



CUADERNOS DE TRABAJO
ESCUELA UNIVERSITARIA DE ESTADÍSTICA

***CALCULANDO LA MATRIZ DE
COVARIANZAS CON LA ESTRUCTURA DE
UNA RED BAYESIANA GAUSSIANA***

Miguel A. Gómez-Villegas

Rosario Susi

Cuaderno de Trabajo número 03/2012



**UNIVERSIDAD
COMPLUTENSE
MADRID**

Los Cuadernos de Trabajo de la Escuela Universitaria de Estadística constituyen una apuesta por la publicación de los trabajos en curso y de los informes técnicos desarrollados desde la Escuela para servir de apoyo tanto a la docencia como a la investigación.

Los Cuadernos de Trabajo se pueden descargar de la página de la Biblioteca de la Escuela www.ucm.es/BUCEM/est/ y en la sección de investigación de la página del centro www.ucm.es/centros/webs/eest/

CONTACTO: Biblioteca de la E. U. de Estadística

Universidad Complutense de Madrid

Av. Puerta de Hierro, S/N

28040 Madrid

Tlf. 913944035

buc_est@buc.ucm.es

Los trabajos publicados en la serie Cuadernos de Trabajo de la Escuela Universitaria de Estadística no están sujetos a ninguna evaluación previa. Las opiniones y análisis que aparecen publicados en los Cuadernos de Trabajo son responsabilidad exclusiva de sus autores.

ISSN: 1989-0567

Calculando la matriz de covarianzas con la estructura de una red Bayesiana Gaussiana

Miguel A. Gómez-Villegas^a, Rosario Susi^b

^a *Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, 28040 Madrid, Spain*

^b *Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, 28040 Madrid, Spain*

Abstract

En este trabajo se introduce una fórmula recursiva que permite calcular la matriz de covarianzas de una red Bayesiana Gaussiana dados los parámetros de la especificación condicionada de la parte cuantitativa del modelo. Además, se determinan las varianzas y las covarianzas del problema considerando los distintos caminos que aparecen en el grafo que recoge la parte cualitativa de la red.

Se aplicarán los conceptos introducidos en este trabajo a un ejemplo real.

Keywords: Red Bayesiana Gaussiana, matriz de covarianzas, normal multivariante.

1. Introduction

Una red Bayesiana (RB) es un modelo gráfico probabilístico que se utiliza para estudiar un conjunto de variables con una estructura de dependencia conocida. Las RBs permiten estudiar problemas del mundo real en una serie

*Corresponding author: rsusi@estad.ucm.es

de ámbitos complejos, como puede ser el diagnóstico médico o de los sistemas dinámicos.

Las RBs han sido estudiadas, entre otros, por Pearl [4], Lauritzen [3], Cowell et al. [1] y Jensen et al. [2].

Una RB tiene dos componentes, como se muestra en la Definición 1. La parte cualitativa es un gráfico acíclico dirigido (DAG) que muestra la estructura de la dependencia entre las variables en estudio. La parte cuantitativa se compone de las distribuciones de probabilidad condicionadas asignadas a las variables del problema dados sus padres en el DAG.

Nuestro modelo objeto de estudio son las redes Bayesianas Gaussianas (RBGs) que se definen como una RB en la que la distribución de probabilidad conjunta es normal multivariante. En particular nuestro objetivo se centra en describir la matriz de covarianzas que determina uno de los parámetros de la parte cuantitativa de la red, dada la estructura del modelo. Como se verá posteriormente en las RBGs, la dependencia entre variables viene fijada por el coeficiente de regresión de cada una de las variables sobre sus padres en el DAG. El coeficiente de regresión junto con la varianza condicionada de cada una de las variables dados sus padres en el DAG, describen la variación del modelo cuando éste se presenta mediante la especificación condicional de la parte cuantitativa de la red.

En este trabajo se introduce una fórmula recursiva que nos permitirá obtener directamente la matriz de covarianzas dados los coeficientes de regresión y las varianzas condicionadas de cada una de las variables del problema dados sus padres en el DAG. Además se obtienen las varianzas y covarianzas teniendo en cuenta los caminos que aparecen en el DAG entre nodos.

La organización del presente trabajo es la siguiente. En la Sección 2 se introduce el modelo objeto de estudio, para ello se definirán inicialmente las RBs y posteriormente las RBGs, detallándose la especificación condicional y la especificación conjunta de la parte cuantitativa. En la Sección 3 se presenta la fórmula recursiva que permite calcular la matriz de covarianzas de la distribución conjunta del modelo y se obtienen las varianzas y covarianzas dados los caminos que se encuentran entre nodos. Una aplicación de las definiciones propuestas se puede ver en la Sección 4. Finalmente, en la Sección 5 se introducen unas breves conclusiones al trabajo presentado.

2. Redes Bayesianas y redes Bayesianas Gaussianas

En esta Sección se presentan las redes Bayesianas Gaussianas. Para ello se introduce inicialmente la definición de red Bayesiana.

Definition 1. *Una red Bayesiana (RB) se define como un par $(\mathcal{G}, \mathcal{P})$ donde \mathcal{G} es un grafo acíclico dirigido DAG, en el que los nodos representan a las variables aleatorias $\mathbf{X} = \{X_1, \dots, X_n\}$ y los arcos reflejan la estructura de dependencias y $\mathcal{P} = \{P(X_1|pa(X_1)), \dots, P(X_n|pa(X_n))\}$ es el conjunto de distribuciones de probabilidad condicionadas, de cada una de las variables X_k dados sus padres $pa(X_k)$ en \mathcal{G} , donde $pa(X_k) \subseteq \{X_1, \dots, X_{k-1}\}$.*

Como se puede observar, se tiene una ordenación para todas las variables del problema, pudiendo ser padre de un nodo X_k cualquiera de los nodos anteriores $\{X_1, \dots, X_{k-1}\}$ e hijos del nodo X_k cualquiera de los nodos posteriores $\{X_{k+1}, \dots, X_n\}$. En la Sección 3 se introducen algunas definiciones de conjuntos de nodos que forman parte de la familia de un nodo cualquiera, dado el DAG.

Las RBs se caracterizan por permitir obtener la distribución de probabilidad conjunta a partir de las probabilidades condicionadas de \mathcal{P} , de forma que

$$P(\mathbf{X}) = \prod_{k=1}^n P(X_k | pa(X_k)) \quad (1)$$

A continuación se definen las redes Bayesianas Gaussianas

Definition 2. *Se define una red Bayesiana Gaussiana (RBG) como una RB donde la distribución de probabilidad conjunta de $\mathbf{X} = (X_1, \dots, X_n)$ es una distribución normal multivariante $N_n(\mu, \Sigma)$ dada por*

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right\}$$

donde μ es el vector de medias n -dimensional y Σ es la matriz de varianzas-covarianzas $n \times n$ definida positiva.

Dada la expresión introducida en (1), en RBGs cada una de las probabilidades condicionadas de \mathcal{P} son normales, tales que

$$X_k | pa(X_k) \sim N_1(x_k | \mu_k + \sum_{j=1}^{k-1} \beta_{jk} (x_j - \mu_j), v_k)$$

donde μ_k es la media de X_k , β_{jk} son los coeficientes de regresión de X_k respecto a $X_j \in pa(X_k)$ y v_k es la varianza condicionada de $X_k | pa(X_k)$.

Los coeficientes β_{jk} miden el grado de relación entre X_j y X_k , por tanto, en RBGs se tiene un arco en el DAG siempre que $\beta_{jk} \neq 0$, esto es

$$\beta_{jk} = 0 \Leftrightarrow X_j \notin pa(X_k)$$

Con los parámetros de la especificación condicional, μ_k , v_k y β_{jk} para todo X_k , es posible determinar la distribución conjunta, o especificación conjunta del modelo, ya que los μ_k son los elementos del vector de medias n -dimensional μ , de forma que $\mu = (\mu_1, \dots, \mu_n)^T$ y las varianzas condicionadas, v_k , y los coeficientes de regresión, β_{jk} , permiten calcular Σ .

Shachter y Kenley [5] proponen la siguiente expresión para calcular la matriz de precisión

$$\Sigma^{-1} = (I - \mathbf{B})\mathbf{D}^{-1}(I - \mathbf{B})^T$$

donde \mathbf{D} es una matriz diagonal dada por $\mathbf{D} = \text{diag}(v_1, \dots, v_n)$ y \mathbf{B} es una matriz triangular superior formada por los coeficientes de regresión β_{jk} con $j = 1, \dots, k - 1$.

Cuando la red tiene muchas variables, es más eficaz computacionalmente calcular la matriz Σ^{-1} o la matriz Σ mediante una expresin recursiva, así, es posible determinar la matriz de precisión Σ^{-1} utilizando la siguiente fórmula recursiva, en la cual $W(k)$ representa la matriz superior izquierda $k \times k$ y β_k es el vector columna dado por los coeficientes de regresin β_{jk} con $j < k$

$$W(k+1) = \begin{pmatrix} W(k) + \frac{\beta_{k+1}\beta_{k+1}^T}{v_{k+1}} & \frac{-\beta_{k+1}}{v_{k+1}} \\ \frac{-\beta_{k+1}^T}{v_{k+1}} & \frac{1}{v_{k+1}} \end{pmatrix}$$

siendo $W(1) = \frac{1}{v_1}$.

La matriz de precisión $\Sigma^{-1} = \mathbf{W}$.

A continuación se introduce un ejemplo real de RBGs para ilustrar los distintos conceptos que se presentan a lo largo de este trabajo.

Example 1. Acuíferos

Supongamos que estamos estudiando el volumen de agua ($100Hm^3$) de un conjunto de acuíferos o almacenamientos de agua subterránea, $\mathbf{X} = \{X_1, \dots, X_9\}$, con filtraciones entre los mismos, como se puede ver en el DAG de la Figura 1.

El volumen de agua de cada acuífero se mide en $100Hm^3$ y se supone que el volumen de agua en cada acuífero se distribuye normalmente.

El volumen medio de agua de cada acuífero, así como la varianza condicionada v_k de X_k dados sus padres $pa(X_k)$ se recogen en la siguiente tabla

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
μ_k	1	2	1	2	3	2	4	3	2
v_k	1	1	1	3	3	2	2	2	2

Los coeficientes de regresión β_{jk} se muestran en el DAG.

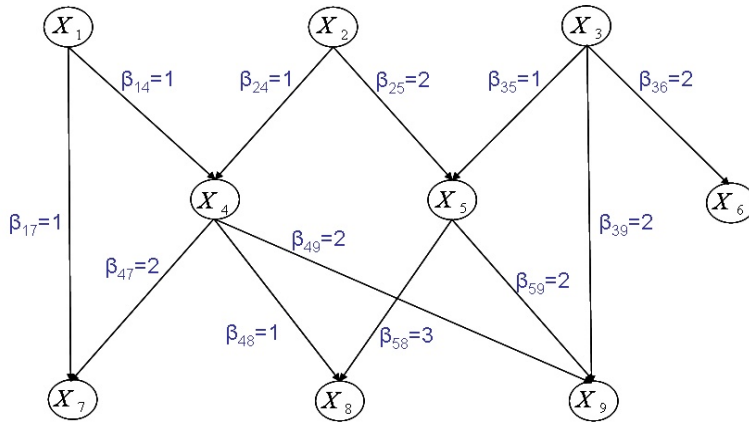


Figure 1: DAG que describe la estructura de los acuíferos

Con los parámetros de cada una de las variables X_k dados sus padres $pa(X_k)$, tenemos que:

$$\begin{aligned}
f(x_1) &\sim N_1(x_1 \mid \mu_1, v_1) \\
f(x_2) &\sim N_1(x_2 \mid \mu_2, v_2) \\
f(x_3) &\sim N_1(x_3 \mid \mu_3, v_3) \\
f(x_4 \mid x_1, x_2) &\sim N_1(x_4 \mid \mu_4 + \beta_{14}(x_1 - \mu_1) + \beta_{24}(x_2 - \mu_2), v_4) \\
f(x_5 \mid x_2, x_3) &\sim N_1(x_5 \mid \mu_5 + \beta_{25}(x_2 - \mu_2) + \beta_{35}(x_3 - \mu_3), v_5) \\
f(x_6 \mid x_3) &\sim N_1(x_6 \mid \mu_6 + \beta_{36}(x_3 - \mu_3), v_6) \\
f(x_7 \mid x_1, x_4) &\sim N_1(x_7 \mid \mu_7 + \beta_{17}(x_1 - \mu_1) + \beta_{47}(x_4 - \mu_4), v_7) \\
f(x_8 \mid x_4, x_5) &\sim N_1(x_8 \mid \mu_8 + \beta_{48}(x_4 - \mu_4) + \beta_{58}(x_5 - \mu_5), v_8) \\
f(x_9 \mid x_3, x_4, x_5) &\sim N_1(x_9 \mid \mu_9 + \beta_{39}(x_3 - \mu_3) + \beta_{49}(x_4 - \mu_4) + \beta_{59}(x_5 - \mu_5), v_9)
\end{aligned}$$

Por tanto, la RBG queda definida mediante su parte cualitativa dada por el DAG de la Figura 1 y su parte cuantitativa mediante la especificación condicional dada por los parámetros μ_k , v_k y β_{jk} para cada variable X_k :

$$\mu_1 = 1 \quad \mu_2 = 2 \quad \mu_3 = 1 \quad \mu_4 = 2 \quad \mu_5 = 3$$

$$\mu_6 = 2 \quad \mu_7 = 4 \quad \mu_8 = 3 \quad \mu_9 = 2$$

$$v_1 = 1 \quad v_2 = 1 \quad v_3 = 1 \quad v_4 = 3 \quad v_5 = 3$$

$$v_6 = 2 \quad v_7 = 2 \quad v_8 = 2 \quad v_9 = 2$$

$$\beta_{14} = 1 \quad \beta_{24} = 1 \quad \beta_{25} = 2 \quad \beta_{35} = 1 \quad \beta_{36} = 2$$

$$\beta_{17} = 1 \quad \beta_{47} = 2 \quad \beta_{48} = 1 \quad \beta_{58} = 3 \quad \beta_{39} = 2$$

$$\beta_{49} = 2 \quad \beta_{59} = 2$$

La especificación conjunta de la parte cuantitativa del modelo viene dada por los parámetros de la distribución conjunta de \mathbf{X} $N_n(\mu, \Sigma)$ tales que

$$\mu = (1, 2, 1, 2, 3, 2, 4, 3, 2)^T$$

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 3 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 & 0 & 2 & 7 & 6 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 3 & 4 \\ 1 & 1 & 0 & 5 & 2 & 0 & 11 & 11 & 14 \\ 0 & 2 & 1 & 2 & 8 & 2 & 4 & 26 & 22 \\ 0 & 0 & 2 & 0 & 2 & 6 & 0 & 6 & 8 \\ 3 & 2 & 0 & 11 & 4 & 0 & 27 & 23 & 30 \\ 1 & 7 & 3 & 11 & 26 & 6 & 23 & 91 & 80 \\ 2 & 6 & 4 & 14 & 22 & 8 & 30 & 80 & 82 \end{pmatrix}$$

3. Cálculo de la matriz de covarianzas

En esta Sección se introduce una fórmula recursiva que permite calcular la matriz de covarianzas dados los parámetros de la especificación condicional, v_k y β_{jk} para todo X_k con $X_j \in pa(X_k)$.

Posteriormente se introducirán algunas definiciones básicas de la teoría de grafos que nos permitirán calcular cada uno de los elementos de la matriz de covarianzas, es decir, las varianzas y covarianzas de las variables de la RBG, teniendo en cuenta la estructura que presenta el DAG de la red.

3.1. Cálculo de la matriz de covarianzas

Para determinar la matriz de covarianzas proponemos la siguiente fórmula recursiva.

Sea β_k el vector columna que recoge los coeficientes de regresión de todos los $pa(X_k)$ con X_k con $pa(X_k) \subseteq \{X_1, \dots, X_{k-1}\}$, es decir, $\beta_k = (\beta_{1k}, \dots, \beta_{k-1k})^T$, entonces

$$\Sigma(k+1) = \begin{pmatrix} \Sigma(k) & \Sigma(k)\beta_{k+1} \\ \beta_{k+1}^T \Sigma(k)^T & v_{k+1} + \beta_{k+1}^T \Sigma(k)\beta_{k+1} \end{pmatrix}$$

con $\Sigma(0) = v_1$.

La matriz de covarianzas es $\Sigma(n)$.

3.2. Cálculo de varianzas y covarianzas con la estructura del DAG

Las varianzas σ_{kk} y covarianzas σ_{jk} con $j < k$, se pueden calcular teniendo en cuenta los caminos del DAG entre los nodos X_j y X_k en el caso de las varianzas, y X_j y X_k para las covarianzas. Para ello es necesario introducir algunas definiciones básicas de la teoría de grafos, dado el DAG que describe la parte cualitativa, es decir, la estructura de dependencia de la RBG.

Así, se definen los siguientes conjuntos de nodos

- *ascendientes del nodo X_k , $as(X_k)$* : conjunto de padres de X_k y padres de padres de X_k , etc., de forma que $as(X_k) \subseteq \{X_1, \dots, X_{k-1}\}$, siendo $pa(X_k) \subseteq as(X_k)$.
- *descendientes del nodo X_k , $de(X_k)$* : conjunto de hijos de X_k e hijos de hijos de X_k , etc., de forma que $de(X_k) \subseteq \{X_{k+1}, \dots, X_n\}$. Los hijos del nodo X_k en el DAG, $hi(X_k)$, son tales que $hi(X_k) \subseteq de(X_k)$.
- *hermanos de X_k , $he(X_k)$* : conjunto de nodos con padres comunes a X_k .
- *familia de X_k , $fa(X_k)$* : conjunto de ascendientes, descendientes y hermanos del nodo X_k en el DAG, incluyendo el propio X_k , es decir,

$$fa(X_k) = \{X_k, as(X_k), de(X_k), he(X_k)\}$$

En el ejemplo de los acuíferos, serán ascendientes del nodo X_7 , los nodos X_1, X_2 y X_4 siendo $as(X_7) = \{X_1, X_2, X_4\}$ y descendientes del nodo X_2 , los nodos X_4, X_5, X_8 y X_9 siendo $de(X_2) = \{X_4, X_5, X_8, X_9\}$. Hermanos de X_4 son $he(X_4) = \{X_5, X_7\}$.

Para el nodo X_5 su familia es $fa(X_5) = \{X_2, X_3, X_4, X_5, X_6, X_8, X_9\}$ siendo los ascendientes $as(X_5) = \{X_2, X_3\}$, los descendientes $de(X_5) = \{X_8, X_9\}$ y los hermanos $he(X_5) = \{X_4, X_6, X_9\}$.

Otras definiciones de interés asociadas a los caminos de un grafo se introducen a continuación

- Se define un *camino del nodo X_k a X_l* con $X_k \in as(X_l)$ siendo $k < l$, como una secuencia de nodos de forma que existe un arco entre un nodo y el siguiente, es decir

$$X_k, \dots, X_p, X_{p+1}, \dots, X_l$$

con $\beta_{p,p+1} \neq 0$ para todo $p = k, \dots, l - 1$

- Se define un *camino del nodo X_k a X_l pasando por X_j* , con $X_j \in as(X_k) \subset as(X_l)$, como una sucesión dada por el camino del nodo X_j a X_k y el camino de X_j a X_l , es decir, dos secuencias de nodos de forma que en cada secuencia existe un arco entre un nodo y el siguiente, tal que

$$\begin{aligned}
& X_j, \dots, X_p, X_{p+1}, \dots, X_k \text{ y } X_j, \dots, X_q, X_{q+1}, \dots, X_l \\
& \text{con } \beta_{p,p+1} \neq 0 \text{ y } \beta_{q,q+1} \neq 0 \\
& \text{para todo } p = j, \dots, k-1 \text{ y } q = j, \dots, l-1
\end{aligned}$$

A continuación, definimos los pesos de los caminos como el producto de los coeficientes de regresión que aparecen en cada uno de los arcos que unen los nodos de un camino y también se define la suma de los pesos de los distintos caminos que hay entre un nodo y otro, así

- *Peso de un camino de X_k a X_l con $X_k \in as(X_l)$ y $k < l$, w_{kl}* : Producto de los coeficientes de regresión que unen los nodos del camino de X_k a X_l ,

$$w_{kl} = \prod_{p=k}^{l-1} \beta_{p,p+1}$$

siendo $X_p \in pa(X_{p+1})$

- *Suma de los pesos de los caminos de X_k a X_l , con $X_k \in as(X_l)$ siendo $k < l$, $S_{w_{kl}}$* :

$$S_{w_{kl}} = \sum w_{kl}$$

- *Peso de un camino del nodo X_k a X_l pasando por X_j , con $X_j \in as(X_k) \subset as(X_l)$, w_{kjl}* : Producto de los coeficientes de regresión que unen los nodos entre X_j y su descendiente X_k y entre X_j y su descendiente X_l , es decir

$$w_{kjl} = w_{jk} w_{jl}$$

siendo con $X_j \in as(X_k) \subset as(X_l)$ con $j < k \leq l$.

- *Suma de los pesos de los caminos de X_k a X_l pasando por X_j , con $X_j \in as(X_k) \subset as(X_l)$ siendo $j < k \leq l$, $S_{w_{kjl}}$:*

$$S_{w_{kjl}} = \sum w_{kjl}$$

Introducidos los conceptos anteriores, proponemos el cálculo de las varianzas y covarianzas de las variables del problema apoyándonos en dichas definiciones. Así, dado el DAG que describe la estructura de una RBG cuya parte cuantitativa se presenta mediante la especificación condicional de la red, las varianzas y covarianzas de las variables del problema se pueden calcular con las expresiones que se introducen en las definiciones 3 y 4 respectivamente.

Definition 3. *Se define la varianza de X_k como su covarianza condicionada v_k mas la suma de la varianza condicionada de cada uno de los ascendientes de X_k , v_j , por la suma de los pesos de los caminos de X_k a X_k pasando por sus ascendientes X_j (con $X_j \in as(X_k)$). Por tanto, para todo X_k con $k = 1, \dots, n$*

$$\sigma_{kk} = v_k + \sum_{X_j \in as(X_k)} v_j S_{w_{kjk}} \quad (2)$$

Definition 4. *Se define la covarianza entre X_k y X_l con $X_k \in as(X_l)$ como la varianza condicionada de X_k por la suma de los pesos de los caminos de X_k a X_l más la suma de la varianza condicionada de cada uno de lo ascendientes*

de X_k , v_j , por la suma de los pesos de los caminos de X_k a X_l pasando por X_j siendo $j < k < l$, esto es

$$\sigma_{kl} = v_k S_{w_{kl}} + \sum_{X_j \in as(X_k)} v_j S_{w_{kjl}} \quad (3)$$

Varianzas y covarianzas de la RBG quedan definidas a partir de las varianzas condicionadas de cada X_k dados sus padres en el DAG, v_k , y los coeficientes de regresión β_{jk} con $X_j \in pa(X_k)$ teniendo en cuenta los caminos entre los nodos del DAG,

4. Aplicación

En este apartado utilizamos las definiciones descritas en la Sección anterior para calcular una varianza, σ_{99} , y una covarianza, σ_{49} , para el ejemplo de los acuíferos, Ejemplo 1. Con figuras se ilustran los distintos caminos que se han de recorrer para calcular la varianza y covarianza de interés.

4.1. Cálculo de σ_{kk}

Para calcular la varianza de X_9 , σ_{99} se utiliza la siguiente expresión $\sigma_{99} = v_9 + \sum_{X_j \in as(X_9)} v_j S_{w_{9j9}}$. En la Figura 2 se muestra el DAG con el nodo X_9 en azul y en rojo los ascendientes del mismo $as(X_9) = \{X_1, \dots, X_5\}$.

En la Figura 3 se señalan todos los posibles caminos de X_9 a X_9 pasando por su ascendiente X_3 . Como se observa en la Figura, se tiene una sucesión dada por el camino del nodo X_3 a X_9 y el camino del nodo X_3 a X_9 , dadas

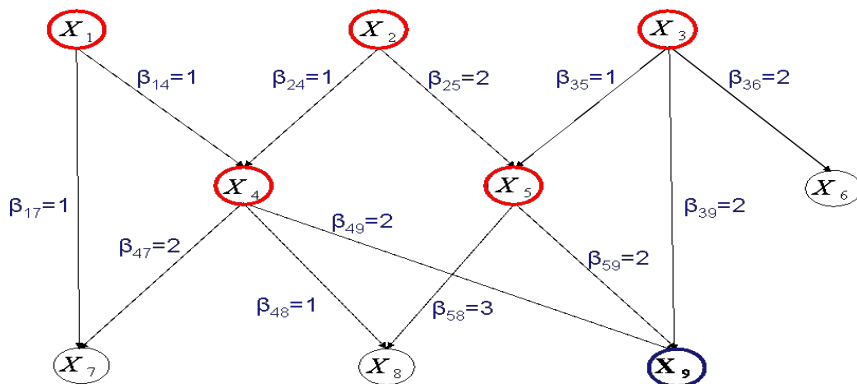


Figure 2: DAG Ejemplo 1. En azul X_9 y en rojo sus ascendientes $as(X_9)$

las direcciones que se muestran en el DAG. En realidad, lo que se tienen son los caminos de X_9 a X_3 y de X_3 a X_9 , pero teniendo en cuenta las direcciones del DAG y la definición de camino del nodo X_k a X_l pasando por X_j donde $k = l$ y $j < k$ la sucesión tenida en cuenta es la dada por el camino del nodo X_3 a X_9 y el camino del nodo X_3 a X_9 . Como en el cálculo de las varianzas y covarianzas lo que se tiene en cuenta es el peso de los caminos, es decir el producto de los β_{jk} que aparecen entre los nodos del camino, en el resto de figuras se obvia la dirección de las aristas y se reflejan los caminos como si se fuera del nodo X_k a su ascendiente X_j y se volviera de X_j a X_l . Como se ha comentado anteriormente esto no modifica ni altera ningún cálculo, simplemente facilita la visión de los caminos.

Detallando cada uno de los posibles caminos de ida y vuelta a X_9 pasando por su ascendiente X_3 y obviando las direcciones del DAG, en las Figuras de la Tabla 1, se puede calcular el peso de cada uno de los caminos de X_9

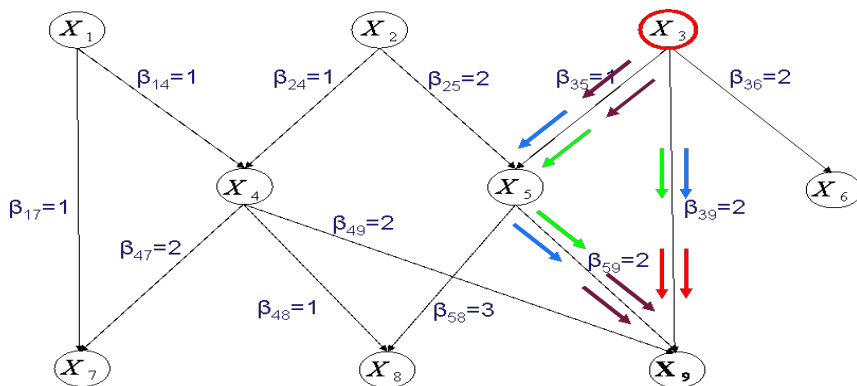


Figure 3: DAG Ejemplo 1. Caminos de X_9 a X_3

a X_9 pasando por su ascendente X_3 como el producto de los coeficientes de regresión de los arcos del grafo.

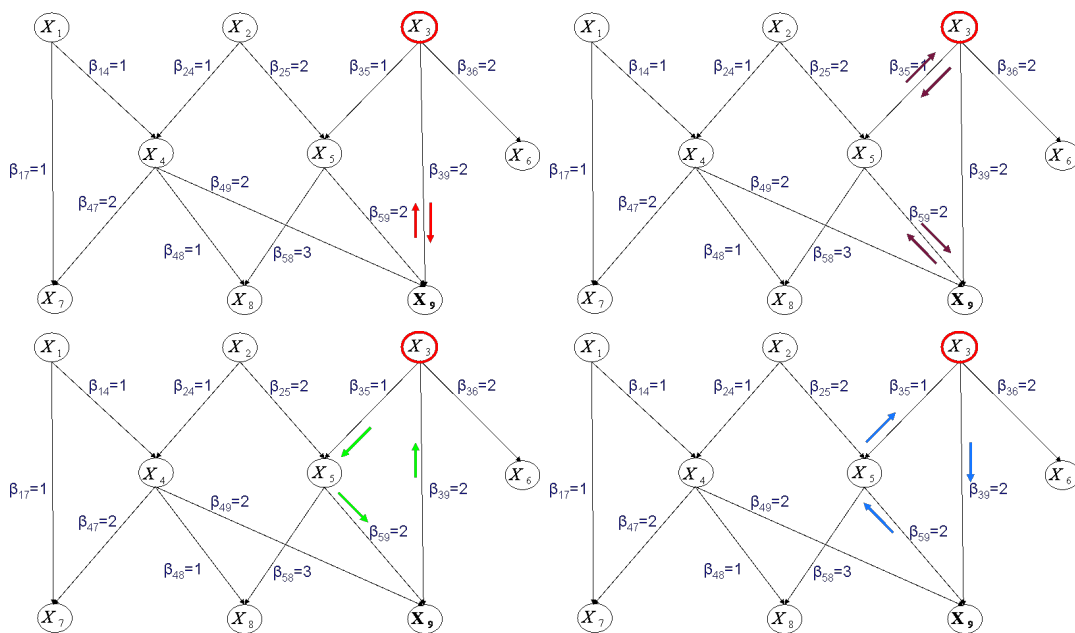


Table 1: DAG Ejemplo 1. Caminos detallados de X_9 a X_9 pasando por su ascendente X_3

En las Figuras 9, 5 y 6 se muestran todos los posibles caminos de ida y vuelta a X_9 pasando por sus ascendientes X_4 , X_5 y X_1 respectivamente. En todos los casos se obvian las direcciones del DAG.

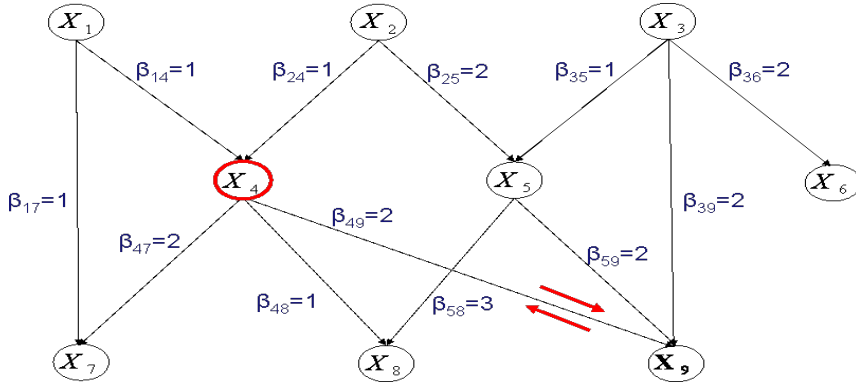


Figure 4: DAG Ejemplo 1. Caminos de X_9 a X_9 pasando por su ascendiente X_4

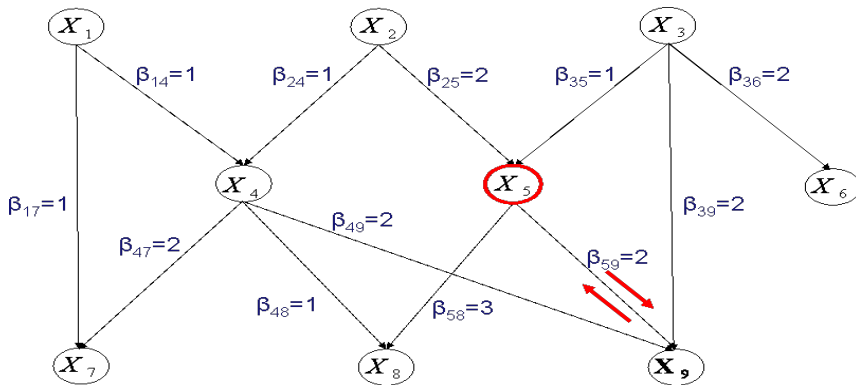


Figure 5: DAG Ejemplo 1. Caminos de X_9 a X_9 pasando por su ascendiente X_5

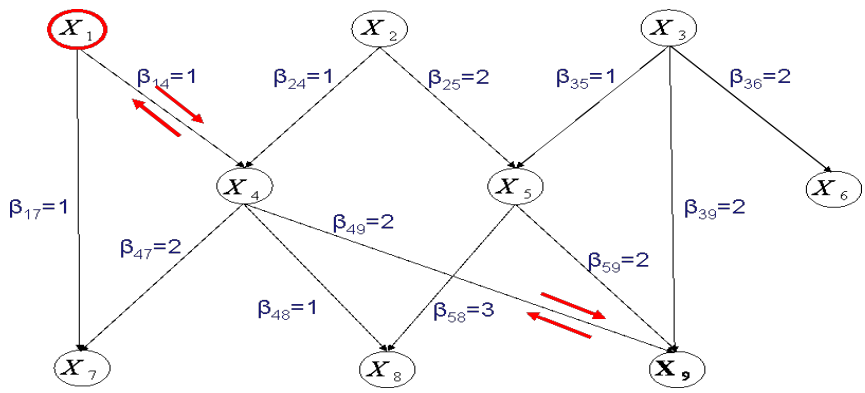


Figure 6: DAG Ejemplo 1. Caminos de X_9 a X_9 pasando por su ascendiente X_1

Finalmente en la Figura 7 se describen los distintos caminos encontrados para ir de X_9 a X_9 pasando por su ascendiente X_2 . Las Figuras recogidas en la Tabla 2 muestran cada uno de los caminos descritos en la Figura 7 de forma desglosada y sin tener en cuenta las direcciones del DAG.

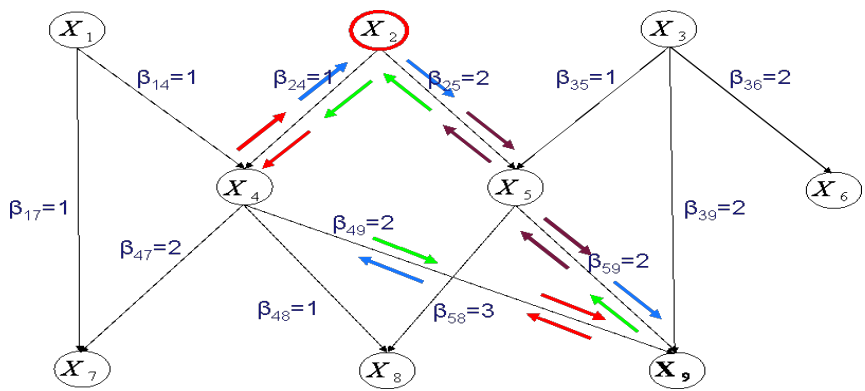


Figure 7: DAG Ejemplo 1. Caminos de X_9 a X_9 pasando por su ascendiente X_2

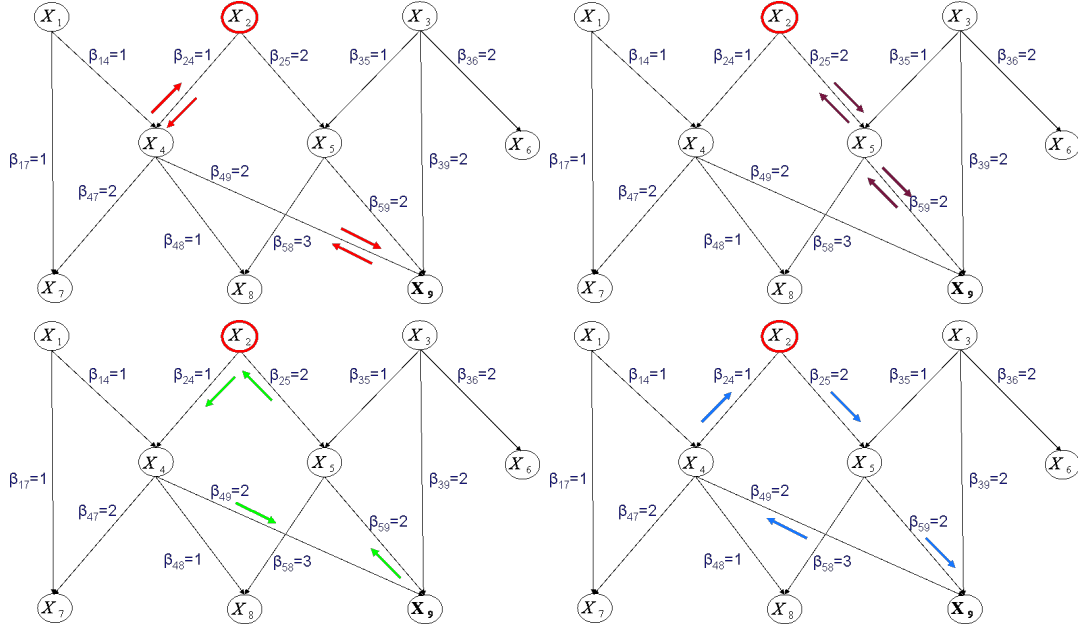


Table 2: DAG Ejemplo 1. Caminos detallados de X_9 a X_9 pasando por su ascendiente X_2

Teniendo en cuenta los pesos de los caminos w_{9j9} con $X_j \in as(X_9)$, la varianza de X_9 , σ_{99} , calculada con la expresión introducida en (3) es tal que

$$\sigma_{99} = v_9 + \sum_{X_j \in as(X_9)} v_j S_{w_{9j9}} = 82$$

4.2. Cálculo de σ_{kl}

Para el cálculo de una covarianza, por ejemplo σ_{49} , primero se ha de tener en cuenta los caminos de X_4 a X_9 y posteriormente los caminos de X_4 a X_9 pasando por los ascendientes de X_4 . En la Figura 8 se muestran en azul los nodos X_4 y X_9 para resaltar inicialmente los caminos de X_4 a X_9 y en rojo

se muestran los ascendientes de X_4 para determinar los caminos de X_4 a X_9 pasando por $\{X_1, X_2\} \in as(X_4)$.

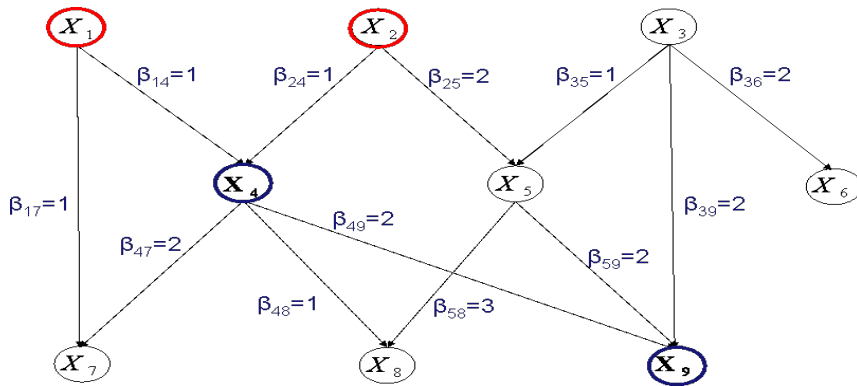


Figure 8: DAG Ejemplo 1. En azul X_4 y X_9 y en rojo los ascendientes de X_4

En la Figura 9 se muestra el único camino existente de X_4 a X_9 .

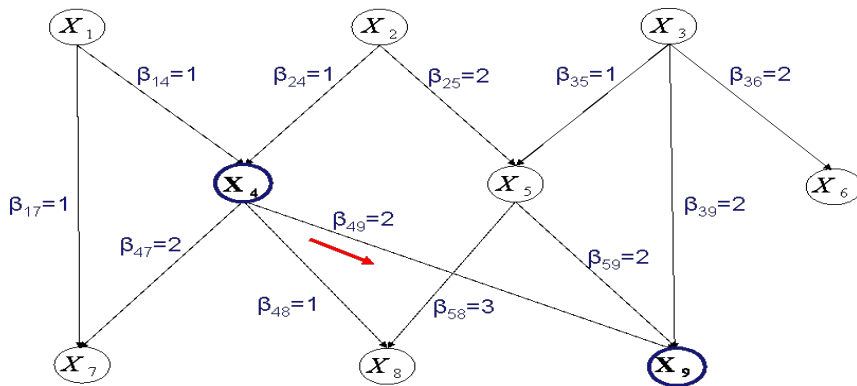


Figure 9: DAG Ejemplo 1. Caminos de X_4 a X_9

En las siguientes Figuras, se describen los distintos caminos de X_4 a X_9 pasando por los ascendientes de X_4 , así en la Figura 10 se muestra el único camino de X_4 a X_9 pasando por X_1 .

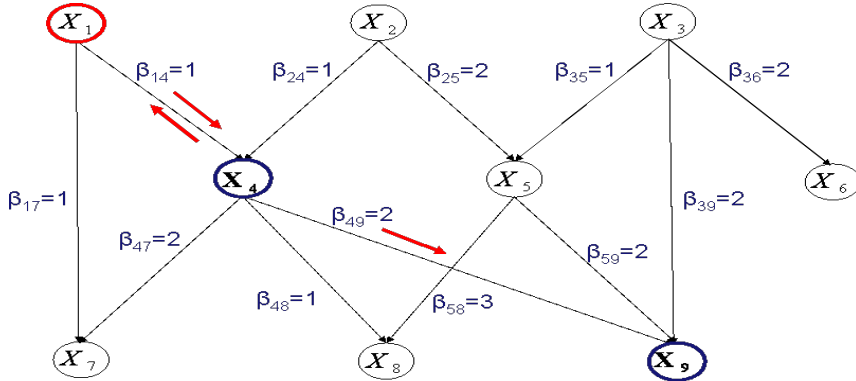


Figure 10: DAG Ejemplo 1. Caminos de X_4 a X_9 pasando por $X_1 \in as(X_4)$

Como los únicos ascendientes de X_4 son los nodos X_1 y X_2 , en la Tabla 3 se muestran los distintos caminos de X_4 a X_9 pasando por el ascendiente $X_2 \in as(X_4)$. Tres son los caminos encontrados para hacer el recorrido de interés tal y como se puede ver en la Tabla 3.

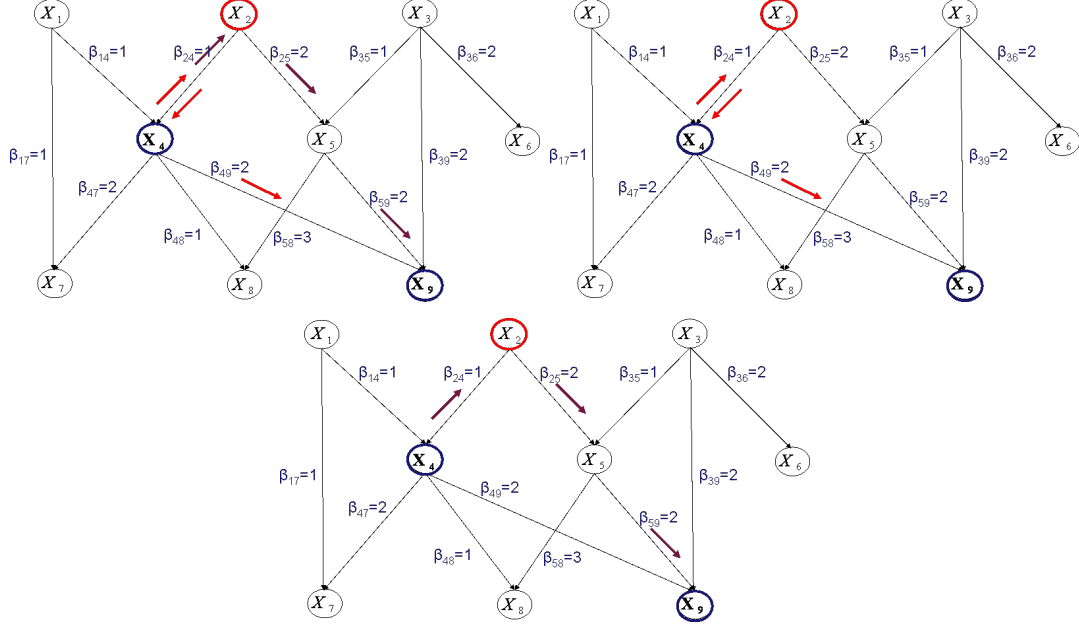


Table 3: DAG Ejemplo 1. Caminos detallados de X_4 a X_9 pasando por $X_2 \in as(X_4)$

Teniendo en cuenta los pesos de los caminos w_{49} y w_{4j9} con $X_j \in as(X_4)$, la covarianza entre X_4 y X_9 , σ_{49} , calculada mediante la expresión (4) es

$$\sigma_{49} = v_4 S_{w_{49}} + \sum_{X_j \in as(X_4)} v_j S_{w_{4j9}} = 14$$

5. Conclusiones y futuros trabajos

En este trabajo se introduce una fórmula recursiva que permite calcular la matriz de covarianzas que describe de forma conjunta la dispersión de la parte cuantitativa de una RBG. Mediante esta expresión es posible determinar la

matriz de covarianzas dados los parámetros que describen la dispersión en la especificación condicional de la red.

Apoyándonos en algunas definiciones de teoría de grafos es posible definir la varianza de una variable en función de los caminos que se muestran en el grafo, de igual forma es posible definir una covarianza entre dos variables. Esto nos permitirá en futuros trabajos estudiar la sensibilidad de una RBG cuando se eliminan arcos en el grafo, cambiándose así los caminos posibles de un nodo a el mismo o a otro nodo.

Un ejemplo real permite la aplicación de los conceptos introducidos en el trabajo y sirve como ilustración de las definiciones propuestas apoyadas en los caminos encontrados en el DAG.

6. Agradecimientos

Este trabajo ha sido posible gracias al Ministerio de Ciencia e Innovación de España, MTM 2008-03282 y a BSCH-UCM GR58/08-A, 910395 - Métodos Bayesianos.

- [1] Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ.1999. *Probabilistic Networks and Expert Systems*. Barcelona: Springer.
- [2] Jensen FV, Nielsen T. 2007. *Bayesian Networks and Decision Graphs.s*, 6th ed. New York : Springer.
- [3] Lauritzen SL. 1996. *Graphical Models*. Oxford: Clarendon Press.
- [4] Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* Morgan Kaufmann, San Mateo, CA.

- [5] Shachter R, Kenley C. 1989. Gaussian influence diagrams. *Management Science*. **35**, 527–550.



Cuadernos de Trabajo

Escuela Universitaria de Estadística

-
- CT02/2012** **What's new and useful about chaos in economic science.**
Andrés Fernández Díaz, Lorenzo Escot and Pilar Grau-Carles
- CT01/2012** **A social capital index**
Enrique González-Arangüena, Anna Khmelnitskaya, Conrado Manuel, Mónica del Pozo
- CT04/2011** **La metodología del haz de rectas para la comparación de series temporales.**
Magdalena Ferrán Aranz
- CT03/2011** **Game Theory and Centrality in Directed Social Networks**
Mónica del Pozo, Conrado Manuel, Enrique González-Arangüena y Guillermo Owen.
- CT02/2011** **Sondeo de intención de voto en las elecciones a Rector de la Universidad Complutense de Madrid 2011**
L. Escot, E. Ortega Castelló y L. Fernández Franco (coords)
- CT01/2011** **Juegos y Experimentos Didácticos de Estadística y Probabilidad**
G. Cabrera Gómez y M^a.J. Pons Bordería
- CT04/2010** **Medio siglo de estadísticas en el sector de la construcción residencial**
M. Ferrán Aranz
- CT03/2010** **Sensitivity to hyperprior parameters in Gaussian Bayesian networks.**
M.A. Gómez-Villegas, P. Main, H. Navarro y R. Susi
- CT02/2010** **Las políticas de conciliación de la vida familiar y laboral desde la perspectiva del empleador. Problemas y ventajas para la empresa.**
R. Albert, L. Escot, J.A. Fernández Cornejo y M.T. Palomo
- CT01/2010** **Propiedades exóticas de los determinantes**
Venancio Tomeo Perucha
- CT05/2009** **La predisposición de las estudiantes universitarias de la Comunidad de Madrid a auto-limitarse profesionalmente en el futuro por razones de conciliación**
R. Albert, L. Escot y J.A. Fernández Cornejo
- CT04/2009** **A Probabilistic Position Value**
A. Ghintran, E. González-Arangüena y C. Manuel
- CT03/2009** **Didáctica de la Estadística y la Probabilidad en Secundaria: Experimentos motivadores**
A. Pajares García y V. Tomeo Perucha
- CT02/2009** **La disposición entre los hombres españoles a tomarse el permiso por nacimiento. ¿Influyen en ello las estrategias de conciliación de las empresas?**
L. Escot, J.A. Fernández-Cornejo, C. Lafuente y C. Poza
- CT01/2009** **Perturbing the structure in Gaussian Bayesian networks**
R. Susi, H. Navarro, P. Main y M.A. Gómez-Villegas

- CT09/2008** **Un experimento de campo para analizar la discriminación contra la mujer en los procesos de selección de personal**
L. Escot, J.A. Fernández Cornejo, R. Albert y M.O. Samamed
- CT08/2008** **Laboratorio de Programación. Manual de Mooshak para el alumno**
D. I. de Basilio y Vildósola, M. González Cuñado y C. Pareja Flores
- CT07/2008** **Factores de protección y riesgo de infidelidad en la banca comercial**
J. M^a Santiago Merino
- CT06/2008** **Multinationals and foreign direct investment: Main theoretical strands and empirical effects**
María C. Latorre
- CT05/2008** **On the Asymptotic Distribution of Cook's distance in Logistic Regression Models**
Nirian Martín y and Leandro Pardo
- CT04/2008** **La innovación tecnológica desde el marco del capital intelectual**
Miriam Delgado Verde, José Emilio Navas López, Gregorio Martín de Castro y Pedro López Sáez
- CT03/2008** **Análisis del comportamiento de los indecisos en procesos electorales: propuesta de investigación funcional predictivo-normativa**
J. M^a Santiago Merino
- CT02/2008** **Inaccurate parameters in Gaussian Bayesian networks**
Miguel A. Gómez-Villegas, Paloma Main and Rosario Susi
- CT01/2008** **A Value for Directed Communication Situations.**
E. González-Arangüena, C. Manuel, D. Gómez, R. van den Brink



UNIVERSIDAD COMPLUTENSE
MADRID