TRABAJOS DE ESTADISTICA
Vol. 5, núm. 1, 1990, pp. 3 a 14

ESTIMACION DE LA FUNCION CUANTIL Y CUANTIL-DENSIDAD MEDIANTE POLINOMIOS DE KANTOROVIC

A. FERNÁNDEZ P. y J. MUÑOZ PÉREZ
Departamento de Estadística e I.O.
Facultad de Matemáticas
Universidad de Sevilla

RESUMEN

En este trabajo se propone un estimador para la función cuantil, basado en polinomios de Kantorovic, como estimador natural, y se prueba que su error absoluto medio es un infinitésimo de orden $n^{-1/2}$. Mediante simulación se pone de manifiesto que dicho estimador conduce a una reducción sustancial del error absoluto medio frente a la función cuantil muestral y, por otra parte, se compara con el estimador basado en polinomios de Bernstein.

Asimismo, se propone un estimador basado en polinomios de Kantorovic para estimar la función cuantil-densidad, que es la derivada del estimador de la función cuantil basado en polinomios de Bernstein.

Palabras clave: quantile function; quantiles; quantile; density; L-estimates.
Clasificación AMS: 62G05, 62G30.

ABSTRACT

In this paper, we propose an estimator to the quantile function based on Kantorovic polynomials which is the natural estimator and it is shown that $E|K_n - Q| \sim O(n^{-1/2})$. By Monte Carlo simulation it is shown that this estimator conduce to substantial reduction for mean absolute error respect to the conventional empirical quantile function and it is compared with the estimator bases on Bernstein polinomials.

Also, we propose an estimator based on Kantorovic polinomial for quantile-density function which is the derivative of estimator based on Bernstein polinomials for quantile function.

Recibido, marzo 1988. Revisado, mayo 1989.

1. INTRODUCCION

Supongamos que $X_1, X_2, ..., X_n$ son variables aleatorias independientes e idénticamente distribuidas con función de distribución F absolutamente continua. La función cuantil $Q(u)$, $0 \leqslant u \leqslant 1$, está definida por:

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geqslant u\}$$

Parzen (1979), señaló la importancia de la función cuantil en el análisis exploratorio de los datos resaltando algunas de sus propiedades. En este mismo trabajo, Parzen propone para la estimación de la función cuantil el estimador tipo núcleo siguiente:

$$T_n(u) = \int_0^1 F_n^{-1}(t) \frac{1}{h(n)} K\left(\frac{t-u}{h(n)}\right) dt =$$

$$= \sum_{i=1}^{n} X_{(i)} \frac{1}{h(n)} \int_{(i-1)/n}^{i/n} k\left(\frac{t-u}{h(n)}\right) dt$$

donde F_n es la función de distribución empírica y K una función núcleo determinada. Este estimador ha sido estudiado posteriormente por Reiss (1980), Falk (1984) y Yang (1985) comparándolo con F_n^{-1}. El éxito de este estimador depende tanto de la función núcleo K elegida como del factor de suavización $h(n)$.

Otros estimadores alternativos a la función cuantil muestral, como estimadores de la función cuantil, fueron propuestos por Reiss (1980), Harrell y Davis (1982), Kaigh y Lachenbruch (1982), Kaigh (1983) y Muñoz-Fernández (1987).

Cuando $Q(u)$ es una función continua sabemos que podemos obtener una buena aproximación de ella mediante su polinomio de Kantorovic,

$$(n+1) \sum_{k=0}^{n} \binom{n}{k} u^k (1-u)^{n-k} \int_{k/(n+1)}^{(k+1)/(n+1)} Q(t)\, dt$$

de manera que si sustituimos en dicha integral $Q(t)$ por el estimador empírico, obtenemos un estimador alternativo para la función cuantil $Q(u)$.

Por otra parte, desde el punto de vista estadístico, como

$$P\{Q(u) \in [X_{(k)}, X_{(k+1)})\} = \binom{n}{k} u^k (1-u)^{n-k}$$

vamos a obtener un L-estimador que coincidirá con el anterior, de manera que la aproximación analítica y estadística van a coincidir.

Como todos los estimadores propuestos para la función cuantil, este estimador también es *sesgado*, pero como veremos primeramente, si la función cuantil $Q(u)$ cumple la condición de Lipschitz, entonces el sesgo es un infinitésimo de orden $O(n^{-1/2})$. Para el estimador tipo núcleo T_n propuesto por Yang (1985), se tiene que el sesgo es un infinitésimo de orden $o(n^{-1/2}) + O(h(n)^2)$ imponiendo seis condiciones, donde $h(n)$ es el parámetro de suavización. Asimismo, bajo ciertas condiciones, Yang (1985) ha probado que $\operatorname{var}(T_n(u)) = o(1/nh(n)^2)$. Nosotros vamos a probar que $E|K_n(u) - Q(u)|$ es un infinitésimo de orden $O(n^{-1/2})$.

2. ESTIMACION DE LA FUNCION CUANTIL POR POLINOMIOS DE KANTOROVIC

Sean $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ los estadísticos de orden de las variables X_i, anteriormente definidas. La función cuantil empírica está definida por:

$$Q_n(u) = X_{(j)}, \quad \text{para} \quad (j-1)/n < u \leqslant j/n \quad j = 1, 2, \ldots, n$$

Para $u = 0$, definimos $Q_n(0) = X_{(0)}$ que se toma igual al mínimo muestral $X_{(1)}$ o a un mínimo cuando sea posible.

Dado que la función cuantil es una función de salto, trataremos de suavizar dicha función, lo mejor posible, mediante una función continua, que va a ser un polinomio de Kantorovic.

Para una función de distribución continua, se sabe que:

$$P(Q(u) \in [X_{(k)}, X_{(k+1)}]) = \binom{n}{k} u^k (1-u)^{n-k}$$

que es independiente de Q. Así un L-estimador natural para $Q(u)$ es

$$Q(u) = \sum_{k=0}^{n} h(X_{(k)}, X_{(k+1)}) \binom{n}{k} u^k (1-u)^{n-k} \qquad u \in [0,1]$$

Para

$$h(X_{(k)}, X_{(k+1)}) = (X_{(k)} + X_{(k+1)})/2$$

obtenemos el estimador

$$K_n(u) = \sum_{k=0}^{n} \left(\frac{X_{(k)} + X_{(k+1)}}{2}\right)\binom{n}{k} u^k (1-u)^{n-k}$$

Por otra parte, si utilizamos la aproximación de $Q(u)$ mediante polinomios de Kantorovic y aproximamos la función cuantil poblacional por la función cuantil modificada de la muestra

$$Q_n(t) = \left[(n+1)(X_{(k+1)} - X_{(k)})\left(t - \frac{k}{n+1}\right) + X_{(k)}\right]$$

para

$$(j-1)/n \leqslant u \leqslant j/n \quad j = 1, 2, \ldots, n$$

obtenemos

$$P_n(Q, u) = (n+1) \sum_{k=0}^{n} \binom{n}{k} u^k (1-u)^{n-k} \int_{k/(n+1)}^{(k+1)/(n+1)} \left[(n+1)(X_{(k+1)} - \right.$$

$$\left. - X_{(k)})\left(t - \frac{k}{n+1} + X_{(k)}\right) dr = \sum_{k=0}^{n} \left(\frac{X_{(k)} + X_{(k+1)}}{2}\right)\binom{n}{k} u^k (1-u)^{n-k}$$

es decir, obtenemos el L-estimador K_n dado anteriormente. Así el estimador K_n obtenido de manera natural basándonos en propiedades estadísticas coincide con la aproximación de $Q(u)$ por polinomios de Kantorovic si se sustituye $Q(t)$ por $Q_n(t)$.

2.1. Insesgadez asintótica

Como todos los estimadores propuestos, K_n es un estimador sesgado, pero como a continuación veremos un sesgo es un infinitésimo de orden $O(n^{-1/2})$.

Teorema 1. Si la función cuantil $Q(u)$ verifica la condición de Lipschitz, entonces

$$|E(K_n(u)) - Q(u)| \leqslant O(n^{-1/2}) \tag{1}$$

Demostración:

Teniendo en cuenta que

$$K_n(u) = \sum_{k=0}^{n} \left(\frac{X_{(K)} + X_{(k+1)}}{2}\right)\binom{n}{k} u^k(1-u)^{n-k} =$$

$$= \frac{1}{2} B_n(u) + \frac{1}{2} B_n^*(u)$$

con

$$B_n(u) = \sum_{k=0}^{n} X_{(k)}\binom{n}{k} u^k(1-u)^{n-k}$$

y

$$B_n^*(u) = \sum_{k=0}^{n} X_{(k+1)}\binom{n}{k} u^k(1-u)^{n-k}$$

Bastará con probar que tanto B_n como B_n^* verifican (1).

Vamos a realizar aquí la demostración para B_n^* y de manera análoga se hace para B_n.

$$E[B_n^*(u)] = \sum_{k=0}^{n} E[X_{(k+1)}]\binom{n}{k} u^k(1-u)^{n-k}$$

y como $E[X_{(r)}] = Q(r/(n+1)) + O(n^{-1/2})$ (según el lema 1 de Muñoz y Fernández (1987)),

$$= \sum_{k=0}^{n} Q\left(\frac{k+1}{n+1}\right)\binom{n}{k} u^k(1-u)^{n-k} + O(n^{-1/2})$$

$$= \bar{B}_n^*(u) + O(n^{-1/2}) \tag{2}$$

El sesgo viene dado por

$$D_n(u) = \bar{B}_n^*(u) - Q(u) + O(n^{-1/2})$$

así

$$|D_n(u)| \leqslant |\bar{B}_n^*(u) - Q(u)| + O(n^{-1/2})$$

Por otra parte, si S_n es una variable aleatoria que sigue una distribución binomial de parámetros n y u, resulta

$$E[Q((S_n+1)/(n+1))] = \sum_{k=0}^{n} Q((k+1)/(n+1))\binom{n}{k}u^k(1-u)^{n-k}$$

y así

$$\begin{aligned} |Q(u) - \bar{B}_n^*(u)| &\leqslant |Q(u) - E[Q(S_n/n)| + \\ &+ |E[Q(S_n/n) - E[Q((S_n + 1)/(n + 1))]| \leqslant \\ &\leqslant E|Q(u) - Q(S_n/n| + E|Q(S_n/n) - Q((S_n + 1)/(n + 1))| \end{aligned}$$

y por la condición de Lipschitz

$$\leqslant M\, E|u - S_n/n| + 1/n$$

Aplicando la desigualdad de Hölder

$$E|u - S_n/n| \leqslant \left\{E\left(\sum_{k=0}^{n}(Y_i - u)/n^2\right\}^{1/2} = [u(1-u)/n]^{1/2} \leqslant [4n]^{-1/2}$$

se concluye la demostración. ■

2.2. Error absoluto medio

Veamos que, bajo ciertas condiciones, el error absoluto medio del estimador K_n es un infinitésimo de orden inferior o igual a $n^{-1/2}$, es decir,

$$E|K_n - Q| \leqslant O(n^{-1/2})$$

Las condiciones que se imponen en el siguiente teorema se pueden considerar bastante restrictivas, sin embargo, resultados similares a éste, se han obtenido bajo condiciones todavía más fuertes para el estimador T_n (ver Yang (1985)).

De todas formas estos resultados se acompañan siempre de estudios mediante simulación de Monte Carlo para comparar el comportamiento de los estimadores en las distribuciones más notables.

Teorema 2. Sea $Q(u)$ una función cuantil continua en $[0, 1]$, que es dos veces diferenciable en $(0, 1)$ con derivadas continuas en $[0, 1]$. Entonces

$$E|K_n - E[K_n]| \leqslant O(n^{-1/2})$$

Demostración:

Como sabemos

$$K_n(u) = \frac{1}{2} B_n(u) + \frac{1}{2} B_n^*(u) \tag{3}$$

de manera que

$$E|K_n - E[K_n]| \leqslant (1/2)E|B_n - E[B_n]| + (1/2)E|B_n^* - E[B_n^*]|$$

basta probar que los dos sumandos de la derecha son equivalentes a infinitésimos de orden $O(n^{-1/2})$.

Sea

$$B_n(u, w) = \sum_{k=0}^{n} X_{(k)}(w)\binom{n}{k} u^k(1 - u)^{n-k} =$$

y como $X_{(k)} = Q[U_{(K)}]$, donde $U_{(k)}$ es k-ésimo estadístico de orden para una distribución uniforme en $(0, 1)$, tenemos

$$= \sum_{k=0}^{n} \left\{Q\left(\frac{k}{n+1}\right) + \left(U_{(k)}(w) - \frac{k}{n+1}\right)Q'\left(\frac{k}{n+1}\right) + \right.$$

$$\left. + \left(U_{(k)}(w) - \frac{k}{n+1}\right)^2 Q''(Z_k(w))\right\}\binom{n}{k} u^k(1 - u)^{n-k}$$

para algún

$$Z_k(w) \in \left[U_{(k)}(w), \frac{k}{n+1}\right] \quad \text{si} \quad U_{(k)}(w) \leqslant \frac{k}{n+1}$$

o

$$Z_k(w) \in \left[\frac{k}{n+1}, U_{(k)}(w)\right] \quad \text{si} \quad U_{(k)}(w) \geqslant \frac{k}{n+1}$$

Por tanto,

$$B_n(u,w) - E[B_n(u)] = \sum_{k=0}^{n} \left\{ \left[U_{(k)}(w) - \frac{k}{n+1} \right] Q'\left(\frac{k}{n+1}\right) + \right.$$

$$\left. + Q\left[U_{(k)}(w) - \frac{k}{n+1} \right]^2 Q''(Z_k(w)) \right\} \binom{n}{k} u^k (1-u)^{n+1-k} + O(n^{-1/2})$$

de donde

$$|B_n(u,w) - E[B_n(u)]| \leqslant \sum_{k=0}^{n} \left| U_{(k)}(w) - \frac{k}{n+1} \right| M \binom{n}{k} u^k (1-u)^{n-k} +$$

$$+ \sum_{k=0}^{n} \left[U_{(k)}(w) - \frac{k}{n+1} \right]^2 M \binom{n}{k} u^k (1-k)^{n-k} + O(n^{-1/2})$$

pues $Q''(u)$ está acotada en $[0,1]$.

Entonces

$$E|B_n(u) - E[B_n(u)]| \leqslant \frac{1}{n+1} \sum_{k=0}^{n} \left[\frac{k(n-k+1)}{n+2} \right]^{1/2} M \binom{n}{k} u^k (1-u)^{n-k} +$$

$$+ \sum_{k=0}^{n} \left[\frac{k(n-k+1)}{(n+2)(n+1)^2} \right] M \binom{n}{k} u^k (1-u)^{n-k} + O(n^{-1/2})$$

así obtenemos

$$E|B_n(u) - E[B_n(u)]| \leqslant An^{-1/2} + O(n^{-1/2}), \qquad A \in R$$

De forma análoga se puede hacer la demostración para B_n^*. ■

Corolario 3. Sea $Q(u)$ una función cuantil dos veces diferenciable en $(0,1)$ con derivadas continuas en $[0,1]$. Entonces

$$E|K_n - Q| \leqslant O(n^{-1/2})$$

Demostración:

Como

$$E|K_n - Q| \leqslant E|K_n - E(K_n)| + |E(K_n) - Q|$$

es una consecuencia de los teoremas 1 y 2. ■

3. ESTUDIO COMPARATIVO DE LOS ESTIMADORES DE BERNSTEIN Y KANTOROVIC POR SIMULACION DE MONTE CARLO

Parzen (1979) clasificó las leyes de probabilidad de acuerdo a su comportamiento límite, cuando u tiende a 0 ó 1, al realizar un examen de sus funciones densidad-cuantil.

Así surgen tres tipos de distribuciones de probabilidad:

— De colas cortas o tipo limitado.
— De colas medianas o tipo exponencial.
— De colas largas o tipo Cauchy.

Seleccionamos para el estudio las distribuciones: Uniforme, Exponencial, Cauchy, Weibull, Pareto y dos distribuciones de la familia de distribuciones Lambda de Ramberg-Schmeiser-Tukey; esta selección fue realizada en base a la clasificación de Parzen antes mencionada.

Para llevar a cabo el estudio comparativo de los estimadores B_n y K_n, donde B_n es el estimador basado en polinomios Bernstein, realizamos 1.000 simulaciones basadas en muestras de tamaño $n = 99$ para diferentes distribuciones y poder comparar sus errores absolutos medios.

TABLA 1

Valores de EARM (valores medios del EAR con 1000 simulaciones)

							Familia Lambda $\lambda_1=0$ $\lambda_2=.2$ $\lambda_3=2$	
	Uniforme	Normal	Exponencial	Cauchy	Weibull	Pareto	$\lambda_2=2$	$\lambda_4=4$
B_n	.9082	.8796	1.1459	.9282	.8881	.8294	.9085	1.019
K_n	.9177	.8701	1.1188	.9444	.9106	.801	.9181	.926

Investigamos el comportamiento de K_n como estimador de la función cuantil, comparando sus valores con los de la función cuantil muestral, mediante la expresión:

$$\mathrm{EAR} = \frac{\sum_{k=1}^{n} \left| Q\left(\frac{k}{n}\right) - K_n\left(\frac{k}{n}\right) \right|}{\sum_{k=1}^{n} \left| Q\left(\frac{k}{n}\right) - X_{(k)} \right|}$$

Estudio similar se efectúa para el estimador B_n (con polinomios de Bernstein). Los resultados del EARM se presentan en la tabla 1.

De la tabla se desprende que el estimador K_n tiene un EARM similar al estimador obtenido por polinomios de Bernstein, y que con cualquiera de estos estimadores se obtiene una reducción sustancial del EARM con respecto a la función cuantil muestral, salvo para el caso de la distribución exponencial negativa. También se puede decir que el estimador K_n es más apropiado que el estimador B_n para distribuciones de cola mediana (exponencial negativa, normal) y para las asimétricas como la familia Lambda con $\lambda_3 = 2$ y $\lambda_4 = 4$ y la exponencial negativa.

4. ESTIMACION DE LA FUNCION CUANTIL-DENSIDAD

La función cuantil-densidad viene dada por la derivada de la función cuantil $q(u) = Q'(u)$ y se puede poner como

$$q(u) = 1/f(Q(u))$$

donde f es la función de densidad.

Tukey (1965) fue el primero que le dio nombre a esta función, llamándola función «sparsity» y Parzen (1979) le concede gran importancia en el análisis estadístico de los datos.

Babu (1986) y Falk (1986) han propuesto estimadores para dicha función.

Como hemos visto, el estimador $K_n(u)$ es el estimador natural de $Q(u)$, tanto desde el punto de vista de la aproximación funcional como desde el punto de vista estadístico; parece razonable estimar la función $q(u)$ basándonos también en su polinomio de Kantorovic,

$$P_n(q, u) = (n+1) \sum_{k=0}^{n} \binom{n}{k} u^k (1-u)^{n-k} \int_{k/(n+1)}^{(k+1)/(n+1)} q(t)\, dt =$$

$$= (n+1) \sum_{k=0}^{n} \binom{n}{k} u^k (1-u)^{n-k} \left[Q\left(\frac{k+1}{n+1}\right) - Q\left(\frac{k}{n+1}\right) \right]$$

Además, desde el punto de vista de la teoría de la aproximación, se tiene que (Maier (1978)):

$$\int_0^1 |P_n(q, t) - q(t)| \, dt = O(1/n)$$

Sustituyendo

$$Q\left(\frac{k+1}{n+1}\right) \quad \text{y} \quad Q\left(\frac{k}{n+1}\right)$$

por los correspondientes valores empíricos

$$Q_n\left(\frac{k+1}{n+1}\right) = X_{(k+1)} \qquad \text{y} Q_n\left(\frac{k}{n+1}\right) = X_{(k)}$$

obtenemos como estimador de $q(u)$

$$(n+1) \sum_{k=0}^{n} \binom{n}{k} u^k (1-u)^{n-k} [X_{(k+1)} - X_{(k)}] =$$

$$= \sum_{k=0}^{n+1} \binom{n+1}{k} X_{(k)} \frac{d}{du} [u^k (1-u)^{n-k+1}] = \frac{d}{du} B_n(u)$$

que corresponde a la derivada del estimador $B_n(u)$ correspondiente a los polinomios de Bernstein y no a la de los polinomios de Kantorovic como cabía esperar. Otra cuestión importante que dejamos abierta es el estudio asintótico de estos estimadores bajo condiciones razonables para la distribución, así como un estudio comparativo de su comportamiento frente a otros estimadores.

Agradecimientos

Los autores expresan su agradecimiento a los referees por todas las sugerencias y observaciones dadas, que han contribuido a una mejor presentación del trabajo.

REFERENCIAS

BABU, G. J. (1986): «Estimation of density quantile function», *Sankhyā,* vol. 48, Series A, Pt. 2, pp. 142-149.

FALK, M. (1984): «Relative deficiency of kernel estimators of quantile», *Ann. Math. Statist.,* vol. 12, n.º 1, pp. 261-268.

FALK, M. (1986): «On the estimation of the quantile density function», *Statistics & Probability Letters,* 4, pp. 69-73. North-Holland.

HARRELL, F. E., y DAVIS, C. E. (1982): «A new distribution-free quantile estimator», *Biometrika,* vol. 69, n.º 3, pp. 635-640.

KAIGH, W. D., y LACHENBRUCH, P. A. (1982): «A generalized quantile estimator», *Commun. Statist. - Theor. Meth.,* vol. 11, n.º 19, pp. 2217-2238.

KAIGH, W. D. (1983): «Quantile interval estimation», *Commun. Statist. - Theor. Meth.,* vol. 12, n.º 21, pp. 2427-2443.

MAIER, V. (1978): «The L_1 saturation class of the Kantorovic operator», *Journal of approximation theory,* vol. 22, pp. 223-232.

MUÑOZ, J., y FERNANDEZ, A. (1987): «Estimating the quantile function by Bernstein polynomials», *Computational Statistics & Data Analysis,* vol. 5, pp. 391-397.

PARZEN, E. (1979): «Nonparametric statistical data modelling», *J. Amer. Statist. Assoc.,* vol. 74, pp. 105-131.

REISS, R. D. (1980): «Estimation of quantiles in certain nonparametric models», *Ann. Math. Statist.,* vol. 8, *n.º* 1, pp. 87-105.

VERAVERBEKE, N. (1987): «A kernel-type estiamtor for generalized quantiles», *Statistics & Probability Letters,* vol. 5, pp. 175-180.

YANG, S. (1985): «A smooth nomparametric estimator of a quantile function», *J. Amer. Statist. Assoc.,* vol. 80, n.º 392, pp. 1004-1011.