*The Annals of Statistics*
1983, Vol. 11, No 2, 372–374

CARL N. MORRIS[1]

*University of Texas at Austin*

We are fortunate that three powerful mathematical statisticians have cooperated here to summarize the current progress of decision theoretic multiparameter estimation for non-normal problems. Paralleling results of Stein and others for the normal distribution, these authors have established in certain discrete settings that the usual estimators can be dominated uniformly in multiparameter settings for weighted sums of squared error loss functions, if the loss function is known.

While these results are a triumph within statistical decision theory, they will affect applied statistics little. Even for the normal distribution, despite Stein's celebrated estimator for the "equal variances case," classical decision theory still rules out good "unequal variances" estimators: in the unequal variances case minimax shrinking coefficients increase with decreasing variance, violating the principle of less shrinking with more information. Thus minimax theory gives the wrong answer for the most prevalent applications. Nor will the rules derived here for the Poisson distribution satisfy applied statisticians. The authors cannot be blamed for this—they have devised ingenious estimators in order to dominate $\delta^0 = X$. Rather, the fault lies in requiring uniform frequentist dominance with respect to the weighted sum of coordinate losses, and that the weights used to define these losses are rarely known in practice. Simpler and more applicable multiparameter estimation shrinking methods are available for these distributional settings, but they emanate from Bayesian or empirical Bayesian viewpoints. The following discussion amplifies these points.

**1. The unequal sample size case.** In many Poisson applications we have $n_i$ independent Poisson observations for estimating the Poisson mean $\lambda_i$, $i = 1, 2, \cdots, p$. This happens, for example, if $X_i$ is the total number of failures of component type $i$ in $n_i$ time periods with failure rate $\lambda_i$, so

$$X_i \overset{\text{ind}}{\sim} \text{Poisson}(n_i\lambda_i), \quad i = 1, 2, \cdots, p \tag{1}$$

and there are $p$ different types of components. In such cases one wishes to estimate $\lambda_i$, and not $\theta_i = n_i\lambda_i$ of the paper. Then $\bar{X}_i \equiv X_i/n_i$ is the unbiased estimate, with variance $\lambda_i/n_i$. The loss function (1.2) of the paper then becomes

$$L_c = \textstyle\sum_1^p c_i(\hat{\lambda}_i - \lambda_i)^2/\lambda_i^{m_i} \tag{2}$$

with $c_i = n_i^{(2-m_i)}$. This choice of $c_i$ has no special appeal, and other $c_i$ also should be considered. In the equal sample size case, however, the losses on $\theta_i$ in (1.2) of the paper and $\lambda_i$ above are equivalent.

Not only do transformations of parameters affect loss functions, but they also affect prior distributions. For example, Table 2 and Table 3 assume exchangeable prior distributions on the $\theta_i$, i.e. $a \le \theta_i \le b$ for various $a$ and $b$. But then the $\lambda_i$ are not exchangeable, because $a/n_i \le \lambda_i \le b/n_i$. In practice, the $\lambda_i$ are more likely to be exchangeable than the $\theta_i$, and in such cases the theory provided does not properly combine sample and apriori information.

Section 3 also covers negative binomial distribution, to which the preceding remarks apply. It is hard to see how the $m_i$ should be chosen for the component losses $(\hat{p}_i - p_i)^2/p_i^{m_i}$ to be meaningful.

**2. Dependence of dominating rules on the loss function.** For each loss function (1.2) a *different* estimation rule is produced that is superior to $\delta^0 = X$. Note that $\delta^0$ emerges

Received December 1982.

[1] Research supported by the National Science Foundation under grant number MCS-8104-250 and the University of Texas, Austin, Texas 78712.