ANNALES
UNIVERSITATIS MARIAE CURIE - SKŁODOWSKA
LUBLIN - POLONIA
VOL. LI. 1, 2 SECTIO A 1997

AGATA BORATYŃSKA and RYSZARD ZIELIŃSKI (Warszawa)

Asymptotic Behavior of Sample Median in a Parametric Model

ABSTRACT. Under rather mild assumptions, sample quantiles are strongly consistent estimators of the appropriate population quantiles. A difficulty in statistical applications of the statement lies in that the convergence is not uniform even in reasonably small statistical models.

1. Introduction. It is well known that if the distribution function F is strictly increasing in a neighbourhood of the point x_p such that $F(x_p) = p$, then the appropriately defined sample quantile of order p is a strongly consistent estimator of the population quantile x_p (Bartoszewicz 1996, Serfling 1980). Strong consistency of the sample quantile is, however, of a rather limited usefulness for statistics because the convergence is not uniform, even in some "small" parametric statistical models.

To demonstrate that, in what follows we consider the problem of estimating the median in the family $\mathcal{F} = \{F_\alpha : 0 < \alpha < 1/2\}$ of distributions F_α on the interval $[0, 1]$, where

$$F_\alpha(x) = \begin{cases} 0, & \text{if } x \le 0; \\ \frac{1}{2}(2x)^\alpha, & \text{if } 0 < x \le \frac{1}{2}; \\ 1 - \frac{1}{2}[2(1-x)]^{1/\alpha}, & \text{if } \frac{1}{2} < x \le 1; \\ 1, & \text{if } x > 1. \end{cases}$$

Observe that for every F_α the median is equal to 1/2. Given a sample $X_1, X_2, \dots, X_n$ and its order statistics $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, let M_n be the sample median

$$M_n = \begin{cases} (X_{k:n} + X_{k+1:n})/2, & \text{if } n=2k, \\ X_{k:n}, & \text{if } n = 2k-1. \end{cases}$$

We show that $E_\alpha M_n - 1/2 = n^{-1/2}C(\alpha) + O(n^{-1})$ and that for every $C > 0$ one can choose $\alpha \in (0, 1/2)$ such that $C(\alpha) > C$. It appears that even in very large samples the error of the estimate of the population median by the sample median might be very large and a statistician is not able to predict it in advance.

2. Results.

Theorem 1. *Let $F_\alpha \in \mathcal{F}$. Then for large n*

$$E_\alpha M_n - \frac{1}{2} = \sum_{i=1}^{\infty} A_i B_i,$$

where

$$A_i = \frac{1}{i!}\left(\frac{1}{\alpha}(\frac{1}{\alpha}-1)\dots(\frac{1}{\alpha}-i+1) - \alpha(\alpha-1)\dots(\alpha-i+1)\right),$$

$$B_1 = -\frac{1}{4\sqrt{k\pi}} + O(k^{-3/2}),$$

$$B_2 = \frac{1}{8(k+\frac{1}{2})} + O(k^{-2})$$

and

$$B_i = \frac{i-1}{2k+i-1} B_{i-2}, \quad i \geq 3.$$

Proof. Let $X_1, \dots, X_n$ be a sample from F_α. Then

$$E_\alpha M_n = 0.5 + I(k, \alpha)$$

where

$$I(k,\alpha) = \frac{1}{2}\frac{\Gamma(2k)}{\Gamma^2(k)} \int_0^{\frac{1}{2}} \left[(2t)^{1/\alpha} - (2t)^\alpha\right] t^{k-1}(1-t)^{k-1} dt$$

$$= \left(\frac{1}{2}\right)^{2k} \frac{\Gamma(2k)}{\Gamma^2(k)} \int_0^1 \left[s^{1/\alpha} - s^\alpha\right] s^{k-1}(2-s)^{k-1} ds.$$

The Taylor expansion of the functions s^α and $s^{1/\alpha}$ at the point $s=1$ gives us

$$s^\alpha = 1+\sum_{i=1}^{\infty}\frac{\alpha(\alpha-1)\ldots(\alpha-i+1)}{i!}(-1)^i(1-s)^i,$$

$$s^{1/\alpha} = 1+\sum_{i=1}^{\infty}\frac{\frac{1}{\alpha}(\frac{1}{\alpha}-1)\ldots(\frac{1}{\alpha}-i+1)}{i!}(-1)^i(1-s)^i.$$

Using the identity $s^{k-1}(2-s)^{k-1}=[1-(1-s)^2]^{k-1}$ we obtain

$$I(k,\alpha)=\left(\frac{1}{2}\right)^{2k+1}\frac{\Gamma(2k)}{\Gamma^2(k)}\sum_{i=1}^{\infty}(-1)^iA_i\int_0^1 s^{(i-1)/2}(1-s)^{k-1}ds$$

with A_i as above. Hence

$$I(k,\alpha)=\sum_{i=1}^{\infty}(-1)^iA_i\left(\frac{1}{2}\right)^{2k+1}\frac{\Gamma(2k)\Gamma(\frac{i+1}{2})}{\Gamma(k)\Gamma(k+\frac{i+1}{2})}.$$

By Stirling's formula in the form

$$i! = i^i\sqrt{2\pi i}e^{-i}e^{\theta/12i},\quad \text{where} \quad \theta\in(0,1)$$

we obtain for i odd

$$B_i=\left(\frac{1}{2}\right)^{2k+1}\frac{\Gamma(2k)\Gamma(\frac{i+1}{2})}{\Gamma(k)\Gamma(k+\frac{i+1}{2})}=\frac{(\frac{i-1}{2})!}{4\sqrt{\pi k^i}}+O\left(k^{-\frac{i+2}{2}}\right).$$

By the formula

$$\Gamma(i+\frac{1}{2})=\frac{(2i)!\sqrt{\pi}}{2^{2i}i!}=\frac{(2i-1)!!\sqrt{\pi}}{2^i}$$

we have for i even

$$B_i=\left(\frac{1}{2}\right)^{2k+1}\frac{\Gamma(2k)\Gamma(\frac{i+1}{2})}{\Gamma(k)\Gamma(k+\frac{i+1}{2})}=\frac{(i-1)!!}{2^{\frac{i}{2}+2}(k+\frac{1}{2})^{\frac{i}{2}}}+O\left(k^{-\frac{i+2}{2}}\right).$$

Substituting these expressions for $i=1$ and $i=2$ to $I(k,\alpha)$ we obtain the assertion.

■

3. A conclusion. As a simple conclusion of the Theorem we obtain that for every $A > 0$ there exists $F \in \mathcal{F}$ such that

$$E_F M_n - m_F = -\frac{A}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

A deeper insight leads us to the conclusion that for every $\varepsilon > 0$ and for every n there exists a distribution $F_\alpha \in \mathcal{F}$ such that

$$E_{F_\alpha} - m_{F_\alpha} > \frac{1}{4} - \varepsilon$$

which is really "very large" when compared with how large is the support of the family $\mathcal{F}$.

References

[1] Bartoszewicz, J., *Wykłady ze statystyki matematycznej*, PWN Warszawa, 1996.

[2] Serfling, R.J., *Approximation theorems of mathematical statistics*, John Wiley & Sons, Inc., 1980.

Instytut Matematyki Stosowanej
Uniwersytet Warszawski
Warszawa, Poland

Instytut Matematyki PAN
Warszawa, Poland

received January 30, 1997