

PROBLEMA TROZADO

Estimación Densidades Diámetro-Altura

Informe de Avance No.5

1. Equipo

Patricio Corvalán (responsable)
Raúl Gouet

2. Objetivos

Estimación de densidades bivariadas Altura-Diámetro a partir de datos de inventarios. Se consideran dos estrategias:

- 1 . Modelo paramétrico basado en la distribución S_{BB} de Johnson y
- 2 . Estimadores no-paramétricos de tipo kernel bivariados, con suavizamiento maximal.

3. Datos

Disponemos de 5 sets de datos reales, proporcionados por la empresa Cía. Forestal de Chile, que serán denominados chile1, chile2, chile3, chile4 y chile5. Estos corresponden a mediciones de altura y DAP para la totalidad de los árboles en las unidades muestrales de 5 inventarios forestales. La información resumida se presenta en los siguientes gráficos.

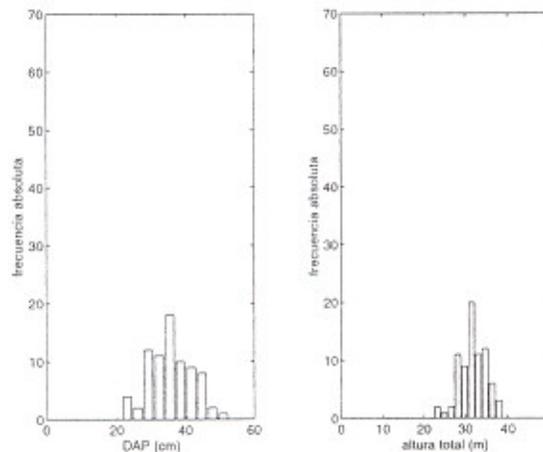


Figura 1. Histograma datos archivo chile1

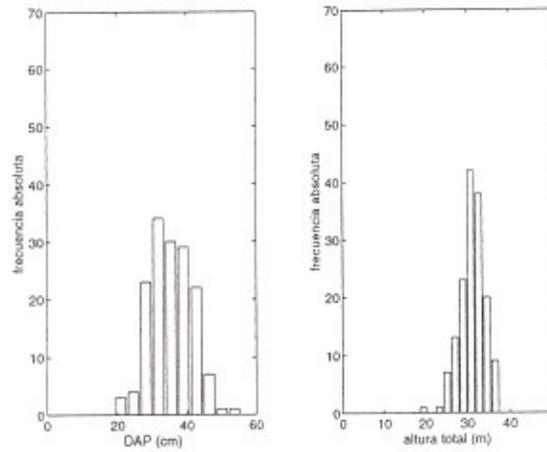


Figura 2. Histograma datos archivo chile2

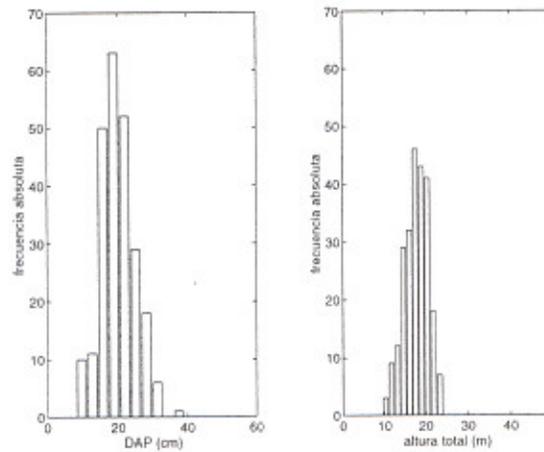


Figura 3. Histograma datos archivo chile3

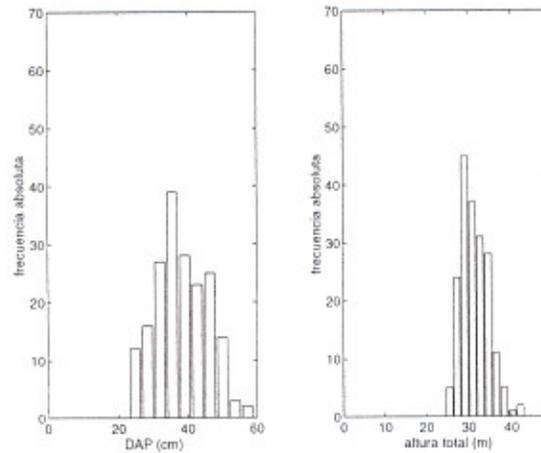


Figura 4. Histograma datos archivo chile4

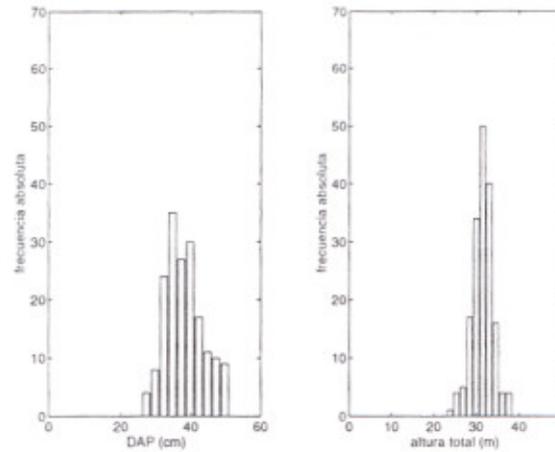


Figura 5. Histograma datos archivo chile5

4. Modelo de Johnson

El modelo S_{BB} de Johnson fue usado por primera vez para ajustar datos de diámetro y altura, por Schreuder y Hafley (S&H) (1977). El vector aleatorio (Y_1, Y_2) tiene ley S_{BB} si

$$Z_1 = \gamma_1 + \delta_1 \log\left(\frac{Y_1}{1 - Y_1}\right), \quad Z_2 = \gamma_2 + \delta_2 \log\left(\frac{Y_2}{1 - Y_2}\right)$$

son v.a. normales standard, con coeficiente de correlación ρ .

Sean D, H las variables diámetro y altura respectivamente. Entonces, se define

$$Y_1 = \frac{D - \xi_1}{\lambda_1}, \quad Y_2 = \frac{H - \xi_2}{\lambda_2}$$

donde ξ_1, ξ_2 son los valores mínimos y λ_1, λ_2 los rangos del diámetro (D) y la altura (H), respectivamente.

Estimación

De acuerdo con S&H, el ajuste del modelo de Johnson se hace como sigue:

- 1 . Estimar los parámetros de las distribuciones marginales.
- 2 . Transformar los pares de observaciones a desvíos normales standard.
- 3 . Estimar el coeficiente de correlación entre los desvíos normales.

Se sabe que no existe solución cerrada para los estimadores máximo verosímil de los parámetros $\xi, \lambda, \gamma, \delta$, pero éstos pueden ser calculados por métodos iterativos.

Es recomendable asignar al parámetro ξ los valores $\xi_1 = 0$ (diámetro mínimo cero) y $\xi_2 = 1.3$ (altura mínima 1.3 mts., donde se mide el DAP). Además, es razonable que

$(\xi_i + \lambda_i)$, $i = 1, 2$ supere al máximo de la variable correspondiente. Así tenemos el siguiente procedimiento: buscar iterativamente el EMV de λ_i usando los EMV condicionales de γ_i y δ_i , de manera que $(\xi_i + \hat{\lambda}_i)$ supere al máximo de la variable que corresponda.

Siguiendo la notación de S&H, los EMV condicionales de los parámetros γ y δ se calculan como sigue:

$$\hat{\gamma}_i = \frac{-\bar{f}_i}{s_i} \quad \hat{\delta}_i = \frac{1}{s_i}$$

donde

$$\bar{f}_i = \frac{1}{n} \sum_{j=1}^n f_{ij}, \quad s_i^2 = \frac{1}{n} \sum_{j=1}^n (f_{ij} - \bar{f}_i)^2, \quad f_{ij} = \log\left(\frac{y_{ij}}{1 - y_{ij}}\right),$$

$$y_{ij} = \frac{(x_{ij} - \xi_i)}{\hat{\lambda}_i}, \quad x_{1j} = D_j, \quad x_{2j} = H_j, \quad i = 1, 2, \quad j = 1, \dots, n$$

Finalmente, ρ se estima calculando la correlación muestral de los valores z . Es decir,

$$\hat{\rho} = \frac{1}{n} \sum_{j=1}^n z_{1j} z_{2j}$$

donde

$$z_{ij} = \hat{\gamma}_i + \hat{\delta}_i \log\left(\frac{y_{ij}}{1 - y_{ij}}\right)$$

Para la estimación de λ se procede a maximizar la función de verosimilitud, que está dada por

$$\begin{aligned} L(\lambda_1, \lambda_2) &= \prod_{i=1}^n p(z_{1j}, z_{2j}; \rho(\lambda_1, \lambda_2)) = \prod_{i=1}^n \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2} \frac{1}{1-\rho^2} (z_{1j}^2 - 2\rho z_{1j} z_{2j} + z_{2j}^2)} \\ &= (2\pi\sqrt{1-\rho^2})^{-n} e^{-\frac{1}{2} \frac{1}{1-\rho^2} \sum_{j=1}^n (z_{1j}^2 - 2\rho z_{1j} z_{2j} + z_{2j}^2)} \\ &= (2\pi\sqrt{1-\rho^2})^{-n} e^{-\frac{1}{2} \frac{1}{1-\rho^2} \sum_{j=1}^n z_{1j}^2 - 2\rho \sum_{j=1}^n z_{1j} z_{2j} + \sum_{j=1}^n z_{2j}^2} \\ &= (2\pi\sqrt{1-\rho^2})^{-n} e^{-\frac{1}{2} \frac{1}{1-\rho^2} (n - 2n\rho^2 + n)} \\ &= (2\pi\sqrt{1-\rho^2})^{-n} e^{-n} \end{aligned}$$

Se observa entonces que la verosimilitud es una simple función creciente de ρ .

Para estudiar como ρ depende de λ , se graficó la superficie correspondiente a la función $\rho(\lambda_1, \lambda_2)$. En la figura 6 se observa que la correlación varía poco con λ_1 y λ_2 , a partir de valores poco superiores a los máximos de las variables respectivas.

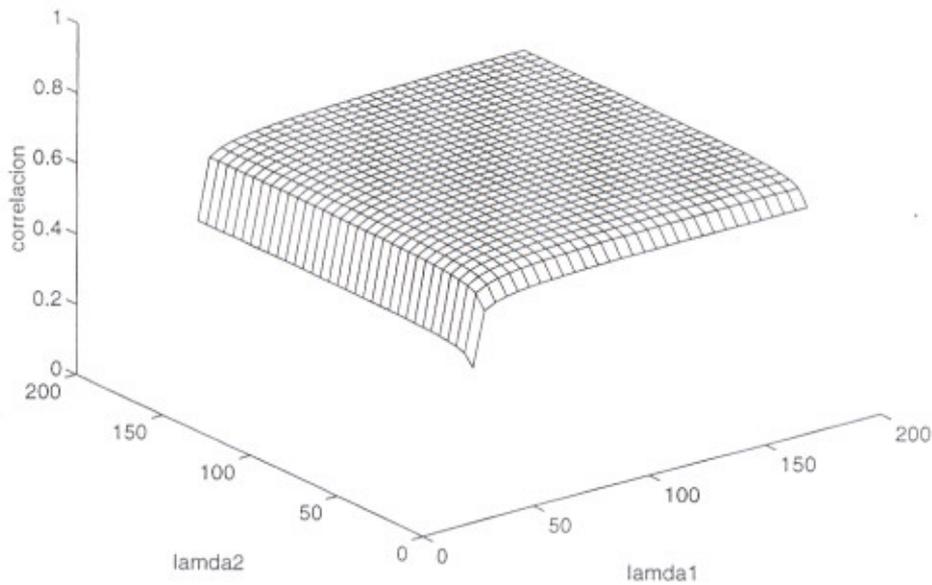


Figura 6. Superficies de correlación

Se puede verificar empíricamente que no tiene sentido maximizar ρ o L como función de λ_1 y λ_2 puesto que los máximos se alcanzan en valores absurdos. Por ejemplo, para el archivo de datos chile 1, la máxima verosimilitud se alcanza en $\lambda_1 = 8346175$ y $\lambda_2 = 198039$; para chile3 se obtiene $\lambda_1 = 104975$ y $\lambda_2 = 131.9$, etc.

Un parámetro importante en el ajuste del modelo S_{BB} es $\phi = \rho\delta_1/\delta_2$. Cuando $0 < \phi < 1$, la derivada de la regresión mediana, definida como

$$(H - \xi_2) = \frac{\lambda_2 \theta}{\left(\frac{\xi_1 + \lambda_1 - D}{D - \xi_1}\right)\phi + \theta}$$

donde $\theta = \exp((\rho\gamma_1 - \gamma_2)/\delta_2)$, no tiene derivadas en los extremos y por lo tanto, el modelo carece de interés práctico. Entonces, para datos de diámetro-altura debería imponerse la restricción $\phi > 1$.

Tenemos así la siguiente estrategia de estimación: calcular los EMV condicionales de los parámetros γ, δ ; estimar ρ por correlación empírica y finalmente, estimar los λ buscando valores cercanos a los diámetros y alturas máximas, tales que $\phi \geq 1$. Las estimaciones de λ_1 y λ_2 se obtienen resolviendo el problema de optimización

$$\min \lambda_1 + \lambda_2 \quad \text{s.a.} \quad \phi(\lambda_1, \lambda_2) \geq 1$$

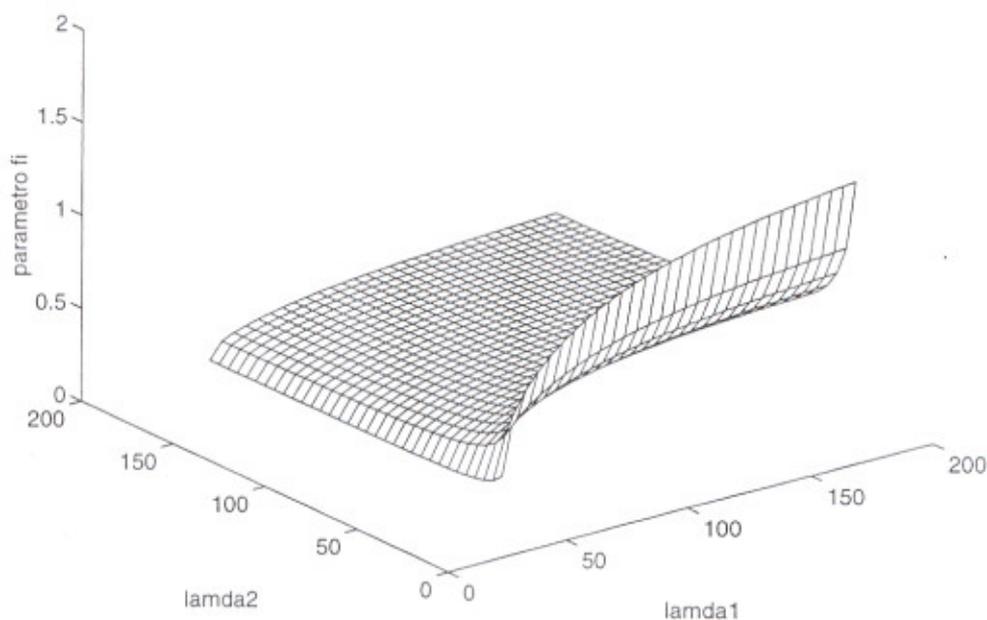


Figura 7. Superficies de valores de ϕ

Resultados

Los resultados finales de estimación de parámetros del modelo de Johnson, para los cinco archivos de datos, se presentan en la tabla 1.

Archivo	$(\hat{\gamma}_1, \hat{\gamma}_2)$	$(\hat{\delta}_1, \hat{\delta}_2)$	$(\hat{\lambda}_1, \hat{\lambda}_2)$	ρ	ϕ
chile1	0.65, -2.03	3.02, 1.27	79.7, 37.6	0.421	1.000
chile2	-0.79, -2.12	2.19, 1.25	60.36, 36.3	0.570	1.000
chile3	-0.10, -1.29	1.68, 1.19	39.31, 22.8	0.709	1.000
chile4	7.10, -1.74	4.24, 1.7	238.6, 41.8	0.406	1.000
chile5	2.16, -2.73	4.36, 1.80	100.8, 37.3	0.413	1.000

Tabla 1. Estimaciones modelo S_{BB}

En la figuras a continuación se puede apreciar gráficamente el resultado de la estimación del modelo S_{BB} . Los puntos representan los pares (DAP, altura total); la línea continua es la función de regresión mediana del modelo S_{BB} ; las cruces (+) corresponden al modelo de regresión usado habitualmente en las empresas: $H = \alpha e^{-\beta/\sqrt{d}}$; las figuras elípticas son las curvas de nivel de la densidad estimada.

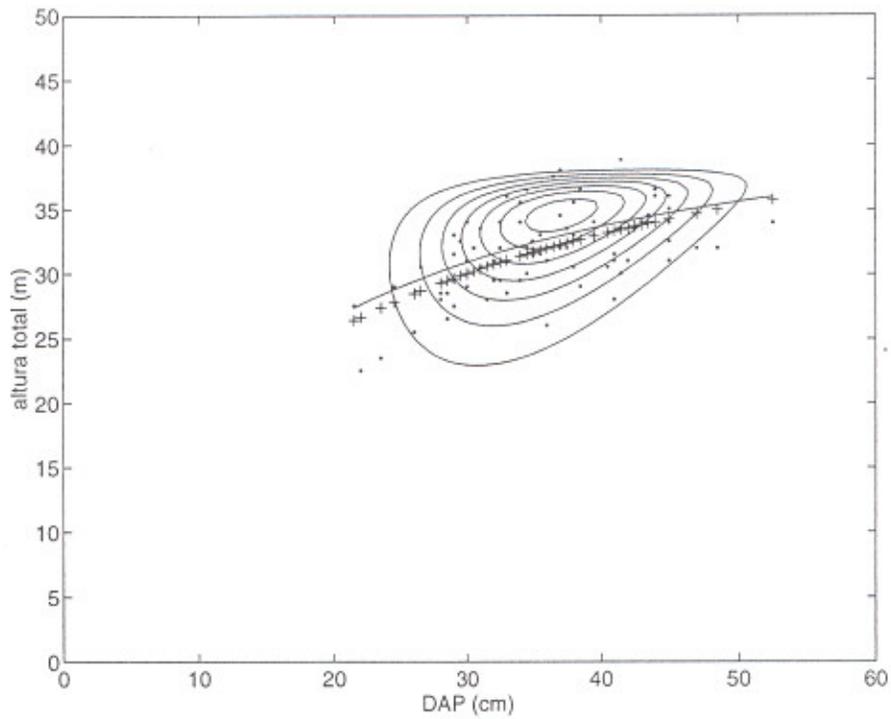


Figura 8. Archivo Chile1. Curvas de Nivel S_{BB} y Regresión Mediana

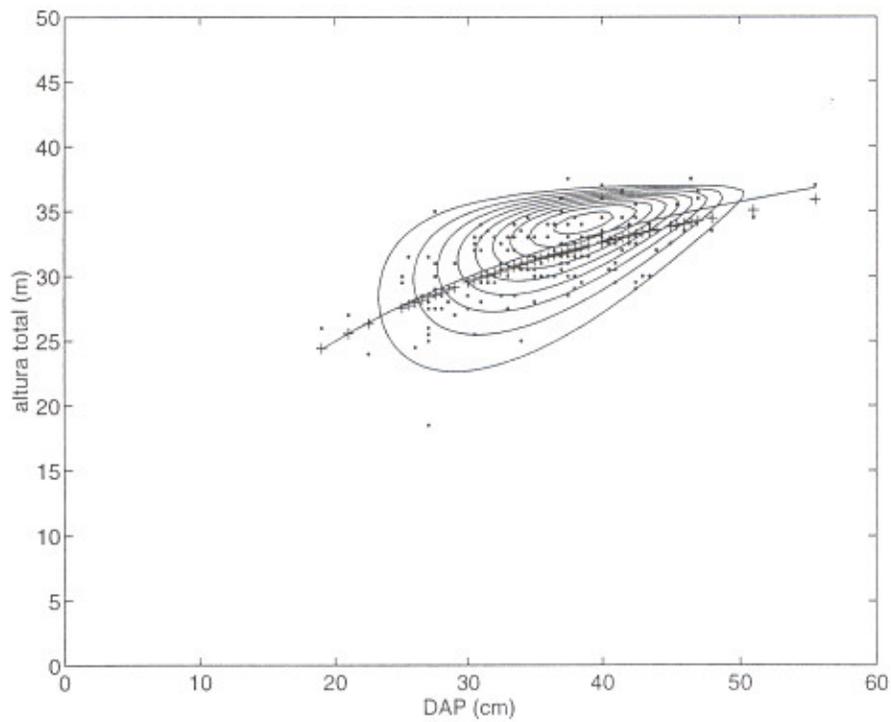


Figura 9. Archivo Chile2. Curvas de Nivel S_{BB} y Regresión Mediana

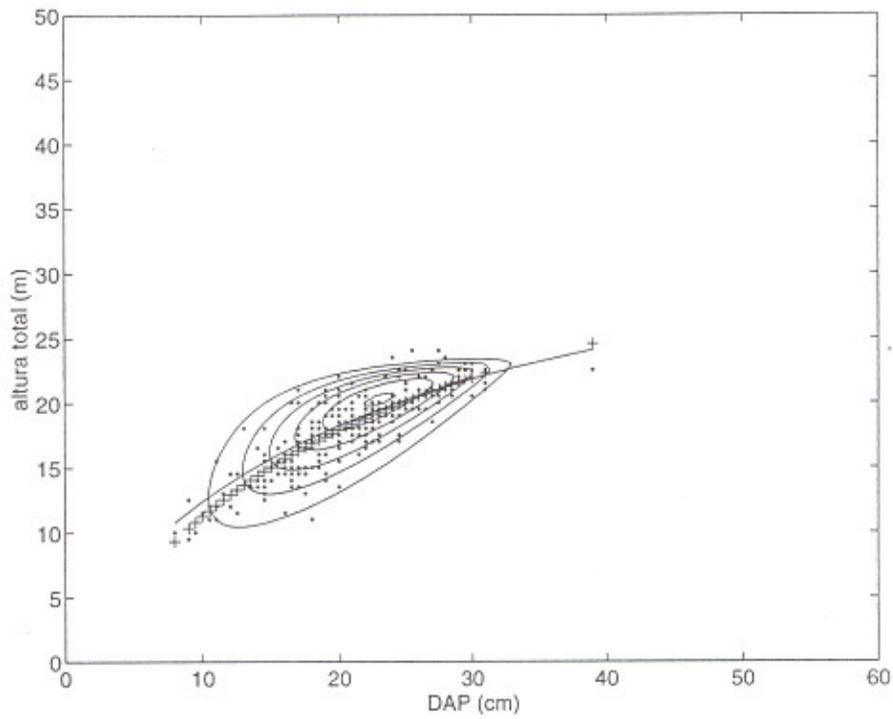


Figura 10. Archivo chile3. Curvas de Nivel S_{BB} y Regresión mediana

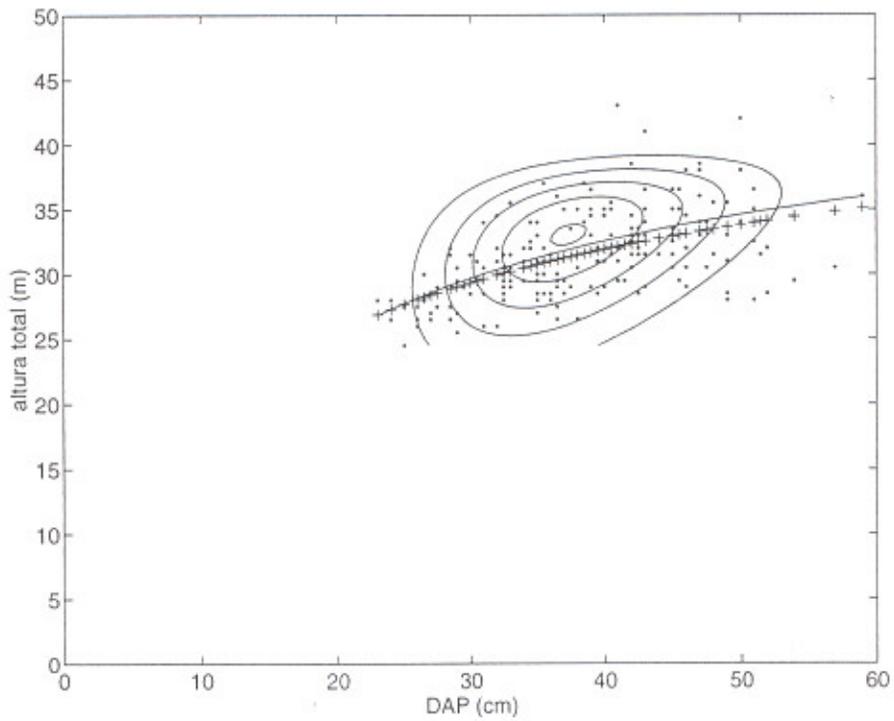


Figura 11. Archivo chile4. Curvas de Nivel S_{BB} y Regresión mediana

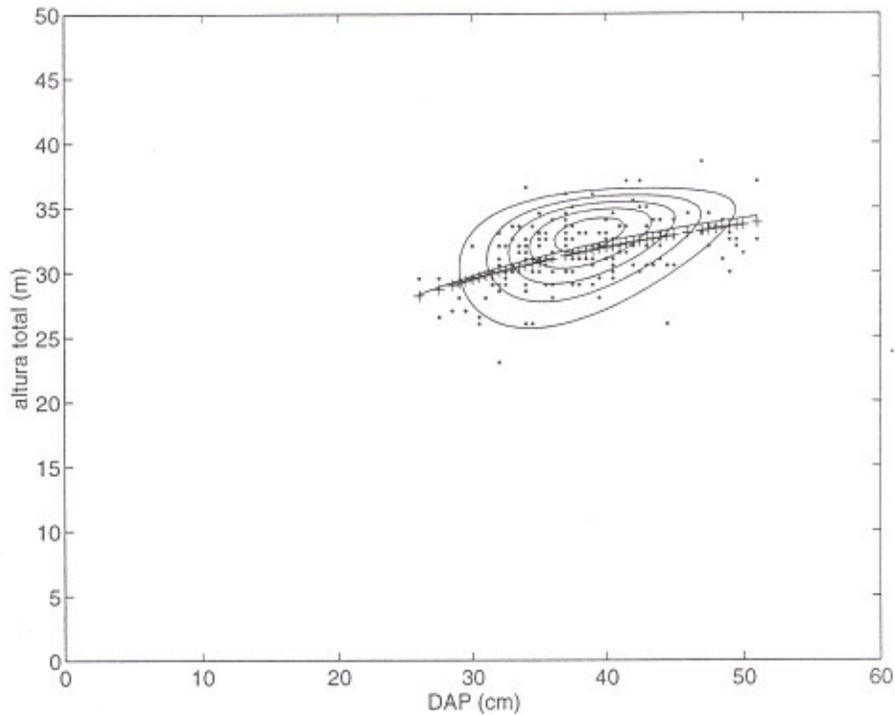


Figura 12. Archivo chile5. Curvas de Nivel S_{BB} y Regresión mediana

5. Modelo no Paramétrico

El objetivo de la estimación no paramétrica de densidades de probabilidad es reconstruir la densidad a partir de datos muestrales, haciendo un mínimo de supuestos sobre su forma funcional. En la práctica, se trata de suavizar la ley empírica usando una familia flexible de funciones.

Estimadores de Kernel

El estimador de kernel, para una densidad de probabilidad sobre \mathbb{R} se define como

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$$

donde X_1, \dots, X_n son los datos muestrales (v.a. iid con densidad f); K es una función llamada kernel o núcleo, que integra a 1 y h es un parámetro de suavizamiento.

En este trabajo interesa la estimación de la densidad de probabilidad conjunta de las variables diámetro-altura.

En el caso bivariado, el estimador de kernel tiene forma general dada por

$$\hat{f}(x, y) = \frac{1}{n|H|} \sum_{i=1}^n K\left(H^{-1}\begin{pmatrix} X_i - x \\ Y_i - y \end{pmatrix}\right)$$

donde $X_i, Y_i, i = 1, \dots, n$ es la información muestral bivariada; H es una matriz de suavizamiento; $|H|$ su determinante y K un núcleo bivariado que integra a 1.

Epanechnikov (1969) encuentra el kernel óptimo en el caso de estimadores del tipo

$$f(x, y) = \frac{1}{nh_1h_2} \sum_{i=1}^n K\left(\frac{X_i - x}{h_1}\right) K\left(\frac{Y_i - y}{h_2}\right)$$

Para ello minimiza el error cuadrático medio integrado. Es decir

$$\int E(\hat{f}(x, y) - f(x, y))^2 dx dy$$

La búsqueda del kernel se reduce a un problema de optimización (cálculo de variaciones con restricciones) cuyo resultado es

$$K(x) = \left(\frac{3}{4\sqrt{5}} - \frac{3x^2}{20\sqrt{5}} \right) \mathbb{1}_{\{|x| \leq \sqrt{5}\}}$$

Sin embargo, existen otros núcleos, ampliamente usados, cuya calidad es muy cercana al óptimo presentado arriba. Tal es el caso del núcleo gaussiano, dado por

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Grado de Suavizamiento

Un problema importante en la estimación no paramétrica de densidades es la elección de los parámetros de suavizamiento. Un alto grado de suavizamiento puede ocultar características finas y por el contrario, poco suavizamiento puede hacer aparecer características espúreas. En general, a mayor tamaño muestral se requiere menos suavizamiento.

Hay mucha literatura sobre la elección automática de los parámetros de suavizamiento. Entre los métodos más importantes se puede citar el uso de familias paramétricas, y las técnicas de validación cruzada. Sin embargo, ambos enfoques tienen serios inconvenientes.

Terrel (1990) propone un nuevo procedimiento, llamado de suavizamiento maximal, que consiste en escoger el mayor grado de suavizamiento, compatible con una estimación de la escala de la densidad.

En general, la teoría de parámetros óptimos de suavizamiento se complica bastante en el caso multidimensional. Sin embargo, el principio de suavizamiento maximal de Terrel conduce a resultados directamente aplicables.

Suavizamiento Maximal

El principio de suavizamiento maximal se describe en los siguientes términos:

1. Se considera un estimador de kernel bivariado \hat{f} , dependiendo de una matriz H de suavizamiento.

2. Se calcula H de manera que se minimice el error cuadrático medio integrado asintótico, obteniéndose como resultado una expresión donde interviene la densidad desconocida f .

3. Se escoge una medida de escala $T(F)$ (por ejemplo la covarianza) asociada a la distribución bivariada F , de densidad f , y se maximiza cierto funcional de H sujeto a $T(F) = T(\hat{F})$, donde \hat{F} es la distribución empírica.

El resultado de la estrategia descrita es el siguiente: la matriz de suavizamiento maximal H satisface la ecuación

$$HH' = C \int K^2(x, y) dx dy \cdot n^{-1/3} \hat{\Sigma}$$

donde C es una constante independiente de K y $\hat{\Sigma}$ es la matriz de covarianzas muestrales.

Si K es esféricamente simétrica, entonces K es función sólo de HH' . Tal es el caso, por ejemplo, del núcleo gaussiano

$$K(x, y) = \frac{1}{2\pi} e^{-\frac{x^2 + y^2}{2}}$$

y del núcleo de tipo Epanechnikov

$$K(x, y) = \frac{2}{\pi a^2} \left(1 - \frac{x^2 + y^2}{a^2} \right) \mathbb{1}_{\{x^2 + y^2 \leq a^2\}}$$

Reducción del Sesgo

Los estimadores de kernel son sesgados pero existen algunas técnicas orientadas a disminuir el sesgo, que en el caso univariado tiene una expansión asintótica formal del tipo

$$\sum_{j=1}^{\infty} \frac{h^{2j} f^{(2j)}(x)}{(2j)!} \int u^{2j} K(u) du$$

Un método usado para reducir sesgo consiste en escoger K tal que $\int u^{2j} K(u) du = 0$ para $j = 1 \dots k - 1$ y $\int u^{2k} K(u) du \neq 0$. Sin embargo, este enfoque produce densidades negativas (lo que puede corregirse) o bien, reduce el sesgo sólo cuando la densidad subyacente f no tiene características interesantes o cuando la muestra es enorme.

Alternativamente, Ruppert y Cline (1994) proponen en un reciente artículo una técnica basada en una transformación de los datos. En lugar de estimar f , se estima la densidad

$$f_Y(y) = f(g^{-1}(y)) \frac{dg^{-1}}{dy}$$

donde g es una transformación creciente y bastante regular. Esto equivale, por cierto, a trabajar con la muestra $Y_i = g(X_i)$ $i = 1, \dots, n$.

Posteriormente, se recupera \hat{f} aplicando la transformación inversa, es decir

$$\hat{f}_g(x) = \hat{f}_Y(g(x)) \frac{dg}{dx}$$

Ruppert y Cline sugieren usar $g = \hat{F}$, donde \hat{F} es un estimador regular de la distribución F . Por ejemplo

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}(t) dt$$

donde \hat{f} es el estimador habitual de kernel de f . La idea del método parece lógica ya que si \hat{F} está cercana a F , entonces las v.a. Y_i serán aproximadamente uniformes en $[0,1]$ y dado que las derivadas de la densidad uniformes son nulas, el sesgo debería ser pequeño. De hecho, los autores demuestran que el error cuadrático pasa de $O(n^{-4/5})$ a $O(n^{-8/9})$.

Actualmente se trabaja en la generalización de los resultados de Ruppert y Cline al caso bivariado.

Aplicación

Los resultados teóricos presentados más arriba son aplicados a los datos contenidos en los archivos chile1 a chile5. Usamos el núcleo de Epanechnikov esféricamente simétrico

$$K(x, y) = \frac{2}{6\pi} \left(1 - \frac{x^2 + y^2}{6} \right) \mathbb{1}_{\{x^2 + y^2 \leq 6\}}$$

con suavizamiento maximal H , dado por

$$HH' = \left(\frac{625}{432n} \right)^{\frac{1}{3}} \int K^2(x, y) dx dy \hat{\Sigma}$$

Resultados

En las figuras 13 a 17 se presentan las curvas de nivel del modelo no paramétrico para los archivos de datos chile1 a chile5. Las cruces corresponden a la función de regresión usada por las empresas: $H = \alpha e^{-\beta/\sqrt{d}}$

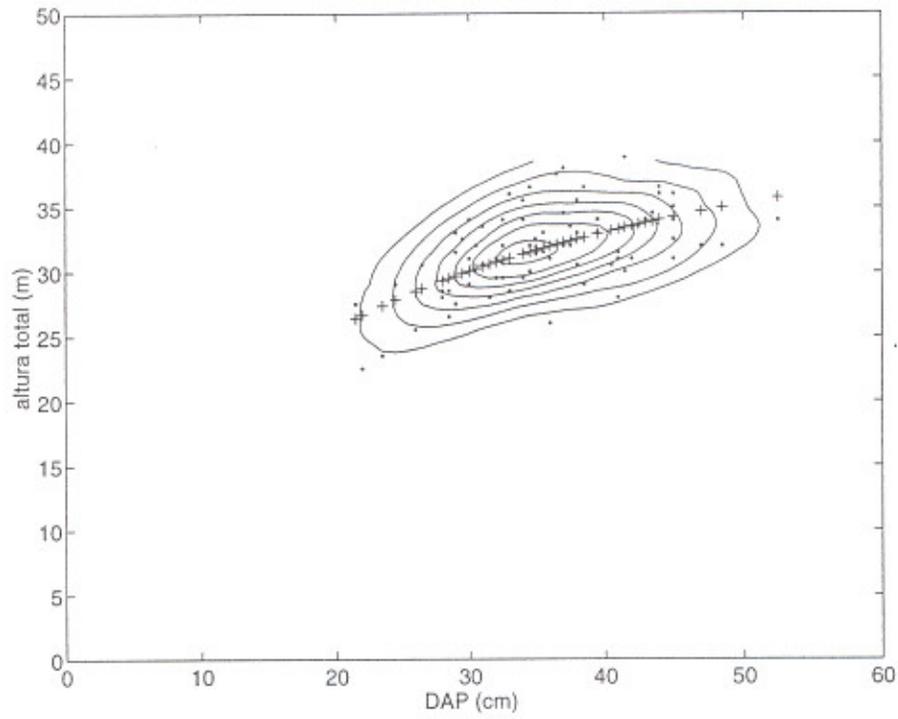


Figura 13. Archivo Chile1. Curvas de Nivel Modelo No Paramétrico

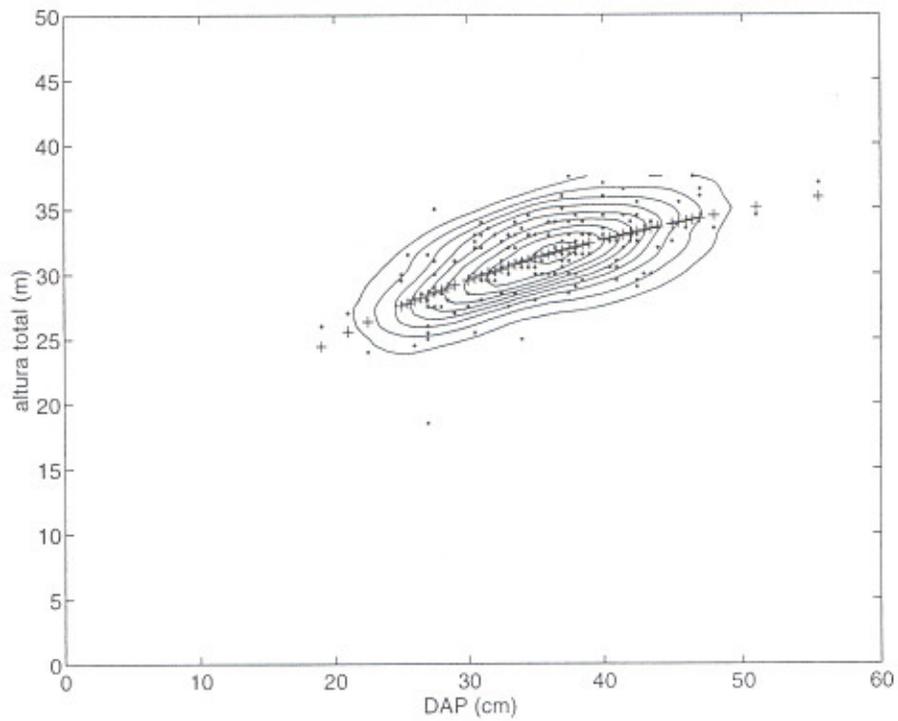


Figura 14. Archivo Chile2. Curvas de Nivel Modelo No Paramétrico

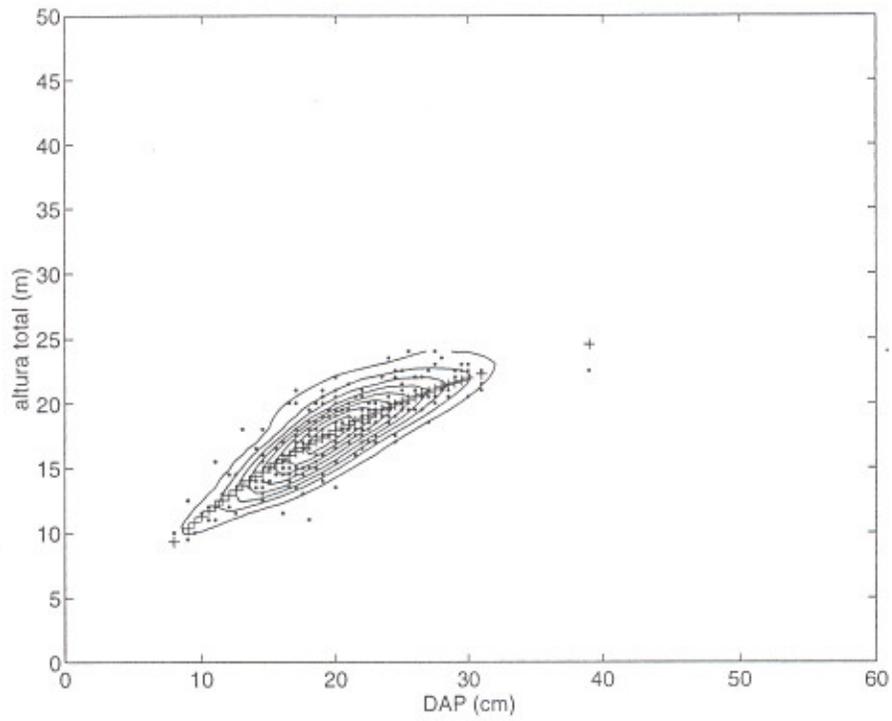


Figura 15. Archivo Chile3. Curvas de Nivel Modelo No Paramétrico

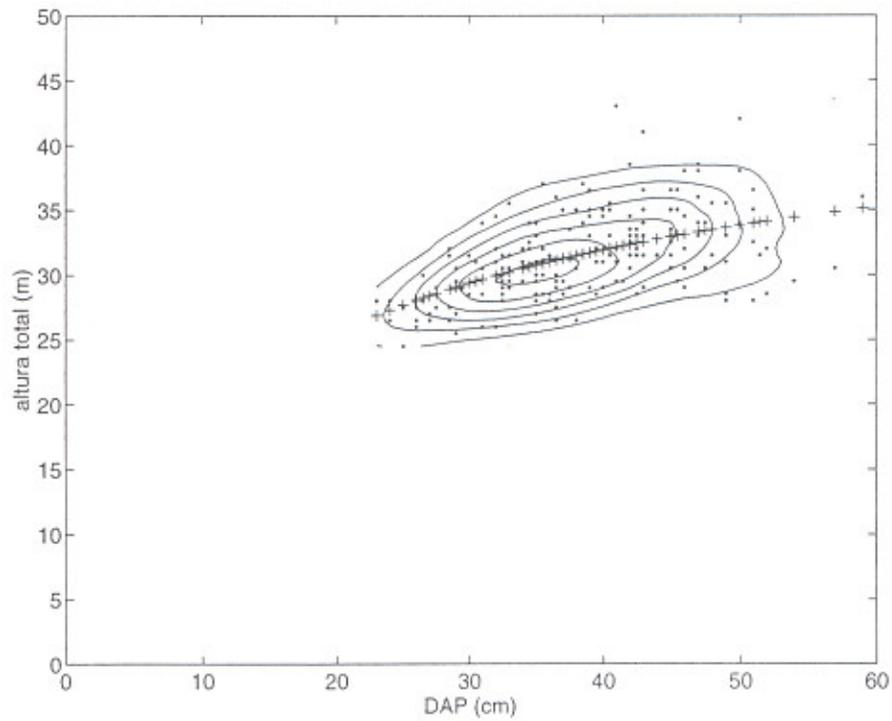


Figura 16. Archivo Chile4. Curvas de Nivel Modelo No Paramétrico

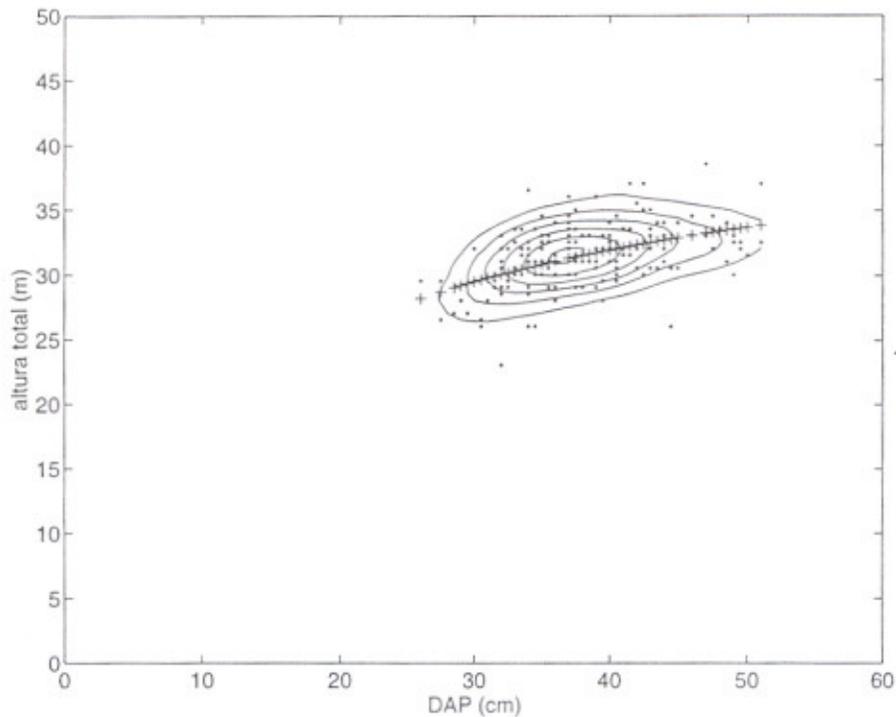


Figura 17. Archivo chile5. Curvas de Nivel Modelo No Paramétrico

Discusión

La comparación visual de las curvas de nivel, para los dos modelos estimados, sugiere un mejor ajuste en el caso no paramétrico. Sin embargo, es necesario disponer de datos que permitan verificar que el modelo no paramétrico produce mejores resultados que los "simuladores de inventario" actualmente en uso. Esperamos contar con dicha información próximamente.

Referencias

- [1] Ruppert D., Cline D. (1994), Bias reduction in kernel density estimation by smooth empirical transformations. *Ann. of Statist.* 22, 1, 185-210.
- [2] Epanechnikov V.A. (1969), Non parametric estimation of a multivariate probability density. *Th. Probab. and its Appl.*, 14, 153-158.
- [3] Schreder H., Hafley W. (1977), A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* 33, 471-478.
- [4] Terrell G. (1990), The maximal smoothing principle in density estimation. *JASA*, 85, 470-477.
- [5] Ward M., Jones M. (1993), Comparison of smoothing parametrization in bivariate kernel density estimation. *JASA*, 88, 520-528.

- [6] Ward M., Marron J., Ruppert D. (1991), Transformation in density estimation, JASA, 86, 343-361.
- [7] Ghong Gu (1993), Smoothing spline density estimation: a dimensionless automatic algorithm. JASA, 88, 495-503.
- [8] Tapia R., Thompson J. (1978), Non Parametric Probability Density Estimation. J. Hopkins Univ. Press.