# MINIMUM VARIANCE ESTIMATION IN DISTRIBUTIONS ADMITTING ANCILLARY STATISTICS

*By* C. RADHAKRISHNA RAO
*Statistical Laboratory, Calcutta*

In a previous paper (Rao, 1952) it was shown that uniformly minimum variance estimates could be obtained if the class of estimates is restricted to functions of a given type such as the class of functions satisfying the property

$$f(x_1+a, x_2+a, \ldots) = a+f(x_1, x_2, \ldots)$$

In this paper, a special class of distributions known to admit ancillary statistics is considered and the properties of minimum variance estimates examined. Ancillary statistics (Fisher, 1934) are those functions of observations whose distibution is independent of unknown parameters $\theta$ under consideration but which together with other statistics are jointly sufficient for $\theta$. In the example of the previous paper, the probability density is of the form $f(x-\theta)$ in which case in a sample $x_1, x_2, \ldots, x_n$ the statistics $y_1 = x_1-x_2$, $y_2 = x_2-x_3, \ldots, y_{n-1} = x_{n-1}-x_n$ are all ancillaries. If to these we add a statistic $U$, which is independent of $y_1, \ldots, y_{n-1}$ then the joint density of $U$ and $y$ can be written

$$P(U, y|\theta) = P(y)P(U|y, \theta)$$

where $P(y)$ is independent of $\theta$. In general the number of ancillaries need not be $(n-1)$ and the number of $U$'s which with the $y$'s are jointly sufficient for $\theta$ need not be one.

Let there exist a statistic $T$ such that its conditional expectation for given $y$ is $\theta$, i.e.

$$E(T|y) = \theta \qquad \ldots \quad (1)$$

We have the following results.

(i) If $I(\theta, y)$ is the conditional information on $\theta$ provided by the sample then

$$V(T|y) \geqslant 1/I(\theta, y)$$

obtained by the general arguments used in (Rao, 1945) when the range of $T$ is independent of $\theta$. Integrating over $y$

$$V(T) \geqslant \int \frac{P(y)}{I(\theta,y)}\, dy$$

The quantity on the right hand side is the appropriate limit to the variance when statistics satisfying the condition (1) are considered. This is generally greater than $1/I(\theta)$ the reciprocal of the total information on $\theta$ provided by the sample.

(ii) Suppose that $g(y)$, a function of the ancillaries only, has zero expectation in which case $T+g$ is also an unbiased estimate of $\theta$. But $V(T+g) = V(T)+V(g)$, covariance between $T$ and $g$ being zero due to the condition (1). This shows that the statistic $T$ cannot be improved upon by adding a function of the ancillaries.

(iii) Consider any other statistic $T'$ unbiased for $\theta$ but not conditionally so. If $E(T'|y) = h(\theta, y)$ then

$$V(T'|y) \geqslant [h'(\theta, y)]^2/I(\theta, y)$$

$$V(T') \geqslant \int \frac{[h'(\theta, y)]^2}{I(\theta, y)} P(y)dy + \int [h(\theta, y) - \theta]^2 P(y)dy$$

$$\geqslant \frac{1}{I(\theta)} + V\{h(\theta, y)\} \qquad \ldots \quad (2)$$

since

$$\int h(\theta, y)P(y)dy = \theta$$
$$\int h'(\theta, y)P(y)dy = 1$$

and therefore

$$\int \frac{[h'(\theta, y)]^2}{I(\theta, y)} P(y)dy \geqslant \frac{1}{\int I(\theta, y) P(y)dy}$$

The result (2) shows that for the limit $1/I(\theta)$ to be attainable $h(\theta, y)$ should be indentically equal to $\theta$. But when $h(\theta, y)$ is indentically equal to $\theta$ it is shown in para (ii) that $V(T) \geqslant E\{1/I(\theta, y)\}$ where the right hand side expression can be equal to $1/I(\theta)$ only when $I(\theta, y)$ is independent of $y$, the ancillary statistics. In large samples it may be possible that $E\{1/I(\theta, y)\} \rightarrow 1/I(\theta)$ in which case the statistic $T$ with conditional variance $1/I(\theta, y)$ is asymptotically the best.

(iv) For a statistic $T$ unbiased for $\theta$ but not conditionally for given $y$,

$$E(T'|y) = h(\theta, y) \neq \theta$$
$$E\{h(\theta, y)\} = \theta$$

Construct the statistic $T' - h(\theta_0, y) + \theta_0$ which is also unbiased for $\theta$. This has necessarily smaller variance than $T'$ at the value $\theta = \theta_0$ provided $\theta_0 \neq h(\theta_0, y)$. There cannot therefore, exist a statistic $T'$ with uniformly minimum variance without its conditional expectation being $\theta$. It does not, however, follw that there exists a statistic with uniformly minimum variance in the class of statistics satisfying the the property (1). Further investigation is needed to compare the relative merits and demerits of statistics such as $T$ and $T'$ in finite samples.

(v) In large samples there seems to be a definite advantage in using statistics satisfying the property (1). One aspect of it when $V(T|y) = 1/I(\theta, y)$ is already considered in para (iii). Let us consider a situation where a single statistic $T$ is jointly sufficient with a number of ancillaries $y_1, y_2, \ldots$ for a parameter $\theta$. Also assume that that $T$ is such that

$$E(T|y) = \theta$$

and $T$ and $y_1, y_2, \ldots$ stochastically converge to $\theta, c_1, c_2, \ldots$ where $c_1, c_2, \ldots$ are necessarily independent of $\theta$. Let there exist another statistic $f(T, y, n)$ which stochastically converges to $\theta$. Then in large samples under some general conditions of the existence of second moments, continuity of the derivatives involved and existence of limiting dis ributions of $n^p(T-\theta)$ and $n^p(y-c)$ for some $p$, the statistics

$$f(T, y, n) - f(\theta, c, n) \text{ and } A(T-\theta) + B(y-c) \qquad \ldots \quad (3)$$

where

$$A = \left.\frac{\partial f}{\partial T}\right|_{T=\theta,\, y=c} = \frac{\partial f(\theta, c, n)}{\partial \theta}$$

$$B = \left.\frac{\partial f}{\partial y}\right|_{T=\theta,\, y=c}$$

tend to have the same asymptotic variance provided

$$\left.\frac{\partial f(T, y, n)}{\partial T}\right|_{\theta, c} \longrightarrow g_1(\theta, c)$$

$$\left.\frac{\partial f(T, y, n)}{\partial y}\right|_{\theta, c} \longrightarrow g_2(\theta, c)$$

as $n \to \infty$ uniformly in $\theta$ and $c$ in the neighbourhood of their true values. The limiting result concerning the statistics in (3) stated as above differs from that found in the literature in that the function $f(T, y, n)$ involves $n$, the sample size, explicitely. Since $f(T, y, n)$ stochastically tends to $\theta$ it follows that $f(\theta, c, n) \to \theta$ as $n \to \infty$. If this convergence is uniform in the neighbourhood of the true value of $\theta$, then

$$\frac{\partial f(\theta, c, n)}{\partial \theta} \to 1 \text{ as } n \to \infty$$

Hence $f(T, y, n)$ has the same asymptotic variance as

$$(T-\theta) + B(y-c)$$

which is

$$V(T) + B^2 V(y)$$

the covariance between $T$ and $y$ being identically zero. In this case $T$ cannot be improved upon. It may however, be possible for $f(T, y, n)$ to exist and possess a smaller asymptotic variance than $T$ when the conditions assumed above do not hold. This needs further investigation.

We consider an example provided by Basu (1952) to show that the maximum likelihood estimates need not have asymptotically the least variance. Let $\xi$ and $\eta$ be the maximum and minimum in a random sample of size $n$ from a rectangular population in the range $\theta$ to $2\theta$. The statistics $\xi$ and $\eta$ are jointly sufficient for $\theta$ and the maximum likelihood estimate of $\theta$ is $\xi/2$. Correcting for bias we may take

$$T_1 = \frac{n+1}{2n+1}\,\xi$$

as an estimate. Its asymptotic variance is $1/4n^2$. Basu proposed an alternative estimate

$$T_2 = \frac{n+1}{5n+4}\,(2\xi+\eta)$$

which has the asymptotic variance $1/5n^2$ smaller than that of $T_1$ the estimate based on the method of maximum likelihood. In fact this is the limiting asymptotic variance attainable, for a regular function of $\xi, \eta$ in large samples has the same asymptotic variance as $a\xi + b\eta$ and the best combination is $(2\xi+\eta)/5$. Now we observe that $t = \eta/\xi$ is an ancillary statistic. The joint distribution of $t$ and $\xi$ is

$$\frac{n(n-1)}{\theta^n}\,\xi^{n-1}\,(1-t)^{n-2}\,d\xi\,dt$$

and that of $t$ alone is

$$(n-1)\left(2^n-\frac{1}{t^n}\right)(1-t)^{n-2}dt$$

The conditional expectation of $\xi$ given $t$ is

$$\theta\frac{n}{n+1}\,\frac{(2t)^{n+1}-1}{(2t)^n-1}\,\frac{1}{t}$$

in which case the statistic

$$T_3=\xi\,t\,\frac{n+1}{n}\,\frac{(2t)^n-1}{(2t)^{n+1}-1}=\frac{n+1}{n}\,\xi\eta\,\frac{(2\eta)^n-\xi^n}{(2\eta)^{n+1}-\xi^{n+1}}$$

has the conditional expectation $\theta$. The expression for its variance is complicated and it is also one of the cases where its asymptotic variance cannot be found by using the result (3). Being a non-regular function of $\xi$ and $\eta$ it is of some interest to examine whether its limiting variance is greater or smaller than $1/5n^2$.

All the estimates $f(\xi, \eta)$ considered above satisfy the property

$$f(\theta\xi, \theta\eta)=\theta f(\xi, \eta)$$

in which case they can be uniformly improved upon from the point of view of variance. Let

$$E\{f(\xi,\eta)\,|\,t\}=\theta h(t)$$
$$E\{f(\xi,\eta)^2\,|\,t\}=\theta^2 g(t)$$

where $t=\eta/\xi$ and $E\{h(t)\}=1$.

Consider the statistic

$$T=\frac{f(\xi,\eta)}{h(t)}\,k(t)$$
$$E(T\,|\,t)=\theta k(t)$$
$$E(T^2\,|\,t)=\theta^2k^2(t)m(t)$$
$$m(t)=g(t)/h^2(t)$$

If $T$ has to be unbiased then

$$E\{k(t)\}=1.$$

Subject to this condition we may choose $k(t)$ to minimise

$$E\{k^2(t)m(t)\}$$

The optimum vlaue of $k(t)$ is obviously $\lambda/m(t)$ where $\lambda$ is chosen such that

$$\int\frac{\lambda}{m(t)}\,P(t)\,dt=1$$

$$\lambda=1\bigg/\int\frac{P(t)}{m(t)}\,dt$$

With this value of $k(t)$, $T$ has a smaller variance than $f(\xi, \eta)$ as an unbiased estimate of $\theta$. We thus obtain a result similar to that derived by Pitman (1941).

REFERENCES

BASU, D. (1952): An example where the maximum likelihood estimate has not the least asymptotic variance (in press).

FISHER, R. A. (1934): Two new properties of mathematical likelihood. *Proc. Roy. Soc. Sc. A.*, 144, 285.

PITMAN, E.J.G. (1938): The estimation of location and scale parameters of a continuous population of any given form. *Biometrika*, 30, 391.

RAO, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37, 81.

RAO, C. R. (1952): Some theorems on minimum variance estimation. *Sankhyā*, 12, Parts 1 & 2.

*Paper received: March, 1952.*