Ann. Inst. Statist. Math.
33 (1981), Part A, 215-223

SIMULTANEOUS ESTIMATION OF SEVERAL POISSON PARAMETERS UNDER SQUARED ERROR LOSS*

KAM-WAH TSUI

(Received Dec. 15, 1978; revised Feb. 21, 1981)

1. Introduction

This paper is devoted to simultaneously estimating the parameters of several independent Poisson random variables. We suppose that X_i, $i=1,\cdots,p$, are independent Poisson random variables with parameters $\lambda_1,\cdots,\lambda_p$. We shall assume that only one observation is taken from each population. Although the maximum likelihood estimator (MLE), $X=(X_1,\cdots,X_p)$, has several nice properties, Peng [2] shows that the MLE is inadmissible under squared error loss by proposing estimators which are better than the MLE uniformly in $\lambda=(\lambda_1,\cdots,\lambda_p)$, as long as $p\geqq 3$.

Basically, Peng's estimators (cf. his Theorems 3.1 and 5.1) pull each component of the MLE towards zero whenever the number of non-zero observations exceeds two. The performance of his estimators is expected to be good when the underlying parameters λ_i are relatively small. When some of the parameters are large, however, very little improvement over the MLE is anticipated. In this situation, some very large observations are likely to occur, and both Peng's estimator (Theorem 3.1) and the MLE give virtually the same estimate. In order to remedy this situation, Peng uses Stein's method [3] to modify his estimator.

If *all* the parameters λ_i are relatively large, none of the estimators proposed by Peng will give noticeable improvement over the MLE. This is essentially due to the fact that those estimators are biased towards the origin, a point far away from the true λ. Estimators that shift the observations towards a point in a neighborhood of the true underlying parameter would be expected to give better estimates in this case. This conjecture is shown to be true in Section 3: for each nonnegative integer k, there is a family of estimators $\hat{\lambda}^{(k)}$ of λ such that $\hat{\lambda}^{(k)}$ dominates the MLE uniformly in λ under the squared error

* This paper is based on part of the author's Ph.D. thesis at the University of British Columbia, supervised by S. J. Press and J. V. Zidek.

loss function. The estimator $\hat{\lambda}^{(k)}$ pulls the MLE towards the integer k whenever the number of observations greater than k is at least three, but otherwise gives the same estimate as the MLE. Peng's estimator is a special case ($k=0$).

Section 2 consists of notation and fundamental results which we shall employ in subsequent sections. In Section 4, we propose estimators that shift the MLE towards a point determined by the data itself. Finally, Section 5 reports the results of a computer simulation designed to quantitatively compare the performance of the MLE with an adaptive estimator and one of Peng's estimators.

2. Notation and fundamentals

Let $x=(x_1,\cdots,x_p)$ be a vector of observations of the random vector $X=(X_1,\cdots,X_p)$, where the X_i's are mutually independent Poisson random variables with parameters $\lambda_1,\cdots,\lambda_p$, respectively. The notation we introduce below is essentially that used by Peng [2].

DEFINITIONS.

(1) $N_j=\#\{x_i\colon x_i=j\}$, i.e., the number of x_i's that are equal to j.

(2) $l=\operatorname*{Max}_{i=1}^{p}\{x_i\}$; $N=(N_0,\cdots,N_l)$.

(3) $J=$the set of all integers; $J^+=$the set of all non-negative integers.

(4) $S=\sum_{i=1}^{p} h^2(x_i)=\sum_{j=0}^{l} N_j h^2(j)$. Here $h: J\to R$ is a real-valued function such that $h(y)=0$ if $y<0$.

(5) $J^p=p$-fold cartesian product of J with itself.

(6) $f_i: J^p\to R$, $i=1,\cdots,p$, are real-valued functions such that

  (i) $f_i(x)=0$ if x has a negative coordinate.

  (ii) $E_\lambda|f_i(X+je_i)|<\infty$ for $j=0,1$, where $e_i=$a p-vector whose ith coordinate is one and whose other coordinates are zero.

(7) $f(x)=(f_1(x), f_2(x),\cdots,f_p(x))$.

(8) $\phi_j(N)\equiv f_i(x)$ if $x_i=j$.

(9) $\phi: J^p\to R$ is a real-valued function satisfying the following properties:

  (i) ϕ is nondecreasing in each argument x_i whenever $x_i>k$.

  (ii) ϕ is nonincreasing in each argument x_i whenever $x_i\leqq k$.

  (iii) There is a real number $B>0$ such that $0\leqq\phi(x)\leqq 2B$ and $\phi(x)\not\equiv 0$.

(10) $\hat{\lambda}=(\hat{\lambda}_1,\cdots,\hat{\lambda}_p)$ is an estimator of λ.

(11) $R(\lambda,\hat{\lambda})$ is the risk function of $\hat{\lambda}$.

We are interested in functions h which satisfy the properties listed in Lemma 2.1 below. Before stating the lemma, we provide a representative example of the functions h desired. We shall adopt the convention that $\sum_{n=u}^{v} W(n) \equiv 0$ if $v<u$, for any real-valued function $W(\cdot)$.

Example 2.1. For any fixed non-negative integer k,

$$h(y)=\begin{cases} 1+\sum_{n=2}^{y-k} 1/(k+n)\,, & \text{if } y \geqq k+1 \\ 0\,, & \text{if } y=k \text{ or } y<0 \\ -b\sum_{n=1}^{k-y} 1/(k+1-n)\,, & \text{if } y=0,\cdots,k-1 \text{ and } k\geqq 1 \end{cases}$$

where b is a positive number to be determined such that (8) of Lemma 2.1 below holds. One such b is $b=3^{1/2}\left(\sum_{n=1}^{k} \frac{1}{k+1-n}\right)^{-1}$ if $k \geqq 2$. When $k=1$, b can be any positive number. The following lemma gives the properties of h. For simplicity, we denote $h_j=h(j)$.

LEMMA 2.1. *Let h be as defined in Example 2.1. Then h satisfies the following properties*:

(1) $h_j^2-h_{j-1}^2$ *is nonincreasing in j for $j>k+1$.* (2) $j[h_j-h_{j-1}]$ *is nondecreasing in j for $j>k$ and* $\lim_{j\to\infty} j[h_j-h_{j-1}]=B$ *for some $B>0$.* (3) $h_j>h_{j-1}$, $j=1,2,\cdots$. (4) $h_k=0$. (5) $h_j>0$ *if* $j\geqq k+1$. (6) *If $k>0$, then $h_j<0$ for $j<k$.* (7) $h_{k+1}\geqq B$. (8) $3B^2>h_1h_0$ *provided $k>0$.* (9) $h_{k+1}\geqq j[h_j-h_{j-1}]$ *for* $j\geqq k+2$. (10) $h_j^2-h_{j-1}^2\leqq h_{j+1}^2-h_j^2$ *for* $1\leqq j<k$ *if* $k>0$.

The proof of the lemma is straightforward and is omitted. The estimators $\hat{\lambda}$ of λ we consider will be of the form $X+f(X)$, where f is as defined before.

In the proofs of our theorems, we employ the important basic identity which was used by Hudson [1] and Peng [2]. We state the result as Lemma 2.2 below.

LEMMA 2.2 (Hudson [1]; Peng [2]). *Suppose X is a random vector with independent Poisson random variables as coordinates, and λ is the corresponding Poisson parameter vector. Then under squared error loss* $L(\lambda,\hat{\lambda})=\sum_{i=1}^{p}(\lambda_i-\hat{\lambda}_i)^2$, *the deterioration in risk D of $\hat{\lambda}=X+f(X)$ as compared to the MLE, X, is*

$$D=R(\lambda,\hat{\lambda})-R(\lambda,X)=\mathrm{E}_\lambda\Delta\,,$$

where

$$(2.1)\qquad \Delta=\sum_{i=1}^{p} f_i^2(X)+2\sum_{i=1}^{p} X_i[f_i(X)-f_i(X-e_i)]\,.$$

In terms of N and $\psi_j(N)$, (2.1) can be rewritten as

$$\Delta=\sum_{j=0}^{l} N_j\psi_j^2(N)+2\sum_{j=1}^{l} A_j\,,$$

where $A_j=jN_j[\psi_j(N)-\psi_{j-1}(N-\delta_j+\delta_{j-1})]$, δ_j is an $(l+1)$ vector with the jth coordinate equal to one and the other coordinates zero. To show that an estimator $\hat{\lambda}$ of λ of the form $X+f(X)$ dominates X under squared loss, it suffices to show that $\Delta(x)\leqq 0$ for all $x\in J^{+p}$ with strict inequality for some $x\in J^{+p}$.

3. Shifting the MLE towards k

We use the notation defined in Section 2 and define

$$(3.1)\qquad f_i(x)=-[r_k\phi(x)h(x_i)]/S\,,\qquad i=1,\cdots,p\,,$$

where $r_k=\left(p-\sum_{n=0}^{k} N_n-2\right)_+$ and $(y)_+=\text{Max}\,\{y,0\}$. We shall show that the estimator $\hat{\lambda}^{(k)}=X+f(X)$ of λ dominates X uniformly in λ under the squared error loss function when $p\geqq 3$. The estimator $\hat{\lambda}^{(k)}$ shifts the coordinates of the MLE towards the integer k provided the number of observations greater than k is at least three.

The proof uses the following series of lemmas. Proofs of the lemmas are straightforward and are omitted.

LEMMA 3.1.

(1) *For* $k\geqq 2$,

$$A_j\leqq\frac{-j[h_j-h_{j-1}]\phi(x)r_kN_j}{S}\left[\frac{S-h_j(h_j+h_{j-1})}{S-h_j^2+h_{j-1}^2}\right]\qquad \textit{if } 1\leqq j<k\,.$$

(2) $A_k\leqq\dfrac{\phi(x)kN_kr_kh_{k-1}}{S+h_{k-1}^2}$ *and* $A_{k+1}=-\dfrac{(k+1)N_{k+1}r_k\phi(x)h_{k+1}}{S}\,.$

(3)
$$\begin{aligned}A_j&\leqq-\frac{j[h_j-h_{j-1}]r_k\phi(x)N_j}{S}\left[\frac{S-h_j(h_j+h_{j-1})}{S-h_j^2+h_{j-1}^2}\right]\\&\leqq-\frac{l[h_l-h_{l-1}]r_k\phi(x)N_j}{S}\left[\frac{S-2h_j^2}{S-h_l^2+h_{l-1}^2}\right]\qquad \textit{for } k+2\leqq j\leqq l\,.\end{aligned}$$

LEMMA 3.2.

$$2h_{k+1}^2\geqq h_l^2-h_{l-1}^2\qquad \textit{where } l\geqq k+2\,.$$

LEMMA 3.3.

$$\left\{N_{k+1}+\sum_{j=k+2}^{l} N_j \frac{S-2h_j^2}{S-h_l^2+h_{l-1}^2}\right\} \geqq r_k\,, \qquad \textit{if } r_k>0\,.$$

THEOREM 3.1. *With $f_i(x)=-r_k\phi(x)h(x_i)/S$, $i=1,\cdots,p$, $p\geqq 3$, and h as described in Lemma 2.1, $\Delta\leqq 0$.*

PROOF. Recall that

$$\Delta=\sum_{j=0}^{l} N_j\phi_j^2(N)+2\sum_{j=1}^{l} A_j\,.$$

Case 1. $r_k=0$. Then $\Delta\equiv 0\leqq 0$.

Case 2. $r_k>0$.

(i) $\sum_{j=0}^{l} N_j\phi_j^2(N)=r_k^2\phi^2(x)/S$

(ii) $2\sum_{j=1}^{l} A_j=2\sum_{j=1}^{k-1} A_j+2A_k+2A_{k+1}+2\sum_{j=k+2}^{l} A_j\,.$

By (1) of Lemma 3.1,

$$\begin{aligned}2\sum_{j=1}^{k-1} A_j &\leqq -\frac{\phi(x)r_k}{S}\sum_{j=1}^{k-1} j[h_j-h_{j-1}]N_j\left[\frac{S-h_j(h_j+h_{j-1})}{S-h_j^2+h_{j-1}^2}\right]\\ &\leqq -\frac{\phi(x)r_k}{S}\sum_{j=1}^{k-1} j[h_j-h_{j-1}]N_j\left[\frac{S-h_j^2-h_1h_0}{S-h_j^2+h_{j-1}^2}\right]\end{aligned}$$

since $h_j<0$ for $j<k$ and $h_j>h_{j-1}$, so $h_1h_0\geqq h_jh_{j-1}$ for $1\leqq j\leqq k-1$. Now $r_k>0$ implies that $S\geqq 3h_{k+1}^2\geqq 3B^2$. By (8) of Lemma 2.1, we have $S-h_j^2-h_1h_0\geqq 0$ for $N_j\neq 0$, and hence $\sum_{j=1}^{k-1} A_j\leqq 0$. By (2) and (3) of Lemma 3.1,

$$\begin{aligned}2\sum_{j=1}^{l} A_j &\leqq \frac{2kN_kr_k\phi(x)h_{k-1}}{S+h_{k-1}^2}\\ &\quad -\frac{2r_k\phi(x)}{S}\left[(k+1)N_{k+1}h_{k+1}+l[h_l-h_{l-1}]\sum_{j=k+2}^{l} N_j\frac{S-2h_j^2}{S-h_l^2+h_{l-1}^2}\right]\\ &\leqq \frac{2kN_kr_k\phi(x)h_{k-1}}{S+h_{k-1}^2}-\frac{2\phi(x)r_kkN_{k+1}h_{k+1}}{S}\\ &\quad -\frac{2r_k\phi(x)}{S}\left[h_{k+1}N_{k+1}+l[h_l-h_{l-1}]\sum_{j=k+2}^{l} N_j\frac{S-2h_j^2}{S-h_l^2+h_{l-1}^2}\right]\\ &\leqq \frac{2kN_kr_k\phi(x)h_{k-1}}{S+h_{k-1}^2}-\frac{2\phi(x)r_kkN_{k+1}h_{k+1}}{S}\\ &\quad -\frac{2r_k\phi(x)l[h_l-h_{l-1}]}{S}\left[N_{k+1}+\sum_{j=k+2}^{l} N_j\frac{S-2h_j^2}{S-h_l^2+h_{l-1}^2}\right]\\ &\qquad\qquad (\text{since } h_{k+1}\geqq l[h_l-h_{l-1}])\end{aligned}$$

$$\leqq \frac{2kN_k r_k \phi(x) h_{k-1}}{S+h_{k-1}^2} - \frac{2\phi(x) r_k k N_{k+1} h_{k+1}}{S} - \frac{2r_k^2 \phi(x) B}{S}$$

(by Lemma 3.3 and since $l[h_l - h_{l-1}] \geqq B$). Consequently,

$$(3.2) \qquad \varDelta \leqq \frac{2kN_k r_k \phi(x) h_{k-1}}{S+h_{k-1}^2} - \frac{2kN_{k+1} r_k \phi(x) h_{k+1}}{S} - \frac{r_k^2 \phi(x)}{S}[2B - \phi(x)] < 0$$

since $0 \leqq \phi(x) \leqq 2B$, $h_{k-1} < 0$, and $h_{k+1} > 0$. Q.E.D.

Notice that the bound (3.2) for the unbiased estimate $\varDelta$ of the deterioration in risk of $\hat{\lambda}^{(k)}$ depends on k, N_k and N_{k+1}. Hence, an appropriate choice of k is likely to result in large savings in risk. One reasonable choice of k would be the prior mean of the λ_i's. The dependency of the bound for $\varDelta$ on k, N_k and N_{k+1} further implies that the estimators $\hat{\lambda}^{(k)}$ for various $k \in J^+$ are competitive; one cannot dominate the other.

Since our estimators depend on h, it is interesting to find more examples of functions h which have the properties in Lemma 2.1. For $k=0$, one example is $h(y) = \ln(ay)$, if $y \geqq 1$, and zero otherwise, where $a \geqq 4$. Another example is $h(j) = \sum_{n=1}^{j} 1/g_n$ for $j = 1, 2, \cdots$, zero otherwise, where $\{g_n\}$ is a sequence of real numbers satisfying (1) $g_1 = 1$, (2) $g_{n+1} - g_n \geqq 1$, for $n = 1, 2, \cdots$, (3) $\{n/g_n\}$ is nonincreasing and $\lim_{j\to\infty}(j/g_j) = B > 0$.

Property (8) of h given in Lemma 2.1 guarantees that, in the proof of Theorem 3.1, $S - h_j(h_j + h_{j-1}) \geqq 0$ for $j < k$, which is a sufficient condition that $\sum_{j=1}^{k-1} A_j \leqq 0$. However, it is not a necessary condition: Theorem 3.1 still holds, for example, if h is as given in Lemma 2.1 but (i) conditions (3) and (8) are replaced by (3)′ $h_j > h_{j-1}$ for $j \geqq k+1$ and (8)′ $h_j = -b < 0$ for $j = 0, \cdots, k-1$, or (ii) condition (8) is replaced by (8)″ $3h_{k+1}^2 > h_1 h_0$.

The estimators $\hat{\lambda}^{(k)}$ derived thus far have the property that if the ith observation is equal to k, then $\hat{\lambda}_i^{(k)} = k$. That is, there is no shifting of the observations having the value k. The next theorem provides an estimator of λ which improves on the MLE but whose estimate of λ_i is not necessarily equal to k if the ith observation is equal to k. The theorem unifies and generalizes Theorems 3.1 and 5.1 of Peng [2]. Its proof is similar to that of Theorem 3.1.

THEOREM 3.2. *Let h be as given in Lemma* 2.1 *but with* (3) *and* (8) *replaced by* (3)′ *and* (8)′ *in* (i) *above, and with* (4) *and* (7) *replaced by* (4)′ $h_k = -b$ *and* (7)′ $h_{k+1} \geqq \mathrm{Max}\{1, B\}$. *Define*

$$f_i(x)=\begin{cases} -r_k\phi(x)h(x_i)/S\,, & \textit{if } x_i>k\,, \\ \phi(x)\,\mathrm{Min}\,\{br_k/S,\, 1-r_kh_{k+1}/S\}\,, & \textit{if } x_i\leqq k \textit{ for } i=1,\cdots,p\,. \end{cases}$$

Suppose $0\leqq\phi(x)\leqq\mathrm{Min}\,\{1, 2B\}$ *if* $x_i<k$, $i=1,\cdots,p$, *and let* $\hat{\lambda}^{(k)\prime}=X+f(X)$. *Then* $\Delta(x)\leqq 0$ *for all* $x\in J^{+p}$.

Remarks.

(1) The special case when $k=0$, $\phi(x)\equiv 1$, $b=1$, and h is as given in Example 2.1 is Peng's [2] Theorem 5.1, which shrinks all non-zero observations towards zero while a possible non-zero estimate of λ_i is given for $x_i=0$.

(2) The case when $k=0$, $\phi(x)\equiv 1$, $b=0$, and h is as given in Example 2.1 is Theorem 3.1 of Peng [2].

(3) If $b=0$ in Theorem 3.2, λ_i will be estimated as zero if $x_i=0$. However, if $b>0$, the estimate for λ_i will be possibly non-zero if $x_i=0$. The choice of a relatively large value of b can be interpreted as reflecting the belief that the λ_i's are non-zero.

4. Adaptive estimators

The estimators $\hat{\lambda}^{(k)}$ of λ suggested in Section 3 pull the MLE towards a prechosen non-negative integer k, and the choice of k is guided by the prior knowledge of the λ_i's. A natural question which arises is: Is there an estimator $\hat{\lambda}$ of λ which shifts the observations towards a point determined by the data itself? Theorem 4.1 below indicates that the answer is affirmative. The proof is similar to that of Theorem 3.1 and is therefore omitted.

THEOREM 4.1. *Let* $m=\mathrm{Min}_{i=1}^{p}\{x_i\}$, *and define* $H_i\colon J^p\to R$, $i=1,\cdots,p$, *as follows*:

$$H_i(x)=1+\sum_{n=2}^{x_i-m} 1/(m+n)\,, \tag{4.1}$$

if $x_i\geqq m+1$ *and* $m\geqq 0$, *and* 0 *otherwise. Let* $\hat{\lambda}^{[m]}=(\hat{\lambda}_1^{[m]},\cdots,\hat{\lambda}_p^{[m]})$ *be such that* $\hat{\lambda}_i^{[m]}=X_i-\left((p-N_m-2)_+\phi(X)H_i(X)\Big/\sum_{j=1}^{p}H_j^2(X)\right)$, $i=1,\cdots,p$, *where* $N_m=\#\{i\colon X_i=m\}$, $p\geqq 4$, $\phi(x)$ *is non-negative and non-decreasing in each* x_i, *and* $\phi(x)\leqq 2$. *Then for all* λ, $\hat{\lambda}^{[m]}$ *dominates* X *under squared error loss.*

Remarks.

(1) Note that $N_m\geqq 1$ and hence the estimators $\hat{\lambda}^{[m]}$ dominate the MLE only when p is at least four instead of three. While Peng's estimator dominates the MLE when $p\geqq 3$, his estimator involves an

implicit choice of $k=0$, towards which the observations are shrunk. If this kind of subjectivity is to be avoided and the shrinkage determined only by the data (the minimum, in this case), one degree of freedom is lost and improvement over the MLE results only when $p\geqq 4$.

(2) It can be shown that the functions H_i satisfy similar properties as those of h described in Lemma 2.1. Other choices of the H_i's are possible.

(3) When all the observations are equal to the same value, the estimator yields the grand mean, an intuitively appealing result.

A further application of Stein's method [3] to $\hat{\lambda}^{[m]}$ yields a more versatile estimator of λ in that, unlike Peng's estimator, it guards against extreme observations and cases in which all the parameters are large or small. In addition, it can be shown that there are estimators which pull the observations toward the jth smallest observation, $x_{(j)}$ ($j\geqq 2$) and still dominate the MLE under squared error loss. An example of the latter result is stated below.

THEOREM 4.2. *Let*

$$H_i^{(j)}(x)=\begin{cases} 1+\sum_{n=2}^{x_i-x_{(j)}}[1/(n+x_{(j)})]\,, & \text{if } x_i\geqq x_{(j)}+1,\ m\geqq 0\,,\\ 0\,, & \text{if } x_i=x_{(j)},\ \text{or } m<0\,,\\ -b\,, & \text{if } m\geqq 0 \text{ and } x_i<x_{(j)} \end{cases}$$

where b is a positive real number. Define

(1) $N_{(j)}=\#\{x_i:\ x_i=x_{(j)}\}$ *and* $N_{(j)+n}=\#\{x_i:\ x_i=x_{(j)}+n\}$, $n=1,2,\cdots$.

(2) $r_{(j)}=\left(p-\sum_{n=1}^{j}N_{(n)}-2\right)_+$; $S=\sum_{i=1}^{p}[H_i^{(j)}(x)]^2$.

(3) $f_i(x)=-r_{(j)}H_i^{(j)}(x)/S,\ i=1,\cdots,p$.

Suppose $p\geqq 5$. For each fixed $j=2,\cdots,(p-3)$, let $\hat{\lambda}^{[j]}=X+f(X)$. Then $\hat{\lambda}^{[j]}$ dominates X under squared error loss and has improvement in risk which exceeds

$$\mathrm{E}_\lambda\{r_{(j)}^2/S+2X_{(j)}N_{(j)+1}r_{(j)}/S+2X_{(j)}N_{(j)}r_{(j)}b/(S+b^2)\}\ .$$

The estimator $\hat{\lambda}^{[j]}$ can be further modified by using Stein's method. The application is straightforward and is therefore not included here.

5. Computer simulation

In this section, we describe the results of a computer simulation used to quantitatively compare the MLE with our estimator $\hat{\lambda}^{[m]}$ in Theorem 4.1 (with $\phi(x)\equiv 1$), and with one of Peng's estimators, $\hat{\lambda}^{(0)}$, whose ith coordinate is $X_i-[(p-N_0-2)_+h(X_i)]\Big/\sum_{j=1}^{p} h^2(X_j)$, $i=1,\cdots,p$, with h as defined in Example 2.1.

The p parameters λ_i were generated randomly within a certain range (c, d), and one observation of each of the p distributions with the parameters thus obtained was generated. Estimates of the parameters were then calculated according to the estimators $\hat{\lambda}^{[m]}$ and $\hat{\lambda}^{(0)}$. Generation of the observations was repeated 2000 times and the risks for the estimators $\hat{\lambda}^{[m]}$ and $\hat{\lambda}^{(0)}$, as well as the MLE were calculated. The percentage of savings in using an estimator $\hat{\lambda}$ as compared to the MLE, $[\{(R(\lambda, X)-R(\hat{\lambda}, \lambda))/R(\lambda, X)\}\cdot 100]\%$, was calculated, and the whole process was then repeated a number of times in order to obtain an average percentage of savings of $\hat{\lambda}^{[m]}$ and $\hat{\lambda}^{(0)}$ over the MLE. The ranges of the parameters λ_i were chosen to be small ($(c, c+4)$, $c=0, 4, 8, 12$) to check the performance of the estimators when the parameters are relatively close to one another.

In most of the cases, the improvement percentage is seen to be an increasing function of p, the number of independent Poisson distributions. Moreover, the improvement percentage generally decreases as the magnitude of the λ_i's increases. When the Poisson parameters are in $(0, 4)$, both estimators have similar percentage improvements in risk (4%, 6%, 8%, and 9% for $p=4$, 5, 8, and 10, respectively). The improvement percentage for $\hat{\lambda}^{(0)}$ over the MLE decreases rapidly as the λ_i's move away from zero. In contrast, the improvement percentages for $\hat{\lambda}^{[m]}$ remain noticeable (5% for $p\geqq 5$) even when the λ_i's are in the interval $(12, 16)$. This supports our conjecture that the estimator $\hat{\lambda}^{[m]}$ is superior to $\hat{\lambda}^{(0)}$ when $p\geqq 4$.

Acknowledgement

The author would like to thank both Professor S. James Press and Professor James V. Zidek for their many helpful suggestions. The referee's comments led to the paper's present concise form.

UNIVERSITY OF BRITISH COLUMBIA

REFERENCES

[1] Hudson, H. Malcolm (1978). A natural identity for exponential families with applications in multiparameter estimation, *Ann. Statist.*, **6**, 473-484.

[2] Peng, James C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distribution, *Technical Report* No. 78, Department of Statistics, Stanford University.

[3] Stein, Charles (1974). Estimation of the parameters of a multivariate normal distribution I: Estimation of the means, *Technical Report* No. 63, Department of Statistics, Stanford University.