APPLICATIONES MATHEMATICAE
30,3 (2003), pp. 251–255

Ryszard Zieliński (Warszawa)

CONSTRUCTING MEDIAN-UNBIASED ESTIMATORS IN ONE-PARAMETER FAMILIES OF DISTRIBUTIONS VIA STOCHASTIC ORDERING

Abstract. If $\theta \in \Theta$ is an unknown real parameter of a given distribution, we are interested in constructing an exactly median-unbiased estimator $\widehat{\theta}$ of θ, i.e. an estimator $\widehat{\theta}$ such that a median $\mathrm{Med}(\widehat{\theta})$ of the estimator equals θ, uniformly over $\theta \in \Theta$. We shall consider the problem in the case of a fixed sample size n (nonasymptotic approach).

1. The model. Let $\mathcal{F}$ be a one-parameter family $\{F_\theta : \theta \in \Theta\}$ of distributions, where Θ is a (finite or infinite) interval on the real line. The family $\mathcal{F}$ is assumed to be a family of distributions with continuous and strictly increasing distribution functions and stochastically ordered by θ so that for every $x \in \operatorname{supp} \mathcal{F} = \bigcup_{\theta\in\Theta} \operatorname{supp} F_\theta$ and for every $q \in (0,1)$, the equation $F_\tau(x) = q$ has exactly one solution in $\tau \in \Theta$. Given a sample $X_1, \ldots, X_n$ from an F_θ, we are interested in median-unbiased estimation of θ; here n is a fixed integer (nonasymptotic approach).

This model covers a wide range of one-parameter families of distributions.

Example 1. The family of uniform distributions on $(\theta, \theta + 1)$, with $-\infty < \theta < \infty$.

Example 2. The family of power distributions on $(0, 1)$ with distribution functions $F_\theta(x) = x^\theta$, $\theta > 0$.

Example 3. The family of gamma distributions with probability distribution functions (pdf) of the form

2000 *Mathematics Subject Classification*: 62F10, 62G05.

Key words and phrases: median-unbiased estimators, stochastic ordering, optimal nonparametric quantile estimators.

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-x}, \qquad x > 0,$$

with $\alpha > 0$.

Example 4. Consider the family of Cauchy distributions with pdf of the form

$$g_\lambda(y) = \frac{1}{\lambda}\, \frac{1}{1+(y/\lambda)^2}, \qquad -\infty < y < \infty,$$

and distribution function of the form

$$G_\lambda(y) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{\lambda}$$

with $\lambda > 0$. The family$\{F_\lambda : \lambda > 0\}$ of distribution functions of $X = |Y|$ with

$$F_\lambda(x) = \frac{2}{\pi} \arctan \frac{x}{\lambda}$$

satisfies the model assumptions so that the problem of estimating λ from a sample $Y_1, \ldots, Y_n$ amounts to estimating λ from the sample $X_1, \ldots, X_n$ with $X_i = |Y_i|$, $i = 1, \ldots, n$.

Example 5. Consider the one-parameter family of Weibull distributions with distribution functions of the form

$$G_\alpha(y) = 1 - e^{-y^\alpha}, \qquad y > 0,\ \alpha > 0,$$

and let $X = \max\{Y, 1/Y\}$. The family $\{F_\alpha : \alpha > 0\}$ of distributions of X with distribution functions

$$F_\alpha(x) = e^{-x^{-\alpha}} - e^{-x^\alpha}, \qquad x > 1,\ \alpha > 0,$$

satisfies the model assumptions.

Example 6 (Estimating the characteristic exponent of a symmetric α-stable distribution). Consider the one-parameter family of α-stable distributions with characteristic functions $\exp\{-t^\alpha\}$, $0 < \alpha \le 2$. The problem is to construct a median-unbiased estimator of α. Some related results can be found in Fama and Roll (1971) and Zieliński (2000). We shall not consider this problem in this note because it needs (and deserves) a special treatment and will be discussed in detail elsewhere.

Generally, every family of distributions F_θ with continuous and strictly increasing F_θ and a location parameter θ satisfies the model assumptions. Similarly, every family of continuous and strictly increasing distributions on $(0, \infty)$ with a scale parameter fits into the model.

2. The method. The method consists in:

1) For a given $q \in (0,1)$, estimating the qth quantile of the underlying distribution in a nonparametric setup; denote the estimator by $\widehat{x}_q$. A restriction is that for a fixed n a median-unbiased estimator of the qth quantile exists iff $\max\{q^n, (1-q)^n\} \le \frac{1}{2}$.

2) Solving the equation $F_\tau(\widehat{x}_q) = q$ with respect to τ. The solution, to be denoted by $\widehat{\theta}_q$, is considered as an estimator of θ. The solution of the equation $F_\tau(x) = q$ with respect to τ will be denoted by $\widehat{\theta}_q(x)$ so that $\widehat{\theta}_q = \widehat{\theta}_q(\widehat{x}_q)$.

In the model, if $\widehat{x}_q$ is a median-unbiased estimator of x_q then, due to monotonicity of $\widehat{\theta}_q(x)$ with respect to x, $\widehat{\theta}_q$ is a median-unbiased estimator of θ. What is more, if $\widehat{x}_q$ is the median-unbiased estimator of x_q the most concentrated around x_q in the class of all median-unbiased estimators which are equivariant with respect to monotone transformations of data (briefly: the *best estimator*) then, due to monotonicity again, $\widehat{\theta}_q$ is the median-unbiased estimator of θ most concentrated around θ in the class of all median-unbiased estimators which are equivariant with respect to monotone transformations of data (briefly: the *best estimator*).

Given $q \in (0,1)$, the best estimator $\widehat{x}_q$ of x_q is given by the formula

$$\text{[E]} \qquad \widehat{x}_q = X_{k:n}\mathbf{1}_{(0,\lambda]}(U) + X_{k+1:n}\mathbf{1}_{(\lambda,1)}(U)$$

where $X_{k:n}$ is the the kth order statistic, $X_{1:n} \le \ldots \le X_{n:n}$, from the sample $X_1, \ldots, X_n$ and

$$\begin{aligned} k &= k(q) \\ &= \text{the unique integer such that } Q(k;n,q) \ge 1/2 \ge Q(k+1;n,q), \end{aligned}$$

$$\lambda = \lambda(q) = \frac{1/2 - Q(k+1;n,q)}{Q(k;n,q) - Q(k+1;n,q)},$$

$$Q(k;n,q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j};$$

here U is a random variable uniformly distributed on $(0,1)$ and independent of the sample $X_1, \ldots, X_n$ (Zieliński 1988).

When estimating θ in a parametric model $\{F_\theta : \theta \in \Theta\}$, the problem is to choose an "optimal" q. To define a criterion of optimality (or "an ordering in the class $\widehat{\theta}_q$, $0 < q < 1$, of estimators"), recall (e.g. Lehmann 1983, Sec. 3.1) that a median-unbiased estimator $\widehat{\theta}$ of a parameter θ is the one for which

$$\text{[K]} \qquad E_\theta|\widehat{\theta} - \theta| \le E_\theta|\widehat{\theta} - \theta'| \quad \text{for all } \theta, \theta' \in \Theta$$

(the estimator is closer to the "true" value $\theta \in \Theta$ than to any other value

$\theta' \in \Theta$ of the parameter). According to this property, we shall choose q_{opt} as the one with minimal risk under the loss function $|\widehat{\theta} - \theta|$, i.e. such that

$$E_\theta|\widehat{\theta}_{q_{\mathrm{opt}}} - \theta| \le E_\theta|\widehat{\theta}_q - \theta|, \quad 0 < q < 1,$$

for all $\theta \in \Theta$, if possible.

Using the fact that $\theta \in \Theta$ generates the stochastic ordering of the family $\{F_\theta : \theta \in \Theta\}$, we shall restrict our attention to finding q_{opt} which satisfies criterion [K] for a fixed θ, for example $\theta = 1$; then the problem reduces to minimizing

$$R(q) = E|\widehat{\theta}_q - 1|$$

with respect to $q \in (0,1)$, where $E = E_1$.

By [E] we obtain

$$R(q) = \lambda(q)E|\widehat{\theta}_q(X_{k(q):n}) - 1| + (1 - \lambda(q))E|\widehat{\theta}_q(X_{k(q)+1:n}) - 1|.$$

It is obvious that q_{opt} satisfies

$$\lambda(q_{\mathrm{opt}}) = 1$$

and

$$E|\widehat{\theta}_{q_{\mathrm{opt}}}(X_{k(q_{\mathrm{opt}}):n}) - 1| \le E|\widehat{\theta}_q(X_{k(q):n}) - 1|, \quad 0 < q < 1.$$

By the very definition of λ, $\lambda(q) = 1$ iff $q \in \{q_1, \ldots, q_n\}$ where q_i satisfies $Q(i; n, q_i) = 1/2$, and the problem reduces to finding the smallest element of the finite set

$$\{E|\widehat{\theta}_{q_i}(X_{i:n}) - 1| : i = 1, \ldots, n\}.$$

If $X_{k:n}$ is the kth order statistic from the sample $X_1, \ldots, X_n$ from a distribution function F, then $U_{k:n} = F(X_{k:n})$ is the kth order statistic from the sample $U_1, \ldots, U_n$ from the uniform distribution on $(0,1)$, which gives us

$$\begin{aligned} E|\widehat{\theta}_{q_i}(X_{i:n}) - 1| &= E|\widehat{\theta}_{q_i}(F^{-1}(U_{i:n})) - 1| \\ &= \frac{\Gamma(n)}{\Gamma(i)\Gamma(n-i+1)} \int_0^1 |\widehat{\theta}_{q_i}(F^{-1}(t)) - 1| t^{i-1}(1-t)^{n-i}\, dt. \end{aligned}$$

The latter can be easily calculated numerically.

3. Applications

Example 1A. In the case of uniform distributions on $(\theta, \theta + 1)$, the solution τ of the equation $F_\tau(\widehat{x}_q) = q$ takes the form $\tau = \widehat{x}_q - q$. For example for $n = 10$ the best estimator is $X_{1:10} - 0.067$ or $X_{10:10} - 0.933$.

Example 2A. In the case of power distributions, the best estimator is the (unique) solution, with respect to τ, of the equation $\widehat{x}_q^\theta = q_{\mathrm{opt}}$; for $n = 10$ the estimator takes the form $-1.81854/\mathrm{Log}[X_{2:10}]$.

Example 3A. In the case of gamma distributions, the best estimator is the (unique) solution, with respect to τ, of the equation $F_\tau(\widehat{x}_q) = q_{\rm opt}$; for $n = 10$ this equation takes the form $F_\tau(X_{3:10}) = 0.2586$.

Example 4A. In the case of Cauchy distributions the solution τ of the equation $F_\tau(\widehat{x}_q) = q$ can be written in the form

$$\tau = \frac{\widehat{x}_q}{\tan\left(\frac{\pi}{2}q\right)}$$

and for $n = 10$ the best estimator is $1.16456 \cdot X_{5:10}$.

Example 5A. In the case of Weibull distributions, the best estimator is the (unique) solution, with respect to τ, of the equation $F_\tau(\widehat{x}_q) = q_{\rm opt}$; for $n = 10$ this equation takes the form $F_\tau(X_{8:10}) = 0.7414$, which gives us the optimal estimator $0.302/\mathrm{Log}(X_{8:10})$.

Acknowledgements. The research was supported by Grant KBN 2 PO3A 033 17.

References

E. F. Fama and R. Roll (1971), *Parameter estimates for symmetric stable distributions*, J. Amer. Statist. Assoc. 66 (334), 331–338.

E. L. Lehmann (1983), *Theory of Point Estimation*, Wiley.

R. Zieliński (1988), *A distribution-free median-unbiased quantile estimator*, Statistics 19, 223–227.

R. Zieliński (2000), *A median-unbiased estimator of the characteristic exponent of a symmetric stable distribution*, ibid. 34, 353–355.

Institute of Mathematics
Polish Academy of Sciences
P.O. Box 137
00-956 Warszawa 10, Poland
E-mail: rziel@impan.gov.pl

Received on 19.11.2002;
revised version on 7.3.2003 (1663)