ON SYMMETRIC ESTIMATORS IN POINT ESTIMATION WITH CONVEX WEIGHT FUNCTIONS

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 chance variables whose joint cumulative distribution function $F(x_1, x_2, \ldots x_n)$ is known to belong to some sub-space Ω of the space of all possible distribution functions F. As for instance it may be known that the X's are independently and indentically distributed so that Ω is the sub-space of all d.f.'s of the form

$$F = G(x_1)G(x_2)\ldots G(x_n), \qquad \ldots \quad (1.1)$$

where $G(x)$ is some one dimensional distribution function.

In point-estimation the problem is to estimate some population characteristic $\theta = \mu(F)$, where $\mu(F)$ is a real valued functional defined for all $F \epsilon \Omega$, with the help of an estimator $t = t(x_1, x_2, \ldots x_n)$ where x_i is a random observation on the chance variable X_i.

Let $W(t, F)$, for any fixed $R \epsilon \Omega$, denote the different weights that the statistician attaches to the different values of t as estimates of $\mu(F)$ and let

$$r(F \mid t) = \int_R W(t, F) dF, \qquad \ldots \quad (1.2)$$

where R is the n-dimensional sample-space, be the risk function associated with the estimator t.. We assume that there exist estimators t for which the integral (1.2) is convergent for all $F \epsilon \Omega$.

If $r(F\ t_1) \leqslant r(F/t_2)$ for all $F \epsilon \Omega$ with the sign of inequality holding for at least one F then t_1 is said to be uniformly more powerful than t_2.

The estimator t_0 will be called admissible if there exists no estimator t uniformly more powerful than t_0,

In this paper we restrict ourselves to only such weight functions as are convex (downwards) functions of t for every $F \epsilon \Omega$. That is

$$W\left(\frac{t_1+t_2}{2}, F\right) \leqslant \tfrac{1}{2} W(t_1, F) + \tfrac{1}{2} W(t_2, F) \qquad \ldots \quad (1.3)$$

for all t_1 and t_2. If the sign of equality holds only when $t_1=t_2$ then the function will be called strictly convex. As for example the following functions are all strictly convex.

(i) $|t-\theta|^{\mu}, \mu>1,$ (ii) $e^{|t-\theta|}-1$

(iii) $a|t-\theta|+b(t-\theta)^2, \quad a\geqslant 0, \quad b>0$

The function $|t-\theta|$ i convex but not strictly so.

2. ADMISSIBILITY AND SYMMETRY*

We prove the following:

Theorem 1: *If every $F\epsilon\Omega$ is symmetric in x_i and x_j and if the weight function $W(t, F)$ be strictly convex then every admissible estimator must be essentially symmetric in x_i, x_j.*

Proof: Let $t=t(x_1, x_2 \ldots x_n)$ be any admissible estimator and let t_1 be obtained from t by interchanging x_i with x_j.

From the symmetry of F in x_i, x_j it follows that t and t_1 are identically distributed for all $F\epsilon\Omega$.

$$\therefore \qquad r(F|t)=r(F|t_1) \text{ for all } F\epsilon\Omega.$$

Now if we define $t_0=\frac{1}{2}(t+t_1)$ then from the convexity of $W(t, F)$ we have

$$r(F|t_0)\leqslant\tfrac{1}{2}r(F|t)+\tfrac{1}{2}r(F|t_1)$$
$$=r(F|t). \qquad \ldots \quad (2.1)$$

Since t is admissible the sign of equality must hold everywhere in (2.1). From the strict convexity of $W(t, F)$ if follows that the set of points where $t\neq t_1$ must be of F-measure zero for all $F\epsilon\Omega$. Thus if the weight function be strictly convex all admissible estimators must by essentially symmetric in x_i, x_j. If the weight function be convex but not strictly so then there may exist admissible estimators which are not essentially symmetric in x_i, x_j. But since corresponding to any such unsymmetric estimator there always exist an estimator symmetric in x_i, x_j and generating the same risk function it follows that we need not go beyond estimators that are symmetric in x_i x_j.

Corollary: *If every $F\epsilon\Omega$ is symmetric in all the x's then for the purpose of estimation with a convex weight function we need restrict ourselves to only symmetric functions of the x's.*

If however we want to restrict our choice of t to a particular class of estimators then, for a particular weight function, the above results can be true without F being completely symmetric in x_i, x_j. For example suppose that $W(t, F)=(t-\theta)^2$ and that we want to restrict our choice of t only to linear functions of the x's. We note that if the first two moments of x_i and its product moments with the other x's are the same as the corresponding moments of x_j then for any linear estimator t

$$r(F|t)=r(F|t_1) \text{ for all } F\epsilon\Omega.$$

*The attention of the auther has been drawn to a paper by Paul R. Halmos entitled "The Theory of Unbiased Estimation" in the *Annals of Mathematical Statistics*, Vol. 17 (1946), where the results proved in this section were partially anticipated.

where t_i is obtained from t by interchanging x_i with x_j and so the proof of the theorem applies.

We now show how the above considerations of symmetry lead to simple proofs of results which are otherwise difficult to obtain.

3. Sample from a finite population

Consider a finite population sith N values $a_1, a_2, \ldots a_N$ and let

$$\alpha = \frac{1}{N} \Sigma a_i \text{ and } \sigma^2 = \frac{1}{N} \Sigma (a_i - \alpha)^2.$$

be the population mean and variance.

Let a random sample $x_1, x_2, \ldots x_n$ of size n be drawn without replacement from the population (x_i is the ith sample drawn). It is clear that the probability

$$P(x_1 = a_{i_1}, \; x_2 = a_{i_2} \ldots x_n = a_{i_n}) = \frac{(N-n)!}{N!}$$

and is the same for all the $(N)/(N-n)!$ possible choices of the indices $i_1, i_2, \ldots i_n$ from the set $1, 2, \ldots N$. Thus it follows that the joint distribution of the chance vector $(x_1, x_2, \ldots x_n)$ is symmetric in all the x's.

Hence in the class of all estimators of the form

$$t = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \qquad \ldots \quad (3.1)$$

we need consider only those for which all the c's are equal *i.e.* $t = c\bar{x}$. Let θ be the population characteristic we want to estimate and let $W(t,F) = (t-\theta)^2$
Then

$$r(F \mid c\bar{x}) = E(c\bar{x} - \theta)^2$$

$$= c^2 V(\bar{x}) + (c\alpha - \theta)^2$$

where
$$V(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}.$$

If in particular $\theta = \alpha$ then

$$r(F \mid c\bar{x}) = c^2 V(\bar{x}) + (c-1)^2 \alpha^2$$

and we observe that $c\bar{x}$ cannot be admissible unless $0 \leqslant c \leqslant 1$. For corresponding to any c_1 outside the interval $0 \leqslant c \leqslant 1$ we can always find another c_0 in the interval such that $c_0 \bar{x}$ is uniformly more powerful than $c_1 \bar{x}$. It is conjectured that for the weight function $(t-\theta)^2$ every $c\bar{x}$ $(0 \leqslant c \leqslant 1)$ is an admissible estimator in the entire class of all possible estimators.

Again in the class of all quadratic estimators we need consider only symmetric estimators of the form

$$t = a \Sigma x_i^2 + b \sum_{i \neq j} x_i x_j + c \, \Sigma \, x_i + d. \qquad \ldots \quad (3.2)$$

If σ^2 be the population characteristic we want to estimate and if we add the further criterion of unbiasedness then from

$$E(t)=an(\sigma^2+\alpha^2)+bn(n-1)\left(-\frac{1}{N-1}\sigma^2+\alpha^2\right)+cn\alpha+d$$

$$\equiv\sigma^2 \text{ for all } \alpha \text{ and } \sigma^2.$$

We have

$$an-b\frac{n(n-1)}{N-1}=1$$

$$an+bn(n-1)=0$$

$$c=d=0.$$

Solving for a and b and substituting in (3.2) we have that in the class of all unbiased quadratic estimators of σ^2 the estimator

$$t=\frac{N-1}{N}\cdot\frac{\Sigma(x_i-\bar{x})^2}{n-1} \quad \ldots \quad (3.3)$$

is uniformly the best estimator provided the weight function $W(t,F)$ is convex (downwards). It is believed that the estimator (3.3) is admissible in the unrestricted class of all estimators.

4. The Markoff set-up

Consider the familiar Markoff set-up where $x_1, x_2, \ldots x_n$ is a set of chance variables with equal variances σ^2 and expected values

$$Ex_i=a_{i1}\tau_1+\ldots+a_{im}\tau_m \qquad i=1, 2, \ldots n, \; m<n$$

where the a_{ij}'s are known constants and the τ's are unknown parameters. Without loss of generality we may assume that the rank of the matrix $(a_{ij})=m$

The problem is to extimate the τ's and σ^2.

At first let us assume that the x's are independently and normally distributed. We shall later on see how far this assumption can be relaxed.

We can always find a unitary orthogonal transformation

$$(z \vdots y)=x(B \vdots C) \quad \ldots \quad 4.1)$$

where $B'=(b_{ij}) \quad i=1,2,\ldots n-m, \; j=1,2,\ldots n$

and $C'=(c_{ij}) \quad i=1,2,\ldots m, \; j=1,2,\ldots n$

such that $Ez_i=0 \quad i=1,2,\ldots n-m$

and $Ey_i=\xi_i\neq 0 \quad i=1,2,\ldots m$

where the ξ's are independent linear functions of the τ's.

From the independence and normality of the x's it follows that the z's and the y's are independent normal variables with the same variance.

Thus if we want to set-up a linear unbiassed estimator of $g_1\xi_1+\ldots+g_m\xi_m$ then from the symmetry of the z's and the condition of unbiassedness it follows that we must choose from

$$t=a(z_1+z_2+\ldots+z_{n-m})+g_1y_1+\ldots+g_my_m. \qquad \ldots \quad (4.2)$$

If further we take our weight function as the square of the error then from the fact that $z_1+z_2+\ldots+z_{n-m}$ is uncorrelated with $g_1y_1+\ldots+g_my_m$ it follows that 'a' must be zero in (4.2).

We now consider the problem of estimating σ^2. In the class of quadratic estimators of σ^2 we must, because of the symmetry of the z's choose from the class

$$t=a\Sigma z_i^2+b\sum_{i\neq j} z_iz_j+\Sigma c_jz_iy_j+\Sigma d_{ij}y_iy_j+e\Sigma z_i+\Sigma f_iy_i+g.$$

From the condition of unbiasedness we have

$$t=\frac{1}{n-m}\Sigma z_i^2+\left\{b\Sigma z_iz_j+\Sigma c_jz_iy_j+e\Sigma z_i\right\}$$

$$=S_0+Q \qquad \ldots \quad (4.3)$$

If we take the weight function as square of the error then from the fact that S_0 is uncorrelated with Q it follows that the minimum variance of t will be attained when $V(Q)=0$, *i.e.* when $Q\equiv0$. Thus S_0 is the minimum variance unbiased quadratic estimator of σ^2. Following the technique of Rao (1952) it can be shown that in the class of all unbiased estimators S_0 is the minimum variance estimator of σ^2 and also that any linear function of the y's is the minimum variance unbiased estimator of its expected value. For proving this the assumption of independence and normality of the x's play an essential role.

If, however, we want to restrict our choice of t to only quadratic functions and take the square of the error as the weight function then it is apparent from the remarks at the end of §2 that the above proof will hold even in the less restricted situation where the moments and the product moments of the z's and the y's up to order four are symmetrical in the z's and further where S_0 be uncorrelated with Q. This will be so if the z's and the y's be mutually uncorrelated up to order four and if the third moments of the of the z's be zeros.

Definition: A set of chance variables $x_1, x_2, \ldots x_n$ are said to be mutually uncorrelated up to order p if

$$Ex_1^{p_1}x_2^{p_2}\ldots x_n^{p_n}=Ex_1^{p_1}.Ex_2^{p_2}\ldots Ex_n^{p_n}$$

for all non-negative integers $p_1, p_2, \ldots p_n$ such that $p_1+p_2+\ldots+p_n=p$.

Let $x_1, x_2, \ldots x_n$ be mutually uncorrelated up to order p and let

$$y_i = a_{i1}x_1 + \ldots + a_{in}x_n \qquad \ldots \quad (4.4)$$
$$i = 1, 2, \ldots n$$

be a linear transformation of the x's.

Under what conditions the y's also will be mutually uncorrelated up to order p?

Since the chance variables $\{a_i x_i + b_i\}$ $i = 1, 2, \ldots n$ will also be mutually uncorrelated up to order p it follows that we can, without any loss of generally, assume that the x's have zero means and unit variances. By adjusting the scales of the y's we can then have that the y's also have zero means and unit variances.

$$\therefore \qquad 1 = V(y_i) = \sum_{r=1}^{n} a^2_{ir} \qquad i = 1, 2, \ldots n$$

and

$$0 = Ey_i Ey_j = Ey_i y_j$$
$$= \sum_{1}^{n} a_{ir} a_{jr} \; (i \neq j, \; i, j = 1, 2, \ldots n)$$

$\therefore$ the transformation (4.4) must be a unitary orthogonal transformation.

Let $c_i(t)$ and $k_i(t)$ be the cumulant generating functions of x_i and y_i respectively znd let $c(t_1, t_2, \ldots t_n)$ and $k(t_1, t_2, \ldots t_n)$ be the joint cumulant generating functions of the x's and the y's respectively. Also let c_{im} and k_{im} be the mth cumulants of x_i and y_i respectively ($i = 1, 2, \ldots n$, $m = 1, 2, \ldots p$). We know that

$$c_{i1} = k_{i1} = 0 \text{ and } c_{i2} = k_{i2} = 1 \qquad i = 1, 2, \ldots n.$$

Now since the x's are uncorrelated up to order p it follows that

$$c(t_1, t_2, \ldots t_n) = c_1(t_1) + \ldots + c_n(t_n) \qquad \ldots \quad (4.5)$$

up to terms with power $\leqslant p$.

Hence

$$k_i(t) = c(a_{i1}t, a_{i2}t, \ldots a_{in}t)$$
$$= c_1(a_{i1}t) + \ldots + c_n(a_{in}t) \qquad \ldots \quad (4.6)$$

up to terms with power $\leqslant p$.

Also

$$k(t_1, t_2, \ldots t_n) = c(\Sigma a_{i1}t_i, \ldots \Sigma a_{in}t_i) \qquad \ldots \quad (4.7)$$
$$= c_1(\Sigma a_{i1}t_i) + \ldots + c_n(\Sigma a_{in}t_i)$$

upto terms with power $\leqslant p$.

A necessary and sufficient condition for the y's to be uncorrelated up to order p is that

$$k(t_1, t_2, \ldots t_n) = k_1/t_1) + \ldots + k_n/t_n) \qquad \ldots \quad (4.8)$$

upto terms with power $\leqslant p$.

Now it is easily seen that if

$$c_{im}=0 \text{ for } m=3, 4,\ldots p \qquad i=1, 2,\ldots n$$

then (4.8) will be satisfied and then

$$k_{im}=0 \text{ for } m=3, 4,\ldots p, \qquad i=1,2,\ldots n.$$

Now if (4.8) be true then we have

$$\begin{aligned} c_i(t) &= k(a_{1i}t, a_{2i}t,\ldots a_{ni}t) \\ &= k_1(a_{1i}t)+\ldots+k_n(a_{ni}t) \end{aligned} \qquad \ldots \quad (4.9)$$

up to terms with power $\leqslant p$.

From (4.6) and (4.9) we have

$$k_{im}=a_{i1}^m c_{1m}+a_{i2}^m c_{2m}+\ldots+a_{in}^m c_{nm}$$

and $\qquad \ldots \quad (4.10)$

$$c_{im}=a_{1i}^m k_{1m}+a_{2i}^m k_{2m}+\ldots+a_{ni}^m k_{nm}$$

$$i=1,2,\ldots n, \quad m=1,2,\ldots p.$$

Hence from a Lemma proved earlier (Basu 1951) it follows that if no $a_{ij}=\pm 1$ then (4.10) can be satisfied if and only if

$$c_{im}=k_{im}=0 \qquad i=1,2\ldots n, \quad m=3,4,\ldots p. \qquad \ldots \quad (4.11)$$

If some $a_{ij}=\pm 1$ then it means that y_i is a function of x_j alone and that no other y involves x_j. Then the uncorrelatedness (up to order p) of y_i with the other y's will follow from the uncorrelatedness (up to order p) of the x's and no further restriction on the cumulants of x_j need be imposed. Thus ignoring such trivial cases we have the following:

Theorem 2: *If $x_1, x_2,\ldots x_n$ be uncorrelated up to order p then a necessary and sufficient condition that there exist non-trivial linear transformations of the x's into $y_1, y_2,\ldots y_n$ such that the y's also are uncorrelated up to order p is that the x's(and therefore the y's) are normal up to order p i.e. $c_{im}=0$, $m=3, 4,\ldots p$, $i=1, 2,\ldots n$.*

Also if the x's have the same variance then any orthogonal transformation will make the y's uncorrelated up to order p.

Thus if in the Markoff set-up we assume that the x's are mutually uncorrelated up to order four and that $\beta_1=0$ and $\beta_2=3$ for all the x's (i.e. the x's are normal up to order four) then the transformation (4.1) will make the z's and the y's uncorrelated and normal up to order four and so in the same way as before we prove that

$$S_0=\frac{1}{n-m}\Sigma z_i^2 \qquad \ldots \quad (4.12)$$

is the minimum variance estimator of σ^2 in the class of all unbiased polynomial estimators of degree not exceeding two.

Hsu (1938) and Rao (1952) proved the same result over a less restricted distribution space Ω but had, therefore, to restrict the scope of the choices for the estimator t.

Hsu, for instance, restricts the choice of t to the class of unbiased quadratic forms xAx' for which $V(xAx')$ is independent of the unknown parameters $\tau_1, \tau_2, \ldots \tau_m$. Rao considers only definite quadratic forms.

REFERENCES

BASU, D. (1951): On the independence of linear functions of independent chance variables. *Proc. Int. Stat. Conf.*, India, 1951 (in press).

HSU, P. L. (1938): On the best unbiased quadratic estimate of the variance. *Stat. Res. Mem.*, 2.

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

WALD, A. (1950): *Statistical Decision Functions*. John Wiley & Sons, New York.