. Dos generadores se utilizan para los tratamientos $E = ABC$ y $F = BCD$ y dos para los bloques $b1 = ACD$ y $b2 = ABD$ (ver Tabla 1). Así el esquema experimental en bloques es $2^{(6+2)-(2+2)}_R$. Los cuatro bloques se obtienen mediante la interacción b1b2.

Referencias

Bisgaard, S. (2000). The Design and Analysis of 2^{k-p}x2^{q-r} Split Plot Experiments. *Journal of Quality Technology.* 32 (1), 39-56.

Box, G.E.P. and Jones, S. (1991). Split Plot Design for Robust Product Experimentation. *Journal of Applied Statistics* 19, 3-26.

Loeppky, J. L. and Sitter, R. R. (2002). Analyzing Unreplicated Blocked or Split Plot Fractional Factorial Designs. *Journal of Quality Technology.* 34 (3), 229-243.

Nair, V. N. (1992). Taguchi's Parameter Desing: A Panel Discussion. *Technometrics* 34, 127-161.

Shoemaker, A. Tsui, K and Wu, C.F. (1991). Economical Experimentation Methods for Robust Desing. *Technometrics.* 33, 415-427.

Taguchi, G. (1986). Introduction to Quality Engineering: Designing Quality Into Products and Processes, Tokyo, Japan: Asian Productivity Organization.

Verosimilitud Perfil y Estimación por Intervalos en un Modelo Normal con Sesgo [1]

Armando Domínguez-Molina
Universidad de Guanajuato
Graciela González-Farías
Centro de Investigación en Matemáticas, A.C.

Resumen

En este trabajo se discute la estimación por intervalos en un modelo estadístico normal sesgado. Se muestra que en dicho modelo la estimación por método de momentos presenta problemas de existencia, mientras que la estimación por el método de máxima verosimilitud presenta problemas de unicidad. Se estudia la función de verosimilitud perfil para el parámetro de sesgo y se muestra como utilizarla para realizar inferencia sobre los parámetros del modelo normal sesgado.

1. Introducción

1.1. Preliminares

Considere un conjunto de observaciones independientes $x = (x_1, ..., x_n)^T$ con función de densidad dada por $f(x;\theta)$, $\theta \in \mathbf{R}^d$, $d \geq 1$, θ fijo y desconocido.

Se define la *función de verosimilitud* de θ como

$$L(\theta; x) = K \prod_{i=1}^{n} f(x_i, \theta),$$

[1]Trabajo apoyado por CONCyTEG: proyecto de investigación 01-16-202-112 y por CONACyT: proyecto de investigación 39017-E.

donde $K > 0$ es una constante arbitraria. La *función de verosimilitud relativa* se define mediante

$$R(\theta; x) = L(\theta; x) / L(\hat{\theta}; x),$$

donde $\hat{\theta} = \hat{\theta}(x)$ es el valor que maximiza $L(\theta; x)$ y es conocido como el *estimador de máxima verosimilitud (EMV)* de θ.

Definimos una *región de verosimilitud* con nivel c para θ como

$$\mathrm{RV}(c, x) = \{\omega \in \Omega : R(\omega; x) \geq c\}, \quad 0 \leq c \leq 1. \tag{1}$$

Cuando θ es escalar las regiones de verosimilitud pueden ser intervalos o unión de intervalos.

Cuando $\theta = (\psi, \chi) \in \Omega = \Omega_\psi \times \Omega_\chi$, es conveniente utilizar la función de verosimilitud perfil para ψ definida como

$$L_p(\psi) = L(\psi, \hat{\chi}_\psi; x),$$

donde $\hat{\chi}_\psi$ es el EMV de χ restringido a un valor específico de ψ. La *función de verosimilitud perfil relativa* para ψ es $R_p(\psi; y) = R(\psi, \hat{\chi}_\psi; y)$. Con R_p definimos las *regiones de verosimilitud perfil* para ψ de manera similar a (1), a saber

$$\mathrm{RVP}(c, x) = \{\psi \in \Omega_\psi : R_p(\psi; x) \geq c\}, \quad 0 \leq c \leq 1.$$

1.2. La distribución normal sesgada

Azzalini (1985) propuso la función de densidad normal sesgada dada por $f(y; \delta) = 2\phi(y)\,\Phi(\delta y)$, $\delta, y \in (-\infty, \infty)$, donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \phi(t)\,dt$, $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$. La versión de la densidad de Azzalini (1985) con parámetros de localización y escala está dada por

$$f(x; \mu, \sigma, \delta) = 2\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\delta\frac{x-\mu}{\sigma}\right), \tag{2}$$

donde $\mu \in \mathbf{R}$ y $\sigma > 0$ son los parámetros de localización y escala, respectivamente y δ es un parámetro real de sesgo, $\delta < 0$ indica sesgo a la izquierda y $\delta > 0$ indica sesgo a la derecha.

Los momentos de la densidad (2) se obtienen de la función generadora de momentos

$$M_X(t) = 2e^{\mu t + \frac{1}{2}\sigma^2 t^2} \Phi\left(\frac{\delta\sigma t}{\sqrt{1+\delta^2}}\right).$$

En Azzalini y Dalla Valle (1996), Gupta, *et al* (2001) y Domínguez, *et al* (2001) se presentan diferentes propuestas para generalizar la densidad (2) al caso multivariado.

2. Estimación

En esta sección mostramos dos procedimientos de estimación: el método de momentos y el método de máxima verosimilitud.

De (2) tenemos que la función de verosimilitud es

$$L(\mu, \sigma, \delta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \prod_{i=1}^{n} \phi\left(\frac{x_i - \mu}{\sigma}\right) \Phi\left(\delta\frac{x_i - \mu}{\sigma}\right). \tag{3}$$

La función de verosimilitud (3) admite múltiples puntos de inflexión y estacionarios y $(\bar{x}, s, 0)$ es un punto silla, donde $\bar{x}$ es la media de las observaciones y $s^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$.

2.1. Estimación de δ

Para μ y σ fijos el estimador de δ está dado por

$$\hat{\delta}_{\mu,\sigma} = \begin{cases} -\infty, & \text{si } X_{(n)} < \mu, \\ \text{un número real}, & \text{si } X_{(1)} < \mu < X_{(n)}, \\ +\infty, & \text{si } X_{(1)} > \mu, \end{cases}$$

obsérvese que σ no influye para que $\hat{\delta}_{\mu,\sigma}$ tome los valores $\pm\infty$. Además se cumple que la probabilidad de que el estimador de δ, para μ fijo, toma los valores $\pm\infty$ con probabilidad

$$\Pr\left(\hat{\delta}_{\mu} = \pm\infty\right) = \left(\tfrac{1}{2} - \tfrac{1}{\pi}\arctan\delta\right)^n + \left(\tfrac{1}{2} + \tfrac{1}{\pi}\arctan\delta\right)^n.$$

2.2. Verosimilitud perfil de δ

Los estimadores de μ y σ existen y son únicos cuando se considera δ fijo. Lo anterior se debe a que la función (1) es fuertemente unimodal para δ fijo. Esto nos permite definir la función de verosimilitud perfil de δ

$$L_p(\delta) = L(\hat{\mu}_\delta, \hat{\sigma}_\delta, \delta),$$

donde $\hat{\mu}_\delta$ y $\hat{\sigma}_\delta$ son los estimadores de máxima verosimilitud restringidos de μ y σ para valores fijos de δ.

La función de verosimilitud perfil de δ no converge a cero cuando $|\delta|$ tiende a ∞, en realidad

$$\lim_{\delta \to \pm\infty} L(\hat{\mu}_\delta, \hat{\sigma}_\delta, \delta) = \left(\frac{1}{2\pi \hat{\sigma}^2_{\text{mín}(x_i)}} \right)^{n/2} e^{-\frac{n}{2}},$$

y

$$\lim_{\delta \to -\infty} L(\hat{\mu}_\delta, \hat{\sigma}_\delta, \delta) = \left(\frac{1}{2\pi \hat{\sigma}^2_{\text{máx}(x_i)}} \right)^{n/2} e^{-\frac{n}{2}},$$

donde $\hat{\sigma}^2_a = \frac{1}{n} \sum (x_i - a)^2$.

2.3. Estimación por el método de momentos

Consideremos $\theta = \frac{\delta}{\sqrt{1+\delta^2}}$ y sea $m_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r$, el método de momentos consiste en resolver la ecuación

$$E[X^r] = m_r,$$

para un conjunto de valores de r que nos den una solución al sistema de ecuaciones generado. Tomando $r = 1, 2$ y 3 tenemos el siguiente sistema de ecuaciones:

$$\mu + \frac{\sqrt{2}}{\sqrt{\pi}}\,\theta\sigma = m_1\ ,$$

$$\sigma^2 + \mu^2 + 2\frac{\sqrt{2}}{\sqrt{\pi}}\,\mu\sigma\theta = m_2\ ,$$

$$\mu^3 + 3\mu\sigma^2 + 3\frac{\sqrt{2}}{\sqrt{\pi}}\,\sigma^3\theta + 3\frac{\sqrt{2}}{\sqrt{\pi}}\,\theta\sigma\mu^2 - \frac{\sqrt{2}}{\sqrt{\pi}}\,\theta^3\sigma^3 = m_3\ ,$$

el cual se puede escribir de forma equivalente como

$$\mu + \frac{\sqrt{2}}{\sqrt{\pi}}\,\theta\sigma = s_1, \quad \left(1 - \frac{2}{\pi}\,\theta^2\right)\sigma^2 = s_2 \quad \text{y} \quad \sqrt{2}\frac{4-\pi}{\left(\sqrt{\pi}\right)^3}\,\theta^3\sigma^3 = s_3, \tag{4}$$

donde $s_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$ y debe sujetarse a $\sigma > 0,\ -1 < \theta < 1$.

El sistema de ecuaciones (4) no siempre tiene solución para cualesquier valores de s_1, s_2 y s_3.

3. Estimación por intervalos o por regiones

Como vimos en la sección anterior, el estimador por el método de momentos para los parámetros del modelo (2) puede no existir. Esto nos impide obtener un intervalo o una región de confianza utilizando el método de momentos. Sin embargo, con la función de verosimilitud siempre es posible construir regiones de verosimilitud, cuya probabilidad de cobertura depende de la distribución de la estadística razón de verosimilitud $R_p(\psi) = L_p(\psi)/L_p\left(\hat{\psi}\right)$, donde ψ puede ser cualquier combinación de los parámetros (μ, σ, δ).

4. Conclusión

Es conveniente seguir estudiando alternativas para la estimación y construcción de regiones de confianza para los parámetros del modelo normal sesgado. El método de momentos

presenta problemas de existencia y el de máxima verosimilitud de unicidad. Sin embargo, para δ fijo, Azzalini (1985) y Arnold, *et al* (1993) entre otros, recomiendan fijar δ en valores adecuados y hacer inferencia para μ y σ con cada δ adecuado. Los valores adecuados de δ pueden obtenerse mediante un intervalo de verosimilitud perfil de δ.

Referencias

Arnold, B.C., Beaver, R.J., Groeneveld, R.A. & Meeker, W.Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58, 3, 471-488.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12,** 171-178.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Domínguez-Molina, J. A., González-Farías, G. and Gupta, A.K. (2001) A General Multivariate Skew Normal Distribution. Department of Mathematics and Statistics, Bowling Green State University, Technical Report No. 01-09.

Gupta, A.K., Gonzáles-Farías, G. and Domínguez-Molina, J.A. (2001). A multivariate skew normal distribution. Department of Mathematics and Statistics, Bowling Green State University, Technical Report No.01-10.

Introducción a las técnicas de captura–marcado–recaptura aplicadas a la pesquería en el municipio de Tenosique, Tabasco

Roger Armando Frias Frias
Universidad Juárez Autónoma de Tabasco,
Universidad Autónoma Agraria Antonio Narro
Emilio Padrón Corral,
Félix de Jesús Sánchez Pérez,
Roberto Coronado Niño
Universidad Autónoma Agraria Antonio Narro
Mario Alfredo Benitez Mandujano
Universidad Juárez Autónoma de Tabasco

1. Introducción

Este proyecto constituye una línea de investigación del **Departamento de Estadística y Cálculo de la Universidad Autónoma Agraria Antonio Narro** (*UAAAN*) con domicilio en Buenavista Saltillo, Coahuila y la **Universidad Juárez Autónoma de Tabasco** (*UJAT*) del Estado de Tabasco. Se presenta como un anteproyecto de tesis de la Maestría en Estadística Experimental. Además forma parte de un programa cooperativo entre las ramas de la Acuacultura y Estadística de ambas Universidades. El interés se centra en realizar un trabajo de equipo entre los profesionales y científicos que laboran en ambas ramas y estructurar un programa concreto al término del mismo. Se espera que este programa pueda ser considerado como un instrumento de obtención de resultados inmediatos y de interés para esta región del Estado de Tabasco.

El sureste mexicano cuenta con cerca de 35 % de los recursos hidrológicos de país, lo cual define la importancia que representa elaborar un programa de muestreos en algunos puntos específicos y evaluar el ambiente ecológico de los mismos. El profesor Benítez Mandujano (1997-1998), Profesor-Investigador de la Extensión Universitaria de los Ríos de la Univer-

sidad Juárez Autónoma de Tabasco, llevo a cabo un estudio titulado **"Evaluación del impacto ecológico por la Introducción de especies ícticas exóticas sobre las especies nativas en la región de los Ríos del Estado de Tabasco"**. Este estudio señala que de acuerdo a los muestreos íctiologicos realizados en el transcurso de un año, se pudo determinar el impacto ecológico que las especies exóticas causan al medio ambiente sobre las especies nativas, ocasionando la desaparición de éstas últimas en un periodo muy corto. En el mismo trabajo, se manifiesta además que la pesquería continental se conforma principalmente por la captura de peces exóticos como la Tilapia y la Carpa. Estas constituyen las especies dominantes debido a sus volúmenes de captura durante ese año de estudio.

En el Estado de Tabasco las cuencas hidrológicas de los ríos Usumacinta, Grijalva y Mezcalapa han originado gran número de lagunas cálidas de tipo temporal y permanente. Estos recursos acuáticos representan un potencial pesquero, acuícola y turístico de gran importancia. Sin embargo, no se tienen registros de su situación ambiental lo cual permite que los lugareños los aprovechen y exploten irracionalmente, provocando en algunos casos la desaparición de especies y en otros más la desaparición por completo de estas lagunas o cuerpos de agua.

2. Area de Estudio

El Estado de Tabasco se encuentra en el sureste de la República Mexicana, localizado entre los 18 36' 60" de latitud Norte y 90 59' -94 08' de Longitud Este-Oeste. Limita al Norte con el Estado de Campeche, al Este con la República de Guatemala y al Oeste con el Estado de Veracruz. Abarca una superficie de $24,959 Km^2$ y ocupa el vigésimo lugar de país por su extensión territorial. Dispone de un litoral hacia el Golfo de México de 192.8 Km y una plataforma continental de $850 Km^2$.

Sus recursos hidrológicos incluyen los ríos más caudalosos de país, entre los que se pueden mencionar al sistema Grijalva-Mezcalapa, el Río de la Sierra y el Usumacinta. El clima predominante es de tipo cálido húmedo con lluvias en verano, presentando temperaturas promedios anuales de $26^0 C$, régimen pluviómetro entre 1800 y 2600mm anuales y precipitación invernal superior al 10 % anual.

Un inventario sobre los cuerpos acuáticos regionales publicado en 1987 por Arredondo y Aguilar señala que el Estado de Tabasco cuenta con 8 sistemas acuáticos que representan 27,276 Ha de superficie lagunar. Sin embargo los estudios sobre estos cuerpos lagunares solo han sido utilizados para establecer estrategias de manejo y administración.

La región de los ríos del Estado de Tabasco es una extensa zona de aproximadamente 10,427 Km^2 que conforman una gran variedad de ambientes lagunares, zonas inundables temporales y permanentes, zonas pantanosas, ríos y arroyos. En esta región las condiciones ambientales son propicias durante todo el año para que poblaciones importantes de peces se establezcan y prosperen rápidamente. Tal es el caso de las especies exóticas (Tilapia y Carpa) presentes en dicha región que no cuentan con un depredador natural constituyendo un verdadero elemento de presión sobre las abatidas poblaciones de especies nativas que conformaban hace algunas décadas una de las pesquería más importante de la región. Las poblaciones de estas especies nativas están disminuidas a tal grado que su pesca es escasa y de talla muy pequeña. Esto es fácil de apreciar ya que en 1987 los registros de captura de especies nativas eran de $3,204{,}3$ toneladas, mientras que el total registrado en 1997 fue tan solo de 939 toneladas.

En cambio, hoy día las pesquerías más importante están conformadas por especies ícticas exóticas como la Tilapia, quien ha ocupado y alterado el hábitat de las especies nativas; sufriendo estas últimas un desplazamiento casi total en algunas zonas del Estado de Tabasco. Por lo que de no tomar medidas urgentes, estas poblaciones estarían en serio peligro de perder completamente sus hábitats naturales.

3. Metodología

Este método de estimación del tamaño de una población llamado **estimación por captura y recaptura** en dos muestras, se basa en las siguientes hipótesis:

1. La población es *cerrada*: ningún pez entra o sale del lago en el intervalo entre las muestras. Esto significa que N es la misma en cada muestra.

2. Cada muestra de peces es una muestra aleatoria simple de la población. Esto significa que cada pez tiene la misma probabilidad de inclusión en una muestra. No ocurre,

por ejemplo, que los más pequeños o menos saludables tengan más posibilidad de ser capturados. Además no existen "peces ocultos" en la población, es decir, imposibles de atrapar.

3. Las dos muestras son independientes. Los peces marcados de la primera muestra se vuelven a mezclar en la población de modo que el estado de marcación de un pez no está relacionado con la probabilidad de que se seleccione en la segunda muestra. Además, los peces en la primera muestra no son "timidos" ni "felices" ante una red, es decir, la probabilidad de que un pez sea atrapado en la segunda muestra no depende de su historia de captura.

4. Los peces no pierden sus marcas y los marcados pueden identificarse como tales.

En esta forma sencilla, la captura y recaptura es un caso particular de la estimación de cocientes del total de una población esto es sea n_1 el tamaño de la primera muestra, n_2 el tamaño de la segunda muestra y m la cantidad de peces atrapados en la segunda muestra que fueron atrapados tambien en la primera.

Referencias

Edwards, W. (1950). *Some Theory of Sampling.* New York: Dover.

Romero, Z. V. y Rodriguez, E. (1998). Evalución Ambiental de Cuatro Lagunas Continentales de Tenosique, Tabasco, México. Tesis.

Dependencia y Análisis de Regresión

José María González Barrios Murguía
Silvia Ruiz Velasco Acosta
IIMAS-UNAM

Uno de los problemas más interesante en el área de Estadística es el de análisis de regresión. Supongamos que tenemos una variable aleatoria Y cuyo comportamiento queremos modelar en término de algún subconjunto de las variables $X_1, X_2, \ldots, X_m$. En general, se propone un modelo en el que existe una función $f : \mathbb{R}^k \to \mathbb{R}$, donde $k \leq m$, y se satisface la relación

$$Y = f(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) + \epsilon, \tag{1}$$

donde $1 \leq i_1 < i_2 < \cdots < i_k \leq m$, y ϵ es un error aleatorio. La función f es, en principio, arbitraria y explica la relación entre la variable Y y las variables $X_1, X_2, \ldots, X_m$. En el caso de una función lineal f existe una amplia literatura en la que se analiza la metodología para encontrar un subconjunto adecuado que pueda explicar la relación lineal entre la variable Y y las $X's$, ver por ejemplo Draper y Smith (1998) o Miller(1990). En el caso de funciones f no lineales también existe literatura, aunque para la selección de variables no existe una metodología que sea aceptada universalmente.

En este trabajo se presenta una solución para selección de variables mediante la noción de dependencia. Primero notemos que si la ecuación (1) es válida, existe una dependencia entre la variable Y y el vector $(X_{i_1}, X_{i_2}, \ldots X_{i_k})$, es decir el comportamiento de Y depende de los valores de las variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, pero el comportamiento de Y es independiente de los valores de las demás $X's$. Asi, si denotamos por $I_m = \{1, 2, \ldots, m\}$, $K = \{i_1, i_2, \ldots, i_k\}$ donde $1 \leq i_1 < i_2 < \cdots < i_k \leq m$ es el subconjunto de índices de I_m que consiste de las variables $X's$ que explican a la variable Y, entonces Y y las variables X_i son independientes si $i \in I_m \backslash K$ y son dependientes en otro caso.

En este trabajo se trata de encontrar cuáles son las variables $X's$ que tienen alguna influencia en el comportamiento de la variable independiente Y, pero no se hace énfasis en encontrar

la función f de tal forma que la relación (1) se cumpla.

En Fernández-Fernández y González-Barrios (2001), se analizó una medida de dependencia multivariada $\delta^m_{X_1,X_2,\ldots,X_n}$ que se basa en la definición de independencia, en la Teoría de Cópulas (ver Nelsen (1999)) y en las funciones de distribución empíricas. Dicha estadística tiene la importante propiedad de que su distribución depende unicamente de la dimensión y del tamaño de muestra, al menos bajo la hipótesis de variables continuas.

En este trabajo se propone una versión modificada de la estadística $\delta^m_{X_1,X_2,\ldots,X_n}$ para medir la dependencia entre una variable Y y un vector $\mathbf{X} = (X_1, X_2, \ldots, X_m)$. Definida de la siguiente manera:

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y sean

$$Y, X_i : (\Omega, \mathcal{F}) \longrightarrow (I\!R, \mathcal{B}(I\!R)) \qquad \text{para} \qquad i \in I_m = \{1, 2, \ldots, m\}$$

$m+1$ variables aleatorias. Definimos

$$\delta_{Y,(X_1,\ldots,X_m)} = \sup_{(y,x_1,x_2,\ldots,x_m)} |F_{y,1,2,\ldots,m}(y, x_1, x_2, \ldots, x_m) - F_y(y)F_{1,2,\ldots,m}(x_1, x_2, \ldots, x_m)|, \quad (2)$$

donde $F_{y,1,2,\ldots,m}$ denota la función de distribución conjunta de Y y de todas las $X'_k s$, F_y es la función de distribución de Y, y $F_{1,2,\ldots,m}$ es la función de distribución conjunta de las $X'_k s$.

La estadística $\delta_{Y,(X_1,\ldots,X_m)}$ satisface las siguientes propiedades:

Teorema 1 *Sea $\delta_{Y,(X_1,\ldots,X_m)}$ como arriba. Entonces*

a) $\delta_{Y,(X_1,\ldots,X_m)} = \delta_{Y,(X_{\sigma(1)},X_{\sigma(2)},\ldots,X_{\sigma(m)})}$ para cada permutación σ de I_m.

b) $\delta_{Y,(X_1,\ldots,X_m)} = 0$ si y sólo si la variable aleatoria Y y el vector aleatorio $(X_1, X_2, \ldots, X_m)$ son independientes.

c) Para cada $x_i \in I\!R$, $i \in I_m$ y $y \in I\!R$

$$-\frac{1}{4} \leq F_{y,1,2,\ldots,m}(y, x_1, x_2, \ldots, x_m) - F_y(y)F_{1,2,\ldots,m}(x_1, x_2, \ldots, x_m) \leq \frac{1}{4}.$$

Por lo tanto $0 \leq \delta_{Y,(X_1,\ldots,X_m)} \leq 1/4$.

d) $0 \leq \delta_{Y,(X_1)} \leq \delta_{Y,(X_1,X_2)} \leq \delta_{Y,(X_1,X_2,X_3)} \leq \cdots \leq \delta_{Y,(X_1,X_2,\ldots,X_{m-1})} \leq \delta_{Y,(X_1,X_2,\ldots,X_m)}$

Además de tener la siguiente propiedad

Lema 3 *Sean* $Y, X_1, X_2, \ldots, X_m, X_{m+1}$, $m+2$ *variables aleatorias definidas en* $(\Omega, \mathcal{F}, P)$. *Si* X_{m+1} *es independiente de* $(X_1, \ldots, X_m)$ *y de* Y, *entonces*

$$\delta_{Y,(X_1,\ldots,X_m)} = \delta_{Y,(X_1,X_2,\ldots,X_m,X_{m+1})}.$$

Supongamos que tenemos una muestra $(m+1)$-dimensional de tamaño j,

$$\mathbf{X}_i = (Y_i, X_{i_1}, X_{i_2}, \ldots, X_{i_m}) \quad \text{para} \quad i = 1, 2, \ldots, j,$$

proveniente de una función de distribución conjunta $F_{\mathbf{X}}(y, x_1, x_2, \ldots, x_m)$ con función de distribución marginal $F_Y(y)$ y función de distribución marginal conjunta $F_{\underline{X}}(x_1, x_2, \ldots, x_m)$. Denotemos por $F_j(y, x_1, \ldots, x_m)$ a la función de distribución conjunta empírica de Y y de las $\mathbf{X}_i's$, sea $F_{j,1}(y)$ la distribución empírica de Y, y $F_{j,1,2,\ldots,m}(x_1, x_2, \ldots x_m)$ la distribución conjunta empírica de $(X_{i_1}, X_{i_2}, \ldots X_{i_m})$ para $i = 1, 2, \ldots, j$. Se propone una versión muestral multidimensional de la medida de dependencia como sigue:

$$\delta^j_{Y,(X_1,X_2,\ldots,X_m)} := \sup_{(y,x_1,x_2,\ldots,x_m)\in\mathbb{R}^{m+1}} |F_j(y, x_1, x_2, \ldots, x_m) - F_{j,1}(y)F_{j,1,2,\ldots,m}(x_1, x_2, \ldots x_m)|. \tag{3}$$

Esta versión muestral imita la versión poblacional. De hecho esta versión muestral tiene sentido pues cuando $j \to \infty$, F_j se aproxima a la distribución conjunta de Y y de las $X's$, y $F_{j,1}$ y $F_{j,1,2,\ldots,m}$ se aproximan a la distribución marginal de Y y la distribución conjunta de $(X_1, X_2, \ldots, X_m)$ respectivamente.

Una importante propiedad que la estadística $\delta^j_{Y,(X_1,X_2,\ldots,X_m)}$ satisface, es el hecho de ser invariante bajo transformaciones estrictamente crecientes sobre las coordenadas del vector X, o bien una transformacion estrictamente monótona de los valores de Y.

Proposición 4 *Sean* $Y_i, X_{i_1}, X_{i_2}, \ldots, X_{i_m}$ *para* $i = 1, 2, \ldots, j$ *una muestra como en la definición de la estadística y sean* $f_l : \mathbb{R} \to \mathbb{R}$ *para* $l = 1, 2, \ldots, m$ *funciones monótonas*

estrictamente crecientes, y $g : \mathbb{R} \to \mathbb{R}$ *una función estrictamente monótona. Entonces*

i) $\delta^j_{Y,(X_1,X_2,\ldots,X_m)} = \delta^j_{g(Y),(f_1(X_1),f_2(X_2),\ldots f_m(X_m))}$ *para variables aleatorias continuas.*

ii) Para cada $m \geq 1$ *y cualquier* $j > m+1$ *y cualesquiera subindices distintos* $i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, m\}$ *con* $k \leq m$

$$\delta^j_{Y,(X_{i_1})} \leq \delta^j_{Y,(X_{i_1},X_{i_2})} \leq \cdots \leq \delta^j_{Y,(X_{i_1},X_{i_2},\ldots,X_{i_k})}.$$

Además se tiene una versión del Teorema 1 para esta estadística, al menos en el caso de variables aletaorias continuas.

Con respecto a la selección de variables, se puede aplicar un prueba de hipótesis de independencia entre la variable respuesta Y y las variables explicativas $(X_1, X_2, \ldots, X_m)$. Siguiendo la misma idea, se puede implementar una serie de pruebas de hipótesis para probar independencia de la variable respuesta Y y cada una de las variables explicativas X_i, para $i = 1, 2, \ldots, m$, quedándose con aquellas $X's$ para las cuales se rechace independencia; después se procedería a probar independencia de la variable respuesta Y y las posibles parejas de variables explicativas (X_i, X_j) con $1 \leq i < j \leq m$, y gracias al hecho de que la distribución de la estadística sólo depende del tamaño de muestra y de la dimensión, utilizar esta distribución para comparar los cuantiles y seleccionar aquella pareja cuyo cuantil muestral sea más alto que los cuantiles de las componentes del vector (X_i, X_j) por separado. Este procedimiento se puede hacer en forma inductiva para poder seleccionar las variables $X's$ que tienen algún efecto en el comportamiento de la variable respuesta Y. Es claro de este procedimiento, que este método sólo detecta las variables $X's$ que influyen en los valores de la respuesta Y, pero no da información acerca de cómo debe ser la función f del modelo (1).

Ejemplo

Los siguientes datos fueron estudiados por Moore (1975) y en el libro de Weisberg (1980) son analizados en datalle. Las variables analizadas son:

Y = Toma de oxígeno en miligramos de oxígeno por minuto.
X_1 = Demanda biologica de oxígeno.
X_2 = Nitrógeno total.
X_3 = Sólidos totales.
X_4 = Sólidos volátiles totales.
X_5 = Demanda química de oxígeno.

Todas las variables explicativas medidas en miligramos por litro. La tabla presenta los valores de la estadística $\delta^j_{Y,(X_1,X_2,\ldots,X_m)}$, el valor de R^2 para los modelos con solo una variable X_i, el valor de R^2 utilizando como varaible respuesta $ln(Y)$ y utilizando como variable respuesta $1/\sqrt{Y}$

δ^{20}_{Y,X_i}	Cuantil	R^2 $Y=\alpha+\beta X_i$	R^2 $ln(Y)=\alpha+\beta X_i$	R^2 $1/\sqrt{Y}=\alpha+\beta X_i$
0.2400	100 %	0.2571	0.5983	0.6512
0.1550	92 %	0.0138	0.0082	0.0013
0.2400	100 %	0.2528	0.6965	0.7216
0.2200	> 99 %	0.2031	0.5054	0.5233
0.2400	100 %	0.4232	0.6926	0.6402

Los valores de la estadística indican una fuerte dependencia entre Y y X_1, X_3, X_4 y X_5, y muy poca entre Y y X_2. Sin embargo, los valores de R^2 para el modelo lineal no la reflejan. Sin embargo los valores de R^2 para los modelos con la variable respuesta transformada reflejan más la dependencia.

1. Referencias

Draper, N.R. and Smith, H. (1998) *Applied Regression Analysis.* (3rd edition). Ed. Wiley & Sons, New York.

Fernández-Fernández, B. and González-Barrios, J.M. (2001) Multidimensional Dependency Measures. Aceptado condicionalmente en *J. of Mult. Anal.*.

Miller, A.J. (1990) *Subset Selection in Regression.* Ed. Chapman-Hall, London.

Nelsen, R.B. (1999) *An Introduction to Copulas.* Lect. Notes in Statist., **139**, Springer-Verlag, New York.

Weisberg, S. (1980) *Applied Linear Regression.* Ed. John Wiley & Sons, New York.

Evaluación de profesores usando Modelos Mixtos

Leticia Gracia Medrano
IIMAS - UNAM
Silvia Ruiz Velasco Acosta
IIMAS - UNAM

1. Introducción

El objetivo de este análisis es conocer las diferencias entre las evaluaciones hechas a un grupo de profesores a nivel Secundaria. Los datos corresponden a 689 evaluaciones que los alumnos dan a 34 diferentes profesores en varios aspectos, por ejemplo: cómo explica en clases, acerca del contenido de las tareas, la actitud frente al grupo, respeto al alumno, manejo de la disciplina, etc. Cada punto es calificado de 1 a 4, siendo el uno la mejor calificación. Cada profesor es evaluado por alumnos que pueden pertenecer a distintos grados y/o grupos. Se cuenta también con el registro de otras características asociadas a los alumnos como sexo, aprovechamiento y esfuerzo.

Como primer paso hicimos un análisis de componentes principales para construir un índice que sirviera como calificación global.

Después modelamos utilizando Modelos de Efectos Mixtos, dado que existen efectos aleatorios como grupo, grado y sexo y efectos fijos como el profesor, nacionalidad.

2. Componentes Principales

Los puntos que los alumnos calificaron fueron:

1. Mi maestro escucha lo que pienso y digo, aún cuando no esté de acuerdo conmigo.
2. Mi maestro explica claramente.

3. Mi maestro me comunica que he hecho algo bien.

4. Mis clases son interesantes y divertidas.

5. Mi maestro tiene una actitud divertida y amigable.

6. Mi maestro es capaz de mantener el orden de manera adecuada.

7. Siento que puedo pedir ayuda al maestro.

8. El maestro respeta mis sentimientos.

9. El maestro prepara tareas interesantes que apoyan mi aprendizaje.

10. El maestro me ayuda a pensar por mí mismo.

11. El tiempo es bien aprovechado en clase.

12. Mi maestro llega a tiempo a clase.

13. El maestro califica y regresa los trabajos a tiempo.

14. El maestro me ayuda a organizarme de manera efectiva.

15. Mi maestro me dice como voy.

Con estos quince puntos y utilizando la matriz de covarianzas se hizo un análisis de componentes principales. El análisis mostró que la primer componentes principal explica el 44.44 % de la variación total, con dos componentes el 52.3 % y con tres componentes 58.3 %.

Se decidió tomar la primer componente como una calificación global para los profesores, ésta es un promedio que pondera principalmente los siguientes puntos: 5 (divertido y amigable), 1 (me escucha), 4 (clase interesante y divertida), 14 (ayuda a organizarme), 9 (tareas interesantes) y 7 (puedo pedir ayuda); siendo alumnos de secundaria no es de extrañar que sea el punto que se refiere a divertido y amigable el que lleve mayor ponderación.

3. Modelos de Efectos Mixtos

Los modelos mixtos se usan principalmente para describir las relaciones entre una variable respuesta (continua) y varias covariables en conjuntos de datos agrupados de acuerdo a uno o varios factores de clasificación. Con los efectos aleatorios comunes dentro de un grupo estos modelos representan la estructura de covarianza que se tiene dentro de los grupos.

Un ejemplo de estos modelos es el siguiente:

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_j + \epsilon_{ij}$$

donde i es el número de individuo, j es el número de repetición . Las $b_{ij} \sim N(0, \sigma_b)$ y las $\beta_i \sim N(0, \sigma)$.

Las β_i llamados efectos fijos describen el comportamiento de la población en general o están asociados con los niveles de un factor.

Las b_{ij} llamadas efectos aleatorios, describen la distribución de un coeficiente dentro de una población o grupo. Los efectos aleatorios están asociados con los individuos que son las unidades experimentales muestreadas de la población.

3.1. El modelo para los Profesores

En este caso, uno de los objetivos de modelar es identificar qué características contribuyen principalmente en la calificación global de los maestros. Se ajustaron varios modelos, en todos la variable respuesta fue la calificación global. Como efectos fijos se consideraron por ejemplo: el departamento al que pertenece el profesor, su nacionalidad y su sexo, mientras que como efectos aleatorios dentro de cada grupo se consideraron el aprovechamiento, el esfuerzo, el sexo del alumno, si el alumno y el profesor son del mismo sexo.

De acuerdo a la tabla 1, donde aparecen los índices Akaike de los modelos ajustados, y siguiendo el criterio de que un menor índice da un mejor modelo, el modelo selecionado fue el siguiente:

$$\text{Calif}_{ij_i} = \overbrace{\text{Nacionalidad+SexProf+Depto}}^{\text{efectos fijos}} + \overbrace{\text{esfuerzo}_{ij_i} + \text{sexoalumno}_{ij_i}}^{\text{efectos aleatorios}} + \epsilon_{ij_i}$$

Donde $i = 1, ..., 10$ es el número de grupos y las j_i toman distintos valores según el tamaño de los grupos.

Los modelos se ajustaron utilizando el método de verosimilitud restringida, y con una estructura general positiva-definida.

Para este modelo se obtuvo una estimación para la varianza de los errores de 2,027. En las tablas 2 y 3 aparecen los resultados para los efectos aleatorios y fijos respectivamente.

Tabla 1. Indice Akaike para los modelos ajustados, todos incluyen los efectos fijos de Nacionalidad, SexProf y Depto.

Akaike	*Efectos Aleatorios Incluidos*
3007,835	sexoalumno+esfuerzo+aprovech.+ sexosiguales
2998,936	sexoalumno+esfuerzo+aprovech.
2998,454	sexoalumno
2995,022	sexoalumno+aprovech.
2993,650	sexoalumno+esfuerzo

Tabla 2. Efectos aleatorios. Desviación estándar y matriz de correlación estimada.

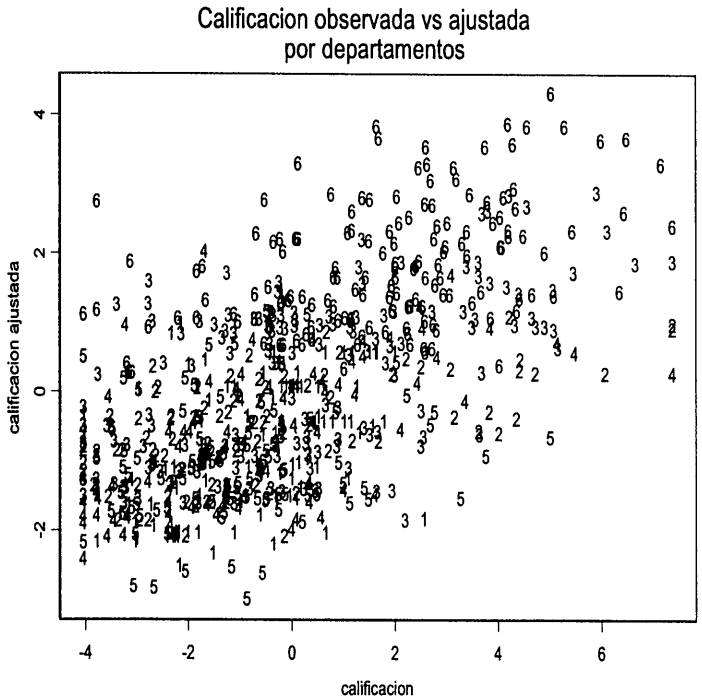
	Error Estándar	Correlación	
(Intercepto)	1,603	(Intrercepto)	sexo
sexo	0,342	0,033	
esfuerzo	0,141	−0,983	0,154

Tabla 3. Coeficientes de los Efectos Fijos.

Coeficiente	Valor	Error	GL	*valor t*	*valor-p*
$DEPT1$	−0,540	0,249	672	−2,166	0,0306
$DEPT2$	−1,038	0,234	672	−4,431	< ,0001
$DEPT3$	0,170	0,172	672	0,989	0,3232
$DEPT4$	−0,490	0,238	672	−2,056	0,0402
$DEPT5$	−1,126	0,262	672	−4,281	< ,0001
$DEPT6$	1,430	0,217	672	6,600	< ,0001
$SEXPROF$	−0,505	0,101	672	−4,999	< ,0001
$NACIONALIDAD$	−0,437	0,119	672	−3,682	0,0003

Las codificaciones de las variables son las siguientes: DEPT1=Inglés, DEPT2= Español y Francés, DEPT3= Humanidades, DEPT4= Matemáticas y Computación, DEPT5=Ciencias, DEPT6= Arte, Música, Educación Física, SEXPROF=1 si el profesor es hombre, 0 en otro caso, NACIONALIDAD = 1 si el profesor es extranjero, 0 en otro caso.

Figura 1: Gráfica de valores observados contra ajustados



4. Conclusiones

En la gráfica de valores ajustados y observados, se puede ver que muchas de las calificaciones ajustadas de los profesores del departamento 6 aparecen con una mayor dispersión que el resto, y en general que al modelo le falta bastante para lograr una buena predicción.

Sin embargo se logra modelar la tendencia de los datos y se puede decir lo siguiente: los efectos fijos resultaron todos ser significativos excepto el departamento 3 (Humanidades). Dado que la mejor calificación es 1 y la peor el 4, los valores menores son las mejores calificaciones, así que los maestros de los departamentos 5 y 2 (Ciencias con $-1{,}12$ y Español-Francés con $-1{,}03$) fueron los mejor evaluados. También ocurre que los profesores hombres tuvieron mejores calificaciones que las mujeres y que los extranjeros estuvieron mejor evaluados que los nacionales. También se ajustó el modelo sin los profesores del departamento 6 y sólo se logró una modesta mejora en el ajuste.

Se encontró que los intervalos de confianza para las desviaciones estándar de los efectos aleatorios no incluyen al cero por lo que quedan incluidos en el modelo.

El modelo podría mejorar si se tuvieran otras variables explicativas como la experiencia del profesor, edad del profesor, pero aún así existen otras variables subjetivas que influyen en las evaluaciones de los alumnos hacia los profesores, pero que son difíciles de cuantificar y por tanto de incluir en el modelo.

Referencias

Pinheiro y Bates (2000). *Mixed-effects Models in S and S-PLUS.* Springer Verlag. New York.

Comparación de Estimadores de Varianza para Diseños de Muestra Bietápicos

Ignacio Méndez Ramírez
Patricia I. Romero Mares
IIMAS, UNAM

1. Introducción

Las encuestas complejas son cada vez más utilizadas en la investigación social. Así, son muy frecuentes las encuestas de opinión, ingresos, salud, violencia, etc. Lo más común es que el diseño de muestra sea complejo, con muestras no autoponderadas, con varias etapas de muestreo, con estratos y estimadores de razón. Para el cálculo de estimadores y sus errores estándar, se usan paquetes estadísticos como SUDAAN, PC-CARP y STATA. Estos paquetes obtienen los estimadores de las varianzas con algunas aproximaciones. En el caso de tener muestras de unidades primarias de muestreo (*UPM*) por aleatorio simple, se efectúan dos tipos de aproximaciones:

1. Ignorar la varianza ocasionada por las unidades secundarias de muestreo (*USM*) y subsecuentes.

2. Suponer que los valores estimados de las *UPM* dentro de estratos son independientes.

Este trabajo se plantea la evaluación empírica del grado de error que se produce con las aproximaciones usadas por esos paquetes comerciales. Para esto se comparan las expresiones de varianza correctas y las que aproximan los paquetes. La comparación se efectúa simulando un muestreo para estimar la tasa y el total de hogares con televisión en localidades rurales del estado de Oaxaca.

2. Estimadores de Varianza

Se presentan los estimadores de totales y de razones, considerando dos grandes tipos de diseños:

- Esquema A de Raj. Hay estratos y varias etapas. Las *UPM* se toman dentro de cada estrato por muestreo aleatorio simple, con cualquier forma, incluso diferente, de muestrear adentro de las *UPM*.
- Esquema B de Raj. Dentro de los estratos las *UPM* se seleccionan con probabilidad proporcional al tamaño con reemplazo. Cada vez que se extrae una *UPM* se realiza el muestreo dentro de ella.

2.1. Estimadores esquema A

El estimador del total Y es: $\widehat{Y} = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} T_{hi} = \sum_{h=1}^{L} \hat{Y}_h$ donde T_{hi} es el estimador del total de la variable Y en la UPM_{hi}. La varianza estimada del estimador del total:

$$\widehat{V}(\widehat{Y}) = \sum_{h=1}^{L} \left[\frac{1}{n_h} - \frac{1}{N_h}\right] N_h^2 S_{bh}^2 \quad + \underbrace{\sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{N_h}{n_h} \left(\widehat{V}(T_{hi})\right)}_{\text{parte despreciada por los paquetes,}}$$

donde: $S_{bh}^2 = \left(\frac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(T_{hi} - \overline{T}_h\right)^2$ y $\overline{T}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} T_{hi}$.

Nótese que se requiere obtener $\widehat{V}(T_{hi})$. Estas varianzas dependen de la forma de muestreo de *USM* y subsecuentes.

El estimador de una razón es $\widehat{R} = \frac{\sum_{h=1}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{X}_h} = \frac{\widehat{Y}}{\widehat{X}}$.

Sea: $\widehat{G} = \sum_{h=1}^{L} \widehat{G}_h = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \widehat{G}_{hi}$, donde $\widehat{G}_{hi}$ es el estimador, según submuestreo, del total de la nueva variable $g_{hijk..} = Y_{hijk..} - \widehat{R}X_{hijk..}$ en las UPM_{hi} en muestra.

El estimador del error cuadrático medio del estimador de R es: $\widehat{ECM}\left(\widehat{R}\right) = \frac{\widehat{V}(\widehat{G})}{\widehat{X}^2}$ con

$$\widehat{V}(\widehat{G}) = \sum_{h=1}^{L} \left[\frac{1}{n_h} - \frac{1}{N_h}\right] N_h^2 S_{bg_h}^2 \quad + \underbrace{\sum_{h=1}^{L}\sum_{i=1}^{n_h} \frac{N_h}{n_h}\left(\widehat{V}(G_{hi})\right)}_{\text{parte despreciada por los paquetes,}}$$

donde: $S_{bg_h}^2 = \left(\frac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(G_{hi} - \overline{G}_h\right)^2$ y $\overline{G}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} G_{hi}$. Nótese que se requiere obtener $\widehat{V}\left(G_{hi}\right)$. Estas varianzas dependen de la forma de muestreo de USM y subsecuentes.

2.2. Estimadores esquema B

El estimador del total Y es:

$$\widehat{Y} = \sum_{h=1}^{L} \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{T_{hi}}{P_{hi}} = \sum_{h=1}^{L} \frac{1}{n_h} \sum_{i=1}^{n_h} Z_{hi} = \sum_{h=1}^{L} \overline{Z}_h.$$

La varianza estimada del estimador del total:

$$\widehat{V}\left(\widehat{Y}\right) = \sum_{h=1}^{L} \frac{1}{n_h} \left(\frac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(\frac{T_{hi}}{P_{hi}} - \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{T_{hi}}{P_{hi}}\right)^2.$$

Nótese que **no** se requieren las $\widehat{V}(T_{hi})$ por lo que esta expresión de varianza es válida para cualquier tipo de submuestreo. El estimador del error cuadrático medio del estimador de R es: $\widehat{ECM}\left(\widehat{R}\right) = \frac{\widehat{V}(\widehat{G})}{\widehat{X}^2}$, con $\widehat{V}\left(\widehat{G}\right) = \sum_{h=1}^{L} \frac{1}{n_h} \left(\frac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(\frac{G_{hi}}{P_{hi}} - \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{G_{hi}}{P_{hi}}\right)^2$.

Nótese que **no** se requieren las $\widehat{V}\left(G_{hi}\right)$, por lo que esta expresión de varianza es válida para cualquier tipo de submuestreo.

2.3. Aproximación *iid* de los paquetes

Los paquetes STATA, PC-CARP y SUDAAN, además de despreciar la contribución a la varianza del submuestreo, efectúan una segunda aproximación, que afecta al esquema A únicamente. Esta aproximación consiste en que suponen que dentro de cada estrato, las *UPM* se muestrean de forma independiente. Equivale a suponer $\frac{n_h}{N_h} = 0$ ó $N_h \to \infty$. A cada UPM_{hi} se le asocia una nueva variable: $d_{hi} = \sum_{j=1}^{m_{hi}} W_{hij} d_{hij}$ donde, d_{hij} es igual a Y_{hij} en el caso de estimar un total, y es igual a $Y_{hij} - \widehat{R} X_{hij}$ en el caso de estimar una razón; y W_{hij} es el factor de expansión correspondiente. Entonces, la varianza estimada de un total estimado, en general, $\widehat{D}$, se expresa como:

$$\widehat{V}\left(\widehat{D}\right) = \widehat{V}\left(\sum_{h=1}^{L}\sum_{i=1}^{n_h} d_{hi}\right)$$

$$\widehat{V}\left(\sum_{h=1}^{L}\sum_{i=1}^{n_h} d_{hi}\right) = \sum_{h=1}^{L}\widehat{V}\left(\sum_{i=1}^{n_h} d_{hi}\right) = \sum_{h=1}^{L} n_h \widehat{V}\left(d_{hi}\right) = \sum_{h=1}^{L} n_h \left(\tfrac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(d_{hi} - \overline{d}_{h.}\right)^2 .$$

2.4. Corrección por finitud

De manera poco justificada, se puede solicitar a los paquetes que efectúen una corrección por finitud. Con esto la expresión para la varianza de un total $\widehat{D}$, en general, es:

$$\widehat{V}\left(\widehat{D}\right) = \sum_{h=1}^{L} n_h \left(1 - \tfrac{n_h}{N_h}\right) \left(\tfrac{1}{n_h - 1}\right) \sum_{i=1}^{n_h} \left(d_{hi} - \overline{d}_{h.}\right)^2 .$$

3. Ejemplo

Se tomaron los datos del Censo del 2000 de Oaxaca y sólo para poblaciones o localidades con menos de 5,000 habitantes. Se conoce el número de habitantes y además la proporción de hogares que tienen televisión. Se consideraron los distritos, que agupan municipios, como estratos, aunque es un criterio que en este caso no es bueno para fines de muestreo. Se tienen entonces:

- Distritos como estratos
- Municipios como *UPM*
- Localidades como *USM*
- Información sobre viviendas totales, viviendas con t.v. y población total en la localidad.

Se simularon 1,000 muestras con diseños de esquemas A y B y con un algoritmo de selección con probabilidad proporcional al tamaño sin reemplazo (aunque los estimadores de varianzas fueron calculados como en el esquema B), las probabilidades fueron proporcionales a la población del municipio. Se ensayaron los siguientes tamaños de muestra:

- Número de *UPM*, $n_h = 2$,4 y 6.
- Número de *USM*, $m_{hi} = 2$ y 4.

4. Resultados

Tabla 1. Comparación de errores de estimación $\frac{1{,}96\sqrt{\widehat{V}_i}}{1{,}96\sqrt{\widehat{V}_1}}$

Y = total viviendas tv	$n_h = 2$ $m_{hi} = 2$	$n_h = 4$ $m_{hi} = 2$	$n_h = 6$ $m_{hi} = 2$	$n_h = 6$ $m_{hi} = 4$
(1)	1	1	1	1
(2)	0.953	0.905	0.847	0.879
(3)	1.015	0.893	0.957	1
(4)	0.953	0.781	0.775	0.798

(1) Varianza completa, (2) Varianza incompleta, (3) iid sin f.c., (4) iid con f.c.

Tabla 2. Comparación del Error de estimación de la razón $\frac{1{,}96\sqrt{\widehat{ECM}(\hat{R}_i)}}{1{,}96\sqrt{\widehat{ECM}(\hat{R}_1)}}$

	$n_h = 2\ m_{hi} = 2$	$n_h = 4\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 4$
(1)	1	1	1	1
(2)	.998	.998	.998	.998
(3)	1.06	.98	1.12	1.12
(4)	0.998	.86	.91	.91

(1) ECM completo, (2) ECM incompleto, (3) iid sin f.c., (4) iid con f.c.

Tabla 3. Comparación de $\hat{V}(\hat{Y})$ en % esquemas A, B y *ppt sr*

	$n_h = 2\ m_{hi} = 2$	$n_h = 4\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 4$
A	100	100	100	100
B	41	43	44	39
ppt sr	43	45	43	41

Tabla 4. Comparación de $\widehat{ECM}(\hat{R})$ en % esquemas A, B y *ppt sr*

	$n_h = 2\ m_{hi} = 2$	$n_h = 4\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 2$	$n_h = 6\ m_{hi} = 4$
A	100	100	100	100
B	74	81	82	81
ppt sr	85	85	84	83

5. Conclusiones

Tablas 1 y 2.

- La eliminación de las varianzas entre *USM* (método 2), produce subestimaciones hasta de 15 % en la estimación de varianzas de totales y casi no las produce para *ECM* de R.
- El supuesto *iid* (método 3), subestima el error de estimación de Y en 11 % con $n_h = 4, m_{hi} = 2$ y es equivalente al método 1 en el resto. Para *ECM* de R subestima en 2 %

con estos mismos tamaños de muestra y sobrestima en el resto entre 6 a 12 %.

- El supuesto *iid* con corrección (método 4) subestima el error de estimación de Y (5 a 22 %) y el *ECM* de R (.2 a 9 %) en todos los casos.

Tablas 3 y 4.

- Tanto para estimar totales como razones, el esquema B produce menores errores de estimación, comparado con el A, más marcada la disminución para totales que para razones.
- El *ppt sr* resultó con ligeramente más variación en los estimadores que el esquema B.

6. Recomendaciones

- Las aproximaciones en la estimación que realizan los paquetes sin factor de corrección por finitud pueden producir sobrestimaciones importantes de los errores de estimación de R, hasta de 12 %. En los errores de estimación de totales puede haber una subestimacón hasta 11 %. La recomendación es hacer la estimación a través de los paquetes; si los errores de estimación están por debajo de la precisión fijada, estas subestimaciones o sobrestimaciones no son importantes, pero si estos errores están alrededor de la precisión fijada se recomienda elaborar los cálculos con las expresiones correctas, programándolas.
- No se recomienda el uso de factores de corrección por finitud en ningún caso.
- El esquema B resulta una mejor alternativa de diseño.

Referencias

Raj, Des (1968). *Sampling Theory.* Mc.Graw-Hill, Inc.

Principales resultados por Localidad. XII Censo General de Población y Vivienda. Disco Compacto. INEGI, 2000.

Shah, B.V., Barnwell, B.G., Bieler, G.S. (1997) *SUDAAN, Software for the Statistical Analysis of Correlated Data. Release 7.5* Research Triangle Park, NC: Research Triangle Institute.

Fuller, W.A., Kennedy, W., Schnell, D., Sullivan, G., Park, H.S. (1986). *PC CARP.* Statistical Laboratory. Iowa State University.

Stata 7 Stata Referemce Manuel Release 7. (2001). Stata Press, College Station, Texas.

Procesos Beta en el Análisis de Supervivencia

Luis E. Nieto-Barajas
Departmento de Estadística, ITAM

1. Introducción

El análisis de supérvivencia es una parte de la estadística que estudia el comportamiento de variables aleatorias no negativas que generalmente tienen que ver con tiempos de vida. Si el tiempo de vida es la duración de una máquina, las técnicas de análsis de supervivencia son igualmente aplicables, pero reciben el nombre de análisis de confiabilidad.

Más formalmente, sea T una variable aletoria no negativa que mide la duración entre dos eventos previamente especificados llamados evento de origen y evento de fallo. Los modelos en un análisis de supervivencia pueden ser tanto en tiempo continuo como en tiempo discreto dependiendo de la naturaleza de la variable aleatoria T.

En muchas ocasiones, los datos de supervivencia se recolectan en intervalos fijos de tiempo. En esos casos es más adecuado utilizar un modelo en tiempo discreto para su análisis. El modelo en tiempo discreto es el siguiente: Sea T una variable aleatoria que toma valores en el conjunto $\{\tau_1, \tau_2, \ldots\}$. Una de las principales cantidades de interés en el análsis de supervicencia es la tasa de riesgo π_k, que se define como la probabilidad instantánea de muerte o fallo dado que el individuo se encontraba vivo en el instante anterior, i.e.,

$$\pi_k = \mathrm{P}(T = \tau_k | T \geq \tau_k).$$

Las funciones de densidad y de distribución de la variable aletoria T se pueden espresar en términos de las tasas de riesgo como:

$$f(\tau_j|\pi_k) = \pi_j \prod_{k=1}^{j-1}(1-\pi_k) \quad \text{y} \quad F(\tau_j|\pi_k) = 1 - \prod_{k=1}^{j}(1-\pi_k).$$

Además de la tasa de riesgo, otra función importante en análisis de supervivencia es la

función de supervivencia, que se define como la probabilidad de sobrevivir un tiempo t, i.e.,

$$S(\tau_j|\pi_k) = \mathrm{P}(T > \tau_j) = 1 - F(\tau_j|\pi_k) = \prod_{k=1}^{j}(1 - \pi_k).$$

En este artículo, se seguirá un enfoque Bayesiano no paramétrico para realizar inferencias sobre el modelo de supervivencia en tiempo discreto. En la Sección 2 se explicará el proceso inicial beta propuesto por Hjort (1990). La Sección 3 trata de un proceso más general llamado proceso de Markov beta introducido por Nieto-Barajas & Walker (2002). Ejemplos ilustrativos de ambos procesos se presentan en la Sección 4 y finalmente se incluye una discusión y comentarios finales en la Sección 5.

2. Proceso beta inicial

Hjort (1990) asume que *a-priori* las tasas de riesgo son independientes con distribución marginal beta. Consideremos la siguiente definición.

Definición 2.1. *(Hjort, 1990) Sean $\pi_1, \pi_2, \ldots$, una sucesión de variables aleatorias independientes tales que*

$$\pi_k \sim Be(\alpha_k, \beta_k),$$

para k=1,2,.... Entonces $\{\pi_k\}$ es un proceso beta.

En otras palabras, la distribución inicial que Hjort asigna a las tasas de riesgo es un proceso estocástico de variables aleatorias independientes con distribución marginal beta.

Si tomamos a n_k como el número de fallas en el momento τ_k y a y_k como el número de individuos al riesgo en τ_k, i.e.,

$$n_k = \sum_{i=1}^{n} I(t_i = \tau_k) \quad \text{y} \quad y_k = \sum_{i=1}^{n} I(t_i \geq \tau_k),$$

entonces, es fácil de demostrar que la distribución final para $\{\pi_k\}$ es un proceso beta de variables aleatorias independientes con

$$\pi_k|t \sim \text{Be}(\alpha_k + n_k, \beta_k + y_k - n_k),$$

para $k = 1, 2, \ldots$. Consecuentemente, si se usa una función de pérdida cuadrática, el estimador Bayesiano de la tasa de riesgo es

$$\widehat{\pi}_k = \frac{\alpha_k + n_k}{\alpha_k + \beta_k + y_k}.$$

Una característica importante de este estimador es que si $\alpha_k \to 0$ y $\beta_k \to 0$ entonces $\widehat{\pi}_k$ se reduce al estimador Nelson-Aalen que es el estimador clásico no paramétrico de la tasa de riesgo.

3. Proceso de Markov beta inicial

Nieto-Barajas & Walker (2002) generalizan el supuesto de independencia entre las tasas de riesgo y asumen que *a-priori* tasas de riesgo sucesivas tienen una relación de dependencia. La idea para introducir la relación de dependencia es mediante un proceso de variables latentes $\{u_k\}$ de la siguiente manera:

$$\pi_1 \to u_1 \to \pi_2 \to u_2 \to \cdots$$

El proceso de Markov beta se define como,

Definición 3.1. *(Nieto-Barajas & Walker, 2002) Sean $\pi_1, \pi_2, \ldots$, una sucesión de variables aleatorias tal que*

$$\pi_1 \sim Be(\alpha_1, \beta_1),$$

$$u_k|\pi_k \sim Bi(c_k, \pi_k) \quad y$$

$$\pi_{k+1}|u_k \sim Be(\alpha_{k+1} + u_k, \beta_{k+1} + c_k - u_k),$$

para k=2,3,.... Entonces $\{\pi_k\}$ es un proceso de Markov beta.

El proceso de Markov beta $\{\pi_k\}$ es un proceso de Markov de orden 1 con esperanza condicional dada por

$$\mathrm{E}[\pi_{k+1}|\pi_k] = \frac{\alpha_{k+1} + c_k\pi_k}{\alpha_{k+1} + \beta_{k+1} + c_k}.$$

Si tomamos $c_k = 0$ entonces el proceso $\{\pi_k\}$ se reduce al proceso beta de independencia de Hjort (1990).

Una propiedad importante de este proceso es que si $\alpha_k = \alpha_1$ y $\beta_k = \beta_1$ para todo k, entonces $\{\pi_k\}$ es un proceso estríctamente estacionario con distribuciones marginales $\pi_k \sim \mathrm{Be}(\alpha_1, \beta_1)$. Más aún, para este proceso estacionario, la estructura de correlación entre variables sucesivas es

$$\mathrm{Corr}(\pi_k, \pi_{k+1}) = \frac{c_k}{\alpha_1 + \beta_1 + c_k}.$$

Tomando las mismas definiciones para n_k y y_k anteriores, Nieto-Barajas & Walker (2002) obtuvieron que la distribución final para $\{\pi_k\}$ es un proceso de Markov beta (de orden 1) que se puede obtener considerando las distribuciones condicionales completas,

$$f(\pi_k|u_{k-1}, u_k, t) = \mathrm{Be}(\alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} - u_{k-1} + c_k - u_k + y_k - n_k) \quad \text{y}$$

$$\mathrm{P}(u_k = u|\pi_k, \pi_{k+1}) \propto \frac{\phi_k^u}{\Gamma(u+1)\Gamma(c_k - u + 1)\Gamma(\alpha_{k+1} + u)\Gamma(\beta_{k+1} + c_k - u)},$$

donde $u = 0, 1, \ldots, c_k$ y

$$\phi_k = \frac{\pi_k\pi_{k+1}}{(1-\pi_k)(1-\pi_{k+1})}.$$

Inferencias finales para este modelo se pueden obtener via un simulador de Gibbs (ver, por ejemplo, Smith & Roberts, 1993).

4. Ejemplos numéricos

Ejemplo 4.1: Duración de remisión en meses de pacientes con leucemia (Klein & Moeschberger, 1997). Los datos corresponden al tiempo de recaida de $n = 21$ pacientes tratados con

placebo. No se observó ningún dato censurado. Se tomó $\alpha_k = 0{,}0001$, $\beta_k = 0{,}0001$ y distintos valores de $c_k = 0, 50$. La Figura 1 muestra los estimadores de las tasas de riesgo (gráfica izquierda) y de la función de supervivencia (gráfica derecha). En ambas gráficas se puede observar que al usar un proceso de Markov beta inicial se obtienen estimadores más suaves y por lo tanto más razonables que al usar un proceso beta (de independencia) inicial.

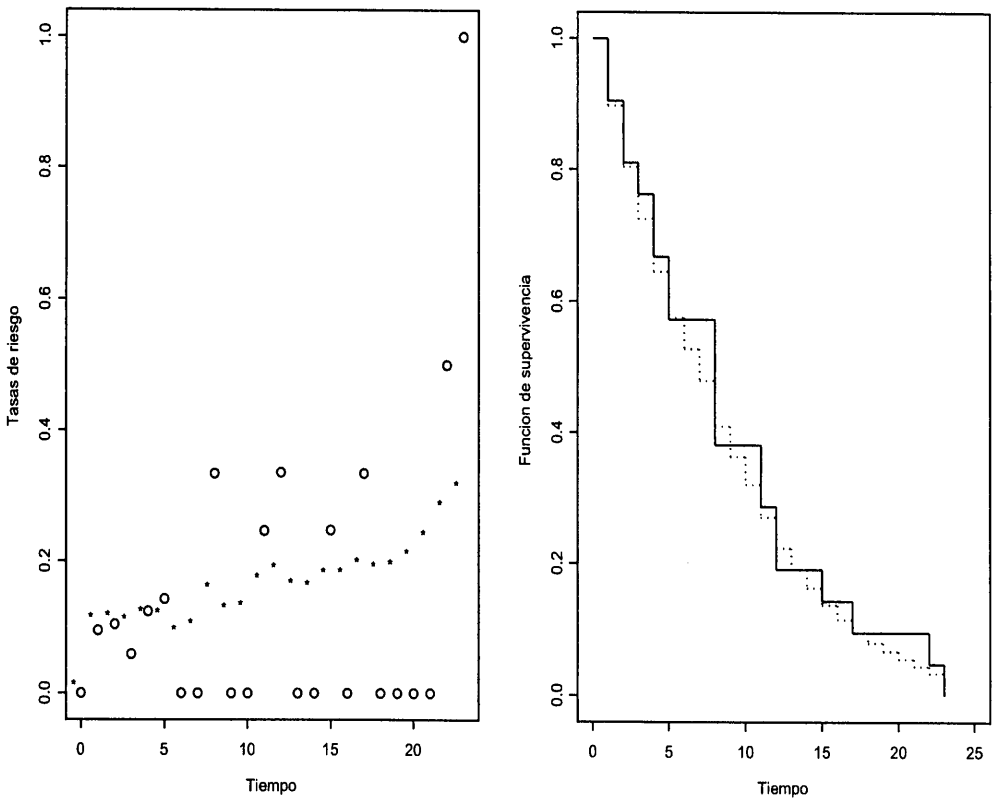


Figura 1: Duración de remisión de leucemia. Izquierda: Estimador de las tasas de riesgo (o) $c_k = 0$ y ($\star$) $c_k = 50$. Derecha: Estimador de la función de supervivencia (——) $c_k = 0$ y (- - -) $c_k = 50$.

Ejemplo 4.2: Datos simulados Po(10). Se simularon $n = 100$ observaciones exactas. Se tomó $\alpha_k = 0{,}0001$, $\beta_k = 0{,}0001$ y valores de $c_k = 0, 150$. La Figura 2 muestra los estimadores de las tasas de riesgo (arriba) y de la función de supervivencia (abajo) al usar los dos procesos beta iniciales. El proceso beta de independencia asigna un valor de cero a las tasas de riesgo en los tiempos donde no hubo observaciones, en cambio los estimadores producidos por el proceso de Markov beta son todos distintos de cero, sin embargo, estos

estimadores no se ajustan al verdadero valor de la tasa de riesgo para valores de τ_k grandes (ésto se debe, tal vez, a un problema numérico). Por otro lado, los estimadores de la función de superviviencia, producidos por ambos procesos iniciales, son bastante razonables, pero si uno observa con cuidado, los estimadores producidos por el proceso de Markov beta se parecen más a la verdadera función de superviviencia.

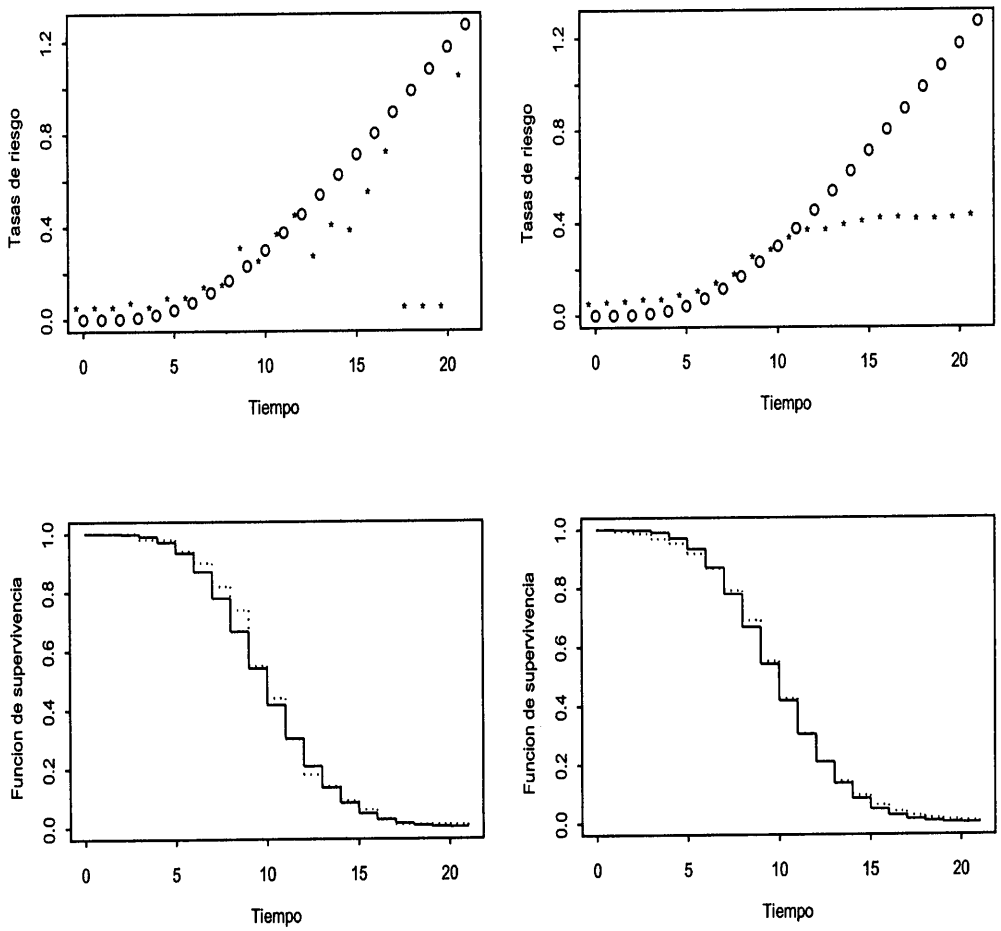


Figura 2: Datos simulados Po(10). Estimador de las tasas de riesgo (o) exacto y ($\star$) estimado, arriba-izquierda ($c_k = 0$), arriba-derecha ($c_k = 150$). Estimador de la función de supervivencia (——) exacto y (----) estimado, abajo-izquierda ($c_k = 0$), abajo-derecha ($c_k = 150$).

En ambos ejemplos se corrió un simulador de Gibbs con 10,000 iteraciones, usando las primeras 1,000 como período de calentamiento.

5. Discusión

Los procesos beta son usados para modelar tasas de riesgo e inducen distribuciónes iniciales no paramétricas. Las distribuciones iniciales inducidas asignan probabilidad uno al espacio de distribuciones discretas.

Las inferencias obtenidas a partir de los procesos beta son sencillas de calcular. En el caso del proceso beta de independencia es posible obtener inferencias analíticamente, mientras que en el caso del proceso beta de Markov es necesario recurrir a un simulador de Gibbs. Sin embargo, el proceso de Markov beta nos da una mayor flexibilidad en el modelado inicial de las tasas de riesgo.

Referencias

Hjort, N.L. (1990). Nonparametric Bayesian estimators based on beta processes in models for time history data. *Annals of Statistics* **18**, 1259-1294.

Klein, J.P. & Moeschberger, M.L. (1997). *Survival analysis (Techniques for censored and truncated data)*. Springer, New York.

Nieto-Barajas, L.E. & Walker, S.G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics.* **29**, 413-424.

Smith, A.F.M. & Roberts, G.O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **55**, 3-23.

Análisis Bayesiano de un Modelo para Datos Circulares

Gabriel Nuñez Antonio
ITAM
Eduardo Gutiérrez-Peña
IIMAS-UNAM

1. Introducción

En muchas aplicaciones se pueden encontrar datos direccionales que no pueden ser tratados adecuadamente con una teoría lineal. Por ejemplo, datos de orientación en biología, datos de pendientes y declinaciones en geología, datos de fluctuaciones en medicina y datos de direcciones de viento en meteorología. La aplicación de técnicas lineales convencionales puede producir paradojas; por ejemplo, la media aritmética de los ángulos 1 y 359 es 180 mientras que por intuición geométrica debería ser 0.

En este trabajo se presenta una introducción al análisis de datos direccionales y se muestran los desarrollos asociados al análisis Bayesiano de un modelo particular para este tipo de datos, basado en la distribución Normal bajo proyecciones.

2. Naturaleza de los datos direccionales

Los datos circulares aparecen en varias formas. Tal vez las dos más relevantes surgen de los más importantes instrumentos de medición circular: *la brújula* y *el reloj*. Observaciones típicas medidas a través de una brújula incluyen direcciones de viento y direcciones de migración de aves. Como ejemplo de observaciones típicas medidas con el reloj se pueden mencionar datos asociados a tiempos de llegada (sobre un reloj de 24 hrs.). Datos similares aparecen como tiempos a lo largo de un año (o tiempos en meses) de la ocurrencia de algún evento.

El interés en desarrollar técnicas para analizar datos direccionales se remonta a la época en la que Gauss desarrolló la teoría de errores para analizar ciertas medidas direccionales

en astronomía. Es un accidente histórico que los errores involucrados fueran suficientemente pequeños para que Gauss usara una aproximación lineal y, como una consecuencia, que desarrollara una teoría lineal en lugar de una teoría direccional de los errores. En muchas aplicaciones, sin embargo, se pueden encontrar datos direccionales que no pueden ser tratados adecuadamente con una teoría lineal.

Como un punto final para entender la naturaleza diferente de los datos circulares con respecto a los datos sobre la línea, se puede ver que el círculo es una curva cerrada pero la línea no, por lo que se pueden anticipar diferencias entre la teoría estadística sobre la línea y sobre el círculo. Por ejemplo, es necesario definir funciones de distribución, funciones características y momentos de tal manera que tomen en cuenta la periodicidad natural del círculo (ver, por ejemplo, Mardia y Jupp (2000)).

3. Modelos para datos circulares

Los modelos para datos circulares se pueden clasificar en tres grandes categorías: modelos generados por proyecciones, modelos "envueltos" (wrapped) y modelos tipo von Mises. Aquí sólo se presenta el análisis de un particular modelo generado por proyecciones: el modelo Normal proyectado. Para una revisión del análisis Bayesiano del modelo de von Mises ver Damien y Walker (1999).

Una forma de obtener distribuciones sobre el círculo es proyectando radialmente las distribuciones sobre el plano. Sea $\mathbf{x}$ un vector aleatorio bidimensional tal que $P(\mathbf{x} = 0) = 0$. Entonces, $\|\mathbf{x}\|^{-1}\mathbf{x}$ es un punto aleatorio sobre el círculo unitario. Un caso importante es cuando $\mathbf{x}$ tiene una distribución Normal bivariada $N_2(\mu, \Sigma)$, en este caso se dice que $\|\mathbf{x}\|^{-1}\mathbf{x}$ tiene una distribución Normal proyectada $NP_2(\mu, \Sigma)$.

Una aplicación típica se presenta en meteorología, cuando la velocidad del viento se modela con una distribución Normal bivariada la correspondiente distribución marginal de la dirección del viento resulta ser una Normal proyectada.

Aunque cualquier distribución continua se puede proyectar radialmente, la distribución Normal proyectada es la más común, por lo que a continuación se presentan algunas de sus

propiedades.

La distribución $NP_2(\mu, \Sigma)$ se reduce a una distribución uniforme si y sólo si $\mu = 0$ y $\Sigma = \sigma^2 I_2$. Cuando $\mu = 0$ se tiene una distribución llamada distribución Gaussiana central angular. Si B es una transformación lineal invertible del plano, entonces

$$\Psi_B(\mathbf{x}) = \frac{1}{\|B\mathbf{x}\|} B\mathbf{x}$$

define una transformación invertible del círculo unitario. De lo anterior, si $\theta = \|\mathbf{x}\|^{-1}\mathbf{x}$ y

$$\theta \sim NP_2(\mu, \Sigma), \text{ entonces } B\mathbf{x} \sim NP_2(B\mu, B\Sigma B^t).$$

La distribución Normal proyectada puede ser unimodal, bimodal y/o simétrica o asimétrica. En la Figura 1 se muestran algunas de estas densidades.

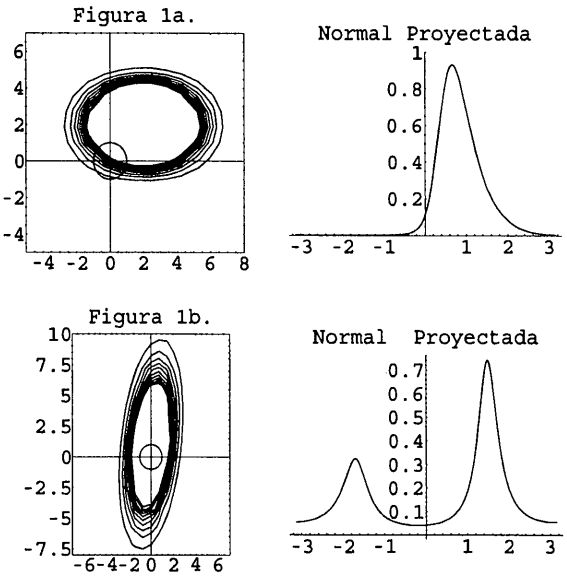


Figura 1: Densidades de Distribuciones Normales proyectadas

4. Análisis Bayesiano del modelo Normal proyectado

El análisis clásico del modelo Normal proyectado se puede revisar, por ejemplo, en Mardia (1972) y Mardia y Jupp (2000). La literatura sobre inferencia Bayesiana es menos extensa debido al problema que representa trabajar con este tipo de distribuciones. A continuación se presenta un análisis Bayesiano de este modelo.

4.1. Especificación del modelo

Función de densidad

$$f_X(x \mid \mu,\ \Lambda) = \frac{|\Lambda|^{1/2}}{2\pi} \exp\{(x-\mu)'\Lambda(x-\mu)\} I_{\mathbb{R}^2}(x) I_{\mathbb{R}^2}(\mu) I_{\gtrneqq}(\Lambda)\ ,$$

con $x = (x_1, x_2)'$.

Si x es el vector correspondiente al ángulo θ, entonces $x_1 = r\ \cos\theta$ y $x_2 = r \operatorname{sen}\theta$ y

$$\begin{aligned} f(\theta, r \mid \mu,\ \Lambda) &= f_X(r\cos\theta, r\operatorname{sen}\theta \mid \mu,\ \Lambda)\ r\ I_{[0,2\pi)}(\theta) I_{(0,\infty)}(r) \\ &= C(\theta) d(\theta)\ r\ \exp\left\{\frac{d(\theta)}{2}[r^2 - 2b(\theta)\ r]\right\} I_{[0,2\pi)}(\theta) I_{(0,\infty)}(r). \end{aligned}$$

A partir de $f(\theta, r \mid \mu,\ \Lambda)$ se puede obtener la distribución Normal proyectada $f(\theta \mid \mu,\ \Lambda)$, al marginalizar con respecto a r. Se debe señalar que $f(\theta \mid \mu,\ \Lambda)$ depende de μ y Λ sólo a través de $\varphi = \Lambda\mu$ y de Λ, por lo que se tiene un problema de identificabilidad. Una propuesta para eliminar este problema es imponer la restricción $||\mu|| = 1$.

El objetivo es realizar inferencias sobre φ y Λ con base en una muestra aleatoria $\theta_1, ..., \theta_n$.

Distribución Inicial Conjugada

Observemos que si se tuviera una muestra aleatoria $(\theta, \mathbf{r}) = \{(\theta_1, r_1), ..., (\theta_n, r_n)\}$ de la densidad $f(\theta, r \mid \mu,\ \Lambda)$, entonces podríamos utilizar la distribución inicial conjugada

$$f(\mu,\ \Lambda \mid \mu_0, \Lambda_0, \alpha_0, \beta_0) = N_2(\mu \mid \Lambda, \mu_0, \Lambda_0\Lambda) Ga_2(\Lambda \mid \alpha_0, \beta_0),$$

con lo cual se simplificaría el análisis. El problema es que sólo se cuenta con una m.a. de ángulos $\theta = \{\theta_1, ..., \theta_n\}$.

4.2. Inferencias vía *Gibbs sampling*

La idea para el análisis mencionado al final de la sección anterior se puede formalizar si se introducen las variables latentes:

$$\mathbf{r} = (r_1, ..., r_n).$$

Con lo anterior se pueden obtener las densidades condicionales completas para μ, Λ y $\mathbf{r}$, las cuales se pueden usar en un Gibb sampler para obtener muestras de la distribución final $f(\mu,\ \Lambda \mid \theta)$.

Se debe mencionar que para simular de la densidad condicional completa de $\mathbf{r}$ se empleó el algoritmo de Metropolis-Hastings.

5. Ejemplo

A continuación se presenta un ejemplo de la metodología expuesta en este trabajo. Los datos fueron tomados de Mardia y Jupp (2000) y consisten en observaciones asociadas a las direcciones que tomaron 76 tortugas después de depositar sus huevos. Cabe mencionar que en Mardia (1972) se considera una mezcla de densidades von Mises para estos datos.

Se obtuvieron simulaciones de la distribución final de $\mu' = (\mu_1, \mu_2)$ y de $\Lambda = Diag(\lambda_1, \lambda_2)$ usando el algoritmo presentado en la sección anterior.

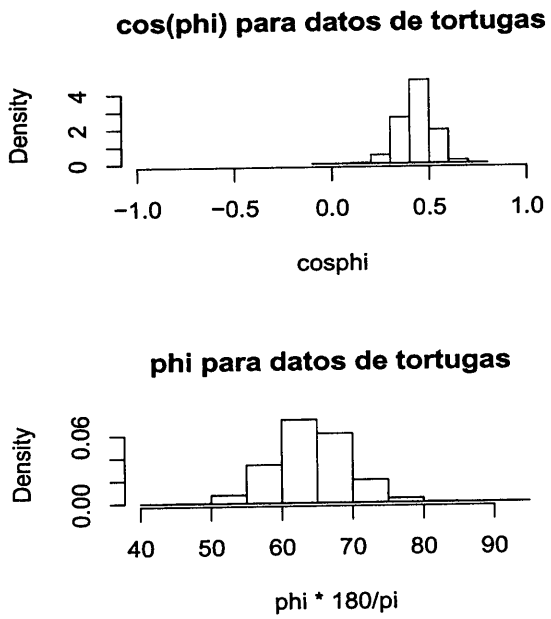


Figura 2: Distribución final de ϕ, con $\phi = \arctan(\mu_1/\mu_2)$

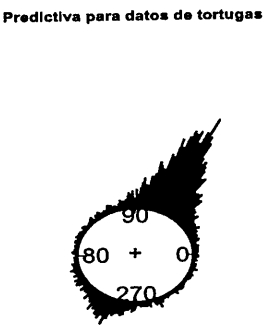


Figura 3: Distribución Predictiva Final de θ

En la Figura 2 se presentan la distribución final de ϕ, donde $\phi = \arctan(\mu_2/\mu_1)$. En la Figura 3 se presenta la distribución predictiva final de θ para estos datos.

Referencias

Damien, P. y Walker, S. (1999). A full Bayesian analysis of circular data using the von Mises distribution, *The Canadian Journal of Statistics*, **2**, 291-298.

Mardia, K. V. (1972). *Statistics of Direccional Data.* Academic Press, London.

Mardia, K. V. y Jupp, E. P. (2000). *Directional Statistics.* John Wiley and Sons Ltd.

Algunos Resultados en Inferencia Fiducial

Federico O'Reilly Togno
IIMAS-UNAM

1. Introducción

Consideremos el problema de inferencia en el cual el parámetro real $\theta \in \Theta = (\underline{\theta}, \overline{\theta})$ es desconocido y T es una estadística suficiente mínima a partir de la cual se llevarán a cabo las inferencias. La función de verosimilitud de θ es proporcional a $g(t;\theta)$, la densidad de T. Si $G(t;\theta)$ es la correspondiente función de distribución, entonces toda la información relevante para hacer inferencias acerca de θ se encuentra también contenida en $G(t;\theta)$.

Supongamos que valores grandes de θ implican valores grandes de T y también que valores pequeños de θ implican valores pequeños de T. Esto significa que un valor grande T, digamos $T = t$, proporciona evidencia en contra de la hipótesis $H_0 : \theta \leq \theta_0$, en el contexto del problema de contrastar con la hipótesis alternativa $H_1 : \theta > \theta_0$.

En la Teoría usual de Contraste de Hipótesis, si $T = t$, la "significancia" asociada a H_0 es simplemente $1 - G(t;\theta_0)$. Dicha significancia es monotóna decreciente como función de t, ya que cuando t se incrementa, la evidencia en contra de H_0 es mucho más fuerte.

Si t (la evidencia) permanece fija y tomamos ahora un valor $\theta_0' < \theta_0$ más "alejado" de la evidencia es lógico esperar que la significancia $1 - G(t : \theta_0')$ de $H_0 : \theta \leq \theta_0'$ sea más pequeña que el *p-value* original asociado a $H_0 : \theta \leq \theta_0$ debido a que ahora la hipótesis nula está más "alejada" de la evidencia. Esto significa que $G(t;\theta)$ debiera ser monótona decreciente como función de θ para cada t fijo.

Más aún, ¿Será cierto que cuando la "distancia entre la hipótesis nula y la evidencia" se hace extrema entonces el *p-value* tiende a cero ? Si θ_0 permanece fijo entonces es claro que si t tiende a su límite superior entonces el *p-value* tiende a 0 ya que $G(t;\theta_0)$ es función de distribución. Si ahora la "evidencia" permanece fija pero la hipótesis se mueve cada vez más,

no es obvio ni en general cierto que el *p-value* tienda a 0. Esta propiedad ocurre si $G(t;\theta) \to 1$ cuando $\theta \to \underline{\theta}$.

Se puede aplicar el mismo razonamiento a $H_0 : \theta \geq \theta_0'$ donde valores pequeños de t representen evidencia en contra de H_0. De este modo definamos formalmente:

1. **Contrastabilidad Monótona.** Cuando $G(t;\theta)$ es monótona en θ para cada t.
2. **Contrastabilidad Extrema.** Cuando

$$G(t;\theta) \to 1 \qquad \text{si} \qquad \theta \to \underline{\theta}\,,$$

$$G(t;\theta) \to 0 \qquad \text{si} \qquad \theta \to \overline{\theta}\,.$$

2. Ejemplos

Ejemplo 1: Parámetro de Localización

Si θ es un parámetro de localización entonces $\Theta = \Re$ y $G(t;\theta) = G_0(t-\theta)$ donde G_0 es una función de distribución conocida. En efecto se satisfacen las propiedades 1 y 2.

Ejemplo 2: Parámetro de Escala

Ahora, si $\theta \in \Theta = \Re^+$ es un parámetro de escala entonces $G(t;\theta) = G_0(t/\theta)$ de nuevo G_0 es una función de distribución conocida. Se satisfacen las propiedades 1 y 2.

Es interesante notar que cuando la propiedad 2 se satisface entonces, si θ tiende a alguno de los extremos del Intervalo Θ, entonces se obtiene una distribución degenerada. De este modo una pregunta interesante es observar que pasa si dicha distribución límite es no degenerada.

Ejemplo 3: (Pedersen, 1978) $X \sim N(\mu, 1)$, se tiene el problema de inferir sobre μ^2. Resulta equivalente al problema de obtener inferencias sobre $\theta = \mid \mu \mid$ usando $T = \mid X \mid$, en tal caso

$$G(t;\theta) = \Phi(t-\theta) - \Phi(-t-\theta)\,, \qquad t > 0\,,$$

se puede verificar que

- $G(t;\theta)$ es monótona decreciente en θ para cada t,
- Los límites

$$\begin{aligned} \lim_{\theta\to\infty} G(t;\theta) &= 0\,, \qquad \text{pero} \\ \lim_{\theta\to 0} G(t;\theta) &= 1-2\Phi(-t) < 1\,, \qquad \forall\, t \geq 0\,. \end{aligned}$$

La interpretación de este límite es que su complemento con respecto a la unidad, $2\Phi(-t)$, es precisamente el *p-value* en el contraste $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$ una vez que se ha observado t. Esto claramente ocurre ya que no se tiene contrastabilidad extrema en $\theta = 0$.

Ejemplo 4:

Considere $X_1, \ldots, X_n$, una muestra aleatoria de una Distribución Gaussiana Inversa con parámetros $\mu, \lambda > 0$. El parámetro de interés es $\theta = \lambda/\mu$ y se utiliza la estadística $T = \hat{\lambda}/\hat{\mu}$, el cociente de los estimadores máximo verosímiles. La función de distribución $G(t;\theta)$ de T está dada por

$$1 - \frac{2^{1-m} e^r r^m}{(m-1)!} \int_{t^*}^{\infty} e^{-ru}(u^2-1)^{m-1} \sum_{i=0}^{m} \frac{2^{-i}(m+i)!}{(m-i)!\, r^i i!} u^{-(i+m)} du\,,$$

donde $t^* = \sqrt{1+1/t}$, $n = 2(m+1)$ y $r = (2m+1)\,\theta$.

No se tiene una prueba analítica de que $G(t;\theta)$ es decreciente como función de θ, pero todos los ensayos numéricos realizados indican que

- G es monótona decreciente en θ y
- Los límites

$$\begin{aligned} \lim_{\theta\to\infty} G(t;\theta) &= 0\,, \\ \lim_{\theta\to 0} G(t;\theta) &< 1\,, \qquad t > 0 \end{aligned}$$

De nuevo se verifica que el "defecto" en G corresponde al *p-value* asociado al contraste

$$H_0 : \theta = 0 \qquad vs. \qquad H_1 : \theta > 0\,,$$

y es de hecho la masa fiducial en $\theta = 0$.

3. Distribución Fiducial

Si $G(t;\theta)$ satisface la condición de **Contrastabilidad Monótona** (decreciente en θ para cualquier t) entonces se define la **Función de Distribución Fiducial** como:

$$H(\theta;t) = 1 - G(t;\theta) \ .$$

Si además se tiene la condición de **Contrastabilidad Extrema** entonces la Distribución Fiducial no tiene "masas" en los extremos del intervalo Θ.

El uso de la distribución fiducial es inmediato y provee de intervalos de confianza (fiduciales) incluso en el caso de tener una masa en algún extremo de Θ.

4. Familia Exponencial Natural

Consideremos a $g(t;\theta)$ miembro de la Familia Exponencial Natural, es decir

$$g(t;\theta) = c(t)e^{\theta t - M(\theta)} \ , \qquad t \in \mathcal{D}_t \ ,$$

donde $\mathcal{D}_t$ es un intervalo abierto que no depende de θ y $\theta \in \Theta$ donde Θ es un intervalo.

Teorema 4.1 *Si $g(t;\theta)$ es un miembro de la Familia Exponencial Natural, entonces $G(t;\theta)$ es monótona decreciente como función de θ para cada $t \in \Re$.*

Este resultado nos dice que si g es miembro de la Familia Exponencial Natural, la derivación de la Distribución Fiducial siempre es posible. Dicha Distribución Fiducial puede o no tener masas en los extremos del intervalo Θ. Si se pretende obtener una Distribución Fiducial absolutamente continua entonces se requiere la condición de Contrastabilidad Extrema. Esta condición es equivalente a la noción de Regularidad en la Familia Exponencial Natural, es decir, $G(t;\underline{\theta})$ y $G(t;\overline{\theta})$ deben ser distribuciones degeneradas.

Teorema 4.2 *Si $g(t;\theta)$ es un miembro de la Familia Exponencial Natural (Regular), entonces $G(t;\theta)$ satisface la condición de Contrastabilidad Extrema.* Es interesante observar que sin la condición de regularidad, el resultado es falso como lo muestra el siguiente ejemplo:

Ejemplo 5: Si en el caso de la Gaussiana Inversa (Ejemplo 4) se reparametriza por medio de $\eta = -\theta^2/2$ donde $\theta = \lambda/\mu$ y λ se supone conocido (digamos $\lambda = 1$) entonces la densidad $g(x;\eta)$ tiene la forma

$$g(x;\eta) = c(x)e^{x\eta - M(\eta)}\,, \qquad M(\eta) = -\sqrt{-2\eta}\,.$$

Si $\eta \to 0$ se obtiene una distribución límite no degenerada. Este es un ejemplo de Familia Exponencial Natural No Regular (*steep*) donde la masa fiducial en $\{0\}$ es el *p-value* asociado al contraste

$$H_0 : \eta = 0 \qquad vs. \qquad H_1 : \eta < 0\,,$$

con $X \sim f(x;\eta)$ habiendo observado $X = x$.

Ejemplo 6: Sea X v.a. con distribución Gamma de parámetro $\alpha > 0$ y función de distribución:

$$F(x;\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^x u^{\alpha-1} e^{-u} du\,,$$

donde la densidad será denotada por $f(x;\alpha)$.

Si $T = \log X$ y $\theta = \alpha$ entonces $M(\theta) = \log\Gamma(\theta)$ de manera que T tiene distribución $G(t;\theta)$ miembro de la Familia Exponencial Natural.

En este caso se tiene además que G es regular pero resulta más interesante demostrar que $F(x;\alpha)$ satisface las condiciones de Contrastabilidad Monótona y Extrema. A saber, cuando $\alpha > 1$, integrando por partes se obtiene $F(x;\alpha) = -f(x;\alpha) + F(x;\alpha-1)$, de manera que si tomamos $\alpha' = \alpha - 1 > 0$ entonces $F(x;\alpha') = F(x;\alpha'+1) + f(x:\alpha'+1)$, de modo que si $\alpha' \to 0$ se obtiene que $F(x;\alpha') \to 1$ si $x > 0$. Recursivamente se tiene que

$$F(x;n+1) = 1 - e^{-x} - \sum_{i=1}^{n} f(x;1+i)\,,$$

de modo que

$$\lim_{n\to\infty} F(x; n+1) = 0 \, .$$

Entonces no existe masa fiducial en $\{0\}$ o en ∞. La Densidad Fiducial para α, una vez observado x es $h(\alpha; x) = -\frac{d}{dx}F(x; \alpha)$ que en este caso es:

$$h(\alpha; x) = \int_0^x (\psi(\alpha) - \log u) f(u; \alpha) du \, ,$$

donde $\psi(\alpha) = -\frac{d}{d\alpha} \log \Gamma(\alpha)$ es la función Digamma. La solución Bayesiana utilizando la distribución inicial de Jeffrey's $\pi(\alpha) \propto \sqrt{\psi'(\alpha)}$ proporciona un resultado muy parecido.

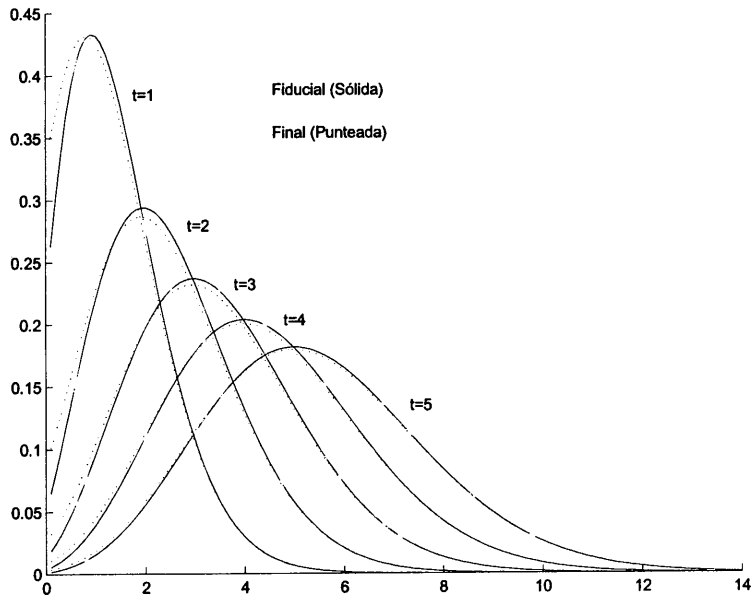


Figura 1: Densidad Fiducial y Final para α

Existe una enorme coincidencia (o cercanía) entre la Fiducial y la Distribución Final con π no informativa, aún para el caso con T discreta y θ continuo.

Nota: Este trabajo es parte de un trabajo más extenso con R. Rueda que está en preparación y agradezco el apoyo de K. Anaya.

Referencias

Brown, L.D., Cai, T.T and A. Das Gupta (2001). Interval estimation for a binomial proportion. *Statistical Science.* **16**(2): 101-133.

Hsieh, H.K. (1990). Inferences on the coefficient of variation of an inverse Gaussian distribution. *Comm. Statist. (Theory and Methods)*, **19**, 1589-1605.

Nádas, A. (1973). Best tests for zero drift based on first passage times in Brownian motion. *Technometrics,* **15**, 125-132.

O'Reilly, F.J. and Rueda, R. (1992). Goodness of fit for the inverse Gaussian distribution. *Can. J. Statist.*, **20**, 387-397.

Patil, S.A. and Kovner, J.L. (1976). On the test and power of the zero drift on first passage times in Brownian motion. *Technometrics*, **18**, 341-342.

Pedersen, J.G. (1978). Fiducial inference. *Int. Stat. Rev.*, **46**, 147-170.

Rueda, R. (1988). Intervalos de confianza y de probabilidad en la distribución Gaussiana inversa. *Aportaciones Matemáticas, Comunicaciones*, **5**, 369-375.

Seshardi, V. and Schuster, J.J. (1974). Exact tests for zero drift based on first passage times in Brownian motion. *Technometrics*, **16**, 133-134.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Annals of Math. Statist.*, **30**, 877-880.

Análisis R/S de Indicadores Financieros

Ernesto Rubio Acosta
IIMAS-UNAM

1. Introducción

En este trabajo se realiza un análisis R/S de ciertos indicadores financieros tales como las series de tiempo asociadas con los rendimientos en CETES, IPC (Bolsa), CPI (inflación) y paridad peso-dólar. El análisis R/S (rango -R- ajustado con la desviación estándar -S-), consiste en la estimación del exponente de Hurst y la longitud del ciclo promedio de cada una de las series de tiempo. El objetivo es determinar si se tienen series con un camino aleatorio, o bien, si se tienen series persistentes. A diferencia de las primeras, estas últimas se caracterizan por tener efectos de memoria. El efecto de memoria se podría explicar sencillamente como "lo que pase ahora influirá en el futuro por un largo tiempo".

2. Series de Tiempo

Una serie de tiempo es un conjunto ordenado de valores numéricos de cualquier variable que cambia con el transcurrir del tiempo. Existen series de tiempo continuas y discretas. Las series de tiempo se pueden caracterizar por medio de discontinuidades, un componente de tendencia, uno o más componentes periódicos y un componente estocástico. Un aspecto importante del componente estocástico es si éste es persistente, aleatorio o antipersistente. Si los valores adyacentes en la serie de tiempo están no correlacionados unos con otros, entonces el componente estocástico es aleatorio. Si los valores adyacentes están positivamente correlacionados, entonces el componente estocástico es persistente. Si los valores adyacentes están negativamente correlacionados, entonces el componente estocástico es antipersistente.

3. Análisis R/S Aplicado a Mercados

Dada una serie de tiempo de precios o ganancias, primero calcúlese la serie de tiempo correspondiente a los rendimientos logarítmicos:

$$Y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

donde Y_t es el rendimiento logarítmico en tiempo t y P_t es el precio o ganancia en tiempo t. El algoritmo para calcular el exponente de Hurst es el siguiente (Fernández (2000), Peters (1994, 1996), Turcotte(1997)):

Paso 1: Se divide la serie de tiempo Y_t de tamaño L, en M intervalos de longitud N, es decir, $MN = L$. Denótese a cada intervalo con I_m donde $m = 1, 2, ..., M$. Denótese a cada elemento del intervalo I_m con $Y_{n,m}$ donde $n = 1, 2, ..., N$. Para cada uno de los M intervalos I_m de longitud N, calcúlese la media de sus elementos:

$$\overline{Y}_m = \frac{1}{N}\sum_{n=1}^{N} Y_{n,m} \tag{2}$$

Paso 2: Para cada uno de los M intervalos I_m de longitud N , calcúlese las N desviaciones acumuladas respecto a la media:

$$D_{n,m} = \sum_{i=1}^{n}\left(Y_{i,m} - \overline{Y}_m\right); \quad \text{para} \quad n = 1, \ldots, N \tag{3}$$

es decir, se obtiene un conjunto $\{D_{n,m}\}_{n=1}^{N}$ para cada uno de los M intervalos.

Paso 3: Para cada uno de los M intervalos I_m de longitud N, calcúlese el rango R_m, el cual se define como la diferencia entre el valor máximo y el valor mínimo de los N valores $D_{n,m}$:

$$R_m = \text{máx}\left[\{D_{n,m}\}_{n=1}^{N}\right] - \text{mín}\left[\{D_{n,m}\}_{n=1}^{N}\right] \tag{4}$$

Paso 4: Para cada uno de los M intervalos I_m de longitud N, calcúlese la desviación estándar

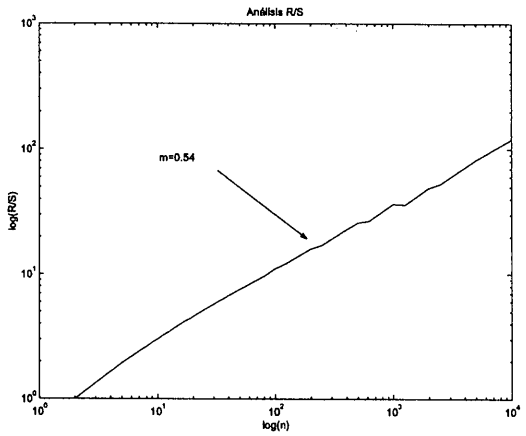


Figura 1: Análisis R/S de una serie de números aleatorios. Exponente de Hurst.

S_m de sus elementos:

$$S_m = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_{n,m} - \overline{Y}_m\right)^2} \tag{5}$$

Paso 5: Para cada uno de los M intervalos I_m de longitud N , calcúlese el cociente R_m/S_m. Ahora calcúlese la media de los M cocientes. Se obtiene el valor medio de R/S para los intervalos de longitud N:

$$\left(\frac{R}{S}\right)_N = \frac{1}{M}\sum_{m=1}^{M}\frac{R_m}{S_m} \tag{6}$$

Paso 6: Repítase los pasos 1 a 5 para cada valor de N tal que N sea divisor de L, es decir, tal que L/N sea entero (igual a M).

Paso 7: Realícese una regresión lineal con $log(N)$ como variable independiente y $log(R/S)_N$ como variable dependiente. La pendiente de dicha recta es el valor del exponente de Hurst, H . Nótese que el valor $H = 0{,}5$ implica aleatoriedad, el intervalo $0{,}5 < \mathrm{H} < 1{,}0$ implica persistencia y, finalmente, el intervalo $0{,}0 < \mathrm{H} < 0{,}5$ implica antipersistencia.

Considérese como ejemplo una serie de tiempo formada por 10000 números aleatorios con distribución normal, media cero y desviación estándar unitaria. El exponente de Hurst calculado es $H = 0{,}54$. Véase figura 1.

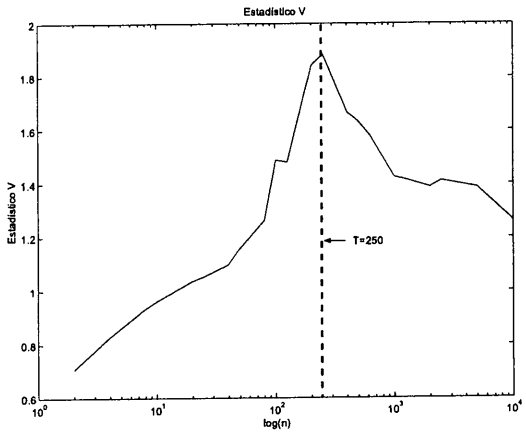


Figura 2: Análisis R/S de una serie con un componente periódico. Estadístico V_N y ciclo promedio.

4. Estadístico V_N

El análisis R/S también se utiliza para detectar ciclos no periódicos y conocer la duración promedio de éstos. Para ello se utiliza el estadístico V_N . El estadístico V_N se define como (2000, Fernández; 1996, 1994, Peters):

$$V_N = \frac{(R/S)_N}{\sqrt{N}} \tag{7}$$

Considérese como ejemplo una serie de tiempo formada por 10000 elementos dada por la siguiente expresión:

$$x\,(n) = cos\,(n) + 5 * std(\cos(n)) * ruido\,(n) \tag{8}$$

El periodo prescrito de la función coseno es de $T = 200$ elementos. El ruido consiste de números aleatorios con distribución normal, media cero y desviación estándar unitaria. El periodo promedio calculado con ayuda del estadístico V_N es de $T = 250$ elementos. Véase figura 2.

5. Casos de Estudio

5.1. Cetes (1 día)

La serie de tiempo CETES (1 día) es una serie diaria que comprende datos disponibles del 1/1/90 al 11/9/98 . El exponente de Hurst calculado es de $H = 0{,}40$. En la figura 3a se muestra la gráfica de su estadístico V_N. Se trata de una serie de tiempo antipersistente.

5.2. IPC (Bolsa)

La serie de tiempo IPC (Bolsa) es una serie diaria que comprende datos disponibles del 1/1/90 al 13/4/98. El exponente de Hurst calculado es de $H = 0{,}61$. En la figura 3b se muestra la gráfica de su estadístico V_N. Se trata de una serie de tiempo persistente.

5.3. CPI (Inflación)

La serie de tiempo CPI (Inflación) es una serie mensual que comprende datos disponibles del 15/1/80 al 15/1/00. El exponente de Hurst calculado es de $H = 0{,}92$. En la figura 3c se muestra la gráfica de su estadístico V_N. Se trata de una serie de tiempo persistente.

5.4. Paridad Peso-Dólar

La serie de tiempo Paridad Peso-Dólar es una serie diaria que comprende datos disponibles del 10/11/94 al 10/4/00. El exponente de Hurst calculado es de $H = 0{,}56$. En la figura 3d se muestra la gráfica de su estadístico V_N. Se trata de una serie de tiempo (casi) aleatoria.

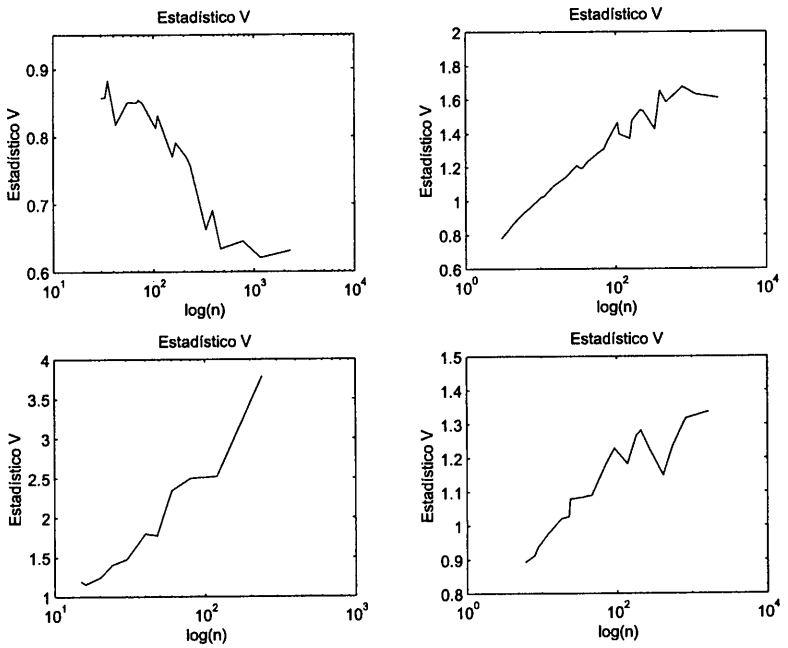


Figura 3: Análisis R/S de indicadores financieros. Estadístico V_N. (3a) -arriba,izquierda-: CETES (1 día); (3b) -arriba,derecha-: IPC (Bolsa); (3c) -abajo,izquierda-: CPI (Inflación); (3d) -abajo,derecha-: Paridad Peso-Dólar.

6. Conclusiones

En la sección 3 se presentó el Análisis R/S y su aplicacion a los mercados financieros. A partir de este tipo de análisis se determinó el exponente de Hurst de una serie de tiempo. Se observó que la serie de tiempo CETES (1 día) es antipersistente, con un exponente de Hurst de $H = 0{,}40$; la serie IPC (Bolsa), persistente con $H = 0{,}61$; la serie CPI (Inflación), persistente con $H = 0{,}92$; y finalmente la serie Paridad Peso-Dólar, (casi) aleatoria con $H = 0{,}56$.

En la sección 4 se presentó el estadístico V_N. A partir de éste se determino el periodo promedio de una serie de tiempo. En general, se requiere que las series de indicadores financieros tengan una duración muy prolongada para poder estimar su periodo promedio. Con los datos disponibles no se aprecia dicho periodo en ningún caso de estudio.

7. Referencias

Fernández, D. (2000). *Dinámica Caótica en Economía.* Madrid: McGraw Hill.

Peters, E. (1994). *Fractal Market Analysis.* New York: John Wiley & Sons.

Peters, E. (1996). *Chaos and Order in the Capital Markets.* New York: John Wiley & Sons.

Turcotte, D. (1997). *Fractals and Chaos in Geology and Geophysics.* New York: Cambridge.

Regresión bajo distribuciones Asimétricas Normales

María Gpe. Russell Noriega
Graciela González Farías
Centro de Investigación en Matemáticas, A.C.

1. Modelo de regresión lineal simple con errores normales sesgados

El énfasis de este trabajo se basa en explorar el potencial de la distribución normal sesgada en aplicaciones a modelos de regresión lineal. En el contexto de regresión lineal existen una gran cantidad de aplicaciones reales, en las cuales los supuestos de normalidad y varianza constante no son factibles. En numerosas situaciones el comportamiento de los datos resulta ser asimétrico y de aquí la inquietud de estudiar el modelo de regresión lineal asumiendo una distribución con características matemáticas similares a la distribución normal y capaz de reproducir el fenómeno de asimetría presentado por los datos. Suponemos entonces que la distribución de los errores es normal sesgada.

Consideremos una variable aleatoria continua X con función de densidad de la forma:

$$f(x;\mu,\sigma,\delta) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left[\delta\left(\frac{x-\mu}{\sigma}\right)\right], \tag{1}$$

con μ el parámetro de localización y σ el parámetro de escala. δ es un número fijo arbitrario conocido como el parámetro de forma, ya que dicho parámetro regula la forma o sesgo de la función de densidad. Finalmente $\phi(x)$ y $\Phi(x)$ denotan la función de densidad y de distribución de una variable normal estándar, respectivamente. Diremos que $X \sim SN(\mu,\sigma^2,\delta)$ siempre que su función de densidad este dada por la ecuación (1).

Usando los resultados de Azzalini (1985), tenemos que: $E(X) = \mu + \sqrt{2/\pi}\sigma\lambda$ y $\mathrm{Var}(X) = \sigma^2\left(1-\frac{2}{\pi}\lambda^2\right)$, con $\lambda = \lambda(\delta) = \delta/(1+\delta^2)^{1/2}$. Para un estudio detallado de la distribución normal sesgada ver Azzalini (1985), Azzalini y Dalla Valle (1996) y Azzalini y Capitanio (1999).

Considere un modelo de regresión donde la distribución de los errores sigue una distribución normal sesgada y ε_i independiente de ε_j para $i \neq j$, tal y como lo proponen en Azzalini y Capitanio (1999). La forma del modelo para el caso univariado es:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim SN(0, \sigma, \delta), i = 1, ..., n, \tag{2}$$

por lo tanto se sigue que los $y_i \sim SN(\alpha + \beta x_i, \sigma, \delta)$.

2. Ecuaciones de verosimilitud

Sea $\theta = (\alpha, \beta, \sigma, \delta)$; la función de verosimilitud para los parámetros está dada por el producto de las densidades de la forma (1), de modo que la correspondiente función de log verosimilitud esta dada por:

$$\ell(\theta; y) = -n \ln \sigma - (1/2\sigma^2) \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 + \sum_{i=1}^{n} \ln \Phi(\delta(y_i - \alpha - \beta x_i)/\sigma). \tag{3}$$

Es posible mostrar que los estimadores de mínimos cuadrados (MC), digamos $\hat{\alpha}_{MC}$, $\hat{\beta}_{MC}$, $\hat{\sigma}_0$ satisfacen las ecuaciones de verosimilitud dadas por las primeras derivadas parciales de $\ell(\theta; y)$ en la ecuación (3) cuando $\delta = 0$. Sin embargo, cabe resaltar que dicho punto, denotado por $\hat{\theta}_0 = (\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$, es de inflexión, *i.e.*, no es ni máximo ni mínimo lo cual se demuestra al probar que la matriz de información de Fisher no es positiva definida ni negativa definida, ver Russell-Noriega y González-Farías (2002).

3. Momentos de los estimadores de mínimos cuadrados

En esta sección calculamos la media y la varianza de los estimadores de MC bajo el supuesto de que los errores se distribuyen con la distribución $SN(0, \sigma, \delta)$. Consideremos el modelo dado en la ecuación (2), note que $E(y_i) = \alpha + \beta x_i + \lambda\sigma\sqrt{\frac{2}{\pi}} = \alpha^* + \beta x_i$ y $\text{Var}(y_i) = \sigma^2\left(1 - \frac{2}{\pi}\lambda^2\right) = \sigma_y^2$, con $\alpha^* = \alpha + \lambda\sigma\sqrt{\frac{2}{\pi}}$ y λ como antes.

De los momentos anteriores obtenemos que: $E(\widehat{\beta}_{MC}) = \beta$ y $E(\widehat{\alpha}_{MC}) = \alpha^*$. Además las expresiones para las varianzas de dichos estimadores son: $\text{Var}(\widehat{\beta}_{MC}) = \sigma_y^2/s_{xx}$; $\text{Var}(\widehat{\alpha}_{MC}) = \sigma_y^2[(1/n + \bar{x}^2/s_{xx}]$, ya que $\text{Cov}(\bar{y}, \widehat{\beta}_{MC}) = 0$.

4. Ejemplo simulado

Suponga el modelo de regresión dado en la ecuación (2), con la finalidad de ilustrar el procedimiento de estimación, para cada uno de los parámetros en el modelo, vía la maximización directa de $\ell(\theta; y)$; así como por el método de MC, se simulan muestras de tamaño 200. El procedimiento de simulación se basa en la siguiente proposición, debida a Henze (1986).

Proposición: Si X_0 y X_1 son variables independientes $N(0,1)$ entonces $Y = (\delta/\sqrt{1+\delta^2})\,|X_0| + (1/\sqrt{1+\delta^2})X_1$ es una variable que distribuye $SN(0,1,\delta)$.

Para simular números aleatorios con distribución $SN(\mu, \sigma, \delta)$ tomamos $Z = \mu + \sigma Y$ con Y simulado a partir de la proposición anterior. El conjunto de datos simulados se generó bajo el siguiente escenario: $\alpha = 1$; $\beta = 0{,}3$; $\delta = 10$; $\sigma = 2$; $n = 200$, de donde obtenemos que: $\alpha^* = \alpha + \lambda\sigma\sqrt{2/\pi} = 2{,}588$ y $\sigma_y^2 = \sigma^2(1 - 2\lambda^2/\pi) = 1{,}4787$.

La tabla siguiente resume las estimaciones de los parámetros por MV y MC.

Parámetros	α	β	δ	σ	α^*	σ_y^2
Simulación	1	0.3	10	2	2.588	1.48
MC o MV (Dist. Normal)		0.299			2.62	1.53
MV (Dist. Normal Sesgada)	1.01	0.299	11.7	2.01	2.6	1.5

En la figura 1 se presenta el comportamiento de los errores en el modelo. La figura 2 presenta la gráfica de dispersión de los datos simulados, con las correspondientes ecuaciones de las rectas ajustadas por MC y MV; recordemos que el procedimiento de MV considera que la distribución de los errores es la normal sesgada.

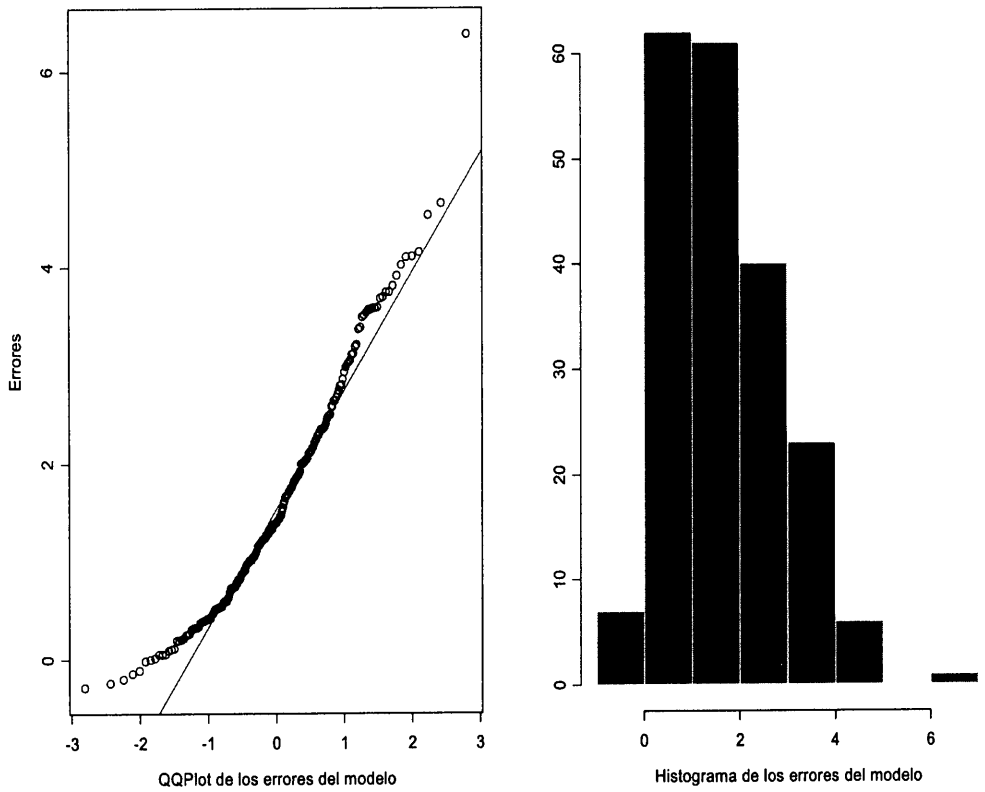


Figura 1: Errores del modelo

En la mayoría de los escenarios simulados, los estimadores de la pendiente por el método de MC y MV resultaron muy parecidos numéricamente, prácticamente las ecuaciones de las rectas estimadas resultaban ser casi paralelas, lo cual se debe a que el estimador de la pendiente por MC resulta ser un estimador insesgado y consistente del parámetro β en el modelo. También se observa que los valores estimados para la varianza son muy cercano al valor verdadero, independientemente del comportamiento del resto de los valores estimados.

5. Conclusiones

En términos generales observamos que las propiedades distribucionales de los estimadores de MC para el modelo de regresión lineal con errores normales sesgados son similares a los

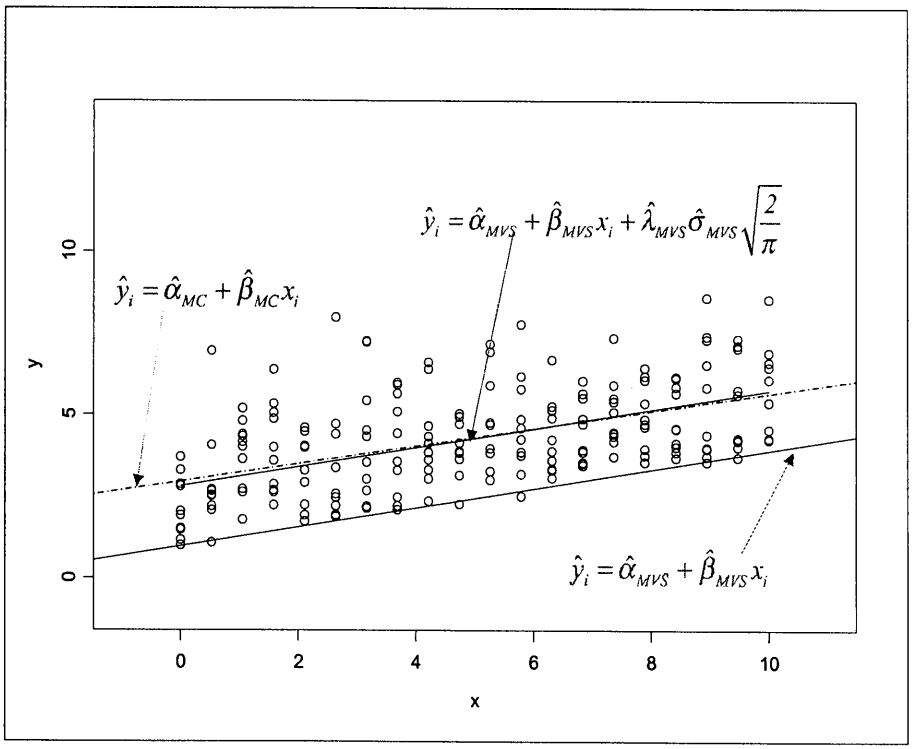


Figura 2: Diagrama de dispersión

del caso tradicional con errores normales. Por ejemplo se observó lo siguiente:

El estimador de la pendiente, $\hat{\beta}_{MC}$, es un estimador insesgado de β. Esta propiedad del estimador de la pendiente puede ser de gran utilidad, por ejemplo dicho valor puede considerarse como valor inicial para el EMV de β, o bien como un valor dado en la función de verosimilitud.

El estimador $\hat{\beta}_{MC}$ es consistente al igual que $\hat{\beta}_{MV}$ propiedad que puede garantizarse si δ es conocido.

La media muestral, $\bar{y}$, no está correlacionada con $\hat{\beta}_{MC}$, *i.e.*, $\text{Cov}(\bar{y}, \hat{\beta}_{MC}) = 0$. Cabe señalar que no son independientes ya que la independencia sólo se tiene en el caso normal, por lo que cuando $\delta = 0$ son independientes.

Los estimadores $\hat{\sigma}^2_{MV}$ y $\hat{\sigma}^2_{MC}$ presentan expresiones similares. Dichos estimadores son iguales

en $\delta = 0$.

El estimador de MC para el intercepto, $\hat{\alpha}_{MC}$, no es insesgado, en realidad $E(\hat{\alpha}_{MC}) = \alpha + \lambda\sigma\sqrt{2/\pi}$, más aún $\hat{\alpha}_{MC} \xrightarrow{p} \alpha + \lambda\sigma\sqrt{2/\pi}$. En este sentido, el estimador de MC para el intercepto nunca será bueno si se desconocen los valores de δ y σ. Sin embargo si δ y σ son conocidos se puede estimar α mediante $\bar{\alpha} = \hat{\alpha}_{MC} - \lambda\sigma\sqrt{2/\pi}$, el cual es consistente.

Dado que los estimadores de MC para α y β son combinaciones lineales de las observaciones, se desprende que su distribución es normal sesgada general. Lo anterior se debe a que la distribución normal sesgada es cerrada bajo combinaciones lineales de rango completo por renglón o por columna como se prueba en Gonzalez-Farías, *et al* (2002).

La función de verosimilitud perfil de δ es muy irregular, puede tener múltiples máximos incluyendo entre ellos a $\delta = \pm\infty$, $\delta = 0$ es siempre una raíz pero no necesariamente un punto que maximice la verosimilitud. En este caso es recomendable seguir las recomendaciones de Azzalini (1985). Dichas recomendaciones consisten en estudiar las propiedades de los parámetros α, β y σ, para valores de δ con altos niveles de verosimilitud perfil. Domínguez-Molina y González-Farías (2002) recomiendan niveles de verisimilitud perfil altos (*e.g.*, $\geq$ 0,5); también observan que la distribución de $R = -2\ln\left(L_p\left(\hat{\delta}\right) - L_p(\delta)\right)$ dista mucho de la distribución ji-cuadrada, esto sucede incluso para muestras de tamaño 500.

6. Referencias

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics.* **12,** 171-178.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society,* B, **61**, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Domínguez-Molina, J. A. and González-Farías, G. (2002). An Optimal procedure for estima-

tion in the univariate skew normal distribution. *Lehmann Simposium, CIMAT, Guanajuato, Gto. 2002.*

González-Farías, G., Domínguez-Molina, J.A. and Gupta, A.K. (2002). Additive properties of skew normal random vectors. Por aparecer en *Journal of Statistical Planning and Inference (JSPI).*

Henze, N. (1986). A probabilitic representation of the "skew-normal" distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.

Russell-Noriega, M. G. y González-Farías, G. (2002), Análisis de Regresión Lineal con Errores Distribuidos Normal Sesgados. *Comunicaciones Técnicas de CIMAT.*

Sobre el Tamaño de Muestra para demostrar la No Inferioridad de un Tratamiento Experimental con respecto a un Estándar usando una Variable Dicotómica

David Sotres-Ramos
Instituto de Socioeconomía, Estadística e Informática del Colegio de Postgraduados, Montecillos, Estado de México

1. Introducción

El modelo estándar para comparar dos tratamientos en base a una variable dicotómica es el de Bernoulli, el cual supone que las observaciones correspondientes al primer tratamiento (estándar) provienen de una muestra aleatoria con una distribución Bernoulli con probabilidad de éxito Π_s [$Bernoulli(\Pi_s)$], y que los datos del segundo tratamiento (experimental) provienen de otra muestra aleatoria con distribución $Bernoulli(\Pi_e)$. Se supone además, que las dos muestras son independientes entre sí. La forma típica para comparar dos tratamientos se formula en términos del contraste de dos hipótesis (nula y alternativa, respectivamente), las cuales tienen la forma:

$$H_0 : \Pi_s = \Pi_e \qquad \text{contra} \qquad H_a : \Pi_s < \Pi_e \tag{1}$$

Esta formulación es adecuada cuando se desea demostrar que el tratamiento experimental es mejor que el estándar. Es decir, el objetivo es demostrar la veracidad de la hipótesis alternativa, ver por ejemplo Fleiss (1981). Sin embargo, en ocasiones el interés no es probar que el tratamiento experimental es mejor que el estándar, sino demostrar que los dos tratamientos son muy similares. Por ejemplo en estudios clínicos de oncología, sida, etc., donde el nuevo tratamiento puede ser menos tóxico, más fácil de administrar o mucho más barato que el estándar, y por lo tanto resulta preferible si es que se prueba que no es inferior en eficacia al tratamiento estándar. Blackwelder (1982) propuso una prueba de hipótesis para demostrar que un tratamiento experimental no es inferior al tratamiento estándar. En este procedimiento el objetivo es demostrar la hipótesis alternativa [$Ha : \Pi_s - \Pi_e < \delta$], dónde

δ es un valor positivo, conocido y pequeño (δ = 0,05 ó 0,10). Cuando esta desigualdad se cumple decimos que el tratamiento experimental no es inferior al tratamiento estándar. Este procedimiento es sencillo de aplicar ya que se dispone de tablas y gráficas para calcular el tamaño de muestra para planear este tipo de estudios, ver por ejemplo Blackwelder (1984) y Machin (1997). Sin embargo, en su artículo Blackwelder (1982) no probó formalmente que el procedimiento propuesto mantiene el nivel de significancia y el poder deseados. El principal objetivo de este trabajo es demostrar formalmente (mediante teoría asintótica) que el procedimiento de Blackwelder si mantiene el nivel de significancia y el poder deseados y también comprobar la fórmula para el cálculo del tamaño de muestra de este procedimiento.

2. Modelo estadístico, formulación y prueba de la hipótesis de no inferioridad.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria con una distribución $Bernoulli(\Pi_s)$ y $Y_1, Y_2, ..., Y_n$ una muestra aleatoria con una distribución $Bernoulli(\Pi_e)$, suponemos además que las dos muestras son independientes entre sí.

Las hipótesis que se desean probar en este caso son:

$$H_0 : \Pi_s - \Pi_e \geq \delta \qquad \text{contra} \qquad H_a : \Pi_s - \Pi_e < \delta \tag{2}$$

dónde δ es una constante positiva, fija y conocida.

La prueba de Blackwelder es rechazar H_0 en favor de H_a si

$$(P_s - P_e - \delta)/s < Z_\alpha \,, \tag{3}$$

dónde P_s y P_e representan la proporción de éxitos muestrales para los tratamientos estándar y experimental respectivamente, Z_α es el percentil de nivel α de la distribución normal estándar y $ds = [P_s(1-P_s)/n + P_e(1-P_e)/n)]^{1/2}$. En las secciones 3 y 4 se demuestra que la prueba en (3) tiene, asintóticamente, un nivel de significancia menor o igual a α y un poder igual a $1-\beta$.

3. Nivel de significancia de la prueba

Resulta claro que el nivel de significancia de la prueba en (3) para contrastar las hipótesis en (2) es:

$$\begin{aligned} \text{Nivel de sig.} &= P\{\text{Rechazar } H_0 \mid H_0 \text{ es cierta }\} \\ &= P\{(P_s - P_e - \delta)/ds < Z_\alpha | \Pi_s - \Pi_e \geq \delta\} \\ &\leq P\{[(P_s - P_e) - (\Pi_s - \Pi_e)]/ds < Z_\alpha\} \end{aligned} \tag{4}$$

En el apéndice 1 se prueba que la estadística de prueba en (3) converge en distribución a la variable aleatoria normal (0,1), la cual denotaremos por N(0,1), es decir,

$$[(P_s - P_e) - (\Pi_s - \Pi_e)]/ds \xrightarrow{L} N(0,1) \tag{5}$$

Por lo tanto, aplicando límites cuando $n \longrightarrow +\infty$ en (4) obtenemos que

$$Lim_{n\longrightarrow+\infty}(\text{Nivel de sig.}) \leq Lim_{n\longrightarrow+\infty} P\{[(P_s - P_e) - (\Pi_s - \Pi_e)]/ds < Z_\alpha\} = \alpha\ ,$$

es decir, asintóticamente el nivel de significancia de la prueba es menor o igual a α.

4. Derivación de la fórmula para el tamaño de muestra

$$\begin{aligned} \text{Poder} &= P\{\text{Rechazar } H_0 \mid H_0 \text{ es falsa } : \Pi_s - \Pi_e < \delta\} \\ &= P\{(P_s - P_e - \delta) < ds\, Z_\alpha \mid [\Pi_s - \Pi_e - \delta] < 0\} \\ &= P\{(P_s - P_e - \delta) - [\Pi_s - \Pi_e - \delta] < ds\, Z_\alpha - [\Pi_s - \Pi_e - \delta]\} \\ &= P\{(P_s - P_e) - (\Pi_s - \Pi_e) < ds\, Z_\alpha - [\Pi_s - \Pi_e - \delta]\} \\ &\quad // \text{ Sea } \Omega = [\Pi_s(1 - \Pi_s)/n + \Pi_e(1 - \Pi_e)/n]^{1/2} // \\ &= P\{\frac{[(P_s - P_e) - (\Pi_s - \Pi_e)]}{\Omega} < \frac{ds}{\Omega} Z_\alpha - [\Pi_s - \Pi_e - \delta]/\Omega\} \\ &\quad // \text{ Ver Apéndice 1 } // \\ &\rightarrow P\{Z < Z_\alpha - [\Pi_s - \Pi_e - \delta]/\Omega\} = 1 - \beta\ . \end{aligned}$$

Al igualar el poder de la prueba con $1-\beta$, se obtiene la ecuación básica para obtener el tamaño de muestra de la prueba:

$$\begin{aligned} Z_\alpha - (\Pi_s - \Pi_e - \delta)/\Omega &= Z_{1-\beta} \\ -(\Pi_s - \Pi_e - \delta)/\Omega &= Z_{1-\beta} - Z_\alpha = Z_{1-\beta} + Z_{1-\alpha} \\ \frac{(\Pi_s - \Pi_e - \delta)^2 n}{[\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]} &= (Z_{1-\beta} + Z_{1-\alpha})^2 \\ n &= \frac{(Z_{1-\beta} + Z_{1-\alpha})^2 [\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]}{(\Pi_s - \Pi_e - \delta)^2} \end{aligned}$$

Apéndice 1. Convergencia de la estadística de prueba

Teorema 1. Sean $X_1, X_2, ..., X_n$ $v^s a^s i^s i^{ds}$ como $Bernoulli(\Pi_s)$, y sean $Y_1, Y_2, ..., Y_n$ $v^s a^s i^s i^{ds}$ como $Bernoulli(\Pi_e)$. Supóngase además que las $X_i's$ y las $Y_i's$ son independientes entre si. Sean $P_s = \sum X_i/n$, $P_e = \sum Y_i/n$, $ds = [P_s(1-P_s)/n + P_e(1-P_e)/n)]^{1/2}$ y $\Omega = [\Pi_s(1-\Pi_s)/n + \Pi_e(1-\Pi_e)/n]^{1/2}$, entonces las siguientes dos proposiciones son verdaderas:

$$[(P_s - P_e) - (\Pi_s - \Pi_e)]/\Omega \xrightarrow{L} N(0,1)\ , \tag{6}$$

$$[(P_s - P_e) - (\Pi_s - \Pi_e)]/ds \xrightarrow{L} N(0,1)\ . \tag{7}$$

Demostración. De los supuestos del teorema es fácil calcular las medias, varianzas y desviaciones estándar de X_i, Y_i y de $(X_i - Y_i)$, para $i = 1, 2, ..., n$,

$$E(X_i) = \Pi_s,\ V(X_i) = \Pi_s(1-\Pi_s),\ DS(X_i) = [\Pi_s(1-\Pi_s)]^{1/2},\ P_s = \sum X_i/n,$$

$$E(Y_i) = \Pi_e,\ V(Y_i) = \Pi_e(1-\Pi_e),\ DS(Y_i) = [\Pi_e(1-\Pi_e)]^{1/2},\ P_e = \sum Y_i/n,$$

$$E(X_i - Y_i) = \Pi_s - \Pi_e,\ V(X_i - Y_i) = \Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e),$$

$$DS(X_i - Y_i) = [\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]^{1/2} = \sigma.$$

Ahora bien, usando la versión estándar del Teorema del Límite Central para las $v^s a^s i^s i^{ds}$ $\{(X_i - Y_i) \mid i = 1, 2, ..., n\}$, ver por ejemplo Serfling (1980), resulta claro que

$$U_n = \frac{\sqrt{n}\,[\sum_{i=1}^{n}(X_i - Y_i)/n - (\Pi_s - \Pi_e)]}{\sigma} \xrightarrow{L} N(0,1)\,,$$

y

$$U_n = \frac{\sqrt{n}[(P_s - P_e) - (\Pi_s - \Pi_e)]}{[\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]^{1/2}} = \frac{[(P_s - P_e) - (\Pi_s - \Pi_e)]}{[\Pi_s(1-\Pi_s)/n + \Pi_e(1-\Pi_e)/n]^{1/2}}\,. \tag{8}$$

Esto comprueba la proposición (6) o primera parte del teorema. Para probar (7), usaremos el llamado Teorema de Convergencia en Distribución, ver por ejemplo Serfling (1980), el cual dice que

$$\text{Si } H_n \xrightarrow{L} N(0,1) \text{ y } J_n \xrightarrow{p} 1 \Rightarrow \frac{H_n}{J_n} \xrightarrow{L} N(0,1)\,. \tag{9}$$

Usando la desigualdad de Tchebychev se prueba que $P_s \xrightarrow{p} \Pi_s$ y $P_e \xrightarrow{p} \Pi_e$ y de aquí se obtiene que:

$$[P_s(1-P_s) + P_e(1-P_e)]^{1/2} \xrightarrow{p} [\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]^{1/2} \tag{10}$$

$$\Rightarrow W_n = \frac{[P_s(1-P_s) + P_e(1-P_e)]^{1/2}}{[\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]^{1/2}} \xrightarrow{p} 1\,. \tag{11}$$

Finalmente, aplicando el Teorema de Convergencia en Distribución en (9) con $H_n = U_n$ y $J_n = W_n$, se obtiene que

$$\frac{U_n}{W_n} = \frac{\frac{\sqrt{n}[(P_s - P_e) - (\Pi_s - \Pi_e)]}{[\Pi_s(1-\Pi_s) + \Pi_e(1-\Pi_e)]^{1/2}}}{\frac{[P_s(1-P_s) + P_e(1-P_e)]^{1/2}}{[\Pi_s(1-\Pi_s)/n + \Pi_e(1-\Pi_e)]^{1/2}}} = \frac{\sqrt{n}[(P_s - P_e) - (\Pi_s - \Pi_e)]}{[P_s(1-P_s) + P_e(1-P_e)]^{1/2}}\,,$$

$$\frac{U_n}{W_n} = \frac{[(P_s - P_e) - (\Pi_s - \Pi_e)]}{[P_s(1-P_s)/n + P_e(1-P_e)/n]^{1/2}} = \frac{[(P_s - P_e) - (\Pi_s - \Pi_e)]}{ds} \xrightarrow{L} N(0,1)\,.$$

Esto comprueba la segunda parte del Teorema 1.

Referencias

Blackwelder, W.C. (1982) "Proving the Null Hypothesis"in clinical trials. Controlled Clin. Trials 3, 345-353.

Blackwelder, W.C. y Chang, M.A. (1984). Sample Size Graphs for "Proving the Null Hypothesis". Controlled Clin. Trials 5, 97-105.

Fleiss, J.L. (1981). Statistical Methods for Rates and Proportions. John Wiley, 2nd edition.

Machin, D. et. al. (1997). Sample Size Tables for Clinical Studies. Blackwell Science, 2nd edition.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley.

Una Estimación del p-value en tablas de doble entrada a través de simulación

José Eliud Vilchis Carrera
Esteban Burguete Hernández
Universidad de las Américas-Puebla

1. Introducción

Las tablas de doble entrada son una herramienta muy útil en diversos campos de la investigación, principalmente en Biología y Medicina. Muchos investigadores hacen uso de ellas para analizar si existe alguna relación entre las variables involucradas en un experimento en particular. Para dicho propósito, el inglés Karl Pearson introdujo la prueba chi cuadrada, la cual no funciona bien para todos los casos por hacer uso de una distribución de probabilidad aproximada. En la literatura se recomienda no utilizar esta prueba cuando en alguna celda se tenga como valor esperado 5 o menos. Esto ha limitado el uso de esta prueba en diferentes trabajos de investigación.

Es por ello que el objetivo del presente trabajo es proponer una prueba en tablas de doble entrada basada en simulación que calcule el p-value de modo más eficiente, y de esta forma superar las limitaciones de las pruebas existentes.

2. Bootstrap Paramétrico

A continuación se describe el bootstrap paramétrico, tomado de Casella y Berger (2002). Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria independiente con función de densidad $f(x \mid \theta)$ conocida, pero θ desconocido (este puede ser un vector de parámetros). Se obtiene $\hat{\theta}_{MV}$ (el estimador de máxima verosimilitud de θ) a partir de la muestra. Substituyendo θ por este estimador, se utiliza $f(x \mid \hat{\theta}_{MV})$ para generar nuevas muestras $x^*_{1i}, x^*_{2i}, \ldots, x^*_{ni}, i = 1, 2, \ldots, b$. Para cada una de las b muestras, se estiman $\hat{\theta}^*_i$, y finalmente se calcula la varianza de $\hat{\theta}$ a

partir de las $\hat{\theta}_i^*$, es decir

$$Var_b^*(\hat{\theta}) = \frac{1}{b-1}\sum_{i=1}^{b}(\hat{\theta}_i^* - \bar{\hat{\theta}}^*)^2, \qquad \text{donde} \qquad \bar{\hat{\theta}}^* = \frac{1}{b}\sum_{i=1}^{b}\hat{\theta}_i^*.$$

3. Estimación del p-value en Tablas de Doble Entrada

Considere una tabla de contingencia con r renglones y c columnas, en la cual se desea probar

$$H_0 : A \text{ y } B \text{ son independientes} \quad vs \quad H_a : A \text{ y } B \text{ no son independientes .}$$

A partir de la tabla observada se calcula el estadístico χ^2_{cal}. La tabla puede expresarse en base a las proporciones que representan los valores observados en las celdas con respecto al total de observaciones, esto es $p_{lm} = \frac{n_{lm}}{n}$, donde n es igual al total de casos en la tabla y n_{lm} el total de casos en la celda ubicada en el renglon l y la columna m. Bajo H_0 cierta, se cumple que $p_{lm} = \frac{p_{l.}}{p_{.m}}$. Se construye la función de distribución de estas proporciones y se generan n números, con los cuales se construye una nueva tabla con los resultados de la simulación. En esta nueva tabla se calcula $\chi^{2,1}_{cal}$ para la primera simulación. El proceso se repite b veces y, a partir de ahí, el p-value aproximado será $\widehat{p-value} = \frac{1}{b}\sum_{i=1}^{b} I(\chi^{2,i} > \chi^2_{cal})$.

Se llevaron a cabo pruebas comparativas entre el p-value de la χ^2 y el de la prueba aquí propuesta. Se tomaron tablas regulares (todos los valores esperados son mayores a 5) e irregulares (al menos un valor esperado menor o igual a 5).

4. Conclusiones

Basándonos en las observaciones para tablas irregulares, se apoya la recomendación de no utilizar la prueba de chi cuadrada cuando al menos una celda tenga como valor esperado 5 o menos. Algunos ejemplos muestran las fallas de la aproximación de la chi cuadrada en el caso mencionado anteriormente

La metodología propuesta en el presente trabajo para estimar el p-value para la prueba de independencia de factores en una tabla de doble entrada es más confiable que las existentes, ya que se basa en la distribución exacta de las frecuencias observadas bajo el supuesto de la hipótesis nula cierta. Consecuentemente, la prueba funciona bien para cualquier tabla dada, con un costo de tiempo que depende del tamaño de la muestra. De esta forma, se recomienda el uso de esta prueba en futuras investigaciones.

Se sugiere para futuras extensiones del presente trabajo:

- Aplicar la presente metodología a otras medidas de las tablas de doble entrada,
- Utilizar otros intervalos de confianza bootstrap, por ejemplo el intervalo basado en los percentiles.

Referencias

Aerts, M. (2002). *Computer Intensive Methods.* Limburgs Universitair Centrum.

Casella, G. y Berger, R.L. (2002). *Statistical Inference.* Cambridge U.P.

Efron, B. y Tibshirani, R.J. (1998). *An Introduction to the Bootstrap.* Chapman & Hall.

Vilchis, J.E. (2003). *Una Estimación del P-value en Tablas de Doble Entrada a Través de Simulación.* Tesis no publicada. Universidad de las Américas-Puebla.

A Comparison of Residual Measures under Configural Frequency Analysis Conditions

Alexander von Eye
Michigan State University

1. Residual Statistics

Cross-classifications can be examined from two perspectives. The first focuses on explaining variable relationships and involves fitting latent variable models or manifest variable models. Examples of methods used under this perspective include latent class analysis and log-linear modeling. When researchers aim at improving a model, residuals can be used to guide model respecification. The second perspective focuses on identifying individual cells or groups of cells that contradict a priori statements about variable relationships. Sample methods used under this perspective include cluster analysis and Configural Frequency Analysis (CFA). In CFA, residuals are not used to make decisions about model improvements. Instead, large residuals are interpreted as indicators of local associations, and researchers look for reasons why individual residuals are that large.

This contribution is concerned with the characteristics of residuals. More specifically, this contribution aims at comparing residual measures that are used either to guide model respecification or to identify those cells that stand out as contradicting assumptions specified for the entire cross-classification. The comparison is based on simulation results.

Consider a log-frequency model of the form $\log m = X\lambda$, where m is an array of expected frequencies, X is the design matrix, and λ is a parameter vector (see Bishop, Fienberg, & Holland, 1975; von Eye, 1988, 2002b). Residuals for this model have been defined in various ways. In this paper, we look at the following four residual measures: (1) the Pearson component X^2; (2) the z-statistic used for the standard normal approximation of the binomial test; (3) Anscombe's z-approximation; and (4) Lehmacher's approximative hypergeometric test. The same four tests have been used in an earlier study (von Eye, 2002a). In the present study, we examine these tests under different conditions. We call these conditions CFA

conditions. Their main characteristic is that the observed frequencies, f_i, are small compared to the sample size, N. Details follow below.

The Pearson test statistic,

$$X_i^2 = \frac{f_i - \hat{m}_i}{\hat{m}_i} ,$$

where f_i is the observed frequency for cell i, is the estimated expected frequency under a base model, and i goes over all cells in the table, is approximately distributed as χ^2 with $df = 1$. The test statistic

$$z_i = \frac{f_i - \hat{m}_i}{\sqrt{\hat{m}_i(1 - \hat{m}_i/N)}} ,$$

where N is the sample size, is a standard normal approximation of the binomial test. If $\hat{m}_i$ is large enough, z_i is standard normally distributed. Anscombe's z-approximation,

$$z_i^A = \frac{3[f_i^{2/3} - (\hat{m}_i - \frac{1}{6})^{2/3}]}{2\hat{m}_i^{1/6}} ,$$

is supposedly more nearly normally distributed than the usual standardized residual, $z_i = (f_i - \hat{m}_i)/\sqrt{\hat{m}_i}$ (Upton, 1973).

Lehmacher (1981) derived the exact variance for the standardized residual. This variance is $\hat{\sigma}_i^2 = Np_i[(1 - p_i - (N-1)(p_i - \tilde{p}_i)]$, with $p_i = \hat{m}_i/N$. To illustrate the estimation of $\tilde{p}_i$ consider a $J \times K \times L$ three-way cross-classification. Then, $\tilde{p}_{ijk}$ can be calculated by

$$\tilde{p}_{ijk} = \frac{(f_{j..} - 1)(f_{.k.} - 1)(f_{..l} - 1)}{(N-1)^3} ,$$

where $f_{i..}$ is the i–th marginal of the first variable, etc. Using $\hat{\sigma}_i^2$ and $\tilde{p}_i$, one can calculate the test statistic

$$z_i^L = \frac{f_i - \hat{m}_i}{\hat{\sigma}}$$

which is normally distributed if the sample size is large. Lehmacher's test requires that the sampling be product-multinomial (also called "independent multinomial", see Agresti, 1996).

2. The Simulation

The simulation reported here uses $2 \times 2 \times 2$ tables. Results for other table formats will be reported elsewhere. A Fortran90 program was written that performed the following tasks:

1. Create frequency distributions for sample sizes from 10 through 1500, in steps that can be selected by the program user. Minimal step size for sample size increments is 2. For each sample size and step size, all frequency distributions are created under the constraints given in ii.

2. Determine for each of the four tests for each cell i for each distribution whether it identifies the observed frequency f_i as significantly larger or smaller than expected; count the significant positive and negative residuals; the log-linear base model was the log-linear main effect model of variable independence; the significance threshold was the Bonferroni-adjusted $a* = 0{,}00625$; this significance level was used to mimic CFA analysis as closely as possible; minimum threshold expected cell frequency was 0,5. In addition, it was determined that the maximum frequency for each cell was $N/4$. This constraint was also introduced to mimic CFA analysis in which individual cell frequencies practically always are of a magnitude that is no more than a fraction of the total sample size.

3. Determine for each distribution the ratio of significant negative to positive residuals.

3. Results

Results will be presented in two parts. The first part describes the relative power of the four residual measures under study. The second part describes the approximation of the four residual measures of their respective sampling distribution.

3.1. The relative power of X^2, z, Anscombe's z, and Lehmacher's z

In the present context, we define *relative power* as the relative frequency with which a residual indicates a significant deviation from a model. Here, we are interested in two aspects of relative power. The first aspect is the rank order: which of the four residual definitions suggests the existence of more significant deviations? The second aspect is symmetry: is the relative frequency of the four residual measures the same for positive and for negative residuals? Figure 1 displays the relative frequencies of positive residuals .

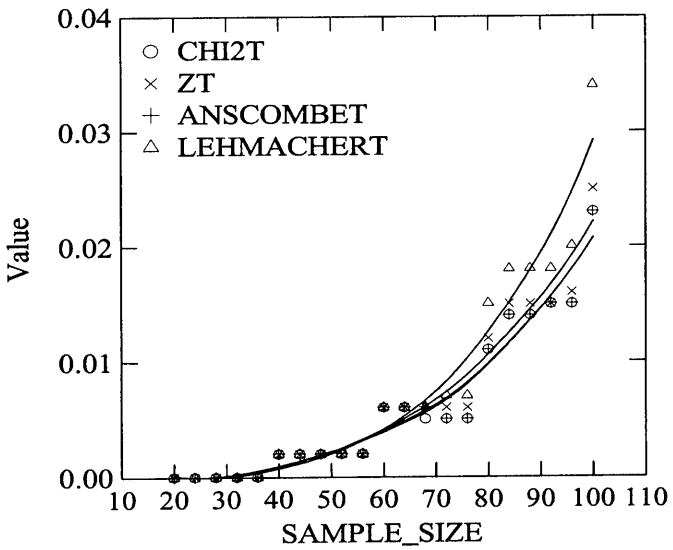


Figure 1: Relative Frequencies of Positive Residuals in Samples Size $10 - 100$ in $2 \times 2 \times 2$ Tables.

Figure 1 shows that the rank order of relative power for positive residuals under the CFA conditions realized in this simulation is Lehmacher's $z >$ $z >$ $X^2 >$ Anscombe's z. Figure 2 displays the relative frequencies for negative residuals.

Figure 2 shows that the rank order of relative power for negative residuals under the CFA conditions realized in this simulation is Lehmacher's $z >$ $z >$ Anscombe's $z >$ X^2. We conclude that the sensitivity of the four tests differs for positive and negative residuals. Figure 3 shows the ratio of antitypes to types for samples of size up to 500.

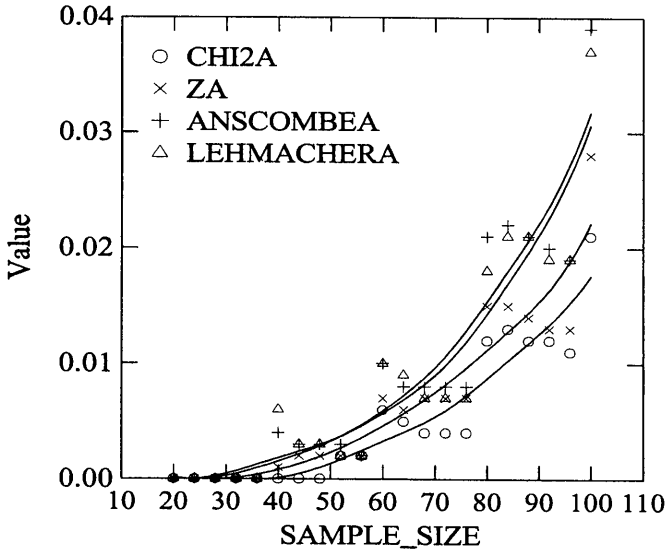


Figure 2: Relative Frequencies of Negative Residuals in Samples Size $10 - 100$ in $2 \times 2 \times 2$ Tables.

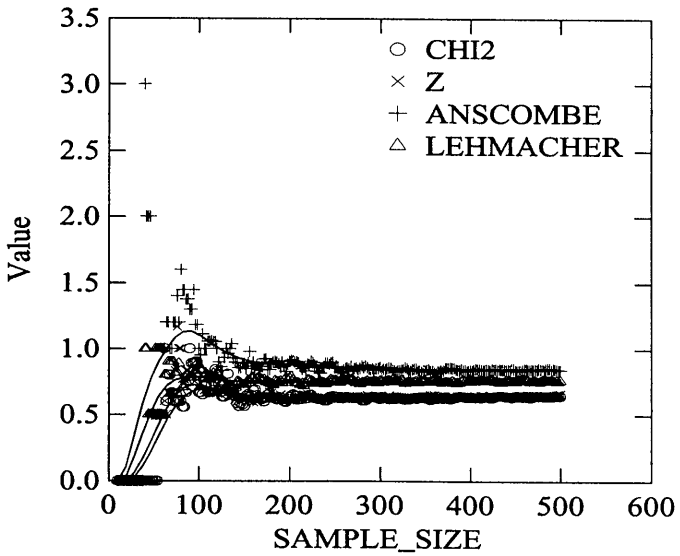


Figure 3: Ratio of Positive to Negative Residuals in Samples Size $10 - 500$ in $2 \times 2 \times 2$ Tables.

Figure 3 suggests that Anscombe's z-approximation comes closest to a $1:1$ ratio of positive to negative residuals, followed by Lehmacher's z, the binomial approximation z, and X^2.

Negative residuals are more frequently identified by all four measures, most extremely so by Pearson's X^2.

3.2. Approximation Characteristics

We now ask how good the chi-square and normal approximations of their corresponding sampling distributions are for the four residual measures. Figure 4 displays the density distributions of the four residual measures.

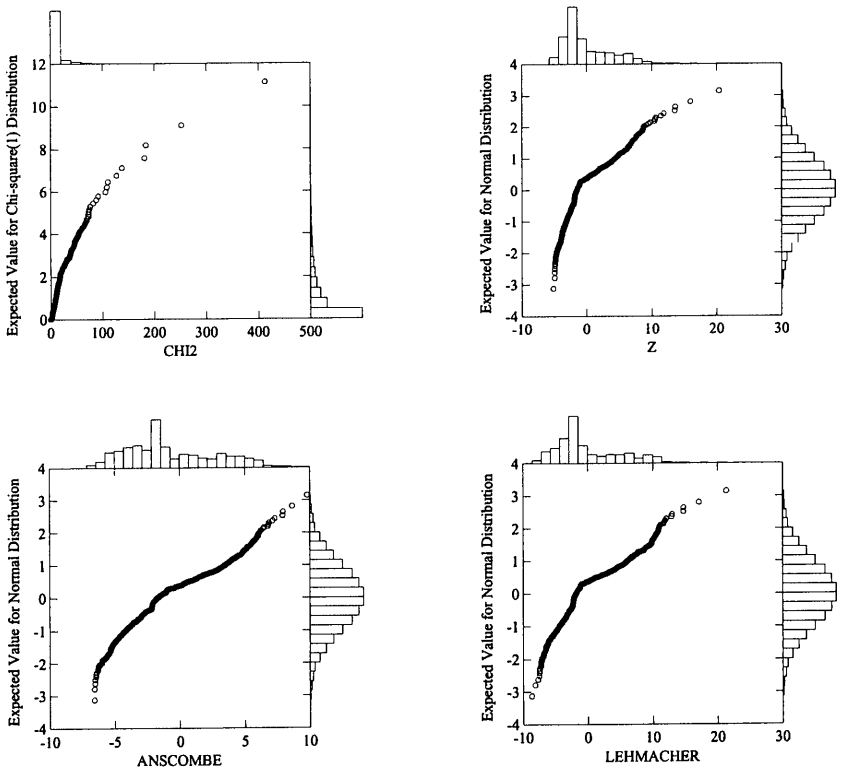


Figure 4: Approximation Characteristics of X^2, z, Anscombe's z, and Lehmacher's z.

We find that, under CFA conditions, (1) none of the approximations of their corresponding sampling distributions is close; (2) each of the distributions that approximate the normal distribution, displays more negative than positive residuals but more extreme positive scores than expected; (3) X^2 has its density mass around zero, but also shows more extreme scores than expected.

4. Conclusions

Under CFA conditions, that is, when the size of the observed cell frequencies is constrained to be small, the four residual measures under study (1) differ in their relative power; (2) differ in their sensitivity to positive and negative residuals; and (3) do not produce distributions that approximate their corresponding sampling distributions well. The first two characteristics are of importance for application in CFA, where large residuals are the basis of interpretation, and positive residuals are interpreted different than large negative residuals. The third characteristic is of importance for significance testing in both log-linear modelling and CFA. It should be noted that the CFA conditions realized in the present simulations are not necessarily representative of conditions in log-linear modeling. In simulations without these constraints, the distributions of the residual measures were much closer to their respective sampling distributions than here (von Eye, 2002a). However, the conditions realized here are representative of many tests performed in the context of CFA. Therefore, methods of testing in CFA may have to be reconsidered.

References

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

Lehmacher, W. (1981). A more powerful simultaneous test procedure in Configural Frequency Analalysis. *Biometrical Journal*, 23, 429 - 436.

Upton, G.J.G. (1978). *The analysis of cross-tabulated data.* Chichester: Wiley.

von Eye, A. (1988). The General Linear Model as a framework for models in Configural Frequency Analysis. *Biometrical Journal*, 30, 59-67.

von Eye, A. (2002). The odds favor antitypes - A comparison of tests for the identification of configural types and antitypes. *Methods of Psychological Research - online*, 7, 1-29. (a)

von Eye, A. (2002). *Configural Frequency Analysis - Methods, Models, and Applications.* Mahwah, NJ: Lawrence Erlbaum. (b)

Primer Analisis de la Consulta Infantil y Juvenil 2000

Marta Zertuche
IIMAS-UNAM

1. Introducción

El Instituto Federal Electoral (IFE) y un grupo importante de instituciones públicas, privadas y sociales llevaron a cabo este proyecto con la finalidad de conocer las percepciones y opiniones de niños y adolescentes sobre valores, prácticas democráticas y problemas públicos relacionados con su vida cotidiana. Esta consulta fue dirigida a niños y adolescentes entre 6 y 17 años de edad, ya que en este rango tienen diferentes maneras de percibir y relacionarse con su entorno, fueron divididos en tres grupos de edad: el primero de 6 a 9 años, el segundo de 10 a 13 años y el tercero de 14 a 17 años. La información fue recabada a través de cuestionarios con respuestas dicotómicas con preguntas fundamentalmente relacionadas con los ámbitos familiar, escolar y comunitario. Todos los cuestionarios fueron contestados el día 2 de julio del 2000, en 14,307 casillas especiales instaladas en todo el territorio nacional, en las que participaron alrededor de 59,500 voluntarios. El objetivo de este trabajo, es presentar los primeros resultados de un análisis de componentes principales para detectar la formación de diferentes grupos y los resultados de diversas regresiones logísticas multivariadas para explicar la formación de dichos grupos.

Debido a la gran cantidad de información, presentaré los resultados del análisis del grupo de 10 a 13 años, ya que son los más claros.

De la base de datos original, se eliminaron los cuestionarios que presentaron más del 10

2. Cuestionario

Para el ámbito familiar y el escolar se formularon las siguientes preguntas, todas con posibles respuestas si o no:

V1: Siento que me entienden y me respetan.

V2: Nos ayudamos unos a otros.

V3: Me tratan con violencia.

V4: Mi opinión cuenta.

V5: Los adultos abusan de su autoridad.

V6: Tengo lugares y tiempo donde jugar.

V7: Se trata igual a las niñas y a los niños.

V8: Mi opinión cuenta para poner las reglas.

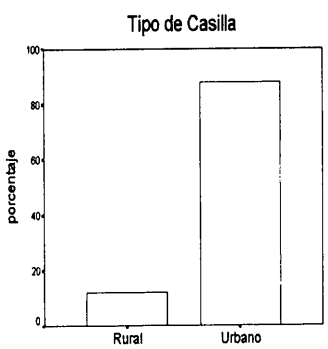
V9: Las reglas y las leyes se aplican igual para todos.

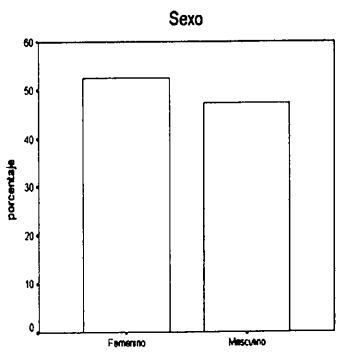
V10: Recibo suficiente información sobre sexualidad.

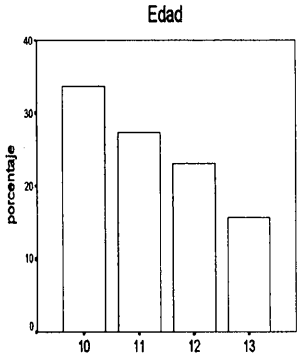
V11: Recibo suficiente información sobre alcohol y drogas.

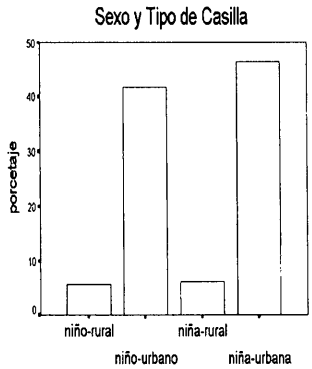
3. Análisis Exploratorio

Se tiene la información de 487,251 cuestionarios y la base de datos se encuentra distribuida de la siguiente manera:









Tipo de Casilla		
	Frecuencia	Porcentaje
Rural	58,226	11.95
Urbano	429,025	88.05
Sexo		
	Frecuencia	Porcentaje
Femenino	254,583	52.25
Masculino	229,578	47.12
Faltantes	3,090	0.63
Edad		
	Frecuencia	Porcentaje
10	164,179	33.69
11	133,728	27.45
12	112,748	23.14
13	76,596	15.72
Sexo y Tipo de Casilla		
	Frecuencia	Porcentaje
niño-rural	27,327	5.61
niño-urbano	202,251	41.51
niña-rural	30,320	6.22
niña-urbana	224,263	46.03
faltantes	3,090	0.63

4. Análisis de Componentes Principales

El objetivo es reemplazar un grupo de variables correlacionadas, a través de una transformación lineal, por un número menor de variables no correlacionadas ordenadas de forma decreciente en términos de su varianza.

La idea es lograr una rotación de ejes principales que transforme a p variables correlacionadas $x_1, \ldots, x_p$ en p variables $z_1, \ldots, z_p$ no correlacionadas. Las coordenadas de los ejes de las nuevas variables, están descritas en los vectores característicos ai que componen a la matriz A que contiene las direcciones de los cosenos (de los nuevos ejes relacionados con los ejes antiguos) utilizados en la transformación:

$$z = A'x$$

Las variables transformadas son llamadas los **componentes principales** de x. El i-ésimo componente principal es : $z_i = \alpha_i' x_i$ que tiene media cero y varianza λ_i, el i-ésimo valor propio.

Suponiendo que la matriz de varianzas-covarianzas de las variables originales es Σ. El primer componente $z_1 = \alpha_1' x_1$ es el de mayor varianza, si se maximiza:

$$\text{Var}(z_1) = \alpha_1' x \alpha_1 \qquad \text{s.a.} \qquad \alpha_1' \alpha_1 = 1.$$

De donde

$$\max_{\alpha} \text{Var}(z_1) = \alpha_1' \lambda \alpha_1 = \lambda.$$

Entonces elijo al valor propio más grande.

Si el procedimiento continúa, se llega a que $\text{Var}(z_i) = \lambda_i$ donde $\text{Var}(z_1) > \text{Var}(z_2) > \cdots > \text{Var}(z_p)$, ie. los valores propios de Σ.

Si es necesario encontrar grupos homogéneos a partir de un conjunto de datos, el Análisis

de Componentes Principales es utilizado para reducir a este conjunto en un menor número de variables transformadas antes de buscar a los grupos.

5. Regresión Logística

Los Modelos Lineales Generalizados están especificados por tres componentes: un **componente aleatorio** que identifica la distribución de probabilidad de la variable de respuesta; un **componente sistemático** que especifica una función lineal de las variables explicativas y que es utilizado como predictor; y una **liga** que describe la relación funcional entre el componente sistemático y el valor esperado del componente aleatorio.

Dentro de estos modelos, el de Regresión Logística se utiliza para una variable de respuesta binaria, por lo que el componente aleatorio se distribuye Bernoulli con

$$E[Y] = \pi(x) \qquad \text{y} \qquad \text{Var}[Y] = \pi(x)(1 - \pi(x))$$

donde $P[Y = 1] = \pi(x)$ para reflejar su dependencia de los valores de las variables explicativas $X = (X_1, \ldots, X_k)$.

Si se supone una relación curvilínea entre x y $\pi(x)$, entonces la función de regresión logística es:

$$\pi(x) = \frac{\exp(\alpha + \beta\, x)}{1 + \exp(\alpha + \beta\, x)}.$$

Para encontrar la función liga, se observa un valor de 1 en la variable de respuesta si los momios son de la forma:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta\, x).$$

Así que los momios se incrementan de forma multiplicativa por $\exp(\beta)$ por cada unidad de

incremento de. Los log-momios muestran la relación lineal:

$$\log\left(\frac{\pi_k(x)}{\sum_{h=1}^{g}\pi_h(x)}\right) = \alpha_k + \beta_k\, x$$

donde g es el número de categorías y $k = 1, \ldots, g$. Este es entonces un **modelo de regresión logística multinomial.**

6. Resultados del Análisis

Primero surgió la necesidad de dividir a las variables de respuesta en dos grupos: familia y escuela.

A continuación presentaré el análisis para las respuestas sobre el ámbito escolar.

6.1. Escuela

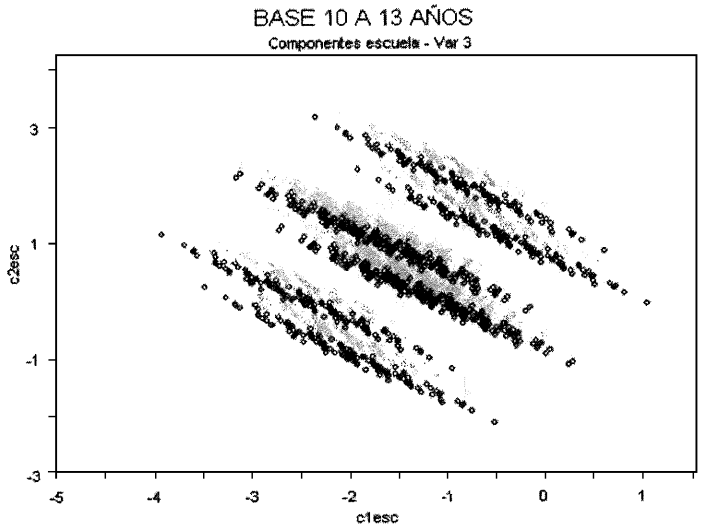
Al aplicar un Análisis de Componentes Principales y graficar el diagrama de dispersión de los dos primeros componentes, las variables que determinan una clara formación de grupos son:

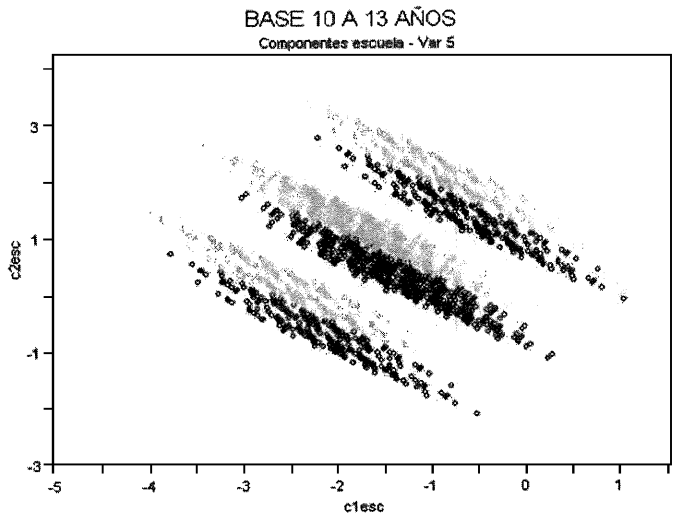
Var 3: me tratan con violencia en mi escuela.

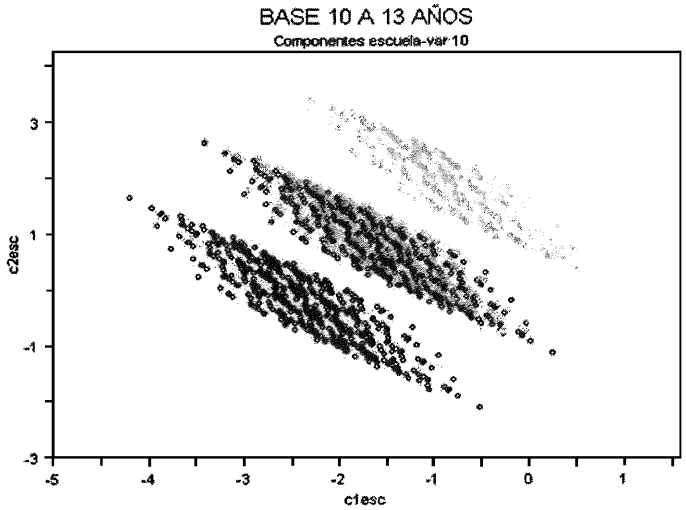
Var 5: los adultos abusan de su autoridad en mi escuela.

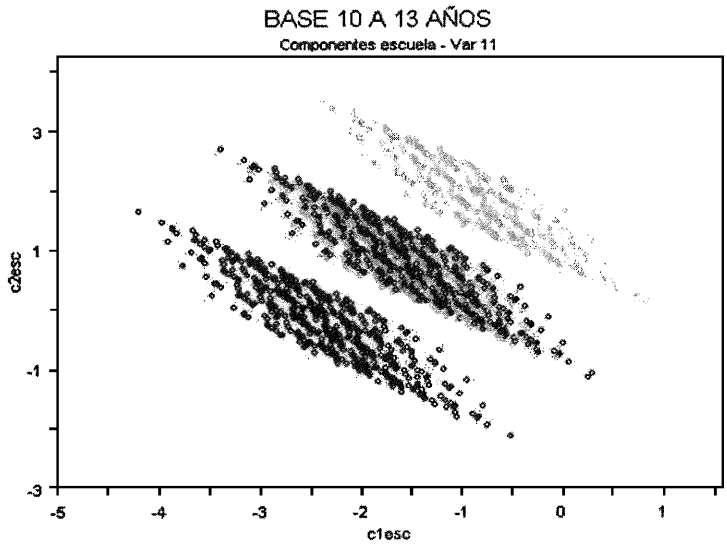
Var 10: recibo suficiente información sobre sexualidad en mi escuela.

Var 11: recibo suficiente información sobre alcohol y drogas en mi escuela.


BASE 10 A 13 AÑOS
Componentes escuela - Var 3
c2esc
c1esc
3
1
-1
-3
-5
-4
-3
-2
-1
0
1


BASE 10 A 13 AÑOS
Componentes escuela - Var 5
c2esc
c1esc
3
1
-1
-3
-5
-4
-3
-2
-1
0
1

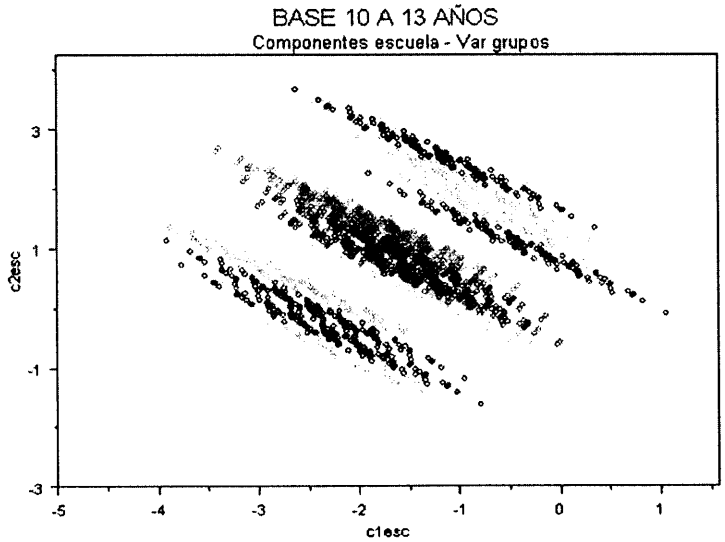

BASE 10 A 13 AÑOS
Componentes escuela-var 10
c2esc
c1esc
3
1
-1
-3
-5
-4
-3
-2
-1
0
1


BASE 10 A 13 AÑOS
Componentes escuela - Var 11
c2esc
c1esc
3
1
-1
-3
-5
-4
-3
-2
-1
0
1

Para cada una de estas variables se observaron grupos bien diferenciados que deben corresponder a las distintas opciones de respuesta al cuestionario. Después, se construyó una nueva variable: **Var grupo** que combina a las posibles respuestas de las variables antes mencionadas, de la siguiente manera:

Var grupo	Var 3	Var 5	Var 10	Var 11
1	si	si	si	si
2	no	si	si	si
3	no	no	no	no
4	si	si	si	no
5	no	si	si	no
6	no	no	si	no
7	si	si	no	no
8	no	no	no	si
9	si	si	no	no

Dichos grupos se ven claramente en la siguiente gráfica de los dos primeros componentes para esta nueva variable:



Para verificar y formalizar la estructura de los grupos, se plantearon modelos de regresión logística multinomial utilizando como componente aleatorio a la variable que clasifica a los diferentes grupos observados en las gráficas anteriores y como variables explicativas al: sexo o tipo de casilla, y las otras variables que corresponden a las preguntas del cuestionario. En estos modelos se hicieron las comparaciones respecto al grupo en el que los niños son tratados con violencia, los adultos abusan de su autoridad y no tienen información sobre sexualidad, alcohol o drogas. Como ejemplo de los modelos de regresión logística multinomial, se tiene al siguiente: Variable explicativa

V1: siento que me entienden y me respetan en mi familia. **0:** No **1:** Si

REGRESION LOGISTICA MULTIVARIADA: VAR GRUPOS = VAR1

VFAGRUPO		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1	Intercept	-2.162	.007	85923.703	1	.000			
	[V1FA=0]	1.223	.035	1215.496	1	.000	3.397	3.171	3.639
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
2	Intercept	-1.614	.006	77034.240	1	.000			
	[V1FA=0]	1.397	.028	2516.430	1	.000	4.042	3.827	4.269
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
3	Intercept	-1.021	.005	49162.041	1	.000			
	[V1FA=0]	.915	.027	1165.077	1	.000	2.498	2.370	2.632
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
4	Intercept	-3.051	.011	74683.988	1	.000			
	[V1FA=0]	2.228	.035	4110.007	1	.000	9.283	8.671	9.937
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
5	Intercept	-1.905	.007	83720.661	1	.000			
	[V1FA=0]	1.712	.028	3780.435	1	.000	5.537	5.243	5.848
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
6	Intercept	-.704	.004	29234.389	1	.000			
	[V1FA=0]	.668	.026	647.174	1	.000	1.951	1.853	2.054
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
7	Intercept	-3.562	.014	62355.405	1	.000			
	[V1FA=0]	2.793	.035	6249.680	1	.000	16.324	15.232	17.494
	[V1FA=1]	0[a]	.	.	0	.	.	.	.
8	Intercept	-2.236	.008	85984.468	1	.000			
	[V1FA=0]	1.948	.029	4573.879	1	.000	7.015	6.630	7.422
	[V1FA=1]	0[a]	.	.	0	.	.	.	.

a. This parameter is set to zero because it is redundant.

En los resultados del modelo se observa que los valores que alcanza $\exp(\beta)$ son siempre mayores a 1, por lo que se concluye que todos los grupos son significativamente mejores que el grupo base, en especial, se tiene que el grupo 4 (niños que si tratan con violencia, los adultos no abusan de su autoridad y si reciben información) y el grupo 7 (a quienes no tratan con violencia, los adultos no abusan de su autoridad y si reciben información) presentan altos valores de $\exp(\beta)$.

7. Conclusiones

Por la naturaleza de esta consulta, se obtuvo una muestra auto seleccionada de niños y jóvenes del país entre 6 y 17 años de edad. Se podría haber pensado en que no acudirían a contestar el cuestionario los que viven bajo condiciones de violencia, injusticia o desinformación. Sin embargo, este grupo destaca claramente en las gráficas surgidas del Análisis de Componentes Principales y éste, junto con ocho grupos más, son identificados a través de los modelos de Regresión Logística Multinomial.

Es importante observar que se identifican exactamente los mismos grupos tanto en el ámbito familiar, el escolar como el del lugar donde viven.

Debido a que se cuenta con una muestra muy grande, es evidente la existencia de significancia estadística en los modelos planteados.

Del análisis de los tres grupos de edades, se observó que al aumentar la edad de los niños, existe mayor diversidad en sus respuestas.

En los tres grupos de edades la representación rural es pobre, por lo que sería de gran interés llevar a cabo otra consulta en la que se logre mayor participación de los niños y jóvenes que viven fuera del medio urbano y así verificar la existencia de este patrón de respuesta.

Se encontraron ciertas combinaciones de respuestas que siguen un patrón específico, por lo que hay que continuar este análisis para estudiar y entender la estructura de dichos patrones.

Esta publicación consta de 410 ejemplares y se terminó de
imprimir en el mes de octubre de 2003 en los talleres gráficos del
Instituto Nacional de Estadística, Geografía e Informática
Av. Héroe de Nacozari Núm. 2301 Sur, Acceso 11, PB
Fracc. Jardines del Parque, CP 20270
Aguascalientes, Ags.
México