A SIMPLE METHOD FOR IMPROVING SOME ESTIMATORS

By Leo A. Goodman[1]

University of Chicago

1. Introduction and Summary. In the past, the principles which have been applied most often in the selection of an estimate are the principles of maximum likelihood and of minimum variance unbiased estimation. Recent statistical literature (e.g., [1]) has pointed out the fact that, while these principles are intuitively appealing, neither of them can be justified very well in a systematic development of statistics. Abraham Wald [2] has indicated a more systematic approach to the problem. One of Wald's ideas may be paraphrased as follows. Consider a random variable X whose distribution depends on an unknown real parameter θ. If the value x of X is observed one makes an estimate, say $f(x)$, and thereby incurs a loss of $W[\theta, f(x)]$. The risk associated with the estimate f is defined to be the expected loss $R_f(\theta) = E\{W[\theta, f(X)] \mid \theta\}$. In choosing between two estimators f_1 and f_2, it seems clear that one would prefer f_1 to f_2 if $R_{f_1}(\theta) \leqq R_{f_2}(\theta)$ for all values of θ, and $R_{f_1}(\theta) < R_{f_2}(\theta)$ for at least one value of θ.

We shall consider only the case where the loss as a function of θ and the estimate $f(x)$ is of the special form

$$W[\theta, f(x)] = \lambda(\theta)(f - \theta)^2; \lambda(\theta) > 0. \tag{1}$$

Reasons for considering this form of $W[\theta, f(x)]$ have been given in [3]. Suppose that we know of an unbiased estimate f whose variance is $K\theta^2$, where K is known. Then, as we shall see, the risk of f is greater than the risk of $f/(K + 1)$. Hence $f/(K + 1)$ is to be preferred to f as an estimator of θ. This result holds for any function $\lambda(\theta) > 0$. Although $f/(K + 1)$ is generally not unbiased in the usual sense, it is unbiased in a certain sense (cf. [4]).

It is seen that a special case of this result is related to the problem of the estimation of the scale parameter of a population whose form is not given but for which the ratio of the first and second moments is known (cf. [5]).

Some special cases and applications are discussed in detail.

2. Results. Let Y be any real-valued random variable whose distribution depends on an unknown real parameter $\theta > 0$.

Theorem 1. *Suppose $\theta E\{Y \mid \theta\}/E\{Y^2 \mid \theta\} = A$ identically in θ, where A is known. Then among all statistics of the form aY, where a is a constant, the risk*

$$R_{aY}(\theta) = E\{\lambda(\theta)(aY - \theta)^2 \mid \theta\}$$

is minimized for each value of θ when $a = A$.

Received 9/29/52.

[1] This report was prepared in connection with research supported by the Office of Naval Research.

PROOF. We have that

$$R_{aY}(\theta) = E\{\lambda(\theta)[a^2Y^2 + \theta^2 - 2aY\theta] \mid \theta\}$$
$$= \lambda(\theta)[a^2E\{Y^2 \mid \theta\} + \theta^2 - 2aAE\{Y^2 \mid \theta\}].$$

The risk is a quadratic function of a which is minimized when

$$\frac{\partial R_{aY}(\theta)}{\partial a} = 2\lambda(\theta)[a - A]E\{Y^2 \mid \theta\} = 0;$$

that is, when $a = A$. Q.E.D.

In this case,

$$R_{AY}(\theta) = \lambda(\theta)[\theta^2 - A^2E\{Y^2 \mid \theta\}].$$

For the function which uniformly minimizes the expected loss, we have that

$$E\{AY \mid \theta\} = \theta[E\{Y \mid \theta\}]^2/E\{Y^2 \mid \theta\} \leqq \theta.$$

Hence, this function will be unbiased only when $E\{Y^2 \mid \theta\} = [E\{Y \mid \theta\}]^2$; that is, when the variance of Y is zero.

Following Lehmann [4], we say that an estimate Y is unbiased with respect to the loss function $W[\hat{\theta}, Y] = \lambda(\hat{\theta})(Y - \hat{\theta})^2$ if for each θ, $E\{W[\hat{\theta}, Y] \mid \theta\}$ is minimized when $\hat{\theta} = \theta$.

THEOREM 2. *Suppose $\theta E\{Y \mid \theta\}/E\{Y^2 \mid \theta\} = A$ identically in θ, where A is known. Then among all statistics of the form aY, where a is a constant, the only one which is unbiased with respect to the loss function $W[\hat{\theta}, Y]$, when $\lambda(\hat{\theta}) = \hat{\theta}^{-2}$, is AY (which uniformly minimizes the risk).*

PROOF. The expected loss function is a quadratic function of $1/\hat{\theta}$ and the minimum of this quadratic function may be computed as in the proof of Theorem 1. We see that this function is minimized when $\hat{\theta} = \theta$ if and only if $a = A$. Q.E.D.

If Y is an unbiased estimate (in the usual sense) of θ, then, $A = \theta^2/E\{Y^2 \mid \theta\}$. The relative improvement in risk obtained by using AY is

$$1 - R_{AY}(\theta)/R_Y(\theta) = 1 + [1 - A]/[1 - 1/A] = 1 - \text{A}.$$

Since the variance of Y is $[(1/A) - 1]\theta^2 = K\theta^2$ we may write AY as $Y/[K + 1]$ and the relative improvement in risk is AK. We have found that $Y/[K + 1]$ is unbiased with respect to the loss function (1) when $\lambda(\theta) = \theta^{-2}$.

Let us now consider the special case where Y is a real-valued random variable whose distribution has the invariance property under a change of scale; that is, the probability function of Y is $\theta^{-1}f(y/\theta)$, $\theta > 0$, where the function $f(y)$ is known, but the parameter θ which determines the scale of the distribution of Y is unknown. Then $E\{Y \mid \theta\} = M\theta$ and $E\{Y^2 \mid \theta\} = N\theta^2$ where $M = E\{Y \mid 1\}$ and $N = E\{Y^2 \mid 1\}$. Since the conditions of Theorems 1 and 2 hold, we see that among all invariant functions $g(Y)$ (i.e., functions having the invariance property $g(cY) = cg(Y)$ for all $c > 0$) there is one which uniformly minimizes the expected loss. This function is $YE\{Y \mid 1\}/E\{Y^2 \mid 1\} = D$. This fact may also be proved using a result due to Pitman ([5], p. 406) which deals with the case where one has a sample from the distribution $\theta^{-1}f(y \mid \theta)$. Pitman showed that the invariant estimator with the smallest mean square error is

$$C = \int_0^\infty \theta^{-3}f\left(\frac{Y}{\theta}\right) d\theta \Big/ \int_0^\infty \theta^{-4}f\left(\frac{Y}{\theta}\right) d\theta.$$

By a change of variables we see that $C = D$. Hence the estimate C may be computed even if the function $f(y)$ is not given but the ratio of the first and second moments is known. In fact, even if the probability distribution of Y/θ is a function of θ, we see that C is the invariant estimate of θ with the smallest mean-squared error when the ratio of the mean and second moment of Y/θ is known. We also have that C is unbiased with respect to the loss function (1) when $\lambda(\hat{\theta}) = \hat{\theta}^{-2}$.

3. Applications. Suppose $x_1, x_2, \cdots, x_n$ is a sample of n from a normal distribution where both the mean μ and variance σ^2 are unknown. Writing $\bar{x} = \sum_{i=1}^n x_i/n$ and $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n-1)$, we find that $s^2(n-1)/(n+1)$ is the invariant estimate of σ^2 which minimizes the risk. Also,

$$s[2/(n-1)]^{1/2}\Gamma(n/2)/\Gamma((n-1)/2)$$

is the invariant estimate of σ which minimizes the risk. When μ is known, writing $t^2 = \sum_{i=1}^n (x_i - \mu)^2/n$, we find that $t^2n/(n+2)$ is the invariant estimate of σ^2 which minimizes the risk. Hodges and Lehmann ([1], p. 17) have shown that $t^2n/(n+2)$ is the unique admissible minimax estimate of σ^2 when $\lambda(\theta) = 1/\theta^2$ They point out the fact that t^2 is neither minimax nor admissible. We also find that $t[2/n]^{1/2}\Gamma((n+1)/2)/\Gamma(n/2)$ is the invariant estimate of σ which minimizes the risk.

Suppose $x_1, x_2, \cdots, x_n$ is a sample of n from a uniform distribution on $(0, p)$, where p is unknown. Writing $y = \max(x_1, x_2, \cdots, x_n)$, we see that $y(n+2)/(n+1)$ is the invariant estimate of p which minimizes the risk (cf. [4], p. 589).

The preceding examples deal with random variables whose distribution have the invariance property under a change of scale. The conditions $E\{Y \mid \theta\} = M\theta$ and $E\{Y^2 \mid \theta\} = N\theta^2$ are weaker than the condition of invariance under change of scale. Hence, Theorems 1 and 2 are stronger than the corresponding invariance theorems. The following example satisfies the conditions of our theorems ("second order invariance"), but the distribution of the random variable does not have the invariance property (in the usual sense) under change of scale.

The distribution of the random variable Y depends on an unknown real parameter θ which may be included in one of two disjoint sets Ω_1 or Ω_2. If θ is in Ω_1, then Y/θ has a Poisson distribution with a mean of 1 (variance also is 1). If θ is in Ω_2, then Y/θ has an exponential distribution with a mean of 1 (variance also is 1). Whether θ is included in Ω_1 or in Ω_2 is unknown, so that the distribution of Y/θ is a function of θ and does not have the invariance property under change of scale. We see that $Y/2$ is the invariant estimate of θ which minimizes the risk and that the risk associated with the unbiased estimator is twice the risk associated with $Y/2$.

Using Theorem 2, we find that all the invariant estimates described in this section which minimize the risk are also unbiased with respect to the loss function (1) when $\lambda(\theta) = \theta^{-2}$. We also see that these invariant estimates C have the optimum properties of minimizing the risk and being unbiased with respect to the loss function (1) with $\lambda(\theta) = \theta^{-2}$ even when the underlying distribution of the variates is not the one specified (normal, uniform, etc.) as long as $E\{C/\theta \mid \theta\} = E\{C^2/\theta^2 \mid \theta\}$.

REFERENCES

[1] J. L. Hodges, Jr. and E. L. Lehmann, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 13–22.

[2] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[3] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates arising from quadratic risk functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.

[4] E. L. Lehmann, "A general concept of unbiasedness," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 587–592.

[5] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of a given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.