STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 362

(SP 95)

Bias in estimation from type I censored samples

W.R. van Zwet

Reprinted from

Statistica Neerlandica 20(1966), p 143 - 148

Bias in estimation from type I censored samples

by W. R. van Zwet *)

UDC 519.21

Summary

Let $\mathscr{P}$ be a family of probability distributions on R^1. This paper raises the question whether a parameter $\theta = \theta(P)$, $P \in \mathscr{P}$, is estimable on the basis of a type I censored sample (i.e. censored on a fixed set C). Two theorems are given that state conditions on θ and C that ensure that θ is not estimable. The results are applied to estimation problems for the normal and POISSON *distributions; it turns out that unbiased estimation is impossible in the majority of practical cases.*

1. Introduction

Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ be independent and identically distributed random variables **) with common distribution P belonging to a family $\mathscr{P}$ of distributions on R^1. Furthermore let C be a non-empty Borel-measurable proper subset of R^1, and suppose that we observe the random variable $\underline{x}_i$ only if it assumes a value $x_i \in C^c = R^1 - C$. If $x_i \in C$ we do not observe the value x_i but only the event $\{x_i \in C\}$. A set of n such observations on $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ is called a type I censored sample (censored on the set C).

Let θ be a real valued parameter for the family $\mathscr{P}$, i.e. a real valued function $\theta = \theta(P)$ defined on $\mathscr{P}$. The question then arises whether θ is estimable (i.e. possesses an unbiased estimate) on the basis of a type I censored sample. This problem does not seem to have received much attention. In fact, most of the work concerning estimation from type I censored samples seems to center around maximum likelihood estimation where the property of small sample unbiasedness does not play an essential part.

Still the question may be of interest for the following reason. J. W. TUKEY [2] and W. L. SMITH [1] have shown that if for a given estimation problem a sufficient statistic exists, then essentially the same statistic is sufficient for the same estimation problem on the basis of a truncated sample. Since this result may easily be carried over to the case of type I censoring of independent random variables, it will in many cases be fairly simple to find a sufficient statistic for the censored case. Therefore, if one could find an unbiased estimator for a type I censored estimation problem, the problem might conceivably be attacked by means of the RAO-BLACKWELL theorem instead of maximum likelihood methods.

*) University of Leiden and Mathematical Centre, Amsterdam.

**) We denote random variables by underlining their symbols.

These hopes will be rudely shattered in what follows. Without even attempting to give a complete analysis of the problem we shall show by a number of very simple remarks that unbiased estimation is impossible in the majority of practical cases.

2. Sufficient conditions for bias

Since $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ are independent and identically distributed estimability of θ would imply the existence of an unbiased estimator of θ that is invariant under permutations of $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$. Obviously this unbiased estimator has finite expectation for all $P \in \mathscr{P}$. We may therefore confine our search for unbiased estimates of θ to estimates of the following form. Let $t_k(x_1, x_2, \cdots, x_k)$ denote the estimate of θ if $(n - k)$ observations are censored and the remaining k variables assume the values $x_1, x_2, \cdots, x_k \, \varepsilon \, C^c$; if all observations are censored the estimate will be t_0. The functions t_k are symmetric in $x_1, x_2, \cdots, x_k$, and integrable with respect to the product measure $P^{(k)}$ over R^k for all $P \in \mathscr{P}$. This estimate for θ is unbiased if and only if for all $P \in \mathscr{P}$

$$\sum_{k=1}^{n} \binom{n}{k} [P(C)]^{n-k} \int_{x_i \in C^c} \cdots \int t_k(x_1, \cdots, x_k) \, dP(x_1) \cdots dP(x_k) + t_0 [P(C)]^n = \theta(P). \quad (2.1)$$

Our conclusions will essentially be based on the following trivial remark:

If for P, $P^* \in \mathscr{P}$, $P(C) = P^*(C) = 1$ and $\theta(P) \neq \theta(P^*)$, then θ is not estimable, since otherwise one would find $t_0 = \theta(P) = \theta(P^*)$. This remark is also intuitively obvious: if all observations are censored with probability 1 for two different values of θ, then there is no way of distinguishing between these parameter values. Loosely speaking, all that we shall do in the sequel will be to show that the above remark continues to apply if $P(C^c)$ as a function of θ possesses one double zero instead of two distinct zeros (a more precise formulation is given in theorem 2.2).

First, however, we extend the argument to limits of sequences of distributions. We consider two infinite sequences $P_j, P_j^* \in \mathscr{P}$, $j = 1, 2, \cdots$, such that

$$\lim_{j \to \infty} \theta(P_j) = \theta_0 \neq \theta_0^* = \lim_{j \to \infty} \theta(P_j^*).$$

Furthermore we suppose that for some Q, $Q^* \in \mathscr{P}$, P_j and P_j^* are absolutely continuous on C^c with respect to Q and Q^* respectively. Then by the RADON-NIKODYM theorem there exist densities $p_j \geqslant 0$ and $p_j^* \geqslant 0$ on C^c satisfying

$$P_j(A) = \int_A p_j(x) \, dQ(x) \text{ and } P_j^*(A) = \int_A p_j^*(x) \, dQ^*(x)$$

for all measurable $A \subseteq C^c$. Let

$$s_j = \sup_{x \in C^c} p_j(x) \text{ and } s_j^* = \sup_{x \in C^c} p_j^*(x).$$

The following result is now easy to prove:

Theorem 2.1

If $\lim_{j\to\infty} s_j = \lim_{j\to\infty} s_j^* = 0$, then θ is not estimable.

Proof

$$\left| \int \cdots \int_{x_i \in C^c} t_k(x_1, \cdots, x_k)\, dP_j(x_1) \cdots dP_j(x_k) \right| \leqslant$$

$$\leqslant s_j^k \cdot \int \cdots \int_{x_i \in C^c} |\, t_k(x_1, \cdots, x_k)\,|\, dQ(x_1) \cdots dQ(x_k) = \mathcal{O}(s_j^k)$$

since t_k is $Q^{(k)}$-integrable for $Q \in \mathscr{P}$. Also

$$P_j(C^c) \leqslant s_j\, Q(C^c) = \mathcal{O}(s_j)$$

and hence by (2.1), in order for θ to be estimable we should have

$$\theta(P_j) = t_0 + \mathcal{O}(s_j) \qquad \text{for } j \to \infty. \tag{2.2}$$

Similarly

$$\theta(P_j^*) = t_0 + \mathcal{O}(s_j^*)$$

and as a result

$$\theta_0 = \lim_{j\to\infty} \theta(P_j) = t_0 = \lim_{j\to\infty} \theta(P_j^*) = \theta_0^*,$$

which contradicts $\theta_0 \neq \theta_0^*$.

Finally we show that the theorem continues to apply in the case of one double zero at θ_0 (in the sence of (2.3)) instead of two distinct zeros at θ_0 and θ_0^*. Considering only the sequence P_j with the properties required above we have

Theorem 2.2

If $\lim_{j\to\infty} s_j = 0$ and either $\theta_0 = \pm\infty$ or

$$\lim_{j\to\infty} \frac{s_j}{\theta(P_j) - \theta_0} = 0, \tag{2.3}$$

then θ is not estimable.

Proof

In the proof of theorem 2.1 it was shown that (2.2) is a necessary condition for θ to be estimable. For $\theta_0 = \pm\infty$ equation (2.2) cannot be satisfied since the right-hand side is bounded. For $\theta_0 \neq \pm\infty$ (2.2) implies $t_0 = \theta_0$ and hence

$$\theta(P_j) - \theta_0 = \mathcal{O}(s_j) \qquad \text{for } j \to \infty$$

which contradicts the hypothesis of the theorem.

3. Examples

We conclude this note by applying the results of section 2 to the estimation of the expectation μ and the standard deviation σ of the normal distribution (denoted by $N(\mu, \sigma^2)$) and the parameter λ of the POISSON distribution (denoted by $Po(\lambda)$). With each example we impose a condition on the set C that ensures that the parameter under discussion is not estimable. By lim is meant $\lim_{j\to\infty}$ throughout this section.

Normal distribution; both μ and σ unknown

Assumption: C contains a non-degenerate open interval $I = (a, b)$.

A) *Estimation of μ*

Choose P_j: $N(\mu, \sigma_j^2)$, P_j^*: $N(\mu^*, \sigma_j^2)$, Q: $N(\mu, 1)$ and Q^*: $N(\mu^*, 1)$, where $\mu, \mu^* \in I$, $\mu \neq \mu^*$, $\sigma_j < 1$ for all j and $\lim \sigma_j = 0$. It is easily verified that $\lim s_j = \lim s_j^* = 0$. Hence by theorem 2.1 μ is not estimable and neither is any function of μ that is not constant on I.

B) *Estimation of σ*

Choose P_j: $N(\mu, \sigma_j^2)$ and Q: $N(\mu, 1)$, where $\mu = \frac{1}{2}(a + b)$, $\sigma_j < 1$ for all j and $\lim \sigma_j = 0$. Then

$$s_j \leqslant \sup_{x \leqslant a; x \geqslant b} p_j(x) = \frac{1}{\sigma_j} \exp\left[-\frac{1}{8}(b-a)^2\left(\frac{1}{\sigma_j^2} - 1\right)\right].$$

Since $\lim s_j = \lim \frac{s_j}{\sigma_j} = 0$,

σ is not estimable (theorem 2.2) and neither is any function of σ that does not tend to a finite limit at least as fast as

$$\frac{1}{\sigma} \exp\left[-\frac{(b-a)^2}{8\sigma^2}\right]$$

tends to zero for $\sigma \to 0$.

Normal distribution; σ known

Assumption: C contains a semi-infinite interval.

Without loss of generality we suppose that C contains the open interval $I = (a, \infty)$. Choose P_j: $N(\mu_j, \sigma^2)$ and Q: $N(a, \sigma^2)$, where $\mu_j > a$ for all j and $\lim \mu_j = \infty$. Then

$$s_j \leqslant \sup_{x \leqslant a; x \geqslant b} p_j(x) = \exp\left[-\frac{1}{2}\left(\frac{\mu_j - a}{\sigma}\right)^2\right].$$

Since $\lim s_j = 0$, μ is not estimable (theorem 2.2) and neither is any function of μ that does not tend to a finite limit at least as fast as

$$\exp\left[-\frac{1}{2}\left(\frac{\mu - a}{\sigma}\right)^2\right]$$

tends to zero for $\mu \to \infty$.

Normal distribution; μ known

Assumption A: C contains a neighbourhood of μ.

By the assumption C contains an open interval $I = (a, b)$ where $\mu = \frac{1}{2}(a + b)$. Following the construction given for the case where both μ and σ are unknown under B) we arrive at the conclusion stated there.

Assumption B: C does not contain a neighbourhood of μ but it contains two disjoint semi-infinite intervals.

Without loss of generality we suppose that $\mu = 0$ and that C contains the intervals $(-\infty, -a)$ and (a, ∞), $a > 0$. Choose P_j: $N(0, \sigma_j^2)$ and Q: $N(0, 1)$, where $\sigma_j > 1$ for all j and $\lim \sigma_j = \infty$. Then

$$s_j \leqslant \sup_{-a \leqslant x \leqslant a} p_j(x) = \frac{1}{\sigma_j} \exp\left[\frac{1}{2} a^2 \left(1 - \frac{1}{\sigma_j^2}\right)\right].$$

Since $\lim s_j = 0$, σ is not estimable (theorem 2.2) and neither is any function of σ that does not tend to a finite limit at least as fast as $\frac{1}{\sigma}$ tends to zero for $\sigma \to \infty$.

POISSON distribution

Assumption A: C contains the set $\{0, 1\}$.

Choose P_j: $Po(\lambda_j)$ and Q: $Po(1)$, where $\lambda_j < 1$ for all j and $\lim \lambda_j = 0$. Then

$$s_j \leqslant \sup_{x \geqslant 2} p_j(x) = \lambda_j^2 \exp(1 - \lambda_j).$$

Since $\lim s_j = \lim \frac{s_j}{\lambda_j} = 0$, λ is not estimable by theorem 2.2.

Assumption B: C contains the set of integers $> a$.

Choose P_j: $Po(\lambda_j)$ and Q: $Po(1)$, where $\lambda_j > 1$ for all j and $\lim \lambda_j = \infty$. Then

$$s_j \leqslant \sup_{x \leqslant a} p_j(x) = \lambda_j^a \exp(1 - \lambda_j).$$

Since $\lim s_j = 0$, λ is not estimable by theorem 2.2 and neither is any function of λ that does not tend to a finite limit at least as fast as $\lambda^a \exp(-\lambda)$ tends to zero for $\lambda \to \infty$.

These examples will suffice to show that in practice non-estimability is the

rule rather than the exception. This is because in actual sampling censoring usually occurs in one or both tails of the distribution, i.e. the set C consists of one or two semi-infinite intervals. In this case we have shown above that for the normal distribution μ and σ (or σ^2) are not estimable with only one exception: If μ is known, censoring is one-sided and C does not contain a neighborhood of μ, we cannot conclude that σ and σ^2 are not estimable. In fact, even for $n = 1$ unbiased estimates for σ and σ^2 are easily found. Suppose, without loss of generality, that $\mu = 0$ and $C = (-\infty, -a)$, $a \geqslant 0$. Then

$$t_0 = 0 \quad , \quad t_1(x) = \begin{cases} 0 & \text{for } -a \leqslant x \leqslant 0 \\ 2x^2 & \text{for } x > 0 \end{cases}$$

is an unbiased estimate of σ^2. Since the information whether or not the observation is censored, together with the value of x^2 when available, constitutes a sufficient and complete statistic (cf. section 1) the minimum variance unbiased estimate of σ^2 is found to be

$$t_0 = 0 \quad , \quad t_1(x) = \begin{cases} x^2 & \text{for } -a \leqslant x \leqslant a \\ 2x^2 & \text{for } x > a. \end{cases}$$

The estimation of σ may be dealt with in a similar fashion.

For the POISSON distribution we found that censoring in one or both tails always ensures that λ is not estimable. However, if for instance $C = \{0, 2\}$ it can be shown that for $n \geqslant 2$, λ is estimable.

References

[1] W. L. SMITH (1957), A note on truncation and sufficient statistics, Ann. Math. Statist. **28**, 247-252.

[2] J. W. TUKEY (1949), Sufficiency, truncation and selection, Ann. Math. Statist. **20**, 309-311.