Sobre las interacciones: distancias e integrales

Vicenç Torra[1]

IIIA, Institut d'Investigació en Intel·ligència artificial,
CSIC, Consejo Superior de Investigaciones Científicas,
Campus Universitat Autònoma de Barcelona s/n,
08193 Bellaterra, Catalunya, Spain

Abstract. La definición de independencia de variables aleatorias se basa en la distribución de probabilidad conjunta. La distribución gausiana multivariante es un tipo de distribución conjunta en el que las variables pueden no ser independientes. Sin embargo son concebibles otros tipos de relaciones diferentes a la distribución gausiana. En un trabajo reciente mostramos que las medidas difusas pueden utilizarse para definir distribuciones de probabilidad con interacciones entre variables. En este trabajo se presenta un resumen de nuestros resultados recientes en este tipo de distribuciones.
Palabras clave: Distribuciones de probabilidad, integral de Choquet, distancia de Mahalanobis, operadores de agregación, independencia entre variables aleatorias.

1 Introducción

La diferencia fundamental entre las integrales difusas [8] y las medias ponderadas es que las primeras permiten combinar la información teniendo en cuenta las interacciones entre las fuentes de información. Estas interacciones no pueden tenerse en cuenta en la media ponderada. En las integrales difusas la interacción se expresa mediante las medidas difusas.

En el mundo de las distancias, la cuestión de la interacción entre variables aparece también en la distancia de Mahalanobis [5]. Esta distancia tiene en cuenta la matriz de covariancias y, por ello, la distancia se ve afectada por las interacciones entre variables.

Así pues, tenemos dos objetos matemáticos que permiten tener en cuenta la interacción, uno es la distancia de Mahalanobis y el otro la integral difusa. El primero expresa las interacciones mediante la matriz de covariancias (y, por tanto, en términos de correlaciones) y el segundo mediante las medidas difusas.

Dada esta situación uno puede plantearse si es posible considerar un marco en el que ambos aspectos estén relacionados. En este trabajo se resumen algunas aportaciones en este sentido.

En la sección 2 repasamos algunas definiciones que necesitamos en este trabajo. En particular, revisamos los conceptos de distancia de Mahalanobis, de medida difusa y de integral de Choquet [2]. En la sección 3 presentamos la distribución de probabilidad gaussiana y una distribución de probabilidad basada

en la integral de Choquet. En la sección 4 presentamos una generalización de la distancia de Mahalanobis y la integral de Choquet, que llamamos el operador Choquet-Mahalanobis. Esta definición nos permite definir una nueva distribución de probabilidad. El artículo acaba con unas conclusiones y algunas líneas de trabajo futuro.

2 Algunas definiciones

Empezamos esta sección definiendo la distancia de Mahalanobis [5]. Esta distancia está basada en la matriz de covariancias.

Sean $Y_1, \ldots, Y_n$ variables aleatorias, y sean $\bar{Y}_i = E(X_i)$ para $i = 1, \ldots, n$; entonces, la matriz de covariancias Σ se define como:

$$\Sigma = \begin{bmatrix} \mathrm{E}[(Y_1 - \bar{Y}_1)(Y_1 - \bar{Y}_1)] & \cdots & \mathrm{E}[(Y_1 - \bar{Y}_1)(Y_n - \bar{Y}_n)] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[(Y_n - \bar{Y}_n)(Y_1 - \bar{Y}_1)] & \cdots & \mathrm{E}[(Y_n - \bar{Y}_n)(Y_n - \bar{Y}_n)] \end{bmatrix}.$$

En este trabajo nos interesa más la notación alternativa siguiente:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_{12} & \cdots & \rho_{1n}\sigma_{1n} \\ \rho_{21}\sigma_{21}^2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_{n1}^2 & \rho_{n2}\sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}.$$

donde σ_i es la desviación estándard de Y_i y $\rho_{i,j}$ es el coeficiente de correlación entre las variables Y_i y Y_j.

Desde un punto de visto matemático, la matriz de covariancias es una matriz semidefinida positiva, y, por tanto, también simétrica. Además, se sabe que cada matriz semi-definida positiva es una matriz de covariancias.

De acuerdo con lo dicho mas arriba, la distancia de Mahalanobis se define en términos de una matriz de covariancias. Damos la definición a continuación.

Definición 1 *Sean $Y_1, \ldots, Y_n$ variables aleatorias; sean a y b dos vectores en $\mathbb{R}^n$, sea Σ la matriz de covariancia de $Y_1, \ldots, Y_n$; entonces, la distancia de Mahalanobis entre a y b se define como:*

$$d_M(a, b) = (a - b)^T \Sigma^{-1} (a - b)$$

A continuación definimos el concepto de medida difusa (medida no aditiva) y de integral de Choquet [2].

Definición 2 *Sea* $X = \{x_1, \ldots, x_n\}$ *un conjunto; entonces, una función* $\mu : 2^X \to [0, \infty)$ *es una medida difusa si satisface los axiomas siguientes:*

(i) $\mu(\emptyset) = 0$ *(condiciones de frontera)*
(ii) $A \subseteq B$ *implica* $\mu(A) \leq \mu(B)$ *(monotonia)*

Definimos a continuación la integral de Choquet, que integra una función respecto de una medida difusa.

Definición 3 *Sea* μ *una medida difusa sobre* X*; entonces, la* integral de Choquet *de una función* $f : X \to \mathbb{R}^+$ *respecto a la medida difusa* μ *se define como*

$$(C)\int f d\mu = \sum_{i=1}^{n} [f(x_{s(i)}) - f(x_{s(i-1)})]\mu(A_{s(i)}), \tag{1}$$

donde $f(x_{s(i)})$ *indica que los índices se han permutado de manera que* $0 \leq f(x_{s(1)}) \leq \cdots \leq f(x_{s(n)}) \leq 1$*, y donde* $f(x_{s(0)}) = 0$ *y* $A_{s(i)} = \{x_{s(i)}, \ldots, x_{s(n)}\}$.

3 Distribuciones de probabilidad: expresando interacciones

La distancia de Mahalanobis está estrechamente relacionada con la distribución gausiana o normal en el caso multivariante. En particular, una distribución gausiana en un espacio multivariante $\mathbb{R}^n$ viene representada mediante un vector de medias $\bar{x}$ y una matriz de covariancias Σ. Entonces, dado cualquier punto $x \in \mathbb{R}^n$ la distribución gausiana se define como

$$P(x) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

Aquí x^T representa la matriz traspuesta de x.

Si consideramos la distancia de Mahalanobis dada en la definición 1 podemos reescribir esta probabilidad como sigue:

$$P_{\bar{x},\Sigma}(x) = \frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}d_\Sigma(x,\bar{x})}$$

Podemos así observar que la distribución normal incluye información sobre la relación entre las variables expresada mediante la matriz de covariancias. De igual forma uno puede considerar si es posible definir una distribución que tenga en cuenta la relación entre variables tal como se representa en una integral de Choquet. En [11, 9] se introdujo una distribución de probabilidad que en lugar de la distancia de Mahalanobis se basaba en la integral de Choquet. Así pues, en este tipo de distribución las interacciones entre variables seran expresadas mediante medidas difusas y no mediante la matriz de covariancias.

Definimos a continuación estas nuevas distribuciones de probabilidad.

$$P_{\bar{x},\mu}(x) = \frac{1}{K} e^{-\frac{1}{2}CI_\mu((x-\bar{x})\otimes(x-\bar{x}))}$$

donde K es una constante que debemos definir de manera que la función sea una probabilidad, y donde $v \otimes w$ denota el vector obtenido como el producto elemento a elemento de los vectores v y w (i.e., $(v \otimes w) = (v_1 w_1 \dots v_n w_n)$).

A esta distribución de probabilidad basada en la integral de Choquet la llamamos *exponential family of Choquet integral based class-conditional probability-density functions*.

En un intento de integrar los dos conceptos de interacción, hemos definido una distribución de probabilidad que incorpora tanto la matriz de *covariancias* Σ como la medida difusa μ. Escribimos aquí covariancias en cursiva porque, como se ha demostrado, la covariancia entre variables aleatorias construidas a partir de la distribución de probabilidad no corresponde a la *covariancia* indicada por la matriz Σ.

4 Integrales y operadores

Como hemos discutido en las secciones precedentes, las medidas difusas y la matriz de covariancias son dos maneras diferentes de expresar interacciones (o dos maneras de expresar interacciones diferentes). En la sección 3 hemos visto como podiamos definir una distribución de probabilidad que incorpora la información de interacción de las medidas difusas.

Dados los dos tipos de interacciones, uno puede plantearse como considerar los dos tipos en un mismo entorno. Hemos definido el operador Choquet-Mahalanobis con este objetivo. Este operador toma un vector de datos y además tanto una medida difusa como una matriz de *covariancias*. Definimos a continuación este operador.

Definición 4 *Sean $Y_1, \dots, Y_n$ variables aleatorias; sea a un vector en $\mathbb{R}^n$, sea Σ la matriz de covariancias de $Y_1, \dots, Y_n$, sea μ una medida difusa sobre el conjunto $\mathcal{Y} = \{Y_1, \dots, Y_n\}$ y sea $\Sigma^{-1} = LL^T$ la descomposición de Cholesky de su inversa. Entonces, definimos el operador de Choquet-Mahalanobis como sigue:*

$$CMI_{\mu,\Sigma}(a) = CI_\mu(v \otimes w)$$

donde v y w son los vectores definidos como $v = a^T L$ y $w = L^T a$ y donde $v \otimes w$ denota el producto elemento a elemento de los vectores v y w (esto es, $(v \otimes w) = (v_1 w_1 \dots v_n w_n)$.

Cuando Σ es una matriz definida positiva, la descomposición de Cholesky es única. Este es el caso cuando utilizamos una matriz de covariancias Σ para generar una distribución de probabilidad. Por tanto, el operador de Choquet-Mahalanobis está bien definido.

En esta definición utilizamos la descomposición de Cholesky porqué es única para matrices definidas positivas.

La definición del operador Choquet-Mahalanobis nos permite introducir una distancia y una distribución de probabilidad. Empezamos con la definición de la distancia.

Definición 5 *Sea μ una medida difusa y sea Σ una matriz definida positiva. Entonces definimos la distancia de Choquet-Mahalanobis como sigue:*

$$d_{\mu,\Sigma}(x,y) = CMI_{\mu,\Sigma}(x-y).$$

La distribución de probabilidad basada en la distancia de Choquet-Mahalanobis sigue:

$$P_{\bar{x},\mu,\Sigma}(x) = \frac{1}{K} e^{-\frac{1}{2} d_{\mu,\Sigma}(x,\bar{x})}$$

donde K es una constante que debemos definir de manera que la función sea una probabilidad.

5 Conclusiones y trabajo futuro

En este trabajo hemos visto algunas aportaciones que permiten integrar en un único marco las interacciones entre variables que encontramos tanto en la matriz de covariancias como en las medidas difusas.

Hemos visto que la construcción de una distribución de probabilidad en términos de la distancia de Mahalanobis puede utilizarse para definir probabilidades en términos de integrales de Choquet y de nuestro nuevo operador.

Recientemente [10] hemos visto como el operador puede generalizarse utilizando las integrales de Choquet que dependen del nivel [3] (*level-dependent Choquet integral*), lo que permite definir nuevas distribuciones de probabilidad.

Hemos estudiado también [10] algunas propiedades de estas distribuciones, como la esperanza y sus covariancias, y si datos siguiendo una distribución basada en la integral de Choquet pasa el test de Mardia [6, 1, 4, 7]. Tenemos previsto estudiar nuevas propiedades de estas distribuciones de probabilidad.

En [9] habiamos motivado este trabajo en relación a las aplicaciones. En particular, vimos como este tipo de distribuciones pueden ser utilizadas para problemas de clasificación. Sin embargo, la aplicación de estos conceptos a problemas reales necesita primero resolver el proceso de identificación de los parámetros de la distribución. Esto es, la medida difusa y, en el caso de la distribución basada en la distancia de Choquet-Mahalanobis, de la matriz Σ. Estas son algunas lineas de trabajo futuro.

Agradecimientos

Se agrade el apoyo de los proyectos del MEC y MINECO ARES (CONSOLIDER INGENIO 2010 CSD2007-00004), eAEGIS (TSI2007-65406-C03-02) y COPRIVACY (TIN2011-27076-C03-03).

References

1. Baringhaus, L.; Henze, N. (1991). Limit distributions for measures of multivariate skewness and kurtosis based on projections. Journal of Multivariate Analysis 38: 51. doi:10.1016/0047-259X(91)90031-V. edit

2. Choquet, G. (1953/54) Theory of capacities, Ann. Inst. Fourier 5 131-295.
3. Greco, S., Matarazzo, B., Giove, S. (2011) The Choquet integral with respect to a level dependent capacity, Fuzzy sets and systems 175 1-35.
4. Kankainen, A., Taskinen, S., Oja, H. (2004). On Mardia's tests of multinormality. In Theory and Applications of Recent Robust Methods. Edited by Hubert, M., Pison, G., Struyf, A., Van Aelst, S. (pp. 153-164). Basel: Birkhäuser.
5. Mahalanobis, P. C. (1936) On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India 2:1 4955.
6. Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications, Biometrika 57:3 519-530.
7. Mecklin, C. J., Mundfrom, D. J. (2004) An appraisal and bibliography of tests for multivariate normality, International Statistical Review 72:1 123-138.
8. Torra, V., Narukawa, Y. (2007) Modeling decisions: information fusion and aggregation operators, Springer.
9. Torra, V., Narukawa, Y. (2012) On a comparison between Mahalanobis distance and Choquet integral: The Choquet–Mahalanobis operator, Information Sciences 190 56–63.
10. Torra, V., Narukawa, Y. (2013) Exponential family of Level dependent Choquet integral based class-conditional probability functions, Proc. of AGOP 2013, in press.
11. Torra, V. (2011) Expressing interactions: Mahalanobis and Choquet operators, Proc. of AGOP 2011.