NOTES

This section is devoted to brief research expository articles on methodology and other short items.

ON SMALL-SAMPLE ESTIMATION

By George W. Brown

Iowa State College

1. Summary. This paper discusses some of the concepts underlying small sample estimation and reexamines, in particular, the current notions on "unbiased" estimation. Alternatives to the usual unbiased property are examined with respect to invariance under simultaneous one-to-one transformation of parameter and estimate; one of these alternatives, closely related to the maximum likelihood method, seems to be new. The property of being unbiased in the likelihood sense is essentially equivalent to the statement that the estimate is a maximum likelihood estimate based on some distribution derived by integration from the original sampling distribution, by virtue of a "hereditary" property of maximum likelihood estimation.

An exposition of maximum likelihood estimation is given in terms of optimum pairwise selection with equal weights, providing a type of rationale for small sample estimation by maximum likelihood.

2. Introduction. In large sample theory of estimation the problems are generally formulated in terms of a random variable $x = (x_1, x_2, \cdots, x_n)$ and a product distribution with, say, a density $g(x|\theta) = f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta)$ where n is permitted to increase without limit. For small sample theory it is sufficient to consider an arbitrary distribution, not necessarily of product form, depending on a parameter θ. For convenience we will assume a distribution density of fixed form $g(x|\theta)$, where x is in Euclidean n-space and θ in Euclidean k-space, $k \leq n$. Granting at the outset that a complete rationale for estimation must be based on considerations like those of Wald [4, 1939] dealing with specified risk functions, it is still a difficult process, in practice, to specify the risk functions and solve the ensuing mathematics problems. It may still be to the point, then, to consider general properties that estimates might be required to have in order to be considered "acceptable", or perhaps even "optimum", over a class of "acceptable" estimates.

In large-sample theory the situation is fairly simple. Consistent estimates have the property that the estimate converges in probability to the true parameter value. "Best" or "optimum" estimates are defined in terms of the order of convergence, or asymptotic variance. All reasonable definitions of "optimum" become asymptotically equivalent, since they all measure essentially the rate of

convergence, so that one might ask for least variance, or least expected absolute deviation, or least expected kth power, without affecting the optimum estimate, in general. Moreover, the consistency property and the optimum properties are in general invariant under simultaneous one-to-one transformation of the parameter and its estimate, i.e., the square of an asymptotically optimum estimate of σ will be an asymptotically optimum estimate of σ^2. Finally, a general estimation method, the method of maximum likelihood, leads to optimum estimates in large samples.

In small samples, on the other hand, the search for corresponding criteria has led to the investigation of best "unbiased" estimates, and the like, where few, if any, of the definitions discussed possess an invariance property under simultaneous one-to-one transformation of the parameter and its estimate.

3. Unbiased estimation. To ensure, in small-sample estimation, that an estimate bears some relation to the parameter it is estimating, it has become the custom to require that an estimate be *unbiased*, which means that the expected value of the estimate agrees with the parameter value. This condition was suggested by the consistency property which is required in large-sample estimation. It ensures, moreover, that the average of a large number of independent estimates made on the same basis will provide a consistent estimate, in the large sample sense. While this consistency property of the average may at times be convenient in practical situations, the fact remains that the problem of estimation from a number of such observations is a different estimation problem, the "best" solution to which need not be the average of the "best" solutions of the original problem corresponding to estimation of θ from a single observation on x, where x has a density $g(x|\theta)$. More to the point, however, is the objection that an unbiased estimate of a parameter does not in general transform into an unbiased estimate when both estimate and parameter are subjected to the same one-to-one transformation. Moreover, one can easily construct situations for which the only acceptable unbiased estimates are clearly inferior from almost any point of view, to estimates which are biased (Girshick, Mosteller and Savage, [1, 1946], and Halmos [2, 1946]).

It may be of interest to consider a few reasonable alternatives to the lack of bias requirement, which seem to accomplish as much as the conventional definition and which, in addition, have an invariance under one-to-one transformation of the parameter and estimate. To avoid confusion, let us attach the qualifying prefix "mean" to the usual unbiased property, so that an estimate will be said to be *mean-unbiased* if its expected value agrees with the parameter value.

Consider as one alternative the following property. An estimate of a one-dimensional parameter θ will be said to be *median-unbiased*, if for fixed θ, the median of the distribution of the estimate is at the value θ, i.e., the estimate underestimates just as often as it overestimates. This requirement seems for most purposes to accomplish as much as the mean-unbiased requirement and has the additional property that it is invariant under one-to-one transformation.

A different alternative requirement which is invariant under transformations is suggested by the definition of unbiased tests of significance (Neyman and Pearson [3, 1936]). Let us say that an estimate is *likelihood-unbiased* if $h(\theta \mid \theta') \leq h(\theta \mid \theta)$, where the estimate $\hat{\theta}$ has probability density $h(\hat{\theta} \mid \theta)$. In other words, an estimation method is likelihood-unbiased if estimates in the neighborhood of a given parameter value θ would occur more frequently when the true value is itself θ than when it differs from θ. On intuitive grounds this seems to be an acceptable kind of requirement, applicable to a very general class of estimation problems. It is evident that the assumption of a density plays no important role here; the situation is analogous to the maximum likelihood situation. The property itself is invariant under simultaneous one-to-one transformations of parameter and estimate for the same reason that maximum likelihood estimates are invariant under such transformations, in fact one can readily see that the likelihood-unbiased condition is equivalent to requiring that $\hat{\theta}$ have such a distribution, as a function of θ, that the maximum likelihood estimate of θ based on $\hat{\theta}$ will be actually equal to $\hat{\theta}$. The obvious implication of this fact is that if a function $\phi(x)$ is given (possibly a sufficient statistic for θ) then there is an essentially unique likelihood-unbiased estimate $\hat{\theta}$ based on ϕ, obtained by finding the maximum likelihood estimate of θ in the distribution of ϕ as a function of θ.

As an example, consider the estimation of σ^2 from a sample of n observations from a normal distribution. Let S^2 be the usual sum of squares, where S^2/σ^2 is distributed like χ^2 on $n - 1$ degrees of freedom. Then the only likelihood—unbiased estimate of σ^2 based on S^2 is $S^2/(n - 1)$. In this case $S^2/(n - 1)$ is also mean-unbiased, a fact which is normally quoted as justification for the division by $n - 1$. Curiously enough, it is customary to estimate σ by $\sqrt{S^2/(n - 1)}$, even though this is a biased estimate of σ, according to the usual notion of "unbiased", referred to here as "mean-unbiased". On the other hand, $\sqrt{S^2/(n - 1)}$ is a perfectly good likelihood-unbiased estimate of σ, by virtue of the invariance under transformations. It might be pointed out, in passing, that the estimate $S^2/(n - 1)$ does not have minimum mean square about σ^2, but that the optimum divisor for minimizing the mean square error about σ^2 is $n + 1$.

The fact that a likelihood-unbiased estimate is the maximum likelihood estimate based on the distribution of the estimate itself suggest further examination of maximum likelihood estimates. If we define a *simple* estimate as one which completely determines a probability distribution for x, then we have as a theorem, the following:

A simple maximum likelihood estimate $\hat{\theta}(x)$ is likelihood-unbiased. What this means is essentially that maximum-likelihood is "hereditary", i.e. if $\hat{\theta}(x)$ maximizes $g(x \mid \theta)$ in a space of n dimensions, and $\hat{\theta}$ has a derived density $h(\hat{\theta} \mid \theta)$ in a space of $k \leq n$ dimensions, then $\theta = \hat{\theta}$ maximizes $h(\hat{\theta} \mid \theta)$. The proof follows readily from the fact that $h(\hat{\theta} \mid \theta)$ is obtained by integration of $g(x \mid \theta)$ over all x such that $\hat{\theta}(x) = \hat{\theta}$.

The example of estimating σ^2, quoted above, shows that the word "simple" cannot be omitted from the statement above. For example, the simple estimate in the parent distribution is the joint estimate $(x, S^2/n)$ of (m, σ^2) and in fact the joint estimate is likelihood-unbiased. On the other hand, S^2/n is not a simple maximum likelihood estimate, and we observe that S^2/n is not likelihood-unbiased. $S^2/(n-1)$ is a simple maximum likelihood estimate of σ^2 based on the distribution of S^2 itself, so that $S^2/(n-1)$ is, as a result, likelihood unbiased.

One can exhibit situations in which the conventional mean-unbiased property is very unnatural, while the likelihood-unbiased property may be quite natural. Consider, for example, the case where σ^2 is to be estimated by use of a χ^2-distributed S^2 with $n-1$ degrees of freedom, but subject to the condition $\sigma^2 \geq \sigma_0^2$, where σ_0^2 is known in advance. Then the estimate $\sigma^2 = \max [S^2/(n-1), \sigma_0^2]$ is certainly biased according to conventional definitions, but is nevertheless, likelihood unbiased. To get a mean-unbiased estimate when σ^2 is near to σ_0^2 is impossible except by admitting estimates less than σ_0^2, which is clearly foolish if it is known that $\sigma^2 \geq \sigma_0^2$.

It may be of interest to include a brief discussion of maximum likelihood estimation in terms of pairwise selection of alternatives, providing a sort of optimum property for maximum likelihood estimation in small samples, in addition to the likelihood-unbiased property. Consider a choice to be made between only two alternative values of θ, say θ_0 and θ_1, by dividing the sample space into two regions S_0 and S_1, such that θ_0 is accepted when x falls in S_0 and θ_1 is accepted when x falls in S_1. Then

$$P_{\theta_0}(S_0) + P_{\theta_0}(S_1) = P_{\theta_1}(S_0) + P_{\theta_1}(S_1) = 1.$$

$P_{\theta_1}(S_0)$ is the probability of making the error of accepting θ_0 when $\theta = \theta_1$ and $1 - P_{\theta_0}(S_0)$ is the probability of making the error of accepting θ_1 when $\theta = \theta_0$. If the two errors are weighted equally, it is evident that a "best" test will choose S_0 so as to minimize $P_{\theta_1}(S_0) + 1 - P_{\theta_0}(S_0)$. It is well known that S_0 will minimize the indicated quantity if S_0 consists of all points x such that $g(x \mid \theta_0) \geq g(x \mid \theta_1)$. Thus we may speak of the region S_0 defined by $g(x \mid \theta_0) \geq g(x \mid \theta_1)$ as an *optimum equal risk acceptance region* for θ_0 against θ_1. Now if we transfer our attention to the general estimation problem we see that the maximum likelihood estimate $\hat{\theta}(x)$ is that value of θ which would be accepted by the optimum equal risk acceptance procedure against all other θ's.

REFERENCES

[1] M. A. Girshick, Frederick Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Annals of Math. Stat.*, Vol. 17 (1946), p. 13.

[2] Paul R. Halmos, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), p. 34.

[3] J. Neyman and E. S. Pearson, "Unbiased critical regions of Type A and Type A_1", *Stat. Res. Mem.*, Vol. 1, p. 1.

[4] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses", *Annals of Math. Stat.*, Vol. 10 (1939), p. 299.