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 λ_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 λ_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 λ_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 λ_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 θ_i in (1.2) of the paper and λ_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 θ_i, i.e. $a \le \theta_i \le b$ for various a and b. But then the λ_i are not exchangeable, because $a/n_i \le \lambda_i \le b/n_i$. In practice, the λ_i are more likely to be exchangeable than the θ_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 δ^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.

as the most broadly applicable estimator of the paper because it alone is considered with each loss function. In practice, statisticians rarely know how to choose among loss functions, and yet *any improvements on* δ^0 *always depend on the loss function.* For example, estimators that dominate δ^0 for L_0 differ sharply from those that dominate for L_1. Do the authors know of a Poisson estimator that dominates δ^0 for both L_0 and L_1? Any possible improvement would be modest because X_i is admissible for each component.

3. Desired extensions. Most estimators suggested by the decision theory approach for multiparameter problems are too complicated for applications. Stein's estimator is reasonably simple, but the rules provided here do not emulate it for large equal λ_i, when they should. Only δ^M shifts δ^0 toward a reasonable center, the geometric mean, but δ^M may not dominate δ^0. For applications, statisticians need simple rules that shift toward a good center (near the mean of the data), that account for unequal sample sizes, and that provide measures of accuracy and interval estimates. While these objectives can be accomplished, the summed loss function approach hinders rather than aids such progress.

4. Bayes and empirical Bayes alternatives. The difficulties just cited occur with the normal distribution too, where empirical Bayes provides a remedy, c.f. Morris (1983b) for a recent summary. To see how empirical Bayes theory applies to the Poisson case, let X_i have the distribution (1), given λ_i, so $\bar{X}_i = X_i/n$ has mean λ_i and variance λ_i/n_i. Also assume that $\{\lambda_i\}$ are independently distributed with common mean $\mu = E(\lambda_i)$ and variance $A = \text{Var}(\lambda_i)$, $i = 1, \cdots, p$. For known μ, A, the best squared error linear Bayes estimator of λ_i is

$$\lambda_i^* = (1 - B_i)\bar{X} + B_i\mu \tag{3}$$

with shrinking coefficient $B_i \equiv (\mu/n_i)/(A + \mu/n_i)$. Note that (3) is the posterior mean $E\lambda_i \mid \bar{X}_i$ if and only if λ_i has the conjugate (gamma) prior distribution. We see that (3) shrinks $\bar{X}_i$ toward the common marginal mean μ and that the amount of shrinkage B_i decreases as n_i increases.

Empirical Bayes theory assumes μ and/or A are unknown, and uses information in the marginal distribution of $\bar{X}_i$ to estimate them, the marginal mean and variance of the independent $\bar{X}_i$ being μ and $A + \mu/n_i$. In the equal sample size case ($n_i \equiv n$) with $p \geq 4$, for example, we can estimate the common value B of the B_i by

$$\hat{B} = \frac{p-3}{p-1}\,\frac{\hat{\mu}/n}{\hat{A} + \hat{\mu}/n} \tag{4}$$

with $\hat{\mu} = \sum \bar{X}_i/p$ and $\hat{A} = \{S/(p-1) - \hat{\mu}/n\}^+$, $S \equiv \sum (\bar{X}_i - \hat{\mu})^2$. Then the empirical Bayes estimator

$$\hat{\lambda}_i \equiv (1 - \hat{B})\bar{X}_i + \hat{B}\hat{\mu} \tag{5}$$

mimics (3). Related empirical Bayes estimators for other natural exponential families appear in Morris (1983a).

The estimator (5), considered as a vector, should outperform the other estimators of Table 2 under L_0 for most exchangeable prior distributions $\mu \geq 0$, $A \geq 0$. The B values for Tables 2 and 3 can be determined from μ and A for the six uniform distributions as $B =$ .60, .82, .88, .91, .33, .45 respectively. These B values are the fractional reductions in risk relative to δ^0 achieved by the linear Bayes rules (3), being directly comparable to the values in Table 2. Accounting for unknown μ, A, the empirical Bayes estimator for $p = 10$ would yield about 70 percent of this or 42, 57, 62, 73, 23, 32 percent respectively. Thus, (5) provides about four times as much improvement over δ^0 as does δ^2, the best minimax estimator of Table 2.

Any improvement on δ^0 must be justified by additional assumptions, e.g. by empirical Bayes or loss function assumptions. Applied statisticians probably can recognize empirical Bayes situations with exchangeable prior distributions more readily than they can specify appropriate loss functions. Theoretical statisticians also need release from the overly stringent demands of the frequentist decision theoretic criteria adopted for multiparameter estimation by Professors Ghosh, Hwang, and Tsui. I hope they will add their impressive talents to the search for a more congenial framework from which to advance the theory and application of multiparameter estimation.

REFERENCES

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **38** 117–130.

MORRIS, C. (1983a). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statistics* **11** June 515–529.

MORRIS, C. (1983b). Parametric empirical Bayes inference: theory and applications (with discussion). *J. Amer. Statist. Assoc.* **48** 47–65.

CARL MORRIS
DEPARTMENT OF MATHEMATICS
INSTITUTE OF STATISTICS
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712