ESTADÍSTICA ESPAÑOLA
Vol. 46, Núm. 155, 2004, págs. 5 a 18

Distribuciones discretas en MDS para datos de disimilaridad generados mediante cuestionarios

por
JUAN ANTONIO M. ORELLANA
JOSÉ FERNANDO VERA
ANDRÉS GONZÁLEZ CARMONA
Departamento de Estadística e I. O
Universidad de Granada

ANTONIO PASCUAL ACOSTA
ANTONIO RUFIÁN LIZANA
Departamento de Estadística
Universidad de Sevilla

RESUMEN

En este trabajo se propone la utilización de distribuciones estadísticas discretas como alternativa a la distribución lognormal para la modelización de datos de disimilaridad provenientes de escalas con pocas modalidades en los modelos confirmatorios MDS de Ramsay y Vera. Los resultados ponen de manifiesto que las distribuciones más apropiadas son las mixturas Poisson-lognormal y Poisson-gaussiana inversa. El análisis de datos simulados confirma la falta de adecuación de un modelo continuo a datos provenientes de las escalas generalmente utilizadas en los procedimientos de sondeo de opinión subjetiva de comparación de pares.

Palabras clave: Multidimensional scaling, Lognormal, Poisson-lognormal, Poisson-gaussiana inversa, Máxima verosimilitud.

Clasificación AMS: 91C15

1. INTRODUCCIÓN

Entre 1977 y 1982, Ramsay desarrolla un modelo probabilístico de MDS donde los datos de disimilaridad se modelizan utilizando la distribución lognormal biparamétrica, siendo estimados los parámetros mediante máxima verosimilitud.

Entre las hipótesis más importantes referentes a los datos en el modelo de Ramsay destacan la continuidad en la escala de medida, el origen determinado en cero y la consiguiente no negatividad de los datos. No obstante, esas asunciones adolecen de varios inconvenientes. Por un lado, la no negatividad y el origen determinado en cero plantean problemas desde el punto de vista práctico, ya que bajo ciertas circunstancias, suele ser habitual encontrar datos negativos o cuyo origen no esté determinado en cero. Por otro lado, la hipótesis de continuidad en la escala de medida presenta un serio problema puesto que los procedimientos habituales de obtención de datos de disimilaridad están basados en escalas de medida discretas, siendo necesarias, como pone de manifiesto Ramsay (1982), un mínimo de siete categorías para poder suponer continuidad en los datos, lo que no resulta frecuente a la hora de diseñar encuestas. Además, es conocido experimentalmente que un individuo no suele emplear de forma homogénea más de cinco modalidades, por lo que aunque las escalas estén formadas por más categorías, una distribución continua no define adecuadamente el modelo.

Los problemas planteados por la eliminación de las hipótesis de no negatividad y origen determinado en cero fueron tratados por Vera (1996) mediante la introducción en el modelo de un parámetro umbral, θ, utilizando una lognormal triparamétrica como distribución subyacente, lo que permite considerar los datos definidos sobre $[\theta, +\infty)$, estimando θ mediante una fase adicional dentro del procedimiento iterativo.

Este trabajo se centra en el problema de considerar una escala de medida continua cuando las modalidades empleadas al emitir un juicio constituyen necesariamente una medida discreta, indicando la necesidad de una distribución alternativa para datos de disimilaridad obtenidos mediante procedimientos directos como los empleados en una encuesta, lo que resultará aplicable tanto al modelo de Ramsay como a la extensión de Vera.

2. PLANTEAMIENTO DEL MODELO

El método propuesto por Ramsay se enmarca dentro de las técnicas confirmatorias de MDS, lo que permite, además de obtener la matriz de configuración, tomar decisiones basadas en contrastes de hipótesis y en regiones de confianza. El método empleado para la estimación de los parámetros que intervienen en el problema es el de máxima verosimilitud, por lo que resulta fundamental determinar adecuadamente la distribución de los datos.

En la versión más elemental del modelo de Ramsay se presuponen observaciones independientes e idénticamente distribuidas, con densidades, $f\left(d_{ij} \mid d_{ij}^{*}, \sigma^2\right)$ donde σ^2 representa el factor de variabilidad del modelo, que es constante para cada par de estímulos (i, j). Para la modelización del comportamiento de los juicios emitidos, se considera un modelo multiplicativo de error de las disimilaridades, d_{ij}, respecto a las distancias reales, d_{ij}^{*}, que viene dado por la expresión $d_{ij} = c \cdot d_{ij}^{*}$, siendo c un error aleatorio, de forma que tomando logaritmos se obtiene el modelo aditivo clásico:

$$\ln d_{ij} = \varepsilon + \ln d_{ij}^{*},$$

donde se ha denotado por $\varepsilon = \ln c$.

Si se asume que ε sigue una distribución normal, $N\left(0, \sigma^2\right)$, entonces la distribución de c será lognormal y por tanto, $d_{ij} \sim \Lambda\left(\ln d_{ij}^{*}, \sigma^2\right)$. La suposición de disimilaridades no negativas y el conocimiento empírico de que la dispersión de los datos psicológicos aumenta con su localización(1), han conducido a proponer la ley lognormal como distribución más adecuada para los datos de disimilaridad.

Para esta distribución, la log-verosimilitud que se debe maximizar para obtener los estimadores de la matriz de configuración y el parámetro de dispersión, puede expresarse de la forma,

$$\ln L_{d_{ij}}\left(\sigma^2, X\right) = -\frac{1}{2}\left(\frac{S}{\sigma^2} + M \ln \sigma^2\right) - \sum_{i \angle j} \ln\left(d_{ij}\right) - M \ln\left(\sqrt{2\pi}\right), \qquad [1]$$

donde,

(1) La ley lognormal es un modelo de error muy usado cuando la razón entre la desviación típica y la media se supone constante.

$$S = \sum_{i \angle j} \ln^2 \left(\frac{d_{ij}}{d_{ij}^*} \right),$$

siendo M el número total de observaciones independientes.

La distribución lognormal posee un estadístico suficiente para el parámetro de dispersión, por lo que la estimación de σ^2 puede realizarse de forma independiente de la de la matriz de configuración, X. De hecho se pospone en el modelo la estimación de X hasta haber obtenido la estimación de σ^2. El estimador máximo-verosímil de σ^2 viene dado por,

$$\hat{\sigma}^2 = \frac{S}{M}$$

Sustituyendo σ^2 por $\hat{\sigma}^2$ en [1], se obtienen las ecuaciones que permiten maximizar respecto de X,

$$\ln L_{d_{ij}} \left(\hat{\sigma}^2, X \right) = - \left(\frac{M}{2} \right) \left(\ln S - 1 - \ln M \right)$$

Las ecuaciones implícitas que se obtienen adoptan la expresión,

$$x_{pq} \sum_j t_{pj} = \sum_j x_{pj}\, t_{pj}, \quad p = 1, \ldots, n, q = 1, \cdots, K$$

donde,

$$t_{pj} = \frac{1}{\left(d_{pj}^* \right)^2} \left[\ln \left(\frac{d_{pj}}{d_{pj}^*} \right) + \ln \left(\frac{d_{jp}}{d_{pj}^*} \right) \right].$$

Estas ecuaciones se resuelven mediante técnicas iterativas obteniendo la configuración máximo-verosímil de los estímulos.

3. MODELIZACIÓN DE DATOS DE DISIMILARIDAD EN ESCALAS DISCRETAS

Para resolver el problema de la falta de continuidad de los datos se plantea en este trabajo la utilización de una distribución discreta adecuada que represente las características de las disimilaridades, en principio, supuestamente positivas y con origen en cero. Por tanto, para la modelización de datos de disimilaridad discretos

en MDS resulta conveniente utilizar distribuciones que, al igual que la lognormal, sean leptocúrticas, presenten asimetría positiva y cuya varianza aumente con la localización. Además, su dominio será en esta primera aproximación N, siendo deseable que, cuando el número de categorías efectivas sea apropiado, se aproximen en lo posible a la distribución lognormal, siendo por tanto esta la extensión natural cuando la escala pueda considerarse continua.

En este trabajo se proponen dos distribuciones como adecuadas a estas características, pudiendo considerarse aceptable su aproximación a la lognormal. Concretamente han sido estudiadas la distribución Poisson-lognormal como alternativa teórica natural y la distribución Poisson-gaussiana inversa como elección apropiada desde el punto de vista computacional.

Para centrar la terminología empleada, se dirá que una variable aleatoria, X, se distribuye según una distribución *Poisson-lognormal* (P-LN) de parámetros μ y σ^2, si lo hace según una distribución P(λ) en la que el parámetro, λ, sigue una distribución lognormal biparamétrica, Λ(μ, σ^2). En este caso, su f.m.p. queda determinada por,

$$P_r = P\{X = r\} = \frac{1}{r!\sqrt{2\pi\sigma}} \int_0^{+\infty} e^{-\lambda}\ \lambda^{r-1} \exp\left\{\frac{-(\ln\lambda - \mu)^2}{2\sigma^2}\right\} d\lambda\,,$$

con $r = 0, 1, 2, \cdots$.

Puede encontrarse un estudio detallado de esta distribución en Shaban (1988). La media y varianza vienen dadas por,

$$E[X] = e^{\mu+\frac{\sigma^2}{2}}\ ;\ Var[X] = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) + e^{\mu+\sigma^2/2}$$

siendo los coeficientes de asimetría y apuntamiento, respectivamente,

$$\gamma_1 = \frac{\alpha^2\,\omega(\omega-1)^2\,(\omega+2) + 3\alpha\sqrt{\omega}\,(\omega-1) + 1}{\alpha\sqrt{\omega}\left[\alpha\sqrt{\omega}\,(\omega-1) + 1\right]^{3/2}} \qquad [2]$$

$$\gamma_2 = \left[\alpha^3\ \omega\sqrt{\omega}\,(\omega-1)^3\left(\omega^3 + 3\omega^2 + 6\omega + 6\right) + 6\alpha^2\,\omega(\omega-1)^2\,(\omega+2) + \right. \qquad [3]$$

$$\left. + 7\alpha\sqrt{\omega}\,(\omega-1) + 1\right] \div \left[\alpha\sqrt{\omega}\left[\alpha\sqrt{\omega}\,(\omega-1) + 1\right]^2\right]$$

donde $\alpha = e^{\mu}$ y $\omega = e^{\sigma^2}$.

Esta distribución ha sido empleada con frecuencia para modelizar la abundancia de especies proporcionando mejores ajustes que la lognormal o la binomial negativa (Anscombe (1950), Holgate (1969)). También ha servido para modelizar poblaciones de plancton o crímenes con baja incidencia (Plassman et al. (2001)) y puede suponérsele cierta convergencia asintótica a la lognormal (Holgate (1969)).

Para su empleo en MDS, la búsqueda de una distribución discreta se ha basado fundamentalmente en la adecuación de la forma de la distribución, analizando la evolución de la razón γ_2 / γ_1^2. Fijando el parámetro de localización, μ, se observa que esta razón es creciente en función de ω, comportamiento que también presenta la distribución lognormal.

Considerando fijo el parámetro, $\mu = 1$, la variación de la razón γ_2 / γ_1^2, calculada cuando σ^2, oscila en el intervalo (0,5], se recoge en la **Tabla 1**.

Tabla 1

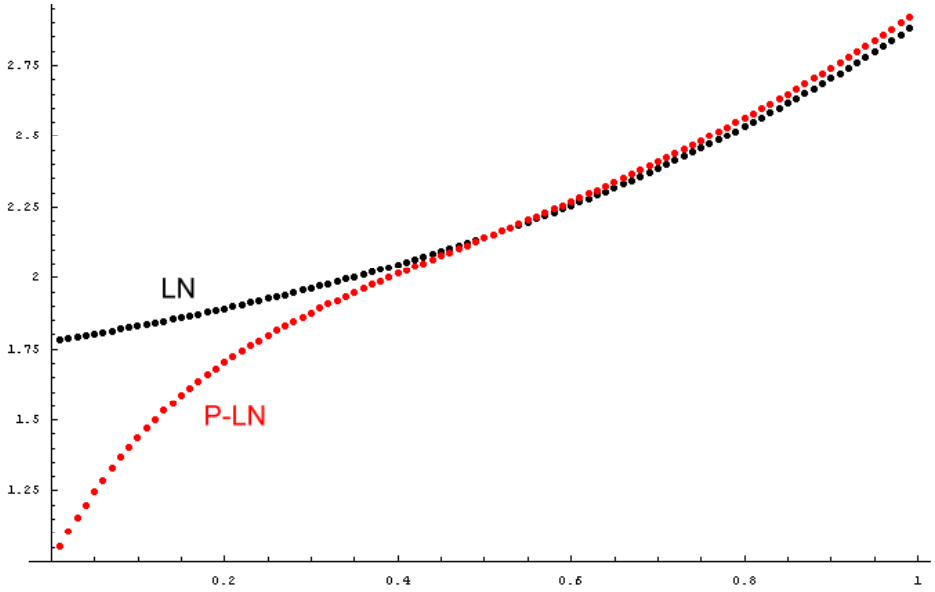
RAZÓN γ_2 / γ_1^2, PARA σ^2 EN (0,5] Y $\mu = 1$ EN LAS DISTRIBUCIONES LOGNORMAL (LN) Y P-LN

σ^2	*0.1*	*0.5*	*1*	*2*	*3*	*4*	*5*
LN	1.830	2.143	2.900	7.005	19.353	53.703	147.453
P-LN	1.438	2.140	2.941	7.071	19.415	53.748	147.482

Gráficamente, puede apreciarse la evolución de los cocientes en función de σ^2 en la Figura 1.

Figura 1

EVOLUCIÓN DE LA RAZÓN γ_2 / γ_1^2



Como se puede apreciar, sólo hay diferencias significativas para valores de σ^2, muy pequeños que hacen que la varianza de la distribución esté muy próxima a su media. Estos valores no son habituales ya que la varianza en las mixturas poissonianas es siempre estrictamente mayor que su media (Teicher (1960)).

La varianza de las dos distribuciones no evoluciona de la misma forma al crecer σ^2 (el crecimiento en P-LN es mayor) por lo que realizamos la misma prueba manteniendo fija la media y permitiendo que la varianza tome un rango de valores equivalente a los obtenidos cuando el parámetro σ^2 en P-LN oscila en el intervalo (0,1]. De nuevo se obtiene una evolución muy similar en ambas distribuciones.

La Tabla 2, recoge algunos valores para las dos distribuciones con media fija igual a 5 y la misma varianza.

Tabla 2

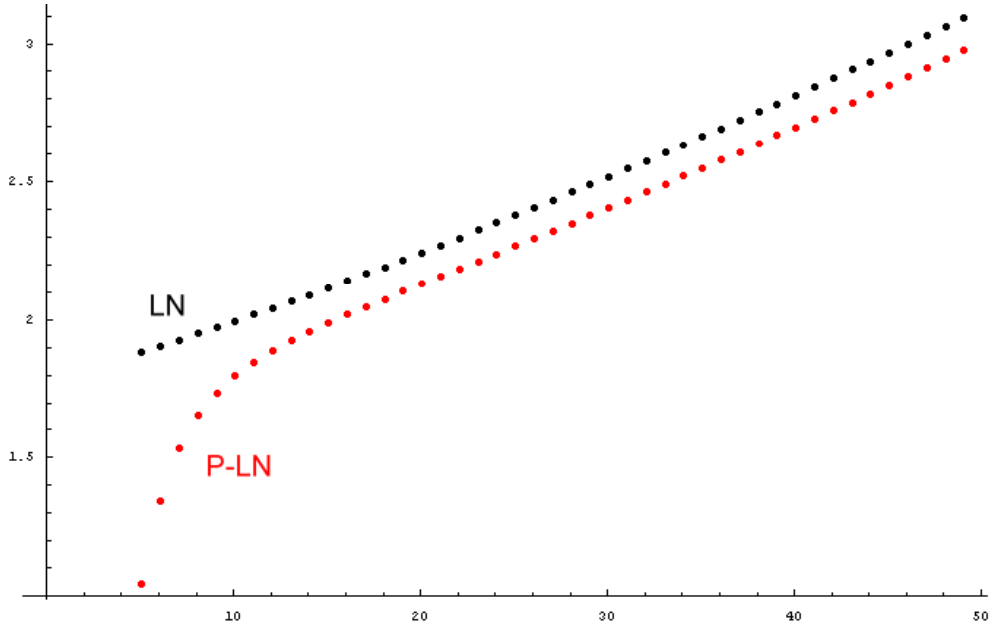
VARIACIÓN DE LA RAZÓN γ_2 / γ_1^2 CON IGUAL VARIANZA Y MEDIA CONSTANTE 5.

Varianza	*5.1*	*10*	*20*	*30*	*40*	*50*	*100*
γ_2 / γ_1^2 (LN)	1.882	1.992	2.240	2.515	2.810	3.120	4.816
γ_2 / γ_1^2 (P – LN)	1.039	1.792	2.126	2.401	2.693	3.003	4.707

Su representación gráfica muestra claramente esa evolución, como puede apreciarse en la Figura 2.

Figura 2

EVOLUCIÓN DE LA RAZÓN γ_2 / γ_1^2



Estas mismas pruebas se han repetido con otros valores tanto de la media como del parámetro μ, obteniéndose las mismas conclusiones, lo que permite afirmar que la distribución P-LN puede utilizarse para modelizar datos de disimilaridad en MDS, sustituyendo adecuadamente a la distribución lognormal y evitando la hipóte-

sis de continuidad en los casos en que la escala realmente utilizada posee menos de siete modalidades.

Considerada esta distribución como alternativa discreta, la siguiente etapa consiste en encontrar los estimadores máximo-verosímiles de los distintos elementos que intervienen en el problema. No obstante, desde el punto de vista computacional resulta extraordinariamente complejo el tratamiento de las ecuaciones de verosimilitud obtenidas, por lo que surge la necesidad de buscar una aproximación de la anterior distribución que mantenga las mismas características y permita el tratamiento analítico del problema de MDS.

La adecuada es la distribución Poisson-gaussiana inversa (P-GI). Shaban (1981) conjetura que esta distribución puede aplicarse en aquellos casos en los que P-LN sea adecuada debido a la similitud de las distribuciones mixturantes (lognormal y gaussiana inversa).

La distribución P-GI es una mixtura poissoniana bi-paramétrica que se obtiene cuando el parámetro de la distribución de Poisson, λ, tiene una distribución gaussiana inversa. Al igual que ocurre con la distribución gaussiana inversa, son múltiples las parametrizaciones que admite esta distribución. Stein, Zuccini y Juritz (1987), proponen la siguiente densidad para la distribución P-GI de parámetros (μ, α), que ha sido la que hemos utilizado.

$$P_X = P\{X = x\} = \left(\frac{(\phi/\alpha)^2}{K_{\frac{1}{2}}(\phi)} \right) \frac{(\mu\phi/\alpha)^x K_{x-\frac{1}{2}}(\alpha)}{x!} =$$

$$= \sqrt{\frac{2\alpha}{\pi}}\, e^{\phi}\, \frac{(\mu\phi/\alpha)^x K_{x-\frac{1}{2}}(\alpha)}{x!}, \quad \mu > 0, \quad \alpha > 0,$$

siendo $\phi = \left(\mu^2 + \alpha^2\right)^{\frac{1}{2}} - \mu$, y $K_\nu(x)$, la función de Bessel modificada de 2ª especie y orden ν.

Entre las ventajas de esta parametrización se incluyen que el parámetro μ es la media poblacional y los estimadores de máxima verosimilitud que se obtienen son más consistentes que los obtenidos con la parametrización original, por lo que es adecuada en la situación que nos ocupa. Si tomamos como medida del error

$$\varepsilon^2 = \sum \frac{\left(p_i - p_i^*\right)^2}{p_i},$$

siendo p_i , el valor de la distribución P-LN y p_i^*, el valor de la aproximación, los errores producidos al emplear P-GI como aproximación de P-LN son casi inapreciables. Algunos de esos errores para una distribución con media m y varianza V se muestran en la **Tabla 3**.

Tabla 3

ERRORES DE APROXIMACIÓN DE LA P-GI A LA P-LN

m	*V*	ε^2
0.1	0.2	0.002600
0.3	1	0.006350
2	3	0.000031
3	6	0.000140
6	10	0.000012
10	92	0.002706

Por tanto, estos resultados muestran la bondad de esta aproximación puntual que permite plantear un modelo probabilístico de MDS sin pérdida significativa de información, bajo una distribución discreta de los datos de disimilaridad, acorde con la escala de medida que desde el punto de vista práctico suele ser derivada.

4. RESULTADOS EXPERIMENTALES

Para comprobar la validez de las distribuciones propuestas en la modelización de los datos de disimilaridad se presenta un experimento de simulación de datos consistente en la generación de 1000 grupos de 30 disimilaridades discretas, aplicando a cada grupo un contraste χ^2 para verificar su ajuste a las distribuciones P-LN, P-GI y LN. Las disimilaridades han sido obtenidas discretizando las perturbaciones de una distancia teórica, d^*, que a su vez ha sido conseguida introduciendo un error multiplicativo con distribución $N(0, \sigma^2)$, generado mediante el método de Box-Müller. Con este procedimiento se simula el comportamiento psicológico de un individuo al responder cuestiones sobre la disimilaridad entre dos estímulos. Los valores de la distancia teórica y de la desviación típica asociada al error, fijos para cada grupo de datos, toman valores entre 1 y 10 y entre 0.4 y 1 respectivamente, considerando únicamente pares de valores (d^*, σ), que den lugar a distribuciones de probabilidad.

Tras calcular el nivel crítico p asociados a cada distribución para cada grupo de datos, los resultados de este experimento proporcionan un 60\% de niveles críticos p asociados a la distribuciones P-LN o P-GI superiores a los asociados a la distribución LN, apreciándose aún más esta tendencia para los conjuntos de datos con mayor diferencia entre media y varianza.

Si se consideran únicamente aquellos niveles críticos p , que rechazan solo una de las distribuciones bajo estudio, el número de grupos de datos para los que el nivel crítico p asociado a la distribución LN es menor que 0.05 y el asociado a P-LN o P-GI es superior a 0.2 es muy superior al número de grupos para el que se rechazaría una distribución P-LN o P-GI.

En la **Tabla 4**, pueden apreciarse algunos de los resultados más sobresalientes de la simulación(2).

Tabla 4

PORCENTAJES OBTENIDOS EN LA SIMULACIÓN

σ	d^*	$P_{PLN} > P_{LN}$	A_{PLN} , R_{LN}	A_{LN} , R_{PLN}
0.7	1.5	76%	13%	-
0.8	1.5	70%	7%	1%
0.6	2	75%	20%	1%
0.8	2	60%	10%	1%
0.6	2.5	68%	20%	1%
0.6	3	60%	10%	1%
0.8	3	60%	3%	1%
0.4	8	52%	4%	-
0.6	8	54%	3%	-

La columna $P_{PLN} > P_{LN}$ indica el porcentaje de niveles críticos p asociados a P-LN que son superiores a los asociados a la LN. La columna A_{PLN} , R_{LN}, indica el porcentaje de conjuntos de datos para los que se obtiene un nivel crítico p superior a 0.2 para P-LN e inferior a 0.05 para LN. Análogamente se define la columna A_{LN} , R_{PLN}.

Como se puede apreciar, a medida que aumenta la distancia teórica, d^* y por tanto, el número de categorías de los datos generados, los resultados para ambas distribuciones son más similares, lo que confirma que la distribución lognormal es la

(2) Los resultados obtenidos al comparar las distribuciones P-GI y LN son prácticamente iguales

extensión natural cuando el número de categorías efectivas permite considerar una escala de medida continua.

Por otra parte hemos considerado la aplicación a unos datos reales, obtenidos experimentalmente mediante encuestación, siguiendo las directrices del ejemplo de Schiffman (1982).

Así pues, se han obtenido 30 matrices de disimilaridad relativas al grado de similaridad entre diez tipos de bebidas de cola (Coca Cola, Coca Cola Light, Coca Cola sin cafeína, Pepsi, Pepsi Light, Pepsi sin cafeína, Casera Cola, Cola Revoltosa, Gold Cola, y Cola Hipercor) clasificándolas en una escala de medida discreta con valores comprendidos entre 0 y 10. El experimento se llevó a cabo con 30 individuos que clasificaron con valores entre 0 y 10 el grado de disimilaridad entre cada uno de los 45 posibles pares de bebidas.

En 20 conjuntos se obtuvo un valor de la varianza menor que el de la media, por lo que no fue necesario realizar el contraste, descartándose el ajuste a estas distribuciones. En los 25 conjuntos restantes, los niveles críticos p asociados a las dos distribuciones discretas fueron siempre superiores a los asociados a la LN, aceptándose el ajuste a esta distribución únicamente en uno de los grupos de disimilaridades, mientras que en 20 ocasiones se obtuvieron niveles críticos p superiores a 0.05 para las distribuciones P-LN y P-GI.

Estos resultados, aunque no resultan tan significativos como los obtenidos para los datos simulados, vuelven a mostrar la necesidad de utilizar distribuciones discretas para la modelización de este tipo de disimilaridades. Por tanto, podemos concluir que los resultados obtenidos, unidos al conocimiento empírico comentado anteriormente, si bien no descartan por completo la utilización de la distribución lognormal para modelizar estos datos de disimilaridad, sí confirman la idea de utilizar estas distribuciones como base de un futuro modelo confirmatorio discreto de MDS.

REFERENCIAS

AITCHISON, J. y BROWN, J. A. C. (1969). «The Lognormal Distribution with special reference to its use in economics». Cambridge at the University Press.

ATKINSON, A. C. y YEH, L. (1982). «Inference for Sichel's Compound Poisson Distribution». *J.A.S.A.*, 77. págs. 153-158.

ANSCOMBE, J.F. (1950). «Sampling Theory of the Negative Binomial and Log Series Distributions». *Biometrika*, 37, págs. 358-382.

FOLKS, J. L. y CHHIKKARA, R. S. (1978). «The Inverse Gaussian Distribution and its Statistical Applications»'. J. R. Statist. Soc., Series B, 40. págs. 263-289.

HAIGHT, F. (1967). «Handbook of the Poisson Distribution».

HOLGATE, P. (1969). «Species Frequency Distributions». *Biometrika*, 56, 3. págs. 651-660

JOHNSON, N. L., KOTZ, S. y KEMP, A. W. (1993). «Univariate Discrete Distributions». John Wiley and Sons Inc.

JOHNSON, N. L., KOTZ, S. y BALAKRISHNAN, N. (1994). «Continuous Univariate Distributions». *Volume 1*. John Wiley and Sons Inc.

ORD, J. K. y WHITMORE, G. A. (1986). «The Poisson-Inverse Gaussian Distribution as a Model for Species Abundance». Commun. Statist. *Theory and Methods*, 15 (3). págs. 853-871.

PLASSMAN, F y TIDEMAN, T. N. (2001). «Does the Right to Carry Concealed Handguns Deter Countable Crimes? Only a Count Analysis Can Say». *Journal of Law and Economics*, 44. págs. 725-746 .

PAUL, S. R. y PLACKETT, R. L. (1978) . «Inference Sensitivity for Poisson Mixtures». *Biometrika*, 65. págs. 591-602.

RAMSAY, J. O. (1977). «Maximum Likelihood in Multidimensional Scaling». *Psicometrika*, 42. págs. 241-266.

RAMSAY, J. O. (1982). «Some Statistical Approaches to Multidimensional Scaling Data». J. R. Statist. Soc., 145, 2. págs. 285-312.

SHABAN, S. A. (1981). «Computation of the Poisson-Inverse Gaussian Distribution». Commun. Statist. - Theory and Methods, A(10). págs. 1389-1399.

SHABAN, S. A. (1988). «Poisson-lognormal Distributions». *Lognormal Distributions. Theory and Application*s. E.L. Crow y K. Shimizu (Editores). Marcel Dekker, Inc. págs. 195-210.

SICHEL, H. S. (1975). «On a Distribution Law for Word Frequencies». *J.A.S.A*., 70. págs. 542-547.

SICHEL, H. S. (1982). «Asymptotic Efficiences of Three Methods of Estimation for the Inverse Gaussian-Poisson Distribution». *Biometrika,* 69. págs. 467-472.

STEIN, G.Z., ZUCCHINI, W. y JURITZ, J. M. (1987). «Parameter Estimation for a Sichel Distribution and its Multivariate Extension». *J.A.S.A*., 82. págs. 938-944.

VERA, J.F. y GONZÁLEZ, A. (1996). «Un procedimiento de MDS a dos vías para el análisis mediante máxima verosimilitud de datos de disimilaridad con origen indeterminado». *QUESTIIO*, Vol 20, 1. págs. 29-43.

WILLMOT, G. (1986). «Mixed Compound Poisson Distributions». *ASTIN Bulletin*, 16. págs. S59-S79.

DISCRETE DISTRIBUTIONS IN MDS FOR DISSIMILARITIES DATA GENERATED BY QUESTIONNAIRES

SUMMARY

We propose here the use of discrete statistical distributions for analysing dissimilarity data deriving from scales containing few categories as an alternative to the lognormal distribution used in Ramsay and Vera's MDS probabilistic models. Our results show that the most suitable distributions are Poisson-lognormal and Poisson-inverse-Gaussian mixtures. An analysis of simulated data confirms the inadequacy of a continuous model for analysing the data within the scales generally used in opinion polls requiring the subjective comparison of pairs of answers.

Key words: multidimensional scaling, lognormal, inverse Poisson-lognormal, inverse poisson-gaussian, maximum likelihood.

AMS Classification: 91C15