RANDOMIZED UNBIASED ESTIMATION OF RESTRICTED PARAMETERS

BY MASAAKI SIBUYA

(Received Jan. 7, 1963; revised Sept. 3, 1963)

1. Introduction

The terminology ' estimator ' is usually understood as a measurable mapping from the sample space into the parameter space. There exist, however, the cases in which traditional statistics (estimators in a loose sense) have the distribution ranges which are not included in the parameter space. Such cases can occur often when the parameter space is ' restricted ', namely when it is certainly known that the parameter is integral, or restricted by inequalities, etc., that is, when the parameter space is a proper subset of the ' natural ' one. We see also the cases when the estimator of vector parameters are obtained by the moment method.

We shall call the estimator of which the range is a subset of the parameter space ' feasible ' in order to emphasize the property. Several authors use the maximum likelihood method to obtain feasible estimators of restricted parameters, finding the parameter value which maximizes the likelihood function within the restricted parameter space. We observe that the estimator's property of feasibility sometimes contradicts its unbiasedness. See the examples of section 3 of this paper and those in [1] where H. Morimoto and the author discuss the UMV unbiased estimator of the restricted selection parameter.

If we require that the estimator is both feasible and unbiased, then we have to rely on randomization. In this note we state a theorem giving the randomized feasible UMV unbiased estimator. We discuss the estimation of a parametric function $\gamma(\theta)$, though it may often be a parameter itself.

The author is grateful to Professor E. W. Barankin for his helpful comments on a previous draft of this paper.

2. Notation and theorem

X is a random variable over a measurable space $(\mathscr{X}, \mathscr{A})$, and $\mathscr{P}=\{P_\theta;\ \theta\in\Omega\}$ is a family of its distributions. For $\{P_\theta\}$, there exists a minimal sufficient statistic T, of which the family of distributions $\{P_\theta^T\}$

is complete. We estimate a real parametric function $\gamma(\theta)$ with the range $G=\{\gamma(\theta),\ \theta\in\Omega\}$, a Borel set on the real line.

A randomized estimator is a family of distributions $\{Q_x,\ x\in\mathscr{X}\}$ labeled by the sample value x on the space $(G, \mathscr{Y})$, where $\mathscr{Y}$ is the class of all Borel measurable subsets of G. This randomized estimator is denoted by Z. We observe the sample value $X=x$, and determine the value of the estimator $Z=z$ according to the distribution $Q_x(\cdot)$ using some random device.

We shall find the randomized estimator Z, which is unbiased;

$$(1)\qquad E_\theta E_{Q_X}(Z)=\int_{\mathscr{X}}\left(\int_G z dQ_x\right)dP_\theta=\gamma(\theta)$$

($E_{Q_x}(Z)$ should be $\mathscr{A}$ measurable) and minimizes the variance

$$(2)\qquad V(Z)=E_\theta E_{Q_X}(Z-\gamma(\theta))^2$$

uniformly in θ.

THEOREM. *We assume the following conditions on $\{P_\theta\}$ and $\gamma(\theta)$: (i) G, the range of the parametric function $\gamma(\theta)$, is a closed set of real line. We denote by $\bar{G}$ the interval of which the end points are the leftmost and rightmost points of G (they may be $\pm\infty$). (ii) There exists such a statistic $\varphi(t)$ based on $T=t(X)$ as*

$$\int \varphi(t)dP_\theta^T=\gamma(\theta)\,,\qquad \textit{for all }\theta\in\Omega$$

and its distribution ranges are subsets of $\bar{G}$ for all $\theta\in\Omega$.

$\bar{G}-G$ is a union of at most countable disjoint open intervals. Denote them by (a_i, b_i), $i=1, 2, \cdots$.

Then the UMV feasible randomized unbiased estimator of $\gamma(\theta)$ is given by

$$(3)\qquad \begin{cases} Z^t=\varphi(t)\,, & \textit{if } \varphi(t)\in G \\ Z^t=\begin{cases} a_i\,, \\ b_i\,, \end{cases} & \begin{array}{l} \textit{with probability } (b_i-\varphi(t))/(b_i-a_i)\,, \\ \textit{with probability } (\varphi(t)-a_i)/(b_i-a_i)\,, \\ \quad \textit{if } \varphi(t)\in(a_i, b_i)\subset\bar{G}-G\,, \end{array} \end{cases}$$

and its variance is

$$(4)\qquad V_\theta(Z^t)=V_\theta(\varphi(T))+\sum_i\int_{\varphi(t)\in(a_i,\, b_i)}(b_i-\varphi(t))(\varphi(t)-a_i)dP_\theta^T\,.$$

Z^t is essentially unique because of the completeness of $\{P_\theta^T\}$.

PROOF. The conditional expectation of a randomized estimator Z

$$E_{Q_x}(Z)=m(x) \tag{5}$$

is the mean of the distribution $Q_x(\cdot)$. The variance of Z, (2), is written as

$$V_\theta(Z)=E_\theta((m(X)-\gamma(\theta))^2+V(Z\,|\,x)) \tag{6}$$

where $V(Z\,|\,x)$ is the conditional variance given $X=x$, or the variance of the distribution $Q_x(\cdot)$;

$$V(Z\,|\,x)=E_{Q_x}(Z-m(x))^2 \,. \tag{7}$$

For a fixed value of $m(x)$, this expression is minimized by the Q_x of the following form;

$$\begin{cases} Z=m(x)\,, & \text{with probability one}\,, \quad \text{if } m(x)\in G \\ Z=\begin{cases} a_i\,, \\ b_i\,, \end{cases} & \begin{array}{l}\text{with probability } (b_i-m(x))/(b_i-a_i)\,, \\ \text{with probability } (m(x)-a_i)/(b_i-a_i)\,, \\ \quad \text{if } m(x)\in(a_i,\, b_i)\,, \end{array} \end{cases} \tag{8}$$

and the minimized variance is

$$\min V(Z\,|\,x)=\begin{cases} 0, & \text{if } m(x)\in G\,, \\ (b_i-m(x))(m(x)-a_i)\,, & \text{if } m(x)\in(a_i,\, b_i)\,. \end{cases} \tag{9}$$

The next stage of the proof is to find such a statistic $m(x)$ as satisfies $E_\theta(m(X))=\gamma(\theta)$ and minimizes the expected value of the 'loss'

$$L(m(x),\,\theta)=\begin{cases} (m(x)-\gamma(\theta))^2, & \text{if } m(x)\in G\,, \\ (m(x)-\gamma(\theta))^2+(b_i-m(x))(m(x)-a_i), & \text{if } m(x)\in(a_i,\, b_i), \end{cases} \tag{10}$$

the conditionally and uniformly minimized variance. The function $L(m,\,\theta)$ is linear for m in $(a_i,\, b_i)$ for all i. In fact, its curve is a segment between the points $(a_i,\,(a_i-\gamma(\theta))^2)$ and $(b_i,\,(b_i-\gamma(\theta))^2)$. So, the function is strictly convex in G. The application of the Rao-Blackwell theorem shows that the essentially unique statistic $\varphi(t(x))$ is what is required.

3. Examples

Example 1. A family of normal distributions $N(\theta,\,\sigma^2)$. Assume that σ^2 is a known constant and θ belongs to $\Omega=(-\infty,\,-b]\cup[b,\,\infty)$, $b>0$, and that a sample of size n is observed. The UMV feasible randomized unbiased estimator of θ is

(11)
$$\begin{cases} Z^t = \bar{x} & \text{if } \bar{x} \in (-\infty, -b] \cup [b, \infty) \\ Z^t = \begin{cases} -b, & \text{with probability } (b-\bar{x})/2b, \\ b, & \text{with probability } (b+\bar{x})/2b, \end{cases} & \text{if } \bar{x} \in (-b, b). \end{cases}$$

Its variance is

(12)
$$V_\theta(Z^T) = \frac{\sigma^2}{n} + \int_{-b}^{b} (b^2 - t^2) \frac{1}{\sigma/\sqrt{n}} \phi\left(\frac{t-\theta}{\sigma/\sqrt{n}}\right) dt$$
$$= \frac{\sigma^2}{n} + \frac{\sigma^2}{n} W\left(\frac{b\sqrt{n}}{\sigma}, \frac{\theta\sqrt{n}}{\sigma}\right),$$

where

(13)
$$W(b, \theta) = (b^2 - \theta^2 - 1)[\Phi(b+\theta) + \Phi(b-\theta) - 1]$$
$$+ (b+\theta)\phi(b-\theta) + (b-\theta)\phi(b+\theta),$$

and $\phi(t)$ and $\Phi(t)$ are respectively the probability density function and the cumulative distribution function of the standard normal distribution. The function $W(b, \theta)$ is illustrated in Figure 1. Further deletion of disjoint open intervals (a_i, b_i), $i=1, 2, \cdots$, from Ω increases the variance $V(Z^T)$ by

$$\frac{\sigma^2}{n} \sum_i W\left(\frac{(b_i - a_i)\sqrt{n}}{2\sigma}, \frac{(b_i + a_i + 2\theta)\sqrt{n}}{2\sigma}\right).$$

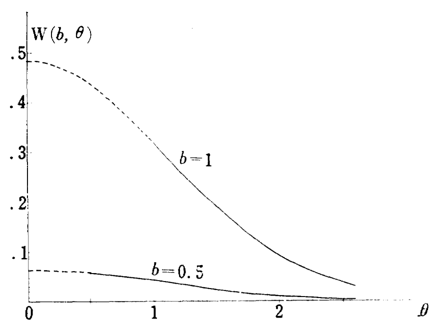

Fig. 1. Increase of variance function

If the sufficient statistic is a true reduction of data as in this example, any ancillary statistic is independent of the sufficient statistic. Here, for example, the statistic $U = \Phi^{-1}((X_1 - X_2)/(\sqrt{2}\,\sigma))$ is distributed uniformly on $(0, 1)$ for all θ. Therefore, U may be used to determine the value of Z. No random device is necessary, and we are able to construct feasible UMV unbiased estimators which are apparently non-randomized.

$\bar{X}$ remains to be a minimal sufficient statistic if the parameter space

Ω contains more than two different points, and remains complete if Ω contains an open interval, however short it may be. If Ω is, for example, the set of all integers $\bar{X}$ is not complete and there exists a non-randomized unbiased estimate $[\bar{X}+1/2]$, rounded integer of $\bar{X}$, which has smaller variance than $\bar{X}$ for large n (see Hammersley [2]). The parameter space $\Omega=[b, \infty)$ may be regarded as the limit of $\Omega=(-\infty, a]\cap[b, \infty)$, $a\to-\infty$. Then the estimator

$$\varphi(t)=\begin{cases} b\,, & \bar{x}<b\,, \\ \bar{x}\,, & \bar{x}\geqq b\,, \end{cases} \tag{15}$$

may be regarded as the limit of a sequence of UMV randomized unbiased estimators, of which the limiting variance is infinitely large.

Example 2. Bernoulli trials. Let $X_1, \cdots, X_n$ be independent of each other, and let X_i's take 1 and 0 with probability p and $q=1-p$ respectively. The parameter space is $[0, 1]-\bigcup_i(a_i, b_i)$. $T=\sum^n X_i/n$ is a minimal sufficient statistic and the UMV unbiased estimator of p. If a sample point $t_k=k/n$ does not belong to the parameter space but to an interval (a_i, b_i), $V(Z^T)$ increases by $\binom{n}{k}p^k q^{n-k}(b_i-k/n)(k/n-a_i)$.

When $n=3$ and $\Omega=\{0, 1/4, 1/2, 3/4, 1\}$, the UMV feasible randomized unbiased estimator is

$$\begin{aligned}
&Z^t=0\,, && t=0\,,\\
&Z^t=\begin{cases}1/4 & \text{with probability } 2/3\,,\\ 1/2 & \text{with probability } 1/3\,,\end{cases} && t=1/3\,,\\
&Z^t=\begin{cases}1/2 & \text{with probability } 1/3\,,\\ 3/4 & \text{with probability } 2/3\,,\end{cases} && t=2/3\,,\\
&Z^t=1\,, && t=1\,,
\end{aligned}$$

and its variance is $3pq/8$. The variance is smaller than that of, for example, a non-randomized feasible unbiased estimator $(X_1+X_2)/2$.

Example 3. Uniform distribution on $[0, \theta]$. Let the parameter space Ω be the set of all positive integers. As discussed in [1] $T=[X_{\max}]$, the integer part of the maximum value, is a minimal sufficient statistic and complete, and the UMV unbiased estimator of θ on $(0, \infty)$ is

$$\varphi(t)=\frac{(t+1)^{n+1}-t^{n-1}}{(t+1)^n-t^n}\ ,$$

and its variance is

$$V_\theta(\varphi(T))=\frac{1}{\theta^n}\sum_{t=0}^{\theta-1}\frac{(t+1)^n t^n}{(t+1)^n-t^n}\sim\left(\frac{\theta-1}{\theta}\right)^n \quad (n\to\infty)\,.$$

As $t+1<\varphi(t)<t+2$, $t=1, 2, \cdots$, the UMV feasible randomized unbiased estimator is

$$Z^t=1\,,\quad t=0\,,$$
$$Z^t=\begin{cases} t+1, \text{ with probability } & ((t+1)^n-2t^n)/((t+1)^n-t^n)\,, \\ t+2, \text{ with probability } & t^n/((t+1)^n-t^n)\,, \end{cases}$$
$$t=1, 2, \cdots, \theta-1\,.$$

Its variance is

$$V_\theta(Z^t)=\frac{2}{\theta^n}\sum_{t=1}^{\theta-1}t^n\sim 2\left(\frac{\theta-1}{\theta}\right)^n \quad (n\to\infty)\,.$$

In this case, we can find feasible non-randomized unbiased estimators as $2[X]+1$, $2[X+1/2]$, etc. Their variances are, of course, larger than $V_\theta(Z^t)$.

Example 4. Hypergeometric distribution.

$$\binom{M}{x}\binom{N-M}{n-x}\Big/\binom{N}{n},\quad \begin{array}{l} M=0, 1, \cdots, N, \\ x=0, 1, \cdots, \max(n, M). \end{array}$$

N and n are given integers, and we estimate M. The non-randomized UMV estimator is XN/n, so, when N/n is not integral, the UMV randomized integral estimator is

$$Z=\begin{cases} [xN/n]\,, & \text{with probability } [xN/n]+1-xN/n\,, \\ [xN/n]+1\,, & \text{with probability } xN/n-[xN/n]\,. \end{cases}$$

$V(Z)$ is larger than $V(XN/n)$ by at most 1/4.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] H. Morimoto and M. Sibuya, "Sufficient statistics and unbiased estimation of restricted selection parameters," to appear.

[2] J. M. Hammersley, "On estimating restricted parameters," *J. R. Statist. Soc.*, (B), Vol. 12 (1950), pp. 192-240 (with discussions).