J. K. GHOSH
Indian Statistical Institute

It appears that Professor Berkson has revived an old debate about the MLE at least partly because of the recent interest in the results of Fisher and Rao on second order efficiency. In view of this it may be worth recording here what second order efficiency does and does not mean for Berkson's famous example from bioassay.

Adopting the notations of Ghosh and Subramanyam (1974 Section 3) (but writing E for the rather pedantic E^U) one may write

$$E(T_n) = \alpha + b(\alpha)/n + o(n^{-1})$$

$$E(\hat{\alpha}_n) = \alpha + b_o(\alpha)/n + o(n^{-1})$$

where T_n is the minimum logit chi-square estimate, $\hat{\alpha}_n$ is the MLE,

$$b(\alpha) = \Sigma\pi_i(1 - \pi_i)(2\pi_i - 1)/I^2 - \Sigma k(2\pi_i - 1)/2I$$

$$b_o(\alpha) = \Sigma\pi_i(1 - \pi_i)(2\pi_i - 1)/2I^2.$$

The corrected MLE may be taken to be a truncated version of

$$\hat{\hat{\alpha}}_n = \hat{\alpha}_n + \{b(\hat{\alpha}_n) - b_o(\hat{\alpha}_n)\}/n.$$

To truncate $\hat{\hat{\alpha}}_n$ one must choose some $d > 0$, such that the true α may be assumed to lie in $(-d, d)$ and then replace $\hat{\hat{\alpha}}_n$ by d or $-d$ according as it exceeds d or falls below $-d$. (The asymptotic theory is insensitive to the choice of d). Let the estimate T_n be truncated in a similar way. *Then the mean squared error of the truncated $\hat{\hat{\alpha}}_n$ is strictly smaller than that of the truncated T_n if terms of $o(n^{-2})$ are neglected.* This result remains true for quite general loss functions, see Ghosh, Sinha and Wieand (1977). As explained in Ghosh and Subramanyam (1974, Section 4) the reason for this is that the MLE approximates Bayes estimates better than its common rivals. When Subramanyam and I started studying second order properties of the MLE we were looking for a Bayes estimate which would be better than the MLE. It came as a surprise to us that this is impossible (up to $o(n^{-2})$ in the mean squared error).

A general result of this sort should make one prefer the MLE to the other BAN estimates commonly used as alternatives provided two rather strong conditions hold. First, the assumption about the form of the likelihood function is correct. Secondly, the terms of $o(n^{-2})$ are negligible for actual samples. It seems to me the first assumption is the more serious one and consequently the main criticism of the MLE should be based on its lack of robustness.

If Berkson's object is to provoke us into a critical reappraisal of the MLE, then I am in complete agreement with him. However, if he means all that he says in his provocative title then we must part ways at some point. Minimum chi-square estimates may be all right in certain forms of data analysis when very little is known about the data. In all other cases the likelihood is too useful a part of the data to be ignored. We should be looking for an estimate which makes use of the likelihood but in a more robust way than the MLE.

I will end by making a few comments on Professor Berkson's examples.

Consider first the bioassay example treated above. The following three statements are easily verified. If all the subjects are killed, i.e., $\Sigma p_i^n = k$, the MLE is ∞. If all the subjects survive, i.e., $\Sigma p_i^n = 0$, the MLE is ∞. In all other cases, i.e., if $0 < \Sigma p_i^n < k$, the MLE is the unique solution of the likelihood equation. Thus, the MLE is a one-one function of the minimal sufficient statistic Σp_i^n, contrary to a claim of Berkson. On the other hand, the definition of the minimum logit chi-square estimate becomes ambiguous if $p_i^n = 0$ or 1 for any i.

In his first example Professor Berkson wants an estimate to be equal to the estimated parameter if all the observations are equal to their expectations. I shall call it Berkson consistency (with respect to the Y's) to distinguish it from Fisher consistency which requires equality of estimate and parameter when the sample distribution function coincides with the true distribution function. (Thus Fisher consistency is Berkson consistency with respect to the sample distribution function.) I will now give an example where no estimate which is admissible with respect to the squared error loss can be Berkson consistent with respect to Y.

Let Y be a single observation from $N(\theta, 1)$ and assume $a < \theta < b$ where $a < b$ are known constants. To be Berkson consistent an estimate $T(Y)$ must equal Y if $a < Y < b$. A standard argument involving analyticity of Bayes estimates then shows T cannot be proper Bayes and hence T is inadmissible. Surely in this example Berkson consistency (with respect to Y) should be repugnant to Bayesians as well an Neyman-Pearsonians.

Here is another example which is instructive in a different way. Consider a single observation Y from $N(\mu, \sigma^2)$ and assume that $\mu = \sigma^2$. In this case $\hat{\mu}$ is not Berkson consistent with respect to Y but it is Berkson consistent with respect to Y^2 which is minimal sufficient. In the example given by Berkson something of this sort happens. If at each dose ΣY as well as ΣY^2 equals its expected value, then the MLE would recover the true values. Of course it is impossible to get such data if one has only one observation for each dose. However, even for Berkson's example the match between y^2 and $\hat{Y}^2$ is no worse than that between y^2 and $\tilde{Y}^2$.

REFERENCES

GHOSH, J. K. and SUBRAMANYAM, K. (1974). Second order efficiency of maximum likelihood estimators. *Sankhyā Ser. A* 36 325—358.

GHOSH, J. K., SINHA, B. K. and WIEAND, H. S. (1977). Second order efficiency of the MLE for any bounded bowl shaped loss function. Unpublished manuscript.