CAPÍTULO 6

MODELOS GAUSSIANOS CONDICIONALES GRÁFICOS

MIGUEL ÁNGEL FAJARDO CALDERA
LIDIA ANDRADES CALDITO

Área de Métodos Cuantitativos[1]
Departamento de Economía Aplicada y Organización de Empresas
Universidad de Extremadura

1. INTRODUCCIÓN

La Modelización Gráfica es una forma de análisis multivariante que utiliza grafos para representar modelos. Aunque sus orígenes pueden derivarse del análisis de caminos (Wright, 1921), han sido claves para su desarrollo actual los artículos de Darroch, Lauritzen, y Speed (1980) y de Lauritzen y Wermuth (1989). Otros textos relevantes son los de Whittaker (1990) y Edwards (1995).

Las características principales de estos modelos son las siguientes:

- En cuanto a su <u>interpretación</u>, los grafos aportan información relacionada con la estructura de independencia condicional existente entre las variables objeto de estudio. Tanto la dependencia como la independencia condicional son las claves teóricas de los modelos gráficos, enlazadas con la propiedad global de Markov que aporta un conjunto de reglas explícitas para interpretar los grafos de independencia.
- <u>Simplificación</u> de las problemáticas objeto de estudio: cualquier procedimiento sistemático para analizar observaciones multivariantes, deberá condensar el conjunto de datos sin eliminar u obscurecer las asociaciones relevantes existentes. Los modelos gráficos abarcan el conjunto de todos los posibles modelos de interacción, siendo estimables e interpretables de un modo sencillo, reduciendo notablemente la complejidad en el proceso de selección de los modelos.

[1] Con la colaboración de Nuria Corrales Dios y Jesús Pérez Mayo

- Unicidad en el tratamiento de datos mixtos: Los modelos gráficos proporcionan un marco uniforme para el análisis estadístico de datos continuos resumidos en una matriz de correlación o de datos discretos resumidos en una tabla de contingencia, y esta uniformidad sugiere su generalización a sistemas de variables mixtas.

Incluso cuando los modelos gráficos más simples no suministren una descripción adecuada del conjunto de datos, el análisis gráfico exploratorio proporcionará información muy útil sobre la importancia relativa de las interacciones, sugiriendo caminos alternativos para proseguir el análisis.

Para conseguir tratar variables discretas y continuas conjuntamente habrá que especificar una familia flexible de distribuciones, suficientemente rica para conjugar los dos tipos de variables y lo más sencilla posible para que su tratamiento no sea excesivamente complicado. Wermuth y Lauritzen, (1990), introdujeron la *familia de distribuciones Gaussiana condicional*, la cual comprende casos específicos de las distribuciones Normal multivariante y Multinomial.

La distribución Gaussiana condicional (GC) es definida a través de los dos términos siguientes: la distribución marginal de las variables discretas es Multinomial y condicionando sobre estas variables, las variables continuas siguen una distribución Normal multivariante. La distribución GC contiene como casos específicos el discreto puro y el continuo puro.

Dada su importancia, iniciaremos este artículo presentando la distribución GC, a partir de la cual podremos analizar la estructura de independencia condicional subyacente a un conjunto de variables. Continuaremos, en el siguiente epígrafe, formulando unas pautas para representar gráficamente esa estructura de independencia condicional, derivada del análisis de la distribución GC. Además se mostrará como enunciar la fórmula que dado un grafo identifica a cada Modelo Gaussiano Condicional. Por último se abordará el problema de la estimación de estos modelos, presentándose una medida de la bondad del ajuste: la desvianza.

2. LA DISTRIBUCIÓN GAUSSIANA CONDICIONAL

Supongamos que tenemos p variables discretas y q variables continuas, denotando el conjunto de variables por Δ y Γ respectivamente. Escribiremos las variables aleatorias correspondientes como (I,Y), y una observación cualquiera como (i,y). Aquí, i será vector p-dimensional que contendrá los valores de la variable discreta, e y es el vector que contiene las variables continuas de dimensión q. Denotaremos por τ al conjunto de todos los i posibles.

Supongamos que la probabilidad de que I=i es p_i y que la distribución de Y dado $I=i$ es normal Multivariante $N(\mu_i, \Sigma_i)$ por lo que la media y la covarianza condicional pueden depender de i. Esto es conocido como la distribución GC. La densidad puede escribirse como:

$$f(i,y) = p_i \left|2\pi\Sigma_i\right|^{-\frac{q}{2}} \exp\left\{-\frac{1}{2}(y-\mu_i)' \Sigma_i^{-1}(y-\mu_i)\right\} \quad (1)$$

Los parámetros $\{p_i, \mu_i, \Sigma_i\}_{\forall i \in \tau}$ son llamados parámetros momentos.

En general, solemos interesarnos en los modelos cuya covarianza es constante sobre i, por lo que $\Sigma_i = \Sigma$. A estos modelos se les denomina homogéneos. Como se verá más tarde, hay generalmente dos tipos de modelos gráficos correspondientes a un grafo dado: un modelo heterogéneo y otro homogéneo.

La ecuación (1) podemos escribirla del siguiente modo:

$$f(i,y) = \exp\left\{\alpha_i + \beta_i' y - \frac{1}{2} y'\Omega_i y\right\} \quad (2)$$

donde α_i es un es un escalar, β_i es un vector de dimensión px1, y Ω_i es una matriz definida positiva, simétrica, de dimensión pxp. Estos son denominados parámetros canónicos. Para pasar de los parámetros momentos a los canónicos utilizamos las siguientes expresiones:

$$\Omega_i = \Sigma_i^{-1}, \quad (3)$$

$$\beta_{\mathrm{i}} = \Sigma_i^{-1}\mu_i, \quad (4)$$

$$\alpha_i = \ln(p_i) - \frac{1}{2}\ln\left|\Sigma_i\right| - \frac{1}{2}\mu_i'\Sigma_i^{-1}\mu_i - \frac{q}{2}\ln(2\pi), \quad (5)$$

y

$$\Sigma_i = \Omega_i^{-1}, \quad (6)$$

$$\mu_{\mathrm{i}} = \Omega_i^{-1}\beta_i, \quad (7)$$

$$\mathrm{p_i} = (2\pi)^{q/2}\left|\Omega_i\right|^{-1/2} \exp\left\{\alpha_i + \frac{1}{2}\beta_i\Omega_i^{-1}\beta_i\right\} \quad (8)$$

Los modelos de interacción jerárquica son construidos restringiendo los parámetros canónicos de un modo similar a los modelos loglineales. Es decir, los parámetros canónicos son desarrollados como sumas de términos de interacción, y los modelos son definidos estableciendo como nulos los términos de interacción de orden superior.

3. FÓRMULA DE LOS MODELOS GAUSSIANOS GRÁFICOS CONDICIONALES (MGGC)

Continuaremos con el caso general. Supongamos que la *fórmula de un modelo* es:

$$d_1,...,d_r / l_1,...,l_s / q_1,...q_t. \tag{9}$$

La *primera* parte se refiere a los *generadores discretos* especificados por la expansión de α_i. La *segunda* parte corresponde a los *generadores lineales* especificados por la expansión de β_i.

Cada generador lineal contiene una variable continua. La expansión para β_i^γ para cualquier $\gamma \in \Gamma$ viene dada por los generadores lineales que contienen γ. La *tercera* la parte cuadrática proporciona la expansión para la matriz de covarianza inversa Ω_i.

Cada *generador cuadrático* contendrá al menos una variable continua. La expansión para $\omega_i^{\gamma\zeta}$ para $\gamma, \zeta \in \Gamma$ viene dada por los generadores cuadráticos contenidos en γ, ζ.

Dos *reglas de sintaxis* restringen los valores posibles de la fórmula que se establecen para asegurar que los modelos no varíen ante cambios de origen y escala de las variables continuas:

a) El generador lineal no debe ser más largo que el generador discreto, es decir, por cada generador lineal l_j debe corresponder a un generador discreto d_k tal que $l_j \cap \Delta \subseteq d_k$.

   Por ejemplo, A,B/ABX/AX no sería admisible ya que aparece un generador lineal ABX pero no aparece el generador discreto conteniendo AB.

b) Los generadores cuadráticos no podrán ser más largos que los generadores lineales correspondientes, es decir, para cada generador cuadrático q_j y cada variable continua $\gamma \in q_j$, debe existir el correspondiente generador lineal l_k tal que $(q_j \cap \Delta) \cup \{\gamma\} \subseteq d_k$.

   Por ejemplo, ABC/AX,BY,CZ/AXY,CZ no sería posible debido a que aparece un generador cuadrático AXY pero ningún generador lineal que contenga AY.

4. RELACIÓN ENTRE LA FÓRMULA DE UN MGGC Y EL GRAFO QUE LO REPRESENTA

Antes de profundizar en la relación entre la fórmula de un modelo y el grafo que lo representa habrá que precisar que entendemos por grafo de independencia condicional.

El concepto de *grafo de independencia condicional* es definido por la Propiedad de Markov para Pares de Variables. Esta propiedad establece que si un par de variables (pertenecientes a un vector distribuido según una distribución GC) son condicionalmente independientes, dado el resto de variables, entonces ese par de variables no serán adyacentes en el grafo de independencia condicional.

Dado un grafo, los MGGC se definen como aquellas familias de distribuciones GC que satisfacen las independencias condicionales implícitas en el grafo de independencia condicional asociado.

Para estudiar la correspondencia entre la fórmula de un modelo y el grafo a través del cual lo representamos, expandiremos la ecuación 2 del siguiente modo:

$$f(i,y)\exp\left\{\alpha_i + \sum_{\gamma\in\Gamma}\beta_i^{\gamma} y_\gamma - \frac{1}{2}\sum_{\gamma\in\Gamma}\sum_{\eta\in\Gamma}\omega_i^{\gamma\eta} y_\gamma y_\eta\right\} \tag{10}$$

y después aplicaremos el criterio de factorización[2] para analizar las independencias por pares de Markov del modelo dado.

Para dos variables discretas del modelo, A y B, $A \perp B|(resto)$ cuando todas las interacciones que involucren a A y B sean siempre cero. Es decir, ninguna de las expansiones de $\alpha_i, \beta_{\mathrm{i}}^{\gamma}$ o $\omega_i^{\gamma\eta}$, para cualquier $\gamma, \eta \in \Gamma$ puede contener la interacción AB. En términos de la fórmula del modelo, implica simplemente que si el generador discreto no contiene AB, por las reglas de sintaxis sabremos que ni los generadores cuadráticos ni los lineales podrán contener AB.

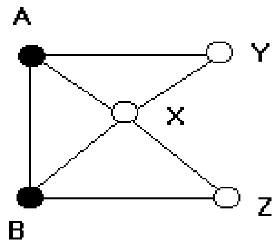
Si A es discreto y X es continuo, y vemos que $A \perp X|(resto)$ significará que cualquier término que involucre la interacción de A con la de X deberá ser cero. Es decir, ninguna expansión de β_i^X ó $\omega_i^{X\eta}$ para cualquier $\eta \in \Gamma$ puede incluir un término de interacción que contenga a A. En relación a la fórmula del modelo, significa que ningún generador lineal puede contener AX, y por las reglas de sintaxis sabremos también que ningún generador cuadrático tampoco contendrá AX.

[2] En relación al criterio de factorización y las propiedades de Markov se ofrece información detallada en el anexo 1

Para dos variables continuas, X e Y, $X \perp Y | (resto)$ conlleva que cualquier ω_i^{XY} será nulo. En términos de la fórmula del modelo significa que ningún generador cuadrático podrá contener XY.

Antes de continuar con la exposición del tema central que nos ocupa, nos parece inevitable parar para hacer una reflexión: repasando las definiciones de independencia condicional expuestas arriba en función de los valores que toman diferentes parámetros, entre ellos los $\omega_i^{\gamma\eta}$*, podemos entrever el importante papel que estos desempeñan a la hora de especificar un modelo concreto, además se puede demostrar que existe una relación directa entre cada* $\omega_i^{\gamma\eta}$ *y el coeficiente de correlación parcial correspondiente. Del tal modo que cuando un* $\omega_i^{\gamma\eta}$ *sea cero, también será nula la correlación parcial existente entre el par de variables que se vea involucrado. De todo esto podemos inferir que un par de variables será condicionalmente independiente dado el resto de variables cuando su correlación parcial sea nula ya que eso estará originado en el hecho de que el correspondiente* $\omega_i^{\gamma\eta}$ *sea cero. Precisamente aquí tenemos la base fundamental de los Modelos Gráficos Gaussianos*[3]*: la medida de la relación entre variables a través del coeficiente de correlación parcial, siendo este mucho más potente que el coeficiente de correlación de Pearson. Por lo tanto, a través del coeficiente de correlación parcial, se modelizarán gráficamente las relaciones existentes entre las variables objeto de estudio.*

Continuando con nuestra argumentación inicial, los resultados enunciados referentes a las relaciones de independencia condicional entre pares de variables continuas, discretas o mixtas, pueden derivarse con facilidad del grafo de independencia construido a partir de la fórmula de un modelo gráfico gaussiano concreto. Por ejemplo, el grafo de $AB / AX, BX, AY, BZ / XY, XZ$ es:



[3] O modelos gráficos que asumen una distribución GC para sus variables.

Para realizar la operación inversa, es decir, encontrar la fórmula de un modelo gráfico a partir del grafo correspondiente *G*, necesitaremos identificar la interacción máxima que sea consistente con *G*. Supongamos que tenemos el grafo anterior, y necesitamos encontrar los cliques G_Δ, es decir, el subgrafo de *G* inducido por las variables discretas. En el grafo representado antes sería AB, luego la primera parte de la fórmula será AB.

Para la parte lineal de la fórmula, deberemos encontrar los cliques de $G_{\Delta\cup\{\gamma\}}$ para cada $\gamma \in \Gamma$. Observamos que estos serán los generadores ABX, AY y BZ, luego la parte lineal será ABX, AY, BZ.

Finalmente para la parte cuadrática, esta dependerá de si estamos interesados en el modelo gráfico homogéneo o en el heterogéneo. Para el modelo homogéneo, necesitaremos identificar los cliques en G_Γ. En este ejemplo en concreto, los cliques de G_Γ son $\{X,Y\}$ y $\{X,Z\}$, por lo que nuestra fórmula adquiriría el siguiente aspecto: $AB / AX, BX, AY, BZ / XY, XZ$.

Para el modelo heterogéneo, habría que encontrar los cliques de *G* que presenten alguna intersección con Γ. En nuestro ejemplo estos serían $\{A, X, Y\}, \{\mathrm{A,B,X}\}$ y $\{\mathrm{B,X,Z}\}$, por lo que la fórmula del modelo quedaría especificada como: $AB / ABX, AY, BZ / AXY, ABX, BXZ$.

Para finalizar, insistir en la idea de que los grafos de independencia condicional clarifican notablemente la estructura de interacción de un conjunto de variables. El grafo pone de manifiesto que conjuntos de variables interactúan, sugiriendo modos en los que el modelo probabilístico puede factorizarse. A través del grafo de independencia de un modelo, podremos conocer si la distribución correspondiente al grafo puede ser factorizada, en distribuciones marginales más simples. Las familias de distribuciones de densidad que dado un grafo de independencia pueden factorizarse completamente son denominadas *modelos descomponibles*[4]. La propiedad característica de sus grafos es la triangulación, pudiéndose demostrar que los grafos triangulados son por su naturaleza una generalización de los grafos que representan cadenas de Markov.

5. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

Los modelos por su naturaleza necesitan datos. Supongamos que tenemos una muestra de *N* observaciones independientes e idénticamente distribuidas

[4] El hecho de que un modelo sea o no descomponible adquirirá gran trascendencia a la hora de estimar un modelo, como ya se mencionará en el epígrafe dedicado a la estimación.

$\left(i^{(k)}, y^{(k)}\right)$ para $k = 1,...N$. Donde *i* representa las p dimensiones de la variable discreta con sus correspondientes niveles, e *y* es un vector q-dimensional. Sean $\left(n_j, t_j \bar{y}_j, SS_j, S_j\right)_{j\in I}$ las frecuencias observadas, las sumas totales de las variables, las medias y las sumas de cuadrados y productos sin corregir, variando estos elementos para cada *j,* es decir,:

$$n_i = \neq \left\{k : i^{(k)} = i\right\}, \quad t_i = \sum_{k:i^{(k)}=i} y^{(k)}, \quad \bar{y}_i = t_i / n_i ,$$

$$SS_i = \sum_{k:i^{(k)}=i} y^{(k)}\left(y^{(k)}\right)',$$

$$S_i = \left[\sum_{k:i^{(k)}=i}\left(y^{(k)} - \bar{y}_i\right)\left(y^{(k)} - \bar{y}_i\right)' \Big/ n_i\right] = \left(SS_i / n_i\right) - \bar{y}_i \bar{y}_i' .$$

También necesitaremos una notación para las magnitudes marginales correspondientes. Para $a \subseteq \Delta$, denotaremos al elemento marginal correspondiente a *i* como i_a y del mismo modo para $d \subseteq \Gamma$, escribiremos el subvector de *y* como y^d . Igualmente, escribiremos las frecuencias marginales como $\left\{n_{i_a}\right\}_{i_a\in I_a}$, los totales marginales de las variables como $\left\{t_{i_a}^d\right\}_{i_a\in I_a}$, y las sumas de cuadrados y productos sin corregir como $\left\{SS_{i_a}^d\right\}_{i_a\in I_a}$.

Supongamos, a continuación, un modelo dado por la fórmula $d_1,...,d_r / l_1,...,l_S / q_1,...,q_t$. A partir de la expresión (2), directamente se demuestra que el conjunto de estadísticos mínimo suficientes viene dado por:

- Un conjunto de frecuencias marginales $\left\{n_{i_a}\right\}_{i_a\in I_a}$ correspondiente a los generadores discretos, es decir, para $a = d_1,...,d_r$.
- Un conjunto de totales marginales de las variables $\left\{t_{i_a}^d\right\}_{i_a\in I_a}$, correspondiente a los generadores lineales, es decir, para $a = l_j \cap \Delta, \gamma = l_j\Gamma, \quad \forall \quad \mathrm{j} = 1,...,\mathrm{s}.$
- Un conjunto de sumas y cuadrados marginales sin corregir $\left\{SS_{i_a}^d\right\}_{i_a\in I_a}$ correspondientes a los generadores cuadráticos, es decir, para $a = q_j \cap \Delta$, y $\mathrm{d} = \mathrm{q_j} \cap \Gamma$ para $j = 1,...,t$.

Como ya se vio, los modelos se construyen restringiendo los parámetros canónicos a través de sus expansiones factoriales. Dado un conjunto de datos, estimaremos por máxima verosimilitud los parámetros de un modelo sujeto a dichas restricciones. A partir de la teoría de la familia exponencial, sabemos que los EMV (estimadores de máxima verosimilitud) se obtienen igualando los estadísticos mínimo suficientes con sus valores esperados. Es decir,

para $a = d_1,...,d_r$, $$\{n_{i_a}\}_{i_a \in I_a} = \{m_{i_a}\}_{i_a \in I_a} \tag{11}$$

para $a \cup \gamma = l_1,...,l_S$ $$\{t_{i_a}^{\gamma}\}_{i_a \in I_a} = \left\{ \sum_{j:j_a=i_a} m_j \mu_j^{\gamma} \right\}_{i_a \in I_a} \tag{12}$$

y para $a \cup d = q_1,...,q_t$ $$\{SS_{i_a}^{d}\}_{i_a \in I_a} = \left\{ \sum_{j:j_a=i_a} m_j \left[\Sigma_j^{dd} + \mu_j^d \left(\mu_j^d\right)' \right] \right\}_{i_a \in I_a} \tag{13}$$

estas son conocidas como las ecuaciones de verosimilitud. Cuando los EMV existan, la solución será única[5] y satisfará todas las restricciones del modelo.

Dependiendo de que estemos o no ante un modelo descomponible, las EMV podrán obtenerse directamente o bien habrá que recurrir a un proceso iterativo para su resolución[6].

Una función de densidad será descomponible cuando pueda expresarse como el producto de varios factores, los cuales serán funciones de densidad obtenidas marginalizando o condicionando sobre la función de densidad conjunta. De este modo, si el modelo fuera descomponible, la distribución conjunta de la muestra, o conjunto de datos observados, podrá reducirse en componentes de dimensiones menores, simplificándose notablemente el problema de la estimación

Si el modelo no fuera descomponible habrá que recurrir a procedimientos iterativos como el algoritmo EPIM (escalamiento proporcional iterativo modificado) descrito por Frydenberg y Edwards (1989).

6. CONTRASTE DE LA BONDAD DE UN MODELO: LA DESVIANZA

La expresión de la desvianza, utilizada para contrastar la bondad de los modelos estimados, se puede escribir como:

$$\ln f(i,y) = \ln p_i - q\ln(2\pi)/2 - \ln|\Sigma_i|/2 - (y-\mu_i)'\Sigma_i^{-1}(y-\mu_i)/2$$

por lo que el logaritmo de la función de verosimilitud muestral $\left(i^{(k)}, y^{(k)}\right)$, $\mathrm{k} = 1,...,\mathrm{N}$ es

[5] Las condiciones generales de existencia y unicidad fueron desarrolladas por Glonek, Darroch, y Speed, (1988), para la familia lineal exponencial, siendo la familia GC en caso específico de estas.

[6] En relación a cuando existen o no las estimaciones directas, y su relación con el concepto de descomponibilidad, ver para más información Leimer (1989), Frydenberg y Lauritzen (1989).

$$l = \sum_i n_i \ln p_i - Nq \ln(\pi)/2 - \sum_i n_i \ln|\Sigma_i|/2 - \sum_{k=1}^N \left(y^{(k)} - \mu_{i^{(k)}}\right)' \Sigma_{i^{(k)}}^{-1} \left(y^{(k)} - \mu_{i^{(k)}}\right)/2$$

el último término puede simplificarse como

$$\sum_i \sum_{k:i^{(k)}=i} \left(y^{(k)} - \mu_{i^{(k)}}\right)' \Sigma_{i^{(k)}}^{-1} \left(y^{(k)} - \mu_{i^{(k)}}\right) = \sum_i \left\{ n_i tr\left(S_i \Sigma_i^{-1}\right) + n_i (\bar{y}_i - \mu_i)' \Sigma_i^{-1} (\bar{y}_i - \mu_i) \right\}$$

Por lo que una expresión alternativa del logaritmo de verosimilitud es:

$$l = \sum_i n_i \ln p_i - Nq \ln(2\pi)/2 - \sum_i n_i \ln|\Sigma_i|/2 - \\ - \sum_i n_i tr\left(S_i \Sigma_i^{-1}\right)/2 - \sum_i n_i (\bar{y}_i - \mu_i)' \Sigma_i^{-1} (\bar{y}_i - \mu_i)/2$$

Para el modelo saturado heterogéneo, a partir de las ecuaciones de verosimilitud (11-13) obtenemos que

$$\hat{p}_i = n_i / N, \ \left(\hat{\mathrm{m}}_\mathrm{i} = N\hat{p}_i\right), \hat{\mu}_\mathrm{i} = \bar{y}_i, \quad \text{y} \ \hat{\Sigma}_\mathrm{i} = S_i,$$

por lo que el máximo del logaritmo de verosimilitud es

$$\hat{l}_f = \sum_i n_i \ln(n_i / N) - Nq \ln(2\pi)/2 - \sum_i n_i \ln|S_i|/2 - Nq/2 \qquad (14)$$

Por tanto, si un modelo tiene EMV $\hat{p}_i, \hat{\mu}_\mathrm{i}, \hat{\Sigma}_\mathrm{i}$, su desvianza respecto al modelo saturado homogéneo será:

$$2\sum_i n_i \ln(n_i/\hat{m}_i) - \sum_i n_i \ln\left|S_i \hat{\Sigma}_i^{-1}\right| + \sum_i n_i \left\{ tr\left(S_i \hat{\Sigma}_i^{-1}\right) - q \right\} + \sum_i n_i (\bar{y}_i - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (\bar{y}_i - \hat{\mu}_i)$$

El modelo saturado homogéneo tiene estimadores:

$$\hat{p}_i = n_i / N, \ \left(\hat{\mathrm{m}}_\mathrm{i} = N\hat{p}_i\right), \ \hat{\mu}_\mathrm{i} = \bar{y}_i, \quad \hat{\Sigma} = \hat{\Sigma}_\mathrm{i} = S,$$

donde $S = \sum_i n_i S_i / N$, por lo que el logaritmo de máxima verosimilitud para este modelo será

$$\hat{l}_f^h = \sum_i n_i \ln(n_i / N) - Nq \ln(2\pi)/2 - N \ln|S|/2 - Nq/2 \qquad (15)$$

y la desvianza de un modelos homogéneo con estimadores de máxima verosimilitud $\hat{p}_i, \hat{\mu}_\mathrm{i}$, y $\hat{\Sigma}$ con respecto al modelo saturado homogéneo se simplifica a la siguiente expresión

$$2\sum_i n_i \ln(n_i/\hat{m}_i) - N \ln\left|\hat{S}\Sigma^{-1}\right| + N\left\{ tr\left(S\hat{\Sigma}^{-1}\right) - q \right\} + \sum_i n_i (\bar{y}_i - \hat{\mu}_i)' \hat{\Sigma}^{-1} (\bar{y}_i - \hat{\mu}_i)$$

Para dos modelos, *M_0 y M_1*, tales que $M_0 \subseteq M_1$, la diferencia de desvianzas sigue bajo M_0 una distribución asintótica χ^2 con unos grados de libertad dados por la diferencia entre los dos modelos del número de parámetros no restringidos.

Para obtener un resumen conciso de las fuentes de variación del conjunto de datos, podemos descomponer la desvianza lo que supone una generalización del análisis clásico de la varianza (Whittaker 1990).

BIBLIOGRAFÍA

DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980) Markov fields and log linear interaction models for contingency tables. *Ann. Stat.*, 8, 522-539

EDWARDS, D.E. (1.995) *Introduction to Graphical Modelling*. Springer-Verlag New York, Inc.

FRYDENBERG Y EDWARDS (1989). A modified iterative proportional scaling algorithm for estimation in regular exponential families. *Comput. Statist. Data anal. to apear*.

LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative.

WERMUTH, N and LAURITZEN, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical Caín models. *J. Roy. Statist. Soc*. B, 52, 1, 21-50.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons Ltd.

WRIGHT, S. (1921). Correlation and causation. *J. Agric. Res*., 20, 557-585.

ANEXO 1

PROPIEDADES DE MARKOV Y CRITERIO DE FACTORIZACIÓN

Las *propiedades de Markov* son tres: la propiedad de independencia global, la de independencia local y la de independencia por pares de variables. El Teorema de Separación establece que la propiedad de independencia por pares implica la propiedad global de Markov.

Propiedad de Markov para Pares de Variables: pares de variables no adyacentes en el grafo de independencia condicional son independientes condicionados al resto de variables

Propiedad Local de Markov: condicionando sólo sobre las variables adyacentes, cualquier variable es independiente de todas las restantes.

Propiedad Global de Markov: subconjuntos de variables cualesquiera separados por una tercera variable son condicionalmente independientes sólo sobre esta tercera variable.

Como ya se ha dicho, los grafos de independencia condicional se basan en la propiedad de Markov de independencia por pares. Las razones que justifican que así sea son en primer lugar que la definición por pares de la relación de independencia condicional se corresponde de un modo natural con la definición teórica gráfica del conjunto de arcos en un gráfico. Y en segundo lugar, que desde el punto de vista de las aplicaciones, el conjunto de requerimientos que se deben de verificar en la construcción de un grafo de independencia para un modelo, dado un conjunto de datos, es el menos riguroso, mientras que la interpretación será la más consistente.

La Propiedad Global de Markov proporciona un resumen preciso de los patrones de interacción existentes entre las variables aleatorias objeto de estudio. Y si aún queremos profundizar un poco más en las conexiones existentes entre el aspecto de un grafo concreto y las propiedades de la función de densidad a el asociado, descubriremos que aplicando el Teorema de Separación a un grafo podrá saberse cuando es posible descomponer la función de densidad en componentes de menor dimensión. Ni que decir tiene las grandes ventajas[7] que ello lleva aparejado.

*El Teorema de Separación dice que: s*i X_a, X_b y X_c son vectores que contienen subconjuntos disjuntos de variables de *X*, y si, en el grafo de

[7] Ejemplo: simplificación de las problemáticas objeto de estudio, reducción de la complejidad de la estimación.

independencia de X, cada vértice de b está separado de los vértices de c por el subconjunto[8] *a,* entonces $X_b \perp X_c | X_a$ **.**

La inversa de este teorema es inmediata: si es cierto que para cualquier a que separe i y j se cumple que $i \perp j | a$, entonces será cierto que para cualquier i y j no adyacentes, $i \perp j | resto$. De cualquier modo, este resultado es trivial porque el "resto" será siempre el conjunto separador para los vértices no adyacentes.

El Teorema de Separación es demostrado, entre otros[9], por Whittaker (1990), utilizando el Criterio de Factorización para la independencia condicional que se expondrá más abajo. Su aplicación queda restringida a variables aleatorias con una función de densidad probabilística conjunta, para poder aplicar la proposición de Independencia de bloques también expuesta a continuación.

Las demostraciones comienzan mostrando como en relación a un grafo, dado por un número finito de vértices, las tres propiedades de Markov son equivalentes.

Para finalizar este anexo expondremos el criterio de factorización y la proposición de independencia de bloques. Pero antes nos resulta imprescindible dar una definición de independencia condicional, a partir de la cual derivaremos dos proposiciones. La proposición 1 es el criterio de factorización. La proposición 2 se refiere a la independencia de bloques.

En adelante, se considerará el espacio probabilístico (Ω, A, P).

Definición

Las variables aleatorias Y y Z son condicionalmente independientes sobre X si y sólo si

$$f_{YZ/X}(y,z;x) = f_{Y/X}(y;x) \cdot f_{Z/X}(z;x)$$

para todos Y y Z y para todo X tales que $f_X(X) > 0$. Se denotará por $Y \perp Z | X$.

Proposición 1: Criterio de Factorización

Las variables aleatorias Y y Z son condicionalmente independientes dada X, $Y \perp Z | X$, si y sólo si existen las funciones g y h tales que

$f_{XYZ}(x,y,z) = g(x,y) \cdot h(x,z)$ para todo $y, z \in R$ y todo x con $f_X(x) > 0$.

Las funciones g y h no tienen por que ser únicas.

[8] Dos vértices, *i* y *j*, están separados por un subconjunto *a* si y sólo si todos los caminos que conecten los dos vértices pasan al menos a través de un miembro del subconjunto. Cuando los vértices, *i* y *j*, no están conectados no existirá ningún subconjunto separador.

[9] Entre quienes han demostrado este teorema destacamos a Speed y Kiiveri (1986) que lo hicieron para el caso concreto en que las funciones de densidad pertenezcan a la familia de la Distribución Normal Multivariante.

Proposición 2: Independencia de bloques

Si (X, Y, Z_1, Z_2) es un vector aleatorio, entonces las siguientes afirmaciones son equivalentes:

a) $Y \perp (Z_1, Z_2) | X$

b) $Y \perp Z_1 | (X, Z_2)$ *y* $Y \perp Z_2 | (X, Z_1)$