

## ON A THEOREM OF HALMOS CONCERNING UNBIASED ESTIMATION OF MOMENTS

H. S. KONIJN

(received 11 August 1963)

### 1. Introduction

In [4] Halmos considers the following situation. Let  $\mathcal{D}$  be a class of distribution functions over a given (Borel) subset  $E$  of the real line, and  $F$  a function over  $\mathcal{D}$ . He investigates which functions  $F$  admit estimates that are unbiased over  $\mathcal{D}$  and what are all possible such estimates for any given  $F$ . In particular he shows that on the basis of a sample (of size  $n$ ) one can always obtain an estimate of the first moment which is unbiased in  $\mathcal{D}$  and that the central moments  $\bar{F}_m$  of order  $m \geq 2$  have estimates which are unbiased in  $\mathcal{D}$  if and only if  $n \geq m$ , provided  $\mathcal{D}$  satisfies the following properties:  $\bar{F}_m$  exists and is finite for all distributions in  $\mathcal{D}$  and  $\mathcal{D}$  includes all distributions which assign probability one to a finite number of points of  $E$ . Halmos also finds that symmetric estimates which are unbiased on  $\mathcal{D}$  are unique<sup>1</sup> and have smaller variances on  $\mathcal{D}$  than unsymmetric unbiased estimates.

He recognizes that his assumptions are too restrictive for most applications and mentions in particular the case where  $\mathcal{D}$  is the class of all normal distributions. The present paper addresses itself to that case.

### 2. Statement of results

If  $\mathcal{D}$  is the class of all nondegenerate univariate normal distributions, then, on the basis of a sample (of size  $n$ ), an estimate of the first moment which is unbiased over  $\mathcal{D}$  exists (and is unique when  $n = 1$ ); and a central moment of order  $2r \geq 2$  has estimates which are unbiased over  $\mathcal{D}$  if and only if  $n \geq 2r$ , and has a unique symmetric unbiased estimate when  $n = 2r$ , but not when  $n > 2r$ .

Specifically, this means the following:

Let  $z_1, \dots, z_n$  be a sample from a normal distribution with mean  $\nu$  and variance  $\omega^2 > 0$ . Let  $\bar{z} = n^{-1} \sum z_i$ ,  $S^2 = \sum (z_i - \bar{z})^2$ . Recall that the even

<sup>1</sup> It will be convenient to call a function on a  $k$ -dimensional Euclidean space the unique function satisfying a certain property if any other function on this space satisfying the property may differ from it only on a set of  $k$ -dimensional Lebesgue measure zero.

central moments  $\bar{F}_{2r}$  equal  $\omega^{2r} 2^{-r} (2r)!/r!$  and the odd ones vanish.

(a) If  $n = 1$ ,  $\bar{z}$  is the unique unbiased estimate of  $\nu$ , and no unbiased estimate of  $\bar{F}_{2r}$  exists for  $r = 1, 2, \dots$ .<sup>2</sup> In [5] this seemingly uninteresting fact turns out to be the key to a quite practical question.

(b) If  $n \geq 2$ ,

$$\bar{f}_{2r} = \frac{\{(n-3)/2\}! (2r)!}{\{(n+2r-3)/2\}! r!} (S/2)^{2r}$$

is an unbiased estimate of  $\bar{F}_{2r}$  ( $r = 1, 2, \dots$ ), and is the unique symmetric unbiased estimate if  $n = 2$ , but not if  $n > 2$ . It then follows from [6] that  $\bar{z}$  and  $\bar{f}_{2r}$

(c) are the unique unbiased estimates of  $\nu$  and  $\bar{F}_{2r}$ , respectively, which depend only on the sufficient statistic  $(\bar{z}, S^2)$  and

(d) have the smallest variance among all unbiased estimates.

Note that  $\bar{z}$  and  $S^2$  are symmetric functions of the observations. The usual symmetric estimate  $\bar{f}'_{2r}$  for  $\bar{F}_{2r}$ , which is unbiased for all distribution functions for which  $\bar{F}_{2r}$  exists, is defined only when  $n \geq 2r$ . When  $r = 1$  it coincides with  $\bar{f}_2$ , when  $r = 2$  it equals [2, 27.6]

$$\bar{f}'_4 = (n!)^{-1} (n-4)! \{n(n^2-2n+3) \sum (z_i - \bar{z})^4 - 3(2n-3)S^4\} \quad (n \geq 4).$$

For any family  $\mathcal{D}$  as first mentioned in the introduction or mentioned in the final section  $\bar{f}'_{2r}$  is the only symmetric estimate which is unbiased for all distributions of  $\mathcal{D}$ . But, if for  $\mathcal{D}$  we take the class of nondegenerate univariate normal distributions, our results imply that the symmetric estimate  $\bar{f}_{2r}$  is also unbiased over this class and has a smaller variance than  $\bar{f}'_{2r}$  for  $r > 1$ .

In the next two sections we prove the parts of (a) and (b) which are not immediate.

### 3. Nonexistence of an unbiased estimate of $\bar{F}_{2r}$ in a sample of one

In this section denote  $z_1$  by  $z$ . If  $h(z)$  is an unbiased estimate of  $\bar{F}_{2r}$  then

$$\int_{-\infty}^{\infty} \{h(z+\nu) - z^{2r}\} \exp(-\frac{1}{2}z^2\omega^{-2}) dz$$

should vanish for all  $\nu$  and all  $\omega > 0$ . This integral can be written as an

<sup>2</sup> It has been remarked that it is obvious that from a sample of one it is not possible to obtain an unbiased estimate of two independent parameters (that is, two functions  $F_1$  and  $F_2$  on a class of distributions such that there exists no function  $g$  in the plane with  $g\{F_1(D), F_2(D)\} = 0$  for all distributions  $D$  in the class). That this is not so is easily shown by an example. Let  $\theta^2 = \nu^2 + \omega^2$ , where  $\nu$  and  $\omega^2$ , the mean and variance, are independent parameters when, e.g., the class is the normal class. Then  $\nu$  and  $\theta^2$  are also independent parameters over that class with unbiased estimates  $z_1$  and  $z_1^2$ .

integral over the positive axis and then we can make the substitution  $u = z^{\frac{1}{2}}$  and obtain, setting  $\omega' = (2\omega^2)^{-1}$ , that

$$\int_0^\infty \{h(-u^{\frac{1}{2}} + \nu) + h(u^{\frac{1}{2}} + \nu) - 2u^r\} u^{-\frac{1}{2}} \exp(-u\omega') du$$

is zero for all  $\nu$  and all  $\omega' > 0$ . This being a Laplace transform of  $u^{-\frac{1}{2}}$  times the expression in brackets, it follows that

$$h(-z + \nu) + h(z + \nu) - 2z^{2r} = 0$$

for all  $\nu$  and almost all positive  $z$ . For all  $\nu$  there is a set  $S_\nu$  on the positive  $z$  axis such that the Lebesgue measure  $l$  of the positive points  $z$  not in  $S_\nu$  is zero and such that the above equality holds on  $S_\nu$ . Denote  $\bigcap_{k=1,2,4,5} S_{c+\frac{ak}{2}}$  by  $T$ .

It is easily shown<sup>3</sup> that there exists a pair of points  $a$  and  $\frac{1}{2}a$  in  $T$ . Choosing  $\nu = a$  and  $2a$  respectively gives for  $z = a$

$$h(0) + h(2a) = 2a^{2r}, \quad h(a) + h(3a) = 2a^{2r},$$

so that

$$h(0) + h(a) + h(2a) + h(3a) = 4a^{2r}.$$

Choosing  $\nu = \frac{1}{2}a$  and  $2\frac{1}{2}a$  respectively gives for  $z = \frac{1}{2}a$

$$h(0) + h(a) = a^{2r}/2^{2r-1}, \quad h(2a) + h(3a) = a^{2r}/2^{2r-1},$$

so that

$$h(0) + h(a) + h(2a) + h(3a) = a^{2r}/2^{2r-2}.$$

Since  $a \neq 0$ , this is a contradiction.

#### 4. Uniqueness of the unbiased symmetric estimate of $\bar{F}_{2r}$ in a sample of two and nonuniqueness in a larger sample

For  $n \geq 2$  (so that  $S^2$  is not identically zero) the sufficiency of the statistic  $(\bar{z}, S^2)$  and the completeness of its distribution imply that  $\bar{f}_{2r}$  is the unique unbiased estimate of its expectation  $\bar{F}_{2r}$  among unbiased estimates depending on  $(\bar{z}, S^2)$  only [5]. Now if  $n = 2$ ,  $(\bar{z}, S^2)$  determines the set  $\{z_1, z_2\}$  of observations, but not their order. Therefore  $\bar{f}_{2r}$  is also the unique unbiased estimate of  $\bar{F}_{2r}$  among unbiased estimates which are symmetric in the observations.

In general, when  $n > 2$ , for any  $a \neq 0$ ,

<sup>3</sup> Let  $a'$  be in  $T$  and let  $0 < b < a'$ . Define the disjoint intervals  $I_i$  from  $ia'$  to  $i(a'+b)$  for  $i = 1, 2$ , which have  $l(I_i T) = ib$ . Denote by  $p_j(I_i T)$  the set of points  $x$  in  $I_i T$  such that  $ix/j$  is in  $I_i T$ ;  $l\{p_j(I_i T)\} = jb$ . Now let

$$T_2 = Tp_2(I_1 T), \quad T_1 = p_1(I_2 T);$$

then, since the  $T_i$  are subsets of  $T$  of measure  $ib$ , there exists  $a > 0$  such that  $\frac{1}{2}ia$  is in  $T_i$  for  $i = 1$  and  $2$ . In fact, there exist  $c$  such that, for almost all  $a$  in  $T$ ,  $\frac{1}{2}a$  is in  $T$ . For brevity use  $c = 0$ .

$$\bar{f}_{2r} + a\{n(n+1) \sum (z_i - \bar{z})^4 - 3(n-1)S^4\}$$

will be an unbiased symmetric estimate of  $\bar{F}_{2r}$  different from  $\bar{f}_{2r}$ , since the mean of  $\sum (z_i - \bar{z})^4$  is  $3n^{-1}(n-1)^2\omega^4$  and the mean of  $S^4$  is  $(n-1)(n+1)\omega^4$ , and since for  $n > 2$  the bracket is not identically equal to zero. For example, if  $n = 3$ ,  $\frac{1}{2} \sum (z_i - \bar{z})^4$  has mean  $\bar{F}_4 + 3\bar{F}_2^2$  and, in the normal case,  $S^4$  has mean  $8\bar{F}_2^2$ , so that  $\frac{1}{2}\{\sum (z_i - \bar{z})^4 - S^4/4\}$  and  $\frac{3}{4} \sum (z_i - \bar{z})^4$  are unbiased estimates of  $\bar{F}_4$  different from  $\bar{f}_4 = 3S^4/8$ .

### 5. Remarks

One could similarly discuss unbiased estimation of other functions over the class of normal distributions.

Fraser [3] adapts Halmos' argument to cases where  $\mathcal{L}$  is a certain class of distributions that have a density. Some cases of this kind have been found by Lehmann and Scheffé; see [1].

The writer is much indebted to T. C. Koopmans and T. N. Srinivasan for helpful suggestions.

### References

- [1] Bell, C. B., Blackwell, D., and Breiman, L., On the Completeness of Order Statistics. *Ann. Math. Statist.*, **31**, (1960), 794-7.
- [2] Cramér, H. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, (1946).
- [3] Fraser, D. A. S., Completeness of Order Statistics. *Canad. J. of Math.*, **6**, (1954), 42-5.
- [4] Halmos, P. R., The Theory of Unbiased Estimation. *Ann. Math. Statist.*, **17**, (1946), 34-43.
- [5] Konijn, H. S., Non-Existence of Consistent Estimator Sequences and Unbiased Estimators: A Practical Example. (To appear)
- [6] Lehmann, E. L., and Scheffé, H., Completeness, Similar Regions, and Unbiased Estimation. *Sankhya*, **10** (1950) 305-40; **15**, (1955) 219-36.

Department of Economics,  
The City College,  
New York,  
and  
Cowles Foundation.