The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1755–1758

ANCILLARY STATISTICS AND ESTIMATION OF THE LOSS IN ESTIMATION PROBLEMS

BY ELSE SANDVED

University of Oslo

Consider a stochastic variable X with probability density $f(x; \theta, \mu, \nu)$ where θ is a scalar whose value we want to estimate, and where μ and ν are nuisance (if they are present). We assume that all parameters can take on values independently of each other. Let the estimator be denoted by $\hat{\theta}$, let the loss be $L(\theta, \hat{\theta})$, and let the risk be $R = EL$.

Suppose that we have chosen the estimator $\hat{\theta}(X)$ (according to one or another principle), and now want a measure of the accuracy of the estimate $\hat{\theta}(x)$. If the loss $L(\theta, \hat{\theta}(x))$ were known, it would have been a perfect measure of the accuracy of $\hat{\theta}(x)$. But $L(\theta, \hat{\theta}(x))$ is of course unknown (except in very trivial cases). *On the other hand, one can estimate the unobservable stochastic variable $L(\theta, \hat{\theta}(X))$ by means of X.*

Let us denote the estimator of $L(\theta, \hat{\theta}(X))$ by $\Lambda(X)$. If $\Lambda(X)$ is close to $L(\theta, \hat{\theta}(X))$ with high probability for all values of the parameters, then it seems reasonable to consider $\Lambda(x)$ as a measure of the accuracy of $\hat{\theta}(x)$, and to present it together with $\hat{\theta}(x)$ as such a measure.

In this paper, we shall consider *best unbiased estimators of L*, defined in the following way:

DEFINITION 1. $\Lambda_0(X)$ is a *best unbiased estimator of L* if it is an unbiased estimator of R, i.e.

$$(1) \qquad E\Lambda_0(X) \equiv R \quad \text{or} \quad (E[\Lambda_0(X) - L(\theta, \hat{\theta}(X))] \equiv 0),$$

and if

$$(2) \qquad E[\Lambda_0(X) - L(\theta, \hat{\theta}(X))]^2 \leqq E[\Lambda(X) - L(\theta, \hat{\theta}(X))]^2$$

for all $\Lambda(X)$ such that (1) is satisfied.

THEOREM 1. *Suppose that T is a complete sufficient statistic, and that L depends on X only through T. Then there is at most one estimator $\Lambda(T)$ of L such that $E\Lambda(T) \equiv R$. If there is one, then it is a best unbiased estimator of L.*

PROOF. Exactly as for the corresponding result for Markov-estimators of parameters.

EXAMPLE 1. $X_1, \cdots, X_n$ are independent and identically normally distributed, $EX_i = \theta$, Var $X_i = \sigma^2$ which is known. Let $\hat{\theta} = \bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and $L = (\hat{\theta} - \theta)^2$. $R = \sigma^2 n^{-1}$. $\Lambda \equiv R$ is of course an unbiased estimator of R, and because of completeness, R is the best unbiased estimator of $(\hat{\theta} - \theta)^2$.

EXAMPLE 2. Let the situation be as in Example 1, except that σ^2 is unknown. Let $\Lambda = (n - 1)^{-1} \sum (X_i - \bar{X})^2$. Since $E\Lambda \equiv R$, Λ is a best unbiased estimator of $(\hat{\theta} - \theta)^2$, because of completeness.

Received 27 September 1967.

In Example 1, the best unbiased estimator of the loss equals Var $\hat{\theta}$, in Example 2, the best unbiased estimator of the loss equals the Markov-estimator of Var $\hat{\theta}$. It is customary to present an estimate $\hat{\theta}$ together with Var $\hat{\theta}$ (or, in case it is unknown, with an estimate of Var $\hat{\theta}$). This Var $\hat{\theta}$ (the estimate of Var $\hat{\theta}$) is then often considered as a measure of the accuracy of the estimate, and this seems reasonable in Examples 1 and 2, since in those examples, Var $\hat{\theta}$ and the Markov-estimator of Var $\hat{\theta}$, respectively, are equal to the best unbiased estimator of the loss.

But Var $\hat{\theta}$ (an estimate of Var $\hat{\theta}$) is not in general usable as a measure of the accuracy of an unbiased estimate. One can often find much better estimators of $(\hat{\theta} - \theta)^2$ than Var $\hat{\theta}$.

THEOREM 2. *Suppose that* $E[L(\theta, \hat{\theta}) \mid Y] \equiv \Lambda(Y)$ *is a statistic. Then* $\Lambda(Y)$ *is uniformly at least as good an estimator of the loss, as the risk* R *is (if* R *is known), and* $\Lambda(Y)$ *is the best estimator (of the loss) based upon* Y.

PROOF.

$$E[L(\theta, \hat{\theta}) - R]^2 \equiv E[L(\theta, \hat{\theta}) - \Lambda(Y)]^2 + E[\Lambda(Y) - R]^2$$
$$\equiv E[\varphi(Y) - L(\theta, \hat{\theta})]^2 - E[\Lambda(Y) - \varphi(Y)]^2 + E[\Lambda(Y) - R]^2$$

for any function $\varphi(Y)$. This is easy to show by multiplying out.

Consequently, if conditional expected loss, given a statistic, itself is a statistic, then it is a better estimator of the loss than the risk is.

EXAMPLE 3. $X_1, \cdots, X_n$ are independent and uniformly distributed $R(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Let $Y = \max X_i - \min X_i$, $Z = \hat{\theta} = \frac{1}{2}(\min X_i + \max X_i)$. Let $L = (\hat{\theta} - \theta)^2$. (Y, Z) is sufficient, and the density of Y equals

$$\begin{array}{ll} n(n-1)y^{n-2}(1-y) & \text{for } 0 < y < 1 \\ 0 & \text{otherwise.} \end{array}$$

The conditional density of Z, given $Y = y$, equals $R(\theta - \frac{1}{2}(1-y), \theta + \frac{1}{2}(1-y))$. $E[(\hat{\theta} - \theta)^2 \mid Y] = \frac{1}{12}(1 - Y)^2 = \Lambda(Y)$ is according to Theorem 2 a better unbiased estimator of the loss than $R = \frac{1}{2}((n+1)(n+2))^{-1}$, $(E(R - L)^2 = \frac{3}{2}n!((n+4)!)^{-1} - \frac{1}{4}(n+1)^{-2}(n+2)^{-2}$, $E(\Lambda - L)^2 = \frac{3}{2}n!((n+4)!)^{-1} - 5n!(6(n+4)!)^{-1}$ and $E(\Lambda - L)^2(E(R - L)^2)^{-1} = 8(n+1)(n+2).(3n(5n+11))^{-1}$.) The model is not complete, so we have not shown that it is a best unbiased estimator of $(\hat{\theta} - \theta)^2$. It is not known whether it is or not.

Intuitively it is reasonable that $\frac{1}{12}(1 - y)^2$ is a good measure of the accuracy of $\hat{\theta}$. If for instance $y = 1$, then we estimate the loss to be zero, and this is completely in accordance with the fact that we then know that we have hit the correct value of θ.

DEFINITION 2. Let (Y, Z, V) be sufficient, where Y and V may be vectors, let $Z = \hat{\theta}$, and let the joint density of (Y, Z, V) be $g(y; \nu)h(z, v \mid y, \theta, \mu)$, where g equals the density of Y. Then Y is *ancillary for* θ. (Here θ is thought of as the parameter of interest; Y is of course also ancillary for μ.) If ν is not present, that is, if the density of Y is completely specified, then Y is *ancillary*.

NOTE. Y is ancillary in Example 3.

To consider an ancillary statistic as given in a statistical inference problem was first recommended by R. A. Fisher. He gave various descriptions of "ancillary statistics", for instance ([2], p. 48) "... ancillary statistics, which themselves tell us nothing about the value of the parameter, but, instead, tell us how good an estimate we have made of it." Basu [1] gives several examples of conditioning on ancillary statistics. Definition 2 above of ancillary statistics (when ν is not present) agrees with the definition of Basu. Lindley [3], p. 58, has a definition of ancillary statistics when there are nuisance parameters. If in Definition 2 above μ is not present, then, in his definition, Y is sufficient for ν and, given ν, Y is ancillary for θ. Definition 2 also agrees with the definitions in [4] and [5] of "an ancillary statistic with respect to a parameter θ."

THEOREM 3. *Let the situation be as in Definition* 2. *If Y is ancillary for θ and complete for the family $g(\cdot\,;\nu)$, then*

(a) $E\hat{\theta} \equiv \theta \Rightarrow E(\hat{\theta} \mid Y = y) \equiv \theta$ *for all* $y \varepsilon A_{\theta,\mu}$, *where* $\int_{A_{\theta,\mu}} g(y;\nu)\,dy \equiv_\nu 1$.

(b) *If $\hat{\theta}$ is a conditional Markov-estimator of θ, given $Y = y$, for every y, then $\hat{\theta}$ is a Markov-estimator of θ.*

(c) $E\Lambda(Y, Z, V) \equiv EL(\theta, \hat{\theta}) \Rightarrow E[\Lambda(Y, Z, V) \mid Y = y] \equiv E[L(\theta, \hat{\theta}) \mid Y = y]$ *for all* $y \varepsilon A_{\theta,\mu}$, *where* $\int_{A_{\theta,\mu}} g(y;\nu)\,dy \equiv_\nu 1$.

(d) *If $\Lambda(Y, Z, V)$ is a conditional best unbiased estimator of $L(\theta, \hat{\theta})$, given $Y = y$ for every y, then $\Lambda(Y, Z, V)$ is a best unbiased estimator of $L(\theta, \hat{\theta})$.*

The proof is straightforward.

EXAMPLE 4. Let $Z_1, \cdots, Z_y$, given $Y = y$, be identically, independently normally distributed, with $EZ_i = \theta$, Var $Z_i = \mu$, and let Y be complete in its distribution. (For instance, the distribution of Y may be completely unknown, under the condition that Y is an integer greater than unity.) Let $Z = \hat{\theta} = Y^{-1} \sum_{i=1}^{Y} Z_i$. Then, for given $Y = y$, $y^{-1} \sum_{i=1}^{y} Z_i$ is the conditional Markov-estimator of θ, and $y^{-1}(y-1)^{-1} \sum_{i=1}^{y} (Z_i - Z)^2$ is the conditional best unbiased estimator of $L = (\hat{\theta} - \theta)^2$. It then follows from Theorem 3 that $\hat{\theta}$ is a Markov-estimator of θ, and that $\Lambda = V = Y^{-1}(Y-1)^{-1} \sum_{i=1}^{Y} (Z_i - Z)^2$ is a best unbiased estimator of $(\hat{\theta} - \theta)^2$.

NOTE. In this example, Λ equals the Markov-estimator of the conditional variance of $\hat{\theta}$, given a statistic which is ancillary for θ.

REFERENCES

[1] BASU, D. (1964). Recovery of ancillary information. *Sankhyā* **26** 3–16.
[2] FISHER, R. A. (1935). The logic of inductive inference. *J. Roy. Statist. Soc.* **48** 39–54.
[3] LINDLEY, D. V. (1965). *Introduction to Probability and Statistics*. Cambridge Univ. Press.
[4] SANDVED, E. (1965). A principle for conditioning on an ancillary statistic. *Skand. Aktuarietidskr.* **49** 39–47.
[5] SANDVED, E. (1967). Ancillary statistics in models without and with nuisance parameters. Statistical Research Report No. 5 1967, Inst. of Math., Univ. of Oslo.