

Especificación de distribuciones condicionadas de variables continuas en redes Bayesianas *

Barry Cobb

School of Business
University of Kansas
1300 Sunnyside Ave. Summerfield Hall
Lawrence, KS 66045-7585
USA
brcobb@ku.edu

Rafael Rumí, Antonio Salmerón

Dept. de Estadística y Mat. Aplicada
Universidad de Almería
Carretera Sacramento s/n
La Cañada de San Urbano 04120, Almería
Spain
{rrumi, Antonio.Salmeron}@ual.es

Resumen

El modelo MTE (mixtura de exponenciales truncadas) se introdujo como una solución general al problema de especificar distribuciones condicionadas para las variables continuas de una red Bayesiana, especialmente como una alternativa a la discretización. En este trabajo comparamos el comportamiento de dos enfoques diferentes a la hora de construir modelos MTE condicionales en un ejemplo tomado del campo empresarial, que es un dominio en el cual las variables normalmente tienen distribuciones condicionadas continuas.

1. Introducción

Una red Bayesiana es un modelo de un dominio con incertidumbre, el cual incluye distribuciones de probabilidad condicionada en su representación numérica. Los métodos para modelar la función de densidad condicionada de las variables continuas de una red Bayesiana incluyen aproximaciones discretas y modelos condicionales lineales Gaussianos (CLG) [5]. Modelar funciones de densidad continuas en redes Bayesianas usando un método que dé lugar a soluciones tratables y en forma cerrada es un problema de investigación en curso.

Recientemente, los potenciales mixturas de exponenciales truncadas (MTE) [6] se introdujeron como una alternativa a la discretización para representar variables continuas en redes Bayesianas. Moral *et al.* [8] sugieren una estructura de árbol mixto para el aprendizaje y la representación de potenciales MTE condicionales. Cobb y Shenoy [2] proponen una serie de operaciones para la inferencia en redes Bayesianas continuas donde las variables pueden ser funciones lineales determinísticas de sus padres y las funciones de densidad se aproximan mediante potenciales MTE. Este enfoque también puede ser implementado para representar la función de densidad condicionada de una variable continua, como demostramos en este trabajo.

El artículo compara los resultados obtenidos usando los cuatro métodos mencionados anteriormente para modelar las funciones de densidad condicionadas de las variables continuas de una red Bayesiana. El resto del artículo se organiza como sigue. En la Sección 2 establecemos la notación y damos la definición del modelo MTE. La Sección 3 contiene la descripción de los modelos de distribución condicionada usados en este trabajo y la Sección 4 informa de la comparación de estos modelos en un ejemplo econométrico. El artículo finaliza con las conclusiones en la Sección 5.

*Este trabajo ha sido subvencionado por el Ministerio de Ciencia y Tecnología, proyecto TIC2001-2973-C05-02 y por fondos FEDER

2. Notación y definiciones

Las variables aleatorias se denotarán mediante letras mayúsculas, e.g., A, B, C . Los conjuntos de variables se denotarán mediante letras mayúsculas resaltadas, e.g., \mathbf{X} . Todas las variables se supone que toman valores en espacios de estados continuos. Si \mathbf{X} es un conjunto de variables, \mathbf{x} es una configuración específica de estados de esas variables. El espacio de estados continuo de \mathbf{X} se denota mediante $\Omega_{\mathbf{X}}$. Los potenciales MTE se denotan mediante letras griegas minúsculas.

En las representaciones gráficas, los nodos discretos se representan mediante óvalos de un borde, los nodos continuos se representan mediante óvalos de doble borde y los nodos que son funciones determinísticas de sus padres se representan mediante óvalos de triple borde.

Un potencial mixtura de exponenciales truncadas (MTE) [6, 9] tiene la siguiente definición.

Definición 1: (Potencial MTE)

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una variable aleatoria n -dimensional. Una función $\phi : \Omega_{\mathbf{X}} \mapsto \mathbb{R}^+$ es un potencial MTE si una de las dos siguientes condiciones se cumple:

1. El potencial ϕ se puede escribir como

$$\phi(\mathbf{x}) = a_0 + \sum_{i=1}^m a_i \exp \left\{ \sum_{j=1}^n b_i^{(j)} x_j \right\} \quad (1)$$

para todo $\mathbf{x} \in \Omega_{\mathbf{X}}$, donde $a_i, i = 0, \dots, m$ y $b_i^{(j)}, i = 1, \dots, m, j = 1, \dots, n$ son números reales.

2. El dominio de las variables, $\Omega_{\mathbf{X}}$, se particiona en hipercubos $\{\Omega_{\mathbf{X}_1}, \dots, \Omega_{\mathbf{X}_k}\}$ tales que ϕ se define como

$$\phi(\mathbf{x}) = \phi_i(\mathbf{x}) \quad \text{si } \mathbf{x} \in \Omega_{\mathbf{X}_i}, \quad i = 1, \dots, k,$$

donde cada $\phi_i, i = 1, \dots, k$ se puede escribir en la forma de la ecuación (1) (i.e. cada ϕ_i es un potencial MTE sobre $\Omega_{\mathbf{X}_i}$).

En la definición anterior, k es el número de trozos y m es el número de términos exponenciales en cada trozo del potencial MTE. Nos referiremos a ϕ_i como el trozo i -ésimo del potencial MTE ϕ y a $\Omega_{\mathbf{X}_i}$ como la porción del dominio de \mathbf{X} aproximado por ϕ_i . En este trabajo todos los potenciales valen cero para regiones no especificadas.

Un potencial MTE f es una *densidad MTE* para \mathbf{X} si integra uno sobre el dominio de \mathbf{X} . En una red Bayesiana, se pueden encontrar dos tipos de funciones de densidad: densidades marginales para los nodos raíz y densidades condicionadas para el resto de nodos. Una *densidad MTE condicionada* $f(x|\mathbf{y})$ es un potencial MTE $f(x, \mathbf{y})$ tal que tras fijar \mathbf{y} a cada uno de sus posibles valores, la función resultante es una densidad para X .

3. Modelos de distribuciones condicionadas

3.1. Aproximaciones discretas

La discretización de distribuciones continuas permite realizar inferencia aproximada en una red Bayesiana con variables continuas. Discretizar variables aleatorias continuas es equivalente a aproximar funciones de densidad (DF) mediante mixturas de distribuciones uniformes. Discretizar con un número pequeño de estados puede conducir a una baja precisión, mientras que discretizar con un número grande de estados puede conducir a un esfuerzo computacional excesivo. Kozlov y Koller [4] mejoran la precisión de la discretización mediante el uso de una partición no uniforme de todas las variables representadas por una distribución y ajustando la discretización para la evidencia. Sin embargo, el aumento de precisión requiere un algoritmo iterativo y aún es problemático para variables continuas cuya DF marginal a posteriori pueda variar ampliamente en función de la evidencia de otras variables relacionadas.

Sun y Shenoy [10] estudian la discretización en redes Bayesianas donde las colas de las distribuciones son especialmente importantes. Encuentran que aumentando el número de

estados durante la discretización siempre aumenta la precisión de la solución; sin embargo, utilizando distribuciones continuas sin discretizar en este contexto proporciona una mejor solución que la mejor aproximación discreta.

3.2. Modelo condicional lineal Gaussiano (CLG)

Sea X un nodo continuo en una red Bayesiana híbrida, sean $\mathbf{Y} = (Y_1, \dots, Y_d)$ sus padres discretos, y sean $\mathbf{Z} = (Z_1, \dots, Z_c)$ sus padres continuos. Los potenciales *condicionales lineales Gaussianos (CLG)* [5] en redes Bayesianas híbridas tienen la forma

$$\mathcal{L}(X | \mathbf{y}, \mathbf{z}) \sim N(w_{\mathbf{y},0} + \sum_{i=1}^c w_{\mathbf{y},i} z_i, \sigma_{\mathbf{y}}^2), \quad (2)$$

donde \mathbf{y} y \mathbf{z} son una combinación de estados continuos y discretos de los padres de X . En esta fórmula, $\sigma_{\mathbf{y}}^2 > 0$, $w_{\mathbf{y},0}$ y $w_{\mathbf{y},i}$ son números reales, y $w_{\mathbf{y},i}$ se define como el i -ésimo componente de un vector de la misma dimensión que la parte continua \mathbf{Z} de las variables padre. Esto supone que la media de un potencial depende linealmente de las variables padre continuas y que la varianza no depende de las variables padre continuas. Para cada configuración de los padres discretos de la variable X , se especifica una función lineal de los padres continuos como la media de la distribución condicionada de X dados sus padres, y se especifica un número real positivo para la varianza de la distribución de X dados sus padres. Los modelos CLG no pueden incorporar variables aleatorias continuas cuya distribución no sea Gaussiana a menos que dicha distribución se aproximada mediante una mixtura de Gaussianas.

3.3. Árboles mixtos de probabilidad

Una densidad condicionada se puede aproximar mediante un potencial MTE usando *árbol de probabilidad mixto* o árbol mixto para acortar. La definición formal es como sigue:

Definición 2: (Árbol mixto)

Decimos que un árbol \mathcal{T} es un *árbol mixto* si se cumplen las siguientes condiciones:

- i. Cada nodo interno representa una variable aleatoria (discreta o continua).
- ii. Cada arco saliente de una variable continua Z se etiqueta con un intervalo de valores de Z , de tal forma que el dominio de Z es la unión de los intervalos correspondientes a los arcos salientes de Z .
- iii. Cada variable discreta tiene un número de arcos salientes igual a su número de estados.
- iv. Cada nodo hoja contiene un potencial MTE definido sobre las variable en el camino de la raíz a esa hoja.

Los árboles mixtos pueden representar potenciales MTE definidos por partes. Cada rama completa en el árbol determina una sub-región del espacio donde el potencial está definido, y la función almacenada en la hoja de una rama es la definición del potencial en la correspondiente sub-región. En [7] se propone un método para aproximar densidades condicionadas mediante árboles mixtos. Se basa en ajustar una densidad MTE en cada hoja del árbol mixto. Por ejemplo, el árbol mixto de la Figura 1 representa la siguiente densidad condicionada $f(x|y)$:

$$\phi(x, y) = \begin{cases} 2,32 \exp\{-2x\} & \text{si } 0 \leq y \leq 1, \quad 0 \leq x \leq 1 \\ 3,33 \exp\{-0,3x\} & \text{si } 0 \leq y \leq 1, \quad 1 < x \leq 1,5 \\ 2 \exp\{-2x\} & \text{si } 1 < y \leq 2, \quad 0 \leq x \leq 1,5 \end{cases} \quad (3)$$

3.4. Relaciones lineales determinísticas

Cobb y Shenoy [2] describen operaciones para la inferencia en redes Bayesiana continuas con variables lineales determinísticas. Dado que la DF conjunta para las variables en una red Bayesiana continua con variables determinísticas no existe, estas operaciones se derivan del método de las convoluciones de la teoría de la probabilidad.

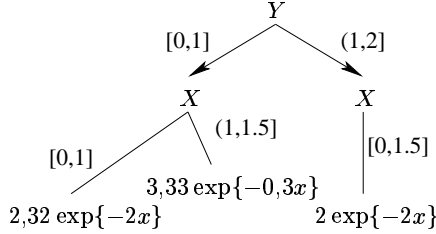


Figura 1: Un árbol de probabilidad mixto que representa el potencial ϕ de la ecuación (3).



Figura 2: La red Bayesiana utilizada en la demostración de las operaciones para relaciones lineales determinísticas.

Consideremos la red Bayesiana de la Figura 2. La variable X tiene una DF representada mediante el potencial MTE, $\phi(x) = 1,287760 - 0,116345 \exp\{1,601731x\}$, donde $\Omega_X = \{x : x \in [0, 1]\}$. La variable Z es una variable aleatoria normal standar, i.e. $\mathcal{L}(Z) \sim N(0, 1)$, representada mediante la 2-trozos, 3-términos aproximación MTE a la DF normal definida en [1] y denotada mediante φ . La variable Y es una función determinística de X y Z , y esta relación se representa mediante la función de masa condicional (CMF), $\alpha(x, y, z) = p_{Y|\{x,z\}}(y) = \mathbf{1}\{y = 3x + z + 2\}$, donde $\mathbf{1}\{A\}$ es la función indicador del evento A .

La DF conjunta para $\{X, Z\}$ es un potencial 2-trozos MTE, definido como

$$\vartheta(x, z) = (\phi \otimes \varphi)(x, z) = \begin{cases} \phi(x) \cdot \varphi_1(z) & \text{si } (-3 \leq z < 0) \cap (0 \leq x \leq 1) \\ \phi(x) \cdot \varphi_2(z) & \text{si } (0 \leq z \leq 3) \cap (0 \leq x \leq 1) \end{cases},$$

donde φ_1 y φ_2 son el primer y el segundo trozo del potencial MTE φ . El símbolo ' \otimes ' denota multiplicación puntual de funciones. La DF conjunta sin normalizar para $\{Y, Z\}$ se obtiene transformando la DF para $\{X, Z\}$ como sigue:

$$\theta(y, z) = \begin{cases} \phi((y - z - 2)/3) \cdot \varphi_1(z) & \text{si } (-3 \leq z < 0) \cap \\ & \cap (0 \leq (y - z - 2)/3 \leq 1) \\ \phi((y - z - 2)/3) \cdot \varphi_2(z) & \text{si } (0 \leq z \leq 3) \cap \\ & \cap (0 \leq (y - z - 2)/3 \leq 1) \end{cases}.$$

Esta es una operación de marginalización donde X se elimina de la combinación de ϑ y α y se denota mediante $\theta = (\vartheta \otimes \alpha)^{-X}$. La función θ sigue siendo un potencial MTE porque la función sustituida por x en θ es lineal en Y y Z . La DF marginal sin normalizar para Y se obtiene integrando el potencial MTE para $\{Y, Z\}$ sobre el dominio de Z como sigue:

$$\eta(y) = \theta^{-Z}(y) = \begin{cases} \int_{-3}^{y-2} \theta_1(y, z) dz & \text{si } -1 \leq y < 2 \\ \int_{y-5}^0 \theta_1(y, z) dz + \int_0^{y-2} \theta_2(y, z) dz & \text{si } 2 \leq y < 5 \\ \int_{y-5}^3 \theta_2(y, z) dz & \text{si } 5 \leq y \leq 8 \end{cases}.$$

Las variables Y y Z son dependientes, de modo que los límites de integración se definen de tal manera que la función se integre sobre el dominio conjunto de Y y Z . El resultado de la operación es un potencial MTE, excepto por el término lineal en el primer y tercer trozo. Este término lineal se reemplaza por un potencial MTE de tal modo que las densidades de la red Bayesiana permanezcan en la clase de potenciales MTE (para más detalles, ver [3]). Las operaciones mencionadas anteriormente se extienden con nueva notación para modelar relaciones lineales determinísticas en redes Bayesianas híbridas en [3].

4. Un ejemplo y comparación

En esta sección comparamos los cuatro métodos descritos en la Sección 3 utilizando para

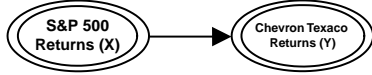


Figura 3: La red Bayesiana del ejemplo para la cotización de Chevron-Texaco.

ello un ejemplo tomado de un modelo económico. Consideremos el modelo de la Figura 3 donde la variación diaria en la cotización de Chevron-Texaco (Y) depende de la del índice bursátil Standard & Poor's (S&P) 500 (X).

Supongamos que S_n es la cotización o valor del índice en el tiempo n . La variación de S entre el tiempo n y el tiempo $n+1$ es el ratio r definido como $S_n \exp\{r\} = S_{n+1}$, asumiendo una composición continua. Así, podemos calcular la variación diaria como $r = \ln(S_{n+1}/S_n)$ si el intervalo de tiempo se supone que es un día. Si suponemos que la cotización o el índice siguen un proceso estocástico de movimiento geométrico Brownian (GBM), la distribución de la cotización o de los índices es una distribución lognormal con parámetros determinados por el movimiento y la volatilidad del proceso GBM. Si las cotizaciones son lognormales, la variación de la cotización se distribuye normalmente dado que el logaritmo de la cotización se distribuye normalmente y que $r = \ln(S_{n+1}) - \ln(S_n)$. Así, la variación de la cotización es una combinación lineal de variables aleatorias normales, que es ella misma una variable aleatoria normal.

En este ejemplo, utilizamos valores de cierre diarios para el S&P 500 y Chevron-Texaco para cada día laborable en el año 2004 para calcular las variaciones diarias. Hay 251 observaciones en la muestra. Seleccionamos aleatoriamente 50 de ellas como conjunto de prueba para testar las distribuciones marginales creadas para las variaciones de Chevron-Texaco (Y) y utilizamos las restantes 201 para parametrizar varios modelos.

La ecuación lineal $y_i = a + b \cdot x_i + \epsilon_i$ viene definida por una regresión mínimo cuadrática del valor de cotización de Chevron-Texaco sobre el índice bursátil S&P 500, donde a es un término independiente, b es la pendiente, y ϵ_i es el término error para la observación i . Estimando

los parámetros para este modelo a partir de los datos surge la ecuación $\hat{y}_i = \hat{a} + \hat{b} \cdot x_i$. Los residuos, calculados como $e_i = y_i - \hat{y}_i$, son una estimación del término error del modelo, y se asume que siguen una distribución normal con media cero y una varianza denotada por σ_Z^2 .

Usando los datos del año 2004 para Chevron-Texaco y el S&P 500 surge el modelo lineal, $\hat{y}_i = 0,083749 + 0,305849 \cdot x_i$, con $\sigma_Z^2 = 1,118700$. Para este modelo, el coeficiente \hat{b} hace referencia a la *beta* de la cotización, el cual es un índice del riesgo sistemático de la cotización, o la sensibilidad de las variaciones de cotización a cambios en las variaciones de los índices de mercado. Este coeficiente es estadísticamente significativo con un t -valor de 2.86 y un p -valor (dos colas) de 0.0043.

Utilizamos los parámetros del modelo de regresión lineal y los datos de las variaciones diarias en la cotización de Chevron-Texaco y el S&P 500 para parametrizar trece modelos de redes Bayesianas y comparar los resultados obtenidos con la distribución real de valores de Chevron-Texaco utilizando el estadístico de contraste KS. Dónde tenga sentido, los métodos se testan utilizando aproximaciones 2, 3, y 4-trozos MTE de la distribución marginal de las variaciones de S&P 500 (X) determinadas utilizando el método descrito en [7]. Nos referiremos a ellas como *aproximaciones marginales* (MAs).

4.1. Aproximaciones discretas

Hemos considerado tres discretizaciones dividiendo el dominio de las variables continuas en 6, 9 y 12 subintervalos respectivamente. Los intervalos han sido determinados de acuerdo con los datos de tal manera que cada uno contiene el mismo número de elementos de la muestra. La probabilidad de cada partición discretizada se calcula mediante máxima verosimilitud.

4.2. Modelo CLG

Usar el modelo CLG para este ejemplo requiere asumir que las variaciones del índice S&P 500 (X) se distribuyen normalmen-

te, i.e. $\mathcal{L}(X) \sim N(\mu_X, \sigma_X^2)$, y que las variaciones de Chevron-Texaco (Y) se distribuyen normalmente con una media función lineal de X y una varianza independiente de X , i.e. $\mathcal{L}(Y | x) \sim N(a\mu_X + b, \sigma_{Y|x}^2)$. La media y varianza de las variaciones del índice S&P 500 obtenidas a partir de los datos son 0.028739 y 0.487697, respectivamente, i.e. $X \sim N(0,028739, 0,487697)$. Utilizamos los resultados del modelo de regresión para definir $Y | x \sim N(0,305849x + 0,083749, 1,18700)$. Dado que $E(Y) = 0,305849 \cdot \mu_X + 0,083749$ y $Var(Y) = (0,305849)^2 \cdot Var(X) + \sigma_{Y|x}^2$, el modelo CLG determina la distribución marginal de Y como $N(0,102797, 1,137741)$.

4.3. Modelo de árbol mixto

Los tres modelos de árbol mixto considerados en este ejemplo se construyen de acuerdo con el método propuesto en [8], particionando el dominio de las variables en 2, 3 y 4 trozos respectivamente. En cada trozo se ajusta una densidad MTE con dos términos exponenciales más una constante, i.e., un potencial MTE de la forma $\phi(y) = a + b \exp\{cx\} + d \exp\{ex\}$. las marginales se calculan usando el algoritmo de propagación Shenoy-Shafer adaptado a MTEs [9].

Durante el cálculo de las marginales, lo que involucra multiplicación de potenciales MTE, el número de términos exponenciales en cada potencial aumenta. Para mantener la complejidad de los potenciales MTE resultantes—medida con el número total del término utilizados—equivalente al número de cortes en la aproximación discreta, empleamos una versión aproximada del algoritmo Shenoy-Shafer que restringe los potenciales a dos términos exponenciales y una constante, con los términos extra podados como se describe en [9]. Nos referimos a este modelo como *árbol mixto podado*.

4.4. Modelo lineal determinístico

El modelo lineal determinístico asume que la variación de cotización de Chevron-Texaco (Y) es una función lineal determinística de la variación de S&P 500 (X) y un término, rui-



Figura 4: La red Bayesiana para el modelo lineal determinístico del ejemplo para la cotización de Chevron-Texaco.

do Gaussiano, (Z), $Y = a + b \cdot X + Z$. Una red Bayesiana representando esta situación se muestra en la Figura 4.

Para parametrizar este modelo, utilizamos potenciales 2, 3, y 4-trozos MTE para X obtenidos mediante la *aproximación marginal* y denotados por ϕ . La variable Z es el ruido Gaussiano que se modeliza mediante la aproximación 2-trozos, 3-términos MTE a la DF normal definida en [1] con $\mu = 0$ y $\sigma^2 = 1,118700$ y denotada mediante φ . La CMF para Y dada $\{X, Z\}$ es $\alpha(x, y, z) = p_{Y|\{x,z\}}(y) = \mathbf{1}\{y = 0,305849x + z + 0,083749\}$. En cada caso de prueba, la DF conjunta $\vartheta = (\phi \otimes \varphi)$ para $\{X, Z\}$ es un potencial MTE. La DF conjunta sin normalizar $\theta(y, z) = (\vartheta \otimes \alpha)^{-X}$ se obtiene sustituyendo $(y - z - 0,083749)/0,305849$ en la DF conjunta para $\{X, Z\}$.

La DF marginal para Y se obtiene integrando la DF conjunta θ para $\{Y, Z\}$ sobre el dominio de Y . Las distribuciones marginales para Y determinadas usando las marginales 2, 3 y 4-trozos MTE para X tienen 8, 11, y 14 trozos, respectivamente.

4.5. Comparación

Los métodos en las Secciones 4.1 hasta 4.4 se comparan usando el estadístico de Kolmogorov-Smirnov (KS), que se define como

$$D(F, G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|, \quad (4)$$

donde F es una distribución objetivo y G es la distribución empírica de la muestra. Este estadístico se puede utilizar para construir un contraste para la hipótesis de que los datos realmente provienen de la distribución objetivo F . Los resultados del test KS (valor del

Cuadro 1: Estadísticos del contraste KS comparando la muestra de prueba con la distribución marginal para cada método.

Método	Interv.	KS Test		Térn.
		D	p -valor	
Aprox. marginal	2	0.099	0.71	6
	3	0.1013	0.6842	9
	4	0.102	0.6758	12
Árbol mixto podado	2	0.1086	0.5969	16
	3	0.0912	0.7996	27
	4	0.0842	0.8702	76
Aprox. discreta	2	0.1098	0.5826	6
	3	0.0974	0.73	9
	4	0.1031	0.6628	12
Modelo lineal det.	6	0.1136	0.5389	6
	9	0.1226	0.4399	9
	12	0.1352	0.3198	12
Modelo lineal det.	2	0.0871	0.8426	64
	3	0.0832	0.8794	85
	4	0.0866	0.8472	110
CLG	1	0.0863	0.85	2

estadístico D y p -valor del test) se muestran en el Cuadro 1.

Los resultados del test para el ejemplo considerado en este trabajo muestran que los métodos basados en modelos MTE dan lugar a mejores resultados que la aproximación discreta. Incluso el caso de árboles mixtos podados mejora la aproximación discreta.

El modelo teórico para este ejemplo es un modelo CLG. Los resultados del experimento muestran que los modelos MTE dan lugar a distribuciones marginales para las variables dependientes que son muy similares al modelo CLG, obteniendo los mejores resultados el modelo lineal determinístico. Este resultado no es sorprendente, ya que las relaciones entre las variables se supone que son lineales con ruido Gaussiano, como en el modelo teórico. Sin embargo, se debe destacar que el caso de cuatro intervalos es favorable al modelo de árbol mixto en términos de p -valor, y utilizando un menor número de términos.

5. Conclusiones

En este trabajo hemos comparado métodos alternativos de especificar distribuciones condicionadas en redes Bayesianas con variables continuas. Los resultados muestran que el mo-

delo MTE es apropiado en problemas donde el modelo teórico tiene distribuciones Gaussianas, lo que es común en problemas econométricos. Utilizamos un ejemplo donde las variables continuas se sabe que son normales para comparar los resultados con el modelo CLG; sin embargo, dado que el modelo MTE se puede usar para aproximar cualquier distribución de probabilidad continua, los resultados se amplían a modelos con densidades no Gaussianas. Usando los resultados de este trabajo, podemos deducir que los resultados obtenidos en modelos más grandes usando los dos enfoques para estimar modelos MTE condicionales son probablemente más precisos; sin embargo, se necesita una más amplia investigación para comparar la complejidad de la propagación en modelos grandes que utilicen estos dos enfoques.

Referencias

- [1] Cobb, B.R. and P.P. Shenoy: Inference in hybrid Bayesian networks with mixtures of truncated exponentials. WP#294, School of Business, University of Kansas, Lawrence, KS (2003).
- [2] Cobb, B.R. and P.P. Shenoy: Inference in hybrid Bayesian networks with deterministic variables. In P. Lucas (ed.): *Proc. of the 2nd European Workshop on Probabilistic Graphical Models (PGM'04)* Leiden, Netherlands (2004) 57–64.
- [3] Cobb, B.R. and P.P. Shenoy: Propagation in hybrid Bayesian networks with linear deterministic variables. WP#314, School of Business, University of Kansas, Lawrence, KS (2005).
- [4] Kozlov, A.V. and D. Koller: Nonuniform dynamic discretization in hybrid networks.

- In D. Geiger and P.P. Shenoy (eds.): *Uncertainty in Artificial Intelligence* **13** (1997) 314–325, Morgan–Kaufman, San Francisco.
- [5] S.L. Lauritzen and F. Jensen: Stable local computation with conditional Gaussian distributions. *Statistics and Computing* **11** (2001) 191–203.
- [6] Moral, S., Rumí, R. and A. Salmerón: Mixtures of truncated exponentials in hybrid Bayesian networks. *Lecture Notes in Artificial Intelligence* **2143** (2001) 156–167.
- [7] Moral, S., Rumí, R. and A. Salmerón: Estimating mixtures of truncated exponentials from data. In J.A. Gámez and A. Salmerón (eds.): *Proc. of the 1st European Workshop on Probabilistic Graphical Models (PGM'02)* Cuenca, Spain (2002) 135–143.
- [8] Moral, S., Rumí, R. and A. Salmerón: Approximating conditional MTE distributions by means of mixed trees. *Lecture Notes in Artificial Intelligence* **2711** (2003) 173–183.
- [9] Rumí, R. and A. Salmerón: Penniless propagation with mixtures of truncated exponentials. ECSQARU'05.
- [10] Sun, L. and P.P. Shenoy: Using Bayesian networks for bankruptcy prediction. WP#302, School of Business, University of Kansas, Lawrence, KS (2003).