The Annals of Statistics
1982, Vol. 10, No. 1, 93–100

SIMULTANEOUS ESTIMATION OF SEVERAL POISSON PARAMETERS UNDER *K*-NORMALIZED SQUARED ERROR LOSS

BY KAM-WAH TSUI AND S. JAMES PRESS

University of British Columbia and University of California, Riverside

In this study, we consider the simultaneous estimation of the parameters of the distributions of p independent Poisson random variables using the loss function $L_k(\lambda, \hat{\lambda}) = \sum (\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k$ for a given positive integer k. New estimators are derived, which include the minimax estimators proposed by Clevenson and Zidek (1975), as special cases. The case when more than one observation is taken from some of the variables is considered.

1. Introduction. Let $X_1, \cdots, X_p$ be p independent Poisson random variables, where X_i has mean λ_i, $i = 1, \cdots, p$. Recently, considerable research has been devoted to the problem of finding better estimators of the λ_i than the Maximum Likelihood Estimator (MLE). Clevenson and Zidek (1975) obtain a class of minimax estimators under normalized squared error loss $L_1(\lambda, \hat{\lambda}) = \sum_{i=1}^p (\lambda_i - \hat{\lambda}_i)^2/\lambda_i$ when $p \geq 2$. Their estimators shrink the MLE towards the origin. A considerable amount of savings in risk as compared to the MLE is expected when the parameters λ_i are relatively small, especially when the λ_i's are close to zero.

In this paper, a more general loss function is considered, namely, k-normalized squared error loss $L_k(\lambda, \hat{\lambda}) = \sum_{i=1}^p (\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k$, where k is a positive integer. New estimators are derived under this loss function. A rationale for these loss functions is given in Section 4.

DEFINITIONS. In order to simplify our expressions, we need the following definitions.
(1) $y^{(k)} = y(y-1) \cdots (y-k+1)$, where k is a positive integer and y is a real number;
(2) e_i = the p-vector whose ith coordinate is one, and whose other coordinates are zero;
(3) $X = (X_1, \cdots, X_p)$, $x = (x_1, \cdots, x_p)$ an observation of X, and $\lambda = (\lambda_1, \cdots, \lambda_p)$;
(4) $Z = \sum_{i=1}^p X_i$, $z = \sum_{i=1}^p x_i$, $S = \sum_{i=1}^p (X_i + k)^{(k)}$, $s = \sum_{i=1}^p (x_i + k)^{(k)}$, $S_i = S - (X_i + k)^{(k)}$, $s_i = s - (x_i + k)^{(k)}$.

Hudson (1974) defines the risk improvement of an estimator $\hat{\lambda}$ over the MLE, $I = R(\lambda, X) - R(\lambda, \hat{\lambda})$, and derives an identity for the unbiased risk improvement estimate U under squared error loss: $I = E_\lambda U(X)$, where U is a function of X only. By means of this identity, Peng (1975) shows that the MLE is inadmissible under squared error loss when $p \geq 3$. In Section 2, we derive a similar identity for the case of loss function L_k. In Section 3, we shall use the identity in the special case when $k = 1$ to show that the estimator $\hat{\lambda} = X - \phi(Z)X/(Z + a)$ dominates the MLE when $p \geq 2$, where $\phi(z)$ is nondecreasing in z, $0 \leq \phi(z) \leq \min\{2(p-1), 2a\}$, with a an arbitrary positive real number. This result includes that of Clevenson and Zidek (1975). Their estimators require $a \geq p - 1$ and $0 \leq \phi(z) \leq 2(p-1)$.

In Section 4, we use the identity derived in Section 2 to show that under loss L_k, the estimator $\hat{\lambda}^{(k)} = (\hat{\lambda}_1^{(k)}, \cdots, \hat{\lambda}_p^{(k)})$, given by

$$\hat{\lambda}_i^{(k)} = X_i - \frac{\phi(Z)(X_i)^{(k)}}{S_i + (X_i)^{(k)}}, \qquad i = 1, \cdots, p,$$

Received June 1978; revised August 1980.
AMS 1970 *subject classifications.* Primary 62C15; secondary 62F10, 62C25.
Key words and phrases. Minimax estimator, Poisson distributions, simultaneous estimation, unbiased risk deterioration estimate.

dominates the MLE when $p \geq 2$. We require that $\phi(z)$ be increasing in z and $0 \leq \phi(z) \leq 2k(p-1)$. Some modified versions of $\hat{\lambda}^{(k)}$ are also proposed.

In order that the performance of some of our estimators might be compared, we report the results of a computer simulation in Section 5. Finally, in Section 6, we extend our results to the case where more than one observation per Poisson random variable is available.

2. Unbiased risk deterioration estimate. Let J be the set of all integers. The following lemma provides an identity which proves to be useful.

LEMMA 1. *Suppose $Y \sim$ Poisson (μ) and $g: J \to R$ is a real-valued function such that $E_\mu|g(Y)| < \infty$ and $g(j) = 0$ if $j \leq 0$. Then $E_\mu\{g(Y)/\mu\} = E_\mu\{g(Y+1)/(Y+1)\}$*

PROOF.

$$E_\mu\{g(Y)/\mu\} = \sum_{y=0}^\infty \{g(y)/\mu\}e^{-\mu}\mu^y/y! = g(0)e^{-\mu}/\mu + \sum_{y=1}^\infty g(y)e^{-\mu}\mu^{y-1}/y!$$
$$= 0 + \sum_{y=0}^\infty g(y+1)e^{-\mu}\mu^y/(y+1)! = E_\mu\{g(Y+1)/(Y+1)\}. \qquad \square$$

The next lemma is an immediate consequence of Lemma 1.

LEMMA 2. *Suppose $Y \sim$ Poisson (μ), k is a positive integer and $g: J \to R$ is a real-valued function such that*

$$E_\mu|g(Y+j)| < \infty, \quad j = 0, \cdots, k-1, \quad \text{and} \quad g(j) = 0 \quad \text{if } j < k.$$

Then

$$E_\mu\{g(Y)/\mu^k\} = E_\mu\{g(Y+k)/(Y+k)^{(k)}\}.$$

PROOF. Induction on k and application of Lemma 1.

Lemma 3 below is a generalization of Lemma 2 to the vector case.

LEMMA 3. *Suppose $X_i \sim$ Poisson (λ_i), $i = 1, \cdots, p$, $p \geq 2$ and the X_i's are mutually independent. Let k be a positive integer. Suppose $f_i: J^p \to R$, $i = 1, \cdots, p$, are functions defined on the p-fold Cartesian product of J, such that $E_\lambda|f_i(X + je_i)| < \infty, j = 1, \cdots, k-1$, and $f_i(x) = 0$, if $x_i < k$. Then*

$$E_\lambda\{f_i(X)/\lambda_i^k\} = E_\lambda\{f_i(X + ke_i)/(X_i + k)^{(k)}\}.$$

PROOF. Condition on $\{X_j : j \neq i\}$ and apply Lemma 2.

Let $\hat{\lambda}(X) = X + f(X)$ be an estimator of λ, where $f(X) = (f_1(X), f_2(X), \cdots, f_p(X))$ and the f_i's satisfy the conditions in Lemma 3. The next lemma gives an unbiased estimate of D_k, the deterioration in risk of $\hat{\lambda}$ as compared to the MLE, X. The proof is straightforward.

LEMMA 4. *Under loss L_k, the deterioration in risk of $\hat{\lambda}$ is*

$$D_k = E_\lambda\{L_k(\lambda, \hat{\lambda}) - L_k(\lambda, X)\} \equiv E_\lambda \Delta_k = R(\lambda, \hat{\lambda}) - R(\lambda, X),$$

where

$$\Delta_k = \sum_{i=1}^p f_i^2(X + ke_i)/(X_i + k)^{(k)} + 2\sum_{i=1}^p (X_i + k)\{f_i(X + ke_i) - f_i(X + (k-1)e_i)\}/(X_i + k)^{(k)}.$$

PROOF. Apply Lemma 3.

From Lemma 4, we see that in order to show that an estimator $\hat{\lambda} = X + f(X)$ dominates X, it is sufficient to show that $\Delta_k(x) \le 0$ for all $x \in J^p$. Application of this technique yields Theorems 1 through 4.

3. Minimax estimators. The usual estimator, X, of λ is minimax under normalized squared error loss L_1. Hence, to show that an estimator $\hat{\lambda}$ of λ is minimax, it suffices to show that the risk of $\hat{\lambda}$ under L_1 is uniformly less than or equal to that of X, i.e. $R(\lambda, \hat{\lambda}) \le R(\lambda, X)$, for all λ. Lemma 4 of the previous section gives us a sufficient condition for such an inequality to hold if the estimator $\hat{\lambda}$ is of the form $X + f(X)$. The condition is

$$\Delta_1 = \textstyle\sum_{i=1}^p f_i^2(x + e_i)/(x_i + 1) + 2\sum_{i=1}^p \{f_i(x + e_i) - f_i(x)\} \le 0 \tag{3.1}$$

for all $x \in J^p$. Using this fact, we proceed to derive a class of minimax estimators of λ which contains the estimators obtained by Clevenson and Zidek (1975, Theorem 2.1).

THEOREM 1. *Suppose $X_i \sim$ Poisson (λ_i), $i = 1, \cdots, p$, $p \ge 2$, and that the X_i's are mutually independent. Then estimators $\hat{\lambda}$ of the following form dominate the* MLE, X, *under L_1:*

$$\hat{\lambda} = X - \phi(Z)X/(Z + a),$$

where $Z = \Sigma x_i$, $a > 0$, $\phi(z)$ is a real-valued function nondecreasing in z such that $0 \le \phi(z) \le \min\{2(p-1), 2a\}$ *and* $\phi(z) \not\equiv 0$.

PROOF. Define

$$f_i(x) = -\phi(z)x_i/(z + a), \quad \text{if } x_i \ge 0$$
$$= 0, \quad \text{if } x_i < 0,$$

$i = 1, \cdots, p$. We see that the f_i's satisfy Lemma 3 of Section 2. From (3.1) we have, using the stated properties of $\phi(\cdot)$,

$$\begin{aligned} \Delta_1 &= \phi^2(z+1)\cdot(z+p)/(z+a+1)^2 \\ &\quad + 2\{-\phi(z+1)(z+p)/(z+a+1) + \phi(z)\cdot z/(z+a)\} \\ &\le \frac{\phi(z+1)}{z+a+1}[\phi(z+1)\cdot(z+p)/(z+a+1) - 2\{(p-1)z + pa\}/(z+a)] \\ &\le \frac{\phi(z+1)}{(z+a+1)^2}[z\{\phi(z+1) - 2(p-1)\} + p\{\phi(z+1) - 2a\}] \le 0. \end{aligned} \tag{3.2}$$

□

Note that the constant a given in the theorem is an *arbitrary* positive real number, while the class of estimators given in Theorem 2.1 of Clevenson and Zidek (1975) requires $a = (p - 1)$ and $0 \le \phi(z) \le 2(p - 1)$. Hence their class of estimators is a subclass of ours.

The estimator $\hat{\lambda}$ shrinks the MLE toward the origin by the amount $\phi(z)x/(z + a)$. For every a, the maximum shrinkage allowable if $\hat{\lambda}$ is to dominate the MLE is $\min\{2(p - 1), 2a\}x/(z + a)$, which in an increasing function of a (coordinatewise) whenever $0 < a \le p - 1$ and a decreasing function whenever $a > p - 1$. Therefore the maximum shrinkage is obtained when $a = p - 1$.

Observe that while $\hat{\lambda}$ gives nonnegative estimates if $\phi(z) \le z + a$ for all z, it may produce negative estimates if the inequality does not hold for some z. In that case, a "plus-rule" version, $\hat{\lambda}^+$, which guarantees nonnegative estimates, should do better. Such a rule is obtained by replacing $\phi(z)$ with $\phi^*(z) = \min\{\phi(z), z + a\}$. That $\hat{\lambda}^+$ always improves upon $\hat{\lambda}$ is immediate from (3.2).

An application of Theorem 1 above gives us some interesting estimators of λ. The result is stated in Corollary 1 below.

COROLLARY 1. *Suppose the X_i's are as given in Theorem* 1. *Then the estimator* $\hat{\lambda} =$

$\{1 - a/(Z + c)\}^t X$ *of* λ *dominates* X *under the loss function* L_1, *provided that* $t \geq 1$, $c > 0$, *and* $0 < a \leq \min\{2(p - 1)/t, c, 2c/t\}$.

PROOF. Rewrite $\hat{\lambda}$ as $\hat{\lambda} = [1 - \{\theta(Z)/(Z + c)\}]X$, where $\theta(z) = \{(z + c)^t - (z + c - a)^t\}/(z + c)^{t-1}$, and check that the conditions of Theorem 1 hold. The estimator $\hat{\lambda}(X) = \{1 - (p - 1)/(Z + p - 1)\}^2 X$, which is an estimator described in the previous corollary with $t = 2$ and $a = c = p - 1$, shrinks more towards the origin than does the estimator $\hat{\lambda}^* = \{1 - (p - 1)/(Z + p - 1)\}X$. Thus, $\hat{\lambda}$ should give a better estimate of λ than $\hat{\lambda}^*$ if the parameters λ_i, $i = 1, \cdots, p$ are close enough to zero. The following argument gives us an interesting insight as to why we might arrive at estimators of the form $\hat{\lambda} = (1 - c/(Z + a))^2 X$.

Let $X_i \sim$ Poisson (λ_i), $i = 1, \cdots, p$, be mutually independent and let $Y_i = 2\sqrt{X_i}$, $\theta_i = 2\sqrt{\lambda_i}$, $i = 1, \cdots, p$. It is approximately true that $Y_i \sim N(\theta_i, 1)$, $i = 1, \cdots, p$, and that the Y_i's are mutually independent. That is, approximately, $Y = (Y_1, \cdots, Y_p) \sim N_p(\theta, I_p)$, where I_p is the $p \times p$ identity matrix and $\theta = (\theta_1, \cdots, \theta_p)$. The James-Stein estimator $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)$ of θ under squared error loss is $\hat{\theta}_i = (1 - r/Y'Y)Y_i$, $i = 1, \cdots, p$. Or, in terms of X and λ, $\sqrt{\hat{\lambda}_i} = (1 - c/Z)\sqrt{X_i}$, $i = 1, \cdots, p$, where $Z = \sum_{i=1}^p X_i$, $c = r/4$, and $\hat{\lambda} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_p)$ is an estimator of λ. We thus have the estimator $\hat{\lambda} = (1 - c/Z)^2 X$ of λ. Since Z has a positive probability of being zero, we are prompted to replace Z by $Z + a$, where a is a positive real number. We thus arrive at the estimator $\hat{\lambda} = (1 - c/(Z + a))^2 X$ of λ. A more detailed heuristic argument to explain a similar estimator is suggested by Brown (1979).

4. Better estimators under k-nsel. The squared error loss function is probably the most extensively studied loss function used in estimation problems. Nevertheless, there are situations in which other loss functions may be more appropriate. In simultaneous estimation of the means of several independent normal random variables $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \cdots, p$, a natural loss function is $\sum_{i=1}^p(\mu_i - \hat{\mu}_i)^2/\sigma_i^2$. On the other hand, if the σ_i^2 are to be estimated,

$$\textstyle\sum_{i=1}^p\{1 - (\hat{\sigma}_i^2/\sigma_i^2)\}^2 = \sum_{i=1}^p (\sigma_i^2 - \hat{\sigma}_i^2)^2/\sigma_i^4$$

is an appropriate loss function. Since the Poisson parameter λ_i is both the mean and the variance of the distribution, the corresponding loss functions would be $\sum_{i=1}^p(\lambda_i - \hat{\lambda}_i)^2/\lambda_i$ and $\sum_{i=1}^p(\lambda_i - \hat{\lambda}_i)^2/\lambda_i^2$, i.e., L_1 and L_2 in our notation.

Theorem 2 gives us a class of estimators $\hat{\lambda}^{(k)} = (\hat{\lambda}_1^{(k)}, \cdots, \hat{\lambda}_p^{(k)})$ uniformly dominating the MLE under the loss function L_k. These estimators have the following properties:

(1) If the observation from the ith population is small (less than k), then the estimator $\hat{\lambda}_i^{(k)}$ of λ_i is the same as the MLE.

(2) If the observation is large (greater than or equal to k), then the estimator $\hat{\lambda}_i^{(k)}$ of λ_i shrinks the MLE towards zero.

Using the definitions given in Section 1, the theorem and its proof are stated as follows:

THEOREM 2. *Suppose that* X_i *are independently Poisson* (λ_i), $i = 1, \cdots, p$, $p \geq 2$, *and that the loss function is* $L_k(\lambda, \hat{\lambda})$. *Then the estimator* $\hat{\lambda}^{(k)}$ *given below dominates the* MLE X *uniformly in* $\lambda = (\lambda_1, \cdots, \lambda_p)$:

$$\hat{\lambda}_i^{(k)} = i\text{th coordinate of } \hat{\lambda}^{(k)} = X_i - \frac{\phi(Z)X_i(X_i - 1) \cdots (X_i - k + 1)}{S_i + X_i(X_i - 1) \cdots (X_i - k + 1)}, \tag{4.1}$$

where $\phi(z)$ *is a real-valued function increasing in* z *such that* $0 \leq \phi(z) \leq 2k(p - 1)$ *and* $\phi(z) \not\equiv 0$.

PROOF. Application of Lemma 4, in the manner of the proof of Theorem 1, yields the bound

$$\Delta_k \leq \phi(z + k)\{\phi(z + k) - 2k(p - 1)\}/S \leq 0. \qquad \square$$

Observe that the estimator given by (4.1) does not always give nonnegative estimates. Thus, the plus-rule $\hat{\lambda}^+$ with coordinates $\hat{\lambda}_i^+ = \max\{\hat{\lambda}_i^{(k)}, 0\}$ $i = 1, \cdots, p$, is expected to dominate (4.1). In fact, the estimator $\hat{\lambda} = X + f(X)$ can always be improved upon by $\hat{\lambda}^+ = X + f^*(X)$, where $f_i^*(X) = -X_i$ or $f_i(X)$ according as $-f_i(X) > X_i$ or not. This can be seen by examining the unbiased estimate Δ_k of the risk given in Lemma 4.

The estimator given by (4.1) can also be improved by using a modified $\phi(z)$, defined to be $z + (p - 1)k$ if $\phi(z) > z + (p - 1)k$, and $\phi(z)$ otherwise. However, this modified estimator is not necessarily a plus-rule type of improvement (i.e., it can give negative estimates). Similar remarks apply to Theorem 3 stated below and Theorem 4 in Section 6.

REMARKS.

(1) When $k = 1$, Theorem 2 is the same as Theorem 2.1 of Clevenson and Zidek (1975).

(2) If $x_i \le k - 1$, then $\hat{\lambda}_i^{(k)}$ is equal to the MLE, i.e., no shrinkage takes place. We see from this that as $k \to \infty$, there is decreasing likelihood that shrinkage will be indicated.

(3) Theorem 3.1 of Clevenson and Zidek (1975) suggests that estimators $\hat{\lambda} = \{1 - \phi(Z)/(Z + p - 1)\}X$ of λ still dominate the MLE under a general loss function $L_K^*(\lambda, \hat{\lambda}) = \sum_{i=1}^p K(\lambda_i)(\lambda_i - \hat{\lambda}_i)^2/\lambda_i$ where $K > 0$ is some non-increasing function. When $K(y) = 1/y^{k-1}$, L_K^* is the loss L_k. However, our estimators do not shrink observations that are less than k; only those observations greater than or equal to k are moved. Therefore, if $\lambda_i \ge k$, our estimators guard against unnecessary shrinkage when the observation happens to be small (i.e. less than k). Since the Clevenson-Zidek estimator shrinks all non-zero observations, we are led to conjecture that our estimators are better than theirs in terms of the percentage in savings compared to the MLE when the λ_i's are relatively large (i.e. when $\min \lambda_i \ge k \ge 2$). Some simulation results which support this conjecture are reported in Section 7.

The next theorem is a generalization of Theorem 1 to the case when loss L_k is used, where k is any positive integer.

THEOREM 3. *Suppose $X = (X_i, \cdots, X_p)$ is as given in Theorem 2 and $\hat{\lambda}^{(k)} = (\hat{\lambda}_1^{(k)}, \cdots, \hat{\lambda}_p^{(k)})$ is an estimator of λ. Let*

$$\hat{\lambda}_i^{(k)} = X_i - \frac{\phi(Z)X_i^{(k)}}{S_i^i + X_i^{(k)} + b}, \quad i = 1, \cdots, p, \tag{4.2}$$

where k is a positive integer, ϕ is nondecreasing and is not identically zero, $0 \le \phi(z) \le \min\left\{2\frac{b + (p - 1)k!}{(k - 1)!}, 2k(p - 1)\right\}$ and $b > -(p - 1)(k!)$. *Then for all λ, $\hat{\lambda}^{(k)}$ dominates X under the loss function L_k.*

PROOF. Again, by Lemma 4,

$$\Delta k \le pk![\phi(z + k) - 2\{b + (p - 1)k!\}/(k - 1)!]\phi(z + k)/(S + b)^2 \le 0. \qquad \square$$

5. Computer simulation. The results of the computer simulation reported in this section are mainly comparisons of the estimators $\hat{\lambda}^{(2)}$ and $\hat{\lambda}^{(1)}$ with the MLE, where $\hat{\lambda}^{(2)}$ is as described in Theorem 2 with $\phi(z) \equiv 2(p - 1)$, and $\hat{\lambda}^{(1)} = \{1 - p/(Z + p)\}X$. The estimator $\hat{\lambda}^{(2)}$ is of considerable appeal because this is the case where a natural loss function $L_2(\lambda, \hat{\lambda}) = \sum_{i=1}^p (1 - \hat{\lambda}_i/\lambda_i)^2$ is used in estimating scale parameters.

The computations reported here were performed both on the IBM 370/168 computer at the University of British Columbia and the Data General NOVA 840 computer at the University of California, Riverside. A FORTRAN program was used in the IBM computer and a BASIC program was used in the NOVA computer. First, the number p of independent Poisson random variables is chosen. Second, p parameters λ_i are generated randomly within a certain range (c, d). Third, one observation of each of the p distributions with the parameters obtained in the second step is generated. Estimates of the parameters are then calculated according to the estimator $\hat{\lambda}$ that we want to test. The third step is repeated 2000 times and the risks under the relevant loss functions for both the estimator and the MLE are calculated. The percentage of the savings in using $\hat{\lambda}$ as compared to the MLE, $\frac{R(\lambda, X) - R(\lambda, \hat{\lambda})}{R(\lambda, X)} \cdot 100\%$, is calculated. The whole process is then repeated a number of times and the average percentage of the savings is calculated.

We chose the range of the parameters λ_i in such a way that we might check the performance of the estimators both when the parameters λ_i fall in a narrow range and when they fall in a wide range. In calculating the percentage of improvement of the various estimators over the MLE, the appropriate loss functions must be used. The loss functions L_1 and L_2 are used for Tables 1 and 2, respectively.

In most of the cases, the improvement percentage is seen to be an increasing function of p, the number of independent Poisson variates. We see that in general, the improvement percentage decreases as the magnitude of the λ_i's increases.

In Table 2, we see that for the ranges considered, the percentage of improvement in risk of $\hat{\lambda}^{(2)}$ over the MLE is considerable when the parameters fall into a narrow interval. For each value of p, $\hat{\lambda}^{(2)}$ performs best when the parameters are in the intervals (0, 4) and (4, 8). The improvement decreases gradually as the magnitude of the λ_i's increases. In contrast, the estimator $\hat{\lambda}^{(1)}$ performs very well only when the parameters are relatively small, with the improvement percentage decreasing dramatically as the magnitude of the λ_i's increases (Table 1). This is as conjectured in Section 4. Although the improvement percentages of $\hat{\lambda}^{(2)}$ over the MLE for the wider ranges of the λ_i's are by no means substantial, they are nevertheless greater than those of $\hat{\lambda}^{(1)}$. Of course, the different loss functions employed for $\hat{\lambda}^{(2)}$ and $\hat{\lambda}^{(1)}$ might contribute to such a difference.

Simulations were performed to determine how much the corresponding plus-rules improve on $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ for the ranges in Tables 1 and 2. The results show that the plus-rules produce the same improvement over the MLE as do $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ except when the parameters are in the ranges (0, 4) and (0, 20), in which cases minimal improvements over $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ are observed.

TABLE 1

Improvement Percentage of $\hat{\lambda}^{(1)}$ *over the* MLE *using loss* L_1

Range of the Parameters λ_i	Percentage of Improvement over the MLE					
	p = 2	p = 3	p = 4	p = 5	p = 8	p = 10
(0, 4)	24	25	27	28	29	30
(4, 8)	5	7	8	10	11	12
(8, 12)	1	4	6	6	7	8
(12, 16)	0+	2	3	4	5	5
(0, 20)	2	4	5	8	7	8
(10, 30)	0+	2	2	4	4	4

TABLE 2

Improvement Percentage of $\hat{\lambda}^{(2)}$ *over the* MLE *using loss* L_2

Range of the Parameters λ_i	Percentage of Improvement over the MLE					
	p = 2	p = 3	p = 4	p = 5	p = 8	p = 10
(0, 4)	23	25	24	34	34	35
(4, 8)	17	23	26	29	31	33
(8, 12)	10	15	20	21	22	24
(12, 16)	7	11	13	15	18	18
(0, 20)	7	8	11	11	11	13
(10, 30)	4	8	9	12	12	12

6. Multiple observations. The estimators derived in the previous sections are based on the situation in which only one observation is taken from each of p independent Poisson populations. Now we consider the case of multiple observation sampling from at least one population.

Suppose $X_{i1}, \cdots, X_{in_i}$ are independent Poisson (λ_i), where $n_i \geq 1$, $i = 1, \cdots, p$. Letting $X_i = \sum_{j=1}^{n_i} X_{ij}$, $i = 1, \cdots, p$, the MLE of $\lambda = (\lambda_1, \cdots, \lambda_p)$ is $(X_1/n_1, \cdots, X_p/n_p)$. In this section, we shall show that there are estimators of λ dominating the MLE under loss L_k. Since the distribution of X_i/n_i is not Poisson if $n_i > 1$, we cannot apply our previous results directly. However, if our interest is to estimate $n_i\lambda_i$, $i = 1, \cdots, p$, then the foregoing theory can be applied because $X_i \sim$ Poisson $(n_i\lambda_i)$ in this case.

The risk function of the MLE under L_k can be rewritten as

$$E_\lambda \sum_{i=1}^p (n_i\lambda_i - X_i)^2 \cdot \frac{n_i^{k-2}}{(n_i\lambda_i)^k}.$$

Since X_i has a Poisson distribution with parameter $n_i\lambda_i$, it is then natural to consider the following problem:

Suppose $X_1, \cdots, X_p$ are independent Poisson random variables with parameters $\lambda_1, \cdots, \lambda_p$, respectively. Suppose one observation is taken from each random variable. We would like to know if there are estimators of λ better than the MLE under the loss function

$$L_k^c(\lambda, \hat{\lambda}) = \sum_{i=1}^p c_i(\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k, \qquad \text{with } c_i > 0.$$

We shall show below that such estimators do exist. As a result, we obtain estimators better than the MLE in the situation where more than one observation is available from some of the Poisson variables.

The following lemma is similar to Lemma 4.

LEMMA 5. *Let X_i be independent Poisson (λ_i), $i = 1, \cdots, p$ and let $f_i: J^p \to R$, $i = 1, \cdots, p$ satisfy the conditions given in Lemma 3. Define $\hat{\lambda} = X + f(X)$. Then, under the loss function $L_k^c(\lambda, \hat{\lambda}) = \sum c_i(\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k$, the deterioration in risk of $\hat{\lambda}$ as compared to X is*

$$R(\lambda, \hat{\lambda}) - R(\lambda, X) = E_\lambda \Delta_k^c,$$

where

$$\Delta_k^c = \sum_{i=1}^p c_i \frac{f_i^2(X + ke_i)}{(X_i + k)^{(k)}} + 2\sum_{i=1}^p c_i(X_i + k)\frac{f_i(X + ke_i) - f_i(X + (k-1)e_i)}{(X_i + k)^{(k)}}.$$

The next theorem supplies estimators that dominate the MLE under the loss function L_k^c.

THEOREM 4. Let X_i be independent Poisson (λ_i), $i = 1, \cdots, p$ and let the loss function be L_k^c. Define

$$f_i(x) = -k(p-1)(c_*/c_i)^{1/2} x_i^{(k)}/(s_i + x_i^{(k)}), \qquad \text{if } x_i \geq 0,$$
$$= 0, \qquad \text{if } x_i < 0,$$

$i = 1, \cdots, p$, where $c_* = \min(c_1, \cdots, c_p)$, and $s_i = \sum_{j \neq 1}^p (x_j + k)^{(k)}$. Let $f(X) = (f_1(X), \cdots, f_p(X))$ and $\hat{\lambda} = X + f(X)$. Then the estimator $\hat{\lambda}$ of λ dominates the MLE X uniformly in λ under the loss function L_k^c.

PROOF. It can be shown that Δ_k^c given in Lemma 5 above satisfies

$$\Delta_k^c \leq -\frac{c_* k^2(p-1)^2}{\sum_{i=1}^p (x_i + k)^{(k)}} \leq 0. \qquad \square$$

The following corollaries provide estimators better than the MLE when there is more than one observation available from some of the Poisson random variables.

COROLLARY 2. *Let X_{ij} be independent Poisson* (λ_i), $i = 1, \cdots, p$, $j = 1, \cdots, n_i$. *Let* $X_i = \sum_{j=1}^{n_i} X_{ij}$, $i = 1, \cdots, p$. *Define* $\hat{\lambda} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_p)$ *by*

$$\hat{\lambda}_i = \frac{X_i}{n_i} - \frac{k(p-1)}{n_i}(n_i/n^*)^{\frac{1}{2}k-1} \cdot X_i^{(k)}/(S_i + X_i^{(k)}),$$

$i = 1, \cdots, p$, where $n^* = \max(n_i, \cdots, n_p)$. Then $\hat{\lambda}$ dominates the MLE $(X_1/n_1, \cdots, X_p/n_p)$ under loss function $L_k(\lambda, \hat{\lambda}) = \sum_{i=1}^{p}(\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k$ with $k = 1$ or 2.

PROOF. Use Theorem 4 with $c_i = n_i^{k-2}$ and note that $c_* = (n^*)^{k-2}$ if $k = 1$ or 2.

COROLLARY 3. *Let X_{ij} be independent Poisson* (λ_i), $i = 1, \cdots, p$, $j = 1, \cdots, n_i$. *Let* $X_i = \sum_{j=1}^{n_i} X_{ij}$, $i = 1, \cdots, p$. *Define* $\hat{\lambda} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_p)$ *by*

$$\hat{\lambda}_i = \frac{X_i}{n_i} - \frac{k(p-1)}{n_i}(n_*/n_i)^{\frac{1}{2}k-1} \cdot X_i^{(k)}/(S_i + X_i^{(k)}),$$

$i = 1, \cdots, p$, *where* $n_* = \min(n_i, \cdots, n_p)$. *Then* $\hat{\lambda}$ *dominates the* MLE $(X_1/n_1, \cdots, X_p/n_p)$ *under loss* L_k *with* $k \geq 3$.

PROOF. Observe that $c_i = (n_i)^{k-2}$ and $c_* = (n_*)^{k-2}$ if $k \geq 3$.

There are, of course, other estimators dominating the MLE under the loss function $L(\lambda, \hat{\lambda}) = \Sigma\, c_i(\lambda_i - \hat{\lambda}_i)^2/\lambda_i^k$. The results will be similar to those derived in Section 4, and we shall therefore not set down the details here. As a final remark, we note that in the squared loss case, there are estimators that dominate the MLE but are different than those described in Theorems 2, 3 and 4. The results in that case can be found in Peng (1975), Hudson (1978) and Tsui (1978).

Acknowledgements. The authors would like to thank a referee for suggesting the plus-rules and modifications of our estimators, and the associate editor for comments which led to improvement in the presentation of the paper.

REFERENCES

BROWN, L. D. (1979). An heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994.

CLEVENSON, L., and ZIDEK, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

HUDSON, H. M. (1974). Empirical Bayes estimation. Stanford Univ. Statist. Tech. Report 58.

HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

PENG, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Stanford Univ. Statist. Tech. Report 78.

TSUI, K. W. (1978). Simultaneous estimation of several Poisson parameters under squared error loss. Univ. Calif., Riverside Statist. Tech. Report 37.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 W. DAYTON ST.
MADISON, WISCONSIN 53706

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
RIVERSIDE, CALIFORNIA 92421