ON A CLASS OF ADMISSIBLE ESTIMATORS OF THE NORMAL VARIANCE

By D. BASU

(*Research Fellow of the National Institute of Sciences*)
Statistical Laboratory, Calcutta

1. Introduction

Let $X_1, X_2, \ldots X_n$ be a set of independent normal variables with mean m and variance v. The problem is to make a point estimate $t(x_1, x_2, \ldots x_n)$ of v on the basis of n random observations $x_1, x_2, \ldots x_n$ on the chance variables $X_1, X_2, \ldots X_n$. Let us first consider the simpler case where the mean m is known. The maximum likelihood estimator of v is S/n where $S=\Sigma(x_i-m)^2$. The amount of information $I(v)$ per unit of sample is $I(v)=1/(2v^2)$ and so from the Cramér-Rao inequality we have that for any unbiassed estimator t of v

$$E(t-v)^2 \geqslant \frac{1}{nI(v)} = \frac{2v^2}{n}. \qquad \ldots \quad (1.1)$$

It is easily verified that the maximum likelihood estimator S/n is unbiased and that its variance is $(2v^2)/n$ so that in the class of all unbiased estimators S/n is essentially the only admissible estimator if we take our loss function proportional to the square of the error committed. That is to say if t be any other unbiased estimator of v then, unless of course $t=S/n$ almost everywhere,

$$E(t-v)^2 \geqslant E(S/n-v)^2$$

with the strict sign of inequality holding for at least one v (as a matter of fact in this case the strict sign of inequality will hold everywhere).

Now consider the class of estimators aS for all values of a. We note that

$$E(a\,S-v)^2 = E\{a(S-nv)+(an-1)v\}^2$$
$$= \{2a^2n+(an-1)^2\}v^2$$

and that the above is minimum at $a=1/(n+2)$, the minimum value being

$$E\left(\frac{S}{n+2}-v\right)^2 = \frac{2v^2}{n+2}. \qquad \ldots \quad (1.2)$$

Thus in the class of estimators of the form aS the only admissible one is $S/(n+2)$. The maximum likelihood estimator S/n, which is also the best unbiased estimator, is not admissible in the sense of Wald. It is very surprising that by introducing a bias in our estimator t we can make the risk function $r(v|t)=E(t-v)^2$ uniformly smaller, than the Cramér-Rao limit (1.1). Such a thing, however, is not possible in every case. Take for instance the case of the normal mean m with known variance, say unity. By the Cramér-Rao inequality if t be any unbiased estimator of m then

$$E(t-m)^2 \geqslant \frac{1}{nI(m)} = \frac{1}{n}.$$

The maximum likelihood estimator $\bar{x}$ attains the above limit for all m. It was proved by the author and, under a more general set-up, by Blyth (1951) that in the class of all estimators that generate continuous risk functions the estimator $\bar{x}$ is admissible. In the next section we proceed to find out a class of admissible estimators for the variance. Throughout in this paper we take the square of the error as our loss function.

2. A CLASS OF ADMISSIBLE ESTIMATORS

Since $S=\Sigma(x_i-m)^2$ is a sufficient statistic for v we have, from the Rao-Blackwell theorem and the convexity of the loss function, that for the purpose of estimating v we need restrict ourselves to only functions of S. The frequency function $p(S|v)ds$ of S is

$$p(S|v)ds=\frac{1}{2^{n/2}\Gamma(n/2)}v^{-n/2}S^{n/2-1}e^{-S/2v}dS$$

$$(0\leqslant S<\infty,\ 0<v<\infty).$$

Now consider the a-priori probability frequency for v

$$p(v)dv=\frac{\lambda^{\mu-1}}{2^{\mu-1}\Gamma(\mu-1)}v^{-\mu}e^{-\lambda/2v}dv \quad (\lambda>0,\ \mu>1) \qquad \ldots \quad (2.1)$$

The joint frequency function of S and v becomes

$$p(S,\ v)dS\,dv=p(v)p(S|v)dS\,dv$$

$$=\frac{\lambda^{\mu-1}}{2^{\frac{n}{2}+\mu-1}\Gamma\left(\frac{n}{2}\right)\Gamma(\mu-1)}S^{\frac{n}{2}-1}v^{-\left(\frac{n}{2}+\mu\right)}e^{-\frac{S+\mu}{2v}}dSdv$$

The marginal frequency of S is

$$p(S)ds=dS\int_0^\infty p(S,\ v)dv$$

$$=\frac{\lambda^{\mu-1}}{B\left(\frac{n}{2},\ \mu-1\right)}\ \frac{S^{\frac{n}{2}-1}}{(S+\lambda)^{\frac{n}{2}+\mu-1}}$$

The a-posteriori frequency function of v, given S, is

$$p(v|S)dv=\frac{p(S,\ v)dv}{p(S)}$$

$$=\frac{(S+\lambda)^{\frac{n}{2}+\mu-1}}{2^{\frac{n}{2}+\mu-1}\Gamma\left(\frac{n}{2}+\mu-1\right)}v^{-\left(\frac{n}{2}+\mu\right)}e^{-\frac{S+\lambda}{2v}}dv$$

The a-posteriori expected value of v given S is

$$E(v|s) = \int_0^\infty vp(v|S)dv$$

$$= \frac{S+\lambda}{n+2\mu-4} = t_{\lambda\mu}(S). \qquad \ldots \quad (2.2)$$

Now consider the estimator $t_{\lambda\mu}$. The risk function generated by $t_{\lambda\mu}$ is

$$r(v|t_{\lambda\mu}) = E[(t_{\lambda\mu}-v)^2|v]$$

$$= \frac{2n+4(\mu-2)^2}{(n+2\mu-4)^2}v^2 - \frac{4\lambda(\mu-2)}{(n+2\mu-4)^2}v + \frac{\lambda^2}{(n+2\mu-4)^2}$$

and the average risk for the a-priori distribution (2.1) is

$$\bar{r}(t_{\lambda\mu}) = \int_0^\infty r(v|t_{\lambda\mu})p(v)dv$$

$$= \frac{\lambda^2}{2(\mu-2)(\mu-3)(n+2\mu-4)} \qquad \ldots \quad (2.3)$$

It should be noted that (2.3) is defined only when $\mu>3$ in (2.1) so that we henceforth restrict ourselves to only such a-priori distributions of the form (2.1) for which $\mu>3$. Thus for every $\lambda>0$ and $\mu>3$ the estimator $t_{\lambda\mu}$ is essentially the only Bayes solution corresponding to the a-priori frequency (2.1) and so every $t_{\lambda\mu}$ $(\lambda>0, \mu>3)$ is admissible. We make the interesting observation that although $t_{\lambda\mu}$ is admissible for every $\lambda>0$ and $\lambda>3$ the limiting Bayes solution $t_{0\mu} = \dfrac{S}{n+2\mu-4}$ obtained by making $\lambda \to 0$ is not admissible for any $\mu>3$. It is conjectured that the limiting Bayes solution $t_{03} = \dfrac{S}{n+2}$ considered in (1.2) is admissible. We also note that there cannot exist any estimator t of v for which the risk function $r(v/t) = E[(t-v)^2|v]$ is a bounded function of v for all v in $0<v<\infty$ For, if possible, let t^0 be an estimator for which $r(v|t^0) \leqslant M$ for all v. Then the average risk, with (2.1) as the a-priori weight function for v, will be $\leqslant M$. But for a sufficiently large λ (and any $\mu>3$) (2.3) will certainly exceed M which contradicts the fact that $t_{\lambda\mu}$ is a Bayes solution. Thus it is clear that there cannot exist any minimax estimator for v and that some other criterion has to be set up for choosing a good estimator from the class (2.2) of admissible estimators.

3. The case when the mean is unknown

So long we considered the case where the mean m was known. Now consider the case where the mean m also is unknown.

Since $\bar{x}$ and $S=\Sigma(x_i-\bar{x})^2$ jointly contain all the information about the parameter point (m, v) it follows from the Rao-Blackwell Theorem that all admissible estimators are essentially (i.e. excepting for a set of Lebesgue measure zero) functions of only $\bar{x}$ and S. If however, we restrict ourselves to only such estimators t for which the risk functions $r(m, v|t) = E[(t-v)^2|m, v]$ are functions of v alone then we can prove the following:

Theorem 3.1: *An admissible estimator for which the associated risk function is independent of m is essentially a function of S alone.*

Proof: Let $t = t(\bar{x}, S)$ be any admissible estimator such that $r(m, v|t)$ is independent of m. Let $t_\lambda = t(\bar{x}+\lambda, S)$. Since m is a location parameter it is readily seen that

$$\begin{aligned} r(m, v|t_\lambda) &= E[(t_\lambda - v)^2|m, v] \\ &= E[(t-v)^2|m+\lambda, v] \\ &= r(m+\lambda, v|t) \\ &= r(m, v|t) \end{aligned}$$

since $r(m, v|t)$ is independent of m. Thus t and t_λ generates the same risk function. Hence from the convexity of the loss function $(t-v)^2$ it follows that the estimator $= \frac{1}{2}(t+t_\lambda)$ will be uniformly more powerful than t unless $t = t_\lambda$ almost everywhere. Since by assumption t is admissible in the class of those estimators that generate the same risk function for all m it follows that, for each real λ, $t(\bar{x}, S) = t(\bar{x}+\lambda, S)$ almost every where in $\bar{x}$, S. And this proves that $t(\bar{x}, S)$ is essentially a function of S alone.

We can prove in the same way as in the previous section that in the sub-class of all those estimators which are essentially functions of S alone all estimators of the form.

$$t_{\lambda\mu} = \frac{S+\lambda}{n-1+2\mu-4} \; (\lambda > 0, \mu > 3) \qquad \dots \quad (3.1)$$

are admissible. It is to be noted that we now have $n-1$ and not n degrees of freedom for S and that is why there is a slight difference in the formulae (2.2) and (3.1). We also note that the best unbiased estimator $S/(n-1)$ and the maximum likelihood estimator S/n are both uniformly less powerful than the limiting form of (3.1) namely $t_{03} = S/(n-1)$.

We now prove that all estimators of the form (3.1) are admissible in the class of all estimators for which the associated risk functions are continuous. It is conjectured that $t_{\lambda\mu}$ is admissible in the unrestricted class of all estimators.

Consider the a-priori probability density function for the parameter point (m,v) namely

$$p(m, v)dmdv = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{m^2}{2\sigma^2}} \frac{\lambda^{\mu-1}}{2^{\mu-1}\Gamma(\mu-1)} v^{-\mu} e^{-\frac{\lambda}{2v}} dmdv$$

$$-\infty < m < \infty, \; 0 < v < \infty, \quad \sigma > 0, \lambda > 0, \mu > 3$$

As in the previous section we take $\mu > 3$ in order that the average risk associated with the Bayes solution may be finite.

The a-posteriori probability density of (m,v) is

$$p(m, v|\bar{x}, S)dmdv = \frac{e^{-\frac{m^2}{2\sigma^2}} v^{-\left(\frac{n}{2}+\mu\right)} e^{-\frac{T}{2v}} dmdv}{\int\int e^{-\frac{m^2}{2\sigma^2}} v^{-\left(\frac{n}{2}+\mu\right)} e^{-\frac{T}{2v}} dmdv}$$

where $$T = n(\bar{x}-m)^2+S+\lambda.$$

Hence the Bayes' solution corresponding to the a-priori density $p(m, v)\, dm\, dv$ is

$$t_{\sigma\lambda_\mu} = E(v\,|\,\bar{x}, S)$$

$$= \int\int vp(m, v\,|\,\bar{x}, S) dm\, dv$$

$$= \frac{\int_{-\infty}^{\infty} e^{-\frac{m^2}{2\sigma^2}}dm \int_0^{\infty} v^{-\left(\frac{n}{2}+\mu-1\right)} e^{-\frac{T}{2v}} dv}{\int_{-\infty}^{\infty} e^{-\frac{m^2}{2\sigma^2}}dm \int_0^{\infty} v^{-\left(\frac{n}{2}+\mu\right)} e^{-\frac{T}{2v}} dv}$$

$$= \frac{1}{n+2\mu-4} \cdot \frac{\int_{-\infty}^{\infty} T^{-\left(\frac{n}{2}+\mu-2\right)} e^{-\frac{m^2}{2\sigma^2}} dm}{\int_{-\infty}^{\infty} T^{-\left(\frac{n}{2}+\mu-1\right)} e^{-\frac{m^2}{2\sigma^2}} dm}$$

As both the numerator and the denominator of the above expression are uniformly convergent for all $\sigma > \delta > 0$ we have

$$\lim_{\sigma\to\infty} t_{\sigma\lambda_\mu}$$

$$= \frac{1}{n+2\mu-4} \quad \frac{\int_{-\infty}^{\infty} T^{-\left(\frac{n}{2}+\mu-2\right)} dm}{\int T^{-\left(\frac{n}{2}+\mu-1\right)} dm}$$

$$= \frac{S+\lambda}{(n-1)+2\mu-4}$$

$$= t_{\lambda\mu}$$

we thus observe that $t_{\lambda\mu}$ is a limiting Bayes solution. This however does not immediately prove the admissibility of $t_{\lambda\mu}$. We sketch below the method of proof.

Since $$e^{-\frac{m^2}{2\sigma^2}} = 1 - \frac{m^2}{2\sigma^2} e^{-\frac{m^2\theta}{2\sigma^2}} \quad 0 < \theta < 1$$

we have

$$t_{\sigma\lambda_\mu} = t_{\lambda\mu} + \frac{1}{\sigma^2} k(\bar{x}, S, \sigma) \qquad \ldots \quad (3.2)$$

where $k(\bar{x}, S, \sigma)$ remains bounded as $\sigma \to \infty$.

$$\therefore \quad r(m, v\,|\,t_{\sigma\lambda_\mu}) = r(m, v\,|\,t_{\lambda\mu}) + \frac{1}{\sigma^2} h(m, v, \sigma) \qquad \ldots \quad (3.3)$$

where $h(m, v, \sigma)$ remains bounded as $\sigma \to \infty$.

Hence, with $p(m, v)\, dm\, dv$ as the a-priori density, the average risk for $t_{\sigma\lambda_n}$ and t_{λ_n} satisfy the following relationship

$$\bar{r}(t_{\sigma\lambda_n}) = \bar{r}(t_{\lambda_n}) - \frac{1}{\sigma^2} g(\sigma) \quad \ldots \quad (3.4)$$

where $g(\sigma)$ remains bounded and non-negative as $\sigma \to \infty$. We omit the detailed discussions regarding convergence that are necessary in (3.2), (3.3), and (3.4).

We now prove that if there exists an estimator t for which

$$r(m, v|t) \leqslant r(m, v|t_{\lambda_n})$$

for all m and v then the set of points where the strict sign of inequality holds must be a set of Lebesgue measure zero.

Let A be the set of points where

$$r(m, v|t) < r(m, v|t_{\lambda_n})$$

If $m(A) > 0$ then we can find a sub-set A_1 of A with positive Lebesgue measure and an $\epsilon > 0$ such that for all (m,v) in A_1

$$r(m, v|t) < r(m, v|t_{\lambda_n}) - \epsilon.$$

Then

$$\bar{r}(t) = \int\int r(m,v/t)\, p(m,v) dm\, dv < \bar{r}(t_{\lambda_n}) - \epsilon \int\int_{A_1} p(m,v) dm\; dv.$$

Since $m(A_1) > 0$ it can be easily seen that

$$\sigma \int\int_{A_1} p(m, v)\, dm dv \to \frac{1}{\sqrt{2\pi}} \int\int_{A_1} p(v) dm\, dv > 0 \quad \text{as } \sigma \to \infty$$

where $p(v) dv$ is the marginal a-priori probability density for v.

Thus we have

$$\bar{r}(t) < \bar{r}(t_{\lambda_n}) - \frac{1}{\sigma} f(\sigma) \quad \ldots \quad (3.5)$$

where $f(\sigma) \to$ a positive constant as $\sigma \to \infty$. Form (3.4) and (3.5) we have that for all sufficiently large σ

$$\bar{r}(t) < \bar{r}(t_{\sigma\lambda_n})$$

which is a contradiction since $t_{\sigma\lambda_n}$ is the Bayes solution corresponding to the a-priori probability density $p(m, v) dm\, dv$.

Hence if t generates a continuous risk function and if $r(m, v|t) \leqslant r(m, v|t_{\lambda_n})$ for all m and v then the sign of equality must hold everywhere for if $r(m, v|t) < (m, v|t_{\lambda_n})$ for at least one (m, v) then, from the continuity of the two risk functions, the strict sign of inequality must hold in a set of positive Lebesgue measure which, as just demonstrated, is a contradiction.

REFERENCES

BASU, D. (1952): On the minimax approach to the problem of estimation. *Proc. Nat. Inst. Sci. India*, **18**, No. 4.

BLYTH, R. C. (1951): On the minimax statistical procedures and their admissibility. *Ann. Math. Stat.* **22**, No. 1.

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

WALD, A. (1950): *Statistical Decision Functions*, John Wiley and Sons.