

CONSTRUCTING MEDIAN-UNBIASED ESTIMATORS
IN ONE-PARAMETER FAMILIES OF DISTRIBUTIONS
via OPTIMAL NONPARAMETRIC ESTIMATION AND STOCHASTIC ORDERING

Ryszard Zieliński

Inst. Math. Polish Acad. Sc.
P.O.Box 137 Warszawa, Poland
e-mail: rziel@impan.gov.pl

ABSTRACT

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

Mathematics Subject Classification: 62F10, 62G05

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

1. THE MODEL

Let \mathcal{F} be a one-parameter family of distributions $\{F_\theta : \theta \in \Theta\}$, where Θ is an (finite or not) 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 $x \in \text{supp } \mathcal{F} = \bigcup_{\theta \in \Theta} \text{supp } F_\theta$ and for every $q \in (0, 1)$, the equation $F_\tau(x) = q$ has exactly one solution in $\tau \in \Theta$. It is obvious that the solution depends monotonically both on x and q . Given a sample X_1, X_2, \dots, X_n from an F_θ , we are interested in a median-unbiased estimation of θ ; here n is a fixed integer (non-asymptotic approach).

The model represents 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 of the form

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

with $\alpha > 0$.

Example 4. Consider the family of Cauchy distributions with probability distribution function of the form

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

and distribution function of the form

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

with $\lambda > 0$. The family of distribution functions $\{F_\lambda, \lambda > 0\}$ 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, Y_2, \dots, Y_n amounts to that from the sample X_1, X_2, \dots, X_n with $X_i = |Y_i|$, $i = 1, 2, \dots, n$.

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

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

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

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

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 $\exp\{-t^\alpha\}$, $0 < \alpha \leq 2$. The problem is to construct a median-unbiased estimator of α . Some related results one can find in Fama and Roll (1971) and Zieliński (2000). We shall not consider the problem in this note because it needs (and it deserves) a special treatment and will be discussed in details elsewhere.

Generally: every family of distributions F_θ with continuous and strictly increasing F_θ and a location parameter θ (i.e. $F_\theta(x) = F_0(x - \theta)$) satisfies the model assumptions. Similarly, every family of continuous and strictly increasing distributions on $(0, +\infty)$ with a scale parameter θ (i.e. $F_\theta(x) = F_1(x/\theta)$) suits the model.

2. THE METHOD

The method consists in

1) for a given $q \in (0, 1)$, estimating the q -th quantile (the quantile of order q) of the underlying distribution in a non-parametric setup; denote the estimator by \hat{x}_q . A restriction is that for a fixed n a median-unbiased estimator of the q -th quantile exists iff $\max\{q^n, (1 - q)^n\} \leq \frac{1}{2}$; in our approach the restriction does not play any role.

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

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

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

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

where $X_{k:n}$ is the the k -th order statistic $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ from the sample X_1, X_2, \dots, X_n and

$$k = k(q)$$

$$= \text{the unique integer such that } Q(k; n, q) \geq \frac{1}{2} \geq Q(k+1; n, q),$$

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

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

here U is a random variable uniformly distributed on $(0, 1)$ and independent of the sample X_1, X_2, \dots, 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 $\hat{\theta}_q$, $0 < q < 1$, of estimators"), let us recall (e.g. Lehmann 1983, Sec. 3.1) that a median-unbiased estimator $\hat{\theta}$ of a parameter θ is that for which

$$[K] \quad E_\theta |\hat{\theta} - \theta| \leq E_\theta |\hat{\theta} - \theta'| \quad \text{for all } \theta, \theta' \in \Theta$$

(the estimator is closer to the "true" value $\theta \in \Theta$ than to any other value $\theta' \in \Theta$ of the parameter). According to the property, we shall choose q_{opt} as that with minimal risk under the loss function $|\hat{\theta} - \theta|$, i.e. such that

$$E_\theta |\hat{\theta}_{q_{opt}} - \theta| \leq E_\theta |\hat{\theta}_q - \theta|, \quad 0 < q < 1$$

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 $\theta = 1$ (if θ is a scale or a shape parameter) or $\theta = 0$ if θ is a location parameter; then the problem reduces to minimizing

$$R(q) = E|\hat{\theta}_q - 1| \quad \text{or} \quad R(q) = E|\hat{\theta}_q|$$

with respect to $q \in (0, 1)$, where $E = E_1$ or $E = E_0$, respectively. Formulas below are given for the case $\theta = 1$; the case of $\theta = 0$ can be treated in full analogy.

By [E] we obtain

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

By the fact that $R(q)$ is a convex combination of two quantities, it is obvious that q_{opt} satisfies

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

and

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

By the very definition of λ , $\lambda(q) = 1$ iff $q \in \{q_1, q_2, \dots, q_n\}$ where q_i satisfies $Q(i; n, q_i) = \frac{1}{2}$, and the problem reduces to finding the smallest element of the finite set

$$\{E|\hat{\theta}_{q_i}(X_{i:n}) - 1|, \quad i = 1, 2, \dots, n\}$$

If $X_{k:n}$ is the k -th order statistic from the sample X_1, X_2, \dots, X_n from a distribution function F , then $U_{k:n} = F(X_{k:n})$ is the k -th order statistic from the sample U_1, U_2, \dots, U_n from the uniform distribution on $(0, 1)$ which gives us

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

The latter can be easily calculated numerically. For numerical integration it should be taken into account that $\hat{\theta}_{q_i}(F^{-1}(U_{i:n})) \geq 1$ iff $0 < t \leq q_i$ or $q_i \leq t < 1$.

3. APPLICATIONS

Example 1A. In the case of uniform distributions on $(\theta, \theta + 1)$, the solution τ of the equation $F_\tau(\hat{x}_q) = q$ takes on the form $\tau = \hat{x}_q - q$. 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 given as the (unique) solution, with respect to τ , of the equation $\hat{x}_q^\theta = q_{opt}$ which, for example in the case with $n = 10$ takes on the form $-1.81854/\text{Log}[X_{2:10}]$.

Example 3A. In the case of gamma distributions, the best estimator is given as the (unique) solution, with respect to τ , of the equation $F_\tau(\hat{x}_q) = q_{opt}$ which, for example in the case with $n = 10$ takes on the form $F_\tau(X_{3:10}) = 0.2586$; we are not able to give an explicit formula for the estimator here.

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

$$\tau = \frac{\hat{x}_q}{\text{tg}(\frac{\pi}{2}q)}$$

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

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

3. ACKNOWLEDGEMENT

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

REFERENCES

Fama, E.F. and Roll, R. (1971): Parameter estimates for symmetric stable distributions, *JASA*, 66(334), 331-338

Lehmann, E.L. (1983): Theory of point estimation. Wiley

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

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