The Annals of Statistics
1978, Vol. 6, No. 3, 473-484

A NATURAL IDENTITY FOR EXPONENTIAL FAMILIES WITH APPLICATIONS IN MULTIPARAMETER ESTIMATION

BY H. M. HUDSON

Macquarie University

A random variable X is said to have distribution in the class $\mathscr{E}_0$ if, for some real valued, positive function $a(\cdot)$, the identity $E\{(X-\mu)g(X)\} = E\{a(X)g'(X)\}$ holds for any absolutely continuous real valued function $g(\cdot)$ satisfying $E|a(X)g'(X)| < \infty$. Examples of a random variable X possessing a distribution in $\mathscr{E}_0$ are (i) X normally distributed with mean μ and standard deviation 1, (ii) X having a gamma density with mean μ and location parameter 1, (iii) $X = 1/Y$ where $Y \sim [(n-2)]^{-1}\chi_n^2$, $n > 2$. Suppose $X_1, \cdots, X_p$, $p \geqq 3$, are independently distributed with distributions in $\mathscr{E}_0$, for some function $a(\cdot)$, and with means $\mu_1, \cdots, \mu_p$. Define $b(x) = \int a(x)^{-1}\,dx$, where the integral is interpreted as indefinite, $B_i = b(X_i)$, $S = \sum_{i=1}^p B_i^2$, $X' = (X_1, \cdots, X_p)$ and $B' = (B_1, \cdots, B_p)$. Then the estimator $X - ((p-2)/S)B$ dominates X if sum of squared error loss is assumed. Similar estimators are obtained, when $p \geqq 4$, which shrink towards an origin determined by the data. There are corresponding results for some discrete exponential families.

1. Introduction. The estimator X of μ in a p-variate normal distribution, with mean μ and identity covariance matrix, has long been known to be inadmissible if $p \geqq 3$ and the loss function is $L(\mu, \delta) = ||\delta - \mu||^2$. In this case James and Stein (1961) introduced the estimator

$$\left(1 - \frac{p-2}{||X||^2}\right)_+ X$$

which is much superior in risk to X in any case in which $||\mu||$ is small.

Berger (1975), Brown (1966), (1975) and other authors included in the bibliography of Berger's paper have extended the inadmissibility results to cover estimation of the location parameter of a wide class of location invariant distributions for a broad range of loss functions. In a very wide class of problems, then, the best invariant estimator is inadmissible. In this more general case there has been very little progress in obtaining estimators showing considerable practical improvement, there being no analogue of the James–Stein estimator.

The aim of this paper is to introduce an identity which affords a straightforward evaluation of the risk of an estimator in an exponential family, and hence permits the extent of possible improvement on the usual estimator to be determined. This identity is an extension to the exponential family of an identity for the normal distribution used by Stein (1974) to obtain a similar unbiased estimator of the risk of an estimator of the mean.

Received June 1975; revised March 1977.

AMS 1970 *subject classifications.* Primary 62C15; Secondary 62F10, 62C25.

Key words and phrases. Multiparameter estimation, James–Stein estimator, squared error loss, empirical Bayes estimation.

2. A natural identity for an exponential family. Let X be a continuous random variable with the exponential density

$$f_\theta(x) = \exp\{\theta x - \psi(\theta)\}k(x) \tag{2.1}$$

with support $R = (-\infty, \infty)$. Let

$$t(X) = -\frac{k'(X)}{k(X)}. \tag{2.2}$$

Then a simple identity holds for any absolutely continuous function g on R such that $E|g'(X)| < \infty$, namely

$$E\{(t(X) - \theta)g(X)\} = E\{g'(X)\}. \tag{2.3}$$

The choice $g(x) \equiv 1$ shows $t(X)$ to be unbiased for θ. Since the exponential family is complete this implies that $t(X)$ is the minimum variance unbiased estimator (MVUE) of θ. The proof of (2.3) is direct;

$$\begin{aligned} Eg'(X) &= \int g'(x)e^{\theta x - \psi(\theta)}k(x)\,dx \\ &= -\int g(X)e^{\theta x-\psi(\theta)}[\theta k(x) + k'(x)]\,dx \\ &= E\{(t(X) - \theta)g(X)\}. \end{aligned}$$

The second equality is obtained, on integrating by parts, since $g(x)e^{\theta x-\psi(\theta)}k(x)$ vanishes when $x \to \pm\infty$ if $E|g'(X)| < \infty$.

If X has as support an interval (c, d), then it is necessary to impose a condition on the form of the density $f_\theta(\cdot)$ at the endpoints c, d if we are to maintain the identity (2.3) and require no more of g than $E|g'(X)| < \infty$. (Equivalently, this condition is required if $t(X)$ is to be unbiased for θ.) The condition required is that $e^{\theta x}k(x) \to 0$ as $x \downarrow c$ or $x \uparrow d$. The appropriate condition is automatically satisfied if $c = -\infty$ or $d = \infty$.

A subclass of the continuous exponential family has particularly simple properties. Consider the class $\mathscr{E}_0$ of probability measures $\{P_\theta : \theta \in \Theta\}$ with densities $f_\theta(\cdot)$ with respect to Lebesgue measure for which

$$E_\theta\{(X - \mu)g(X)\} = E_\theta\{a(X)g'(X)\} \tag{2.4}$$

for some function $a: R \to R$ and for all absolutely continuous functions g such that $E|a(X)g'(X)| < \infty$; where X is any random variable with distribution P_θ, and $E_\theta X = \mu$. The density $f_\theta(\cdot)$ may then be shown to be of the form

$$f_\theta(x) = \exp\{\mu \textstyle\int a(x)^{-1}\,dx - \chi(\theta)\}a(x)^{-1}\exp\{-\textstyle\int xa(x)^{-1}\,dx\}, \tag{2.5}$$

where the integrals are to be interpreted as indefinite integrals, provided simple regularity conditions are satisfied.

It suffices for the purposes of this paper to note that if X has the density (2.5), if $a \geqq 0$ and if $\int a(x)^{-1}\,dx$ exists in the interior of the domain of X, then (2.4) is an immediate consequence of (2.3). For, setting $b(x) = \int a(x)^{-1}\,dx$, $B = b(X)$ has density of the form

$$f_\theta(b) = \exp\{\mu b - \psi(\mu)\}k(b). \tag{2.6}$$

Applying (2.3)

$$E_\theta\{(t(B) - \mu)h(b)\} = E\{h'(B)\},$$

provided $E|h'(B)| < \infty$. But $t(B) = X$, since X is an unbiased estimator of μ and a sufficient statistic, and $t(B)$ is the MVUE of μ. Rewriting the above identity in terms of X, by setting $g(X) = h(B)$, so that $h'(B) = g'(X)(dX/dB) = a(X)g'(X)$, the result (2.4) is obtained. Note that if $E|a(X)| < \infty$ then, on substituting $g(x) = x$ in (2.4),

$$\sigma^2 = E\{(X - \mu)^2\} = E\{a(X)\}$$

so that $a(X)$ is an unbiased estimator of the variance of X.

Some examples of distributions satisfying (2.4) follow.

EXAMPLE 1. If $X \sim N(\mu, 1)$ then, from (2.3),

$$E\{(X - \mu)g(X)\} = E\{g'(X)\}$$

whenever $E|g'(X)| < \infty$, so $\mathscr{E}_0$ contains the normal distributions.

EXAMPLE 2. If $X \sim$ gamma $(\mu, 1)$, so that

$$f_\mu(x) = \frac{1}{\Gamma(\mu)} x^{\mu-1}e^{-x}, \qquad x > 0, \tag{2.7}$$

then $f_\mu(x)$ is given by (2.5) when $a(x) = x$. $B = \log X$ has range $(-\infty, \infty)$ so (2.3) is satisfied. Thus the gamma $(\mu, 1)$ distribution is contained in $\mathscr{E}_0$.

EXAMPLE 3. If $a(x) = x^2/(n/2 - 1)$, $x > 0$, in (2.5) then

$$\begin{aligned} f_\theta(x) &\propto \exp\left\{-\left(\frac{n}{2} - 1\right)\mu x^{-1}\right\} x^{-2} \exp\left\{-\left(\frac{n}{2} - 1\right)\log x\right\} \\ &= x^{-n/2-1} \exp\left\{-\left(\frac{n}{2} - 1\right)\mu x^{-1}\right\} \qquad x > 0. \end{aligned}$$

Now $B = (n/2 - 1)(1 - X^{-1})$ has range $(-\infty, 1)$. Here $k(b) = 1 - b$, and $e^{\theta b}(1 - b) \to 0$ as $b \uparrow 1$, so (2.3) holds.

If X has this density then $Y = 1/X$ has density

$$\begin{aligned} f(y) &\propto y^{n/2+1} \exp\left\{-\left(\frac{n}{2} - 1\right)\mu y\right\} y^{-2} \qquad y > 0 \\ &= y^{n/2-1} \exp\left\{-\left(\frac{n}{2} - 1\right)\mu y\right\} \qquad y > 0 \end{aligned} \tag{2.8}$$

i.e., $Y \sim [(n - 2)\mu]^{-1}\chi_n^2$.

A similar identity for a discrete exponential family may be obtained. Let X, taking values in $N = \{0, 1, 2, \cdots\}$, have the probability function

$$f_\theta(x) = \exp\{\theta x - \psi(\theta)\}k(x). \tag{2.9}$$

Let $\phi = e^\theta$. Then, for $g: N \to R$ satisfying $E|g(X)| < \infty$,

$$\phi Eg(X) = E\{t(X)g(X - 1)\}, \tag{2.10}$$

where

(2.11) $$t(x) = 0 \qquad \text{for} \quad x = 0$$
$$= \frac{k(x-1)}{k(x)} \qquad \text{for} \quad x = 1, 2, \cdots$$

is the minimum variance unbiased estimator of ϕ.

3. Applications to continuous multiparameter estimation. Suppose random samples are selected from each of p populations, and it may be assumed that the distribution of the ith sample is given by the density

$$f_{\theta_i}(x) = \exp\{\theta_i T(x) - \psi(\theta_i)\}k_0(x) \qquad i = 1, \cdots, p\,.$$

Then in estimating any function of θ_i, $i = 1, \cdots, p$, it is enough to consider a sufficient statistic X_i, $i = 1, \cdots, p$, obtained from the ith sample. It is well known that X_i itself follows an exponential distribution.

A common problem is that of the estimation of the mean $\mu = (\mu_1, \cdots, \mu_p)'$ of $X = (X_1, \cdots, X_p)'$. Here the alternative form of the identity (2.3) may be used to obtain an estimator which dominates the usual estimator X of μ for the subclass $\mathscr{E}_0$ of the continuous exponential family.

Let $X_1, \cdots, X_p$ be independent, and X_i satisfy the identity

$$E_\mu\{(X_i - \mu)g(X_i)\} = E_\mu\{a(X_i)g'(X_i)\}\,,$$

for some nonnegative function $a(\cdot)$, for any function g such that $E\{|a(X)g'(X)|\} < \infty$. Consider the difference in risk, ΔR, for the estimators X and $X + g(X)$ under squared error loss

$$\Delta R = E_\mu L(\mu, X) - E_\mu L(\mu, X + g(X))$$
$$= E_\mu\{\sum (X_i - \mu_i)^2\} - E_\mu\{\sum (X_i + g_i(X) - \mu_i)^2\}$$

where $g_1, \cdots, g_p$ are the components of $g: R^p \to R$. Then, on expanding the squares,

$$\Delta R = E_\mu\{-2(X_i - \mu_i)g_i(X) - \sum g_i^2(X)\}$$
$$= E_\mu\{-2 \sum a(X_i)g_{ii}(X) - \sum g_i^2(X)\}$$

where $g_{ii}(X) = (\partial/\partial X_i)g_i(X)$. This last equality is obtained by conditioning on $\{X_j : j \neq i\}$ and applying (2.4). The necessary assumptions regarding the function $g(\cdot)$ will be checked later.

Set

$$g_i(X) = -\frac{(p-2)}{S} b(X_i) \qquad i = 1, \cdots, p$$

where

(3.1) $$b(x) = \int a(x)^{-1}\,dx\,, \qquad S = \sum_{i=1}^p b^2(X_i)\,.$$

Then

(3.2) $$g_{ii}(X) = \frac{p-2}{S} a(X_i)^{-1} + \frac{(p-2)}{S^2} b(X_i)\frac{\partial S}{\partial X_i}$$
$$= -\frac{p-2}{S} a(X_i)^{-1} + \frac{2(p-2)}{S^2} b^2(X_i)a(X_i)^{-1}$$

so that

$$\Delta R = E_\mu\left\{\frac{2(p-2)}{S}p - \frac{4(p-2)}{S^2}\sum b^2(X_i) - \frac{(p-2)^2}{S^2}\sum b^2(X_i)\right\} \tag{3.3}$$

$$= E_\mu\left\{\frac{(p-2)^2}{S}\right\}.$$

If $p \geqq 3$ then (3.3) indicates that the improvement in risk of the estimator

$$X_i - \frac{(p-2)}{S}b(X_i) \tag{3.4}$$

is necessarily positive, since S is positive. Thus the estimator (3.4) dominates the unbiased estimator X of the mean vector μ when $p \geqq 3$.

It remains only to check the conditions for the identity (2.4). From (3.2)

$$|a(X_i)g_{ii}(X)| = \left|-\frac{p-2}{S} + \frac{2(p-2)}{S^2}b^2(X_i)\right|$$

$$< \left|\frac{p-2}{S}\right| + \left|\frac{2(p-2)S}{S^2}\right|$$

$$= \frac{3(p-2)}{S}.$$

Thus, provided $E\{S^{-1}\} < \infty$, the condition is satisfied.

Now note that, from (3.1) and (2.6),

$$E\{S^{-1}\} = \int \cdots \int \frac{1}{\sum_i t_i^2} \prod_{j=1}^p \exp\{\mu_j t_j - \psi(\mu_j)\}k(t_j)\,dt_j\,.$$

It clearly suffices to show the finiteness of the integrand in the sphere $A = \{t: ||t|| \leqq \delta\}$. The result is then obvious on transforming to polar coordinates $(r, \boldsymbol{\phi})$ recalling that the Jacobian of the transformation is bounded by cr^{p-2} for some constant c, and that k is bounded in A.

COROLLARY 1 (James–Stein (1962), Stein (1974)). *Let X have a p-variate normal distribution with mean vector μ and variance matrix the identity. Suppose $p \geqq 3$, then, with squared error loss*

$$L(\mu, \delta) = \sum_{i=1}^p (\delta_i - \mu_i)^2, \tag{3.5}$$

the unbiased estimator X is dominated by the estimator

$$\left(1 - \frac{p-2}{S}\right)X,$$

which has risk

$$p - E_\mu\left\{\frac{(p-2)^2}{S}\right\}.$$

COROLLARY 2. *Let $X_1, \cdots, X_p$ be independent random variables having gamma densities of the form* (2.7) *with means $\mu_1, \cdots, \mu_p$. If $p \geqq 3$, the unbiased estimator*

X is dominated by the estimator

$$X - \frac{p-2}{S} B ,$$

where

$$B_i = \log X_i$$
$$S = \sum_{i=1}^{p} \log^2 X_i .$$

The risk of this estimator is

$$R(\mu, X) - E\left\{\frac{(p-2)^2}{S}\right\} .$$

COROLLARY 3. *Let $X_1, \cdots, X_p$ be independent and $Y_i = X_i^{-1}$ have the distribution*

$$Y_i \sim [(n-2)\mu_i]^{-1}\chi_n^2 , \qquad n > 4 .$$

Then the unbiased estimator X of its mean vector μ is dominated by the estimator

$$X_i - \frac{(p-2)}{(n/2-1)\sum(1-X_j^{-1})^2}(1 - X_i^{-1}) , \qquad i = 1, \cdots, p ,$$

when $p \geqq 3$ and $L(\mu, \delta)$ is given by (3.5).

The improvement term, $E\{(p-2)^2/S\}$, is quite substantial in favourable circumstances. For the normal distribution example S has a noncentral χ^2 distribution and the improvement term can be calculated exactly, being $(p-2)$ when $\mu = 0$. This compares favourably with the risk p of the unbiased estimator. In Corollary 3 one may apply Jensen's inequality to ΔR and obtain a lower bound for the improvement, when $\mu_1 = \cdots = \mu_p = 1$, relative to the risk R_0 of the unbiased estimator. Then $\Delta R/R_0 \geqq ((p-2)^2/p^2)((n-4)/(n+2))$, if $n > 4$. Jensen's inequality may be applied in Corollary 2 also. Then $\Delta R/R_0 \geqq (p-2)^2/\nu p^2$ when $\mu_1 = \cdots = \mu_p = 1$ and $\nu = E \log^2 X$, where X has a negative exponential distribution with mean 1.

Note that the estimators obtained in (3.4) shift all estimates towards an arbitrary origin in the domain of X (chosen in Corollary 1 to be 0 and in Corollaries 2 and 3 to be 1), the choice of origin being determined by the actual choice of the indefinite integral. This suggests that there is an analogue of the usual refinement of the James–Stein estimator

$$\bar{X} + \left(1 + \frac{p-3}{S}\right)(X_i - \bar{X}) \qquad i = 1, \cdots, p ,$$

where $\bar{X} = \sum_{i=1}^{p} X_i/p$, $S = \sum_{i=1}^{p}(X_i - \bar{X})^2$, which shifts towards an origin determined by the data.

4. Some refinements of the multiparameter estimator. In this section a theorem is obtained showing the method of shifting towards an origin determined by the data.

THEOREM 1. *Given $X_1, \cdots, X_p$ independently distributed as in Section* 3, *and* *squared error loss, set* $B_i = b(X_i)$, *with* $b(\cdot)$ *as defined in* (3.1), $\bar{B} = \sum B_i/p$, $S = \sum (B_i - \bar{B})^2$. *Then for* $p \geqq 4$ *the estimator*

$$X_i - \frac{(p-3)}{S}(B_i - \bar{B})$$

dominates X *and has risk function*

$$R(\mu, X) - E_\mu\left\{\frac{(p-3)^2}{S}\right\}.$$

PROOF. The estimator is $X + g(X)$ where $g_i(X) = -((p-3)/S)(B_i - \bar{B})$, $i = 1, \cdots, p$. Since

$$\partial\bar{B}/\partial X_i = \frac{1}{p}a(X_i)^{-1}, \qquad \partial S/\partial X_i = 2(B_i - \bar{B})a(X_i)^{-1}$$

$$g_{ii}(X) = \frac{\partial g_i}{\partial X_i} = -\frac{(p-3)}{S}\left(a(X_i)^{-1} - \frac{1}{p}a(X_i)^{-1}\right) + \frac{2(p-3)}{S^2}(B_i - \bar{B})^2 a(X_i)^{-1}$$

and therefore

$$-\sum a(X_i)g_{ii}(X) = \frac{(p-3)}{S}(p-1) - \frac{2(p-3)}{S^2}\sum (B_i - \bar{B})^2$$
$$= \frac{(p-3)^2}{S}.$$

Thus, proceeding as before,

$$\Delta R = E_\mu\{\sum (X_i - \mu_i)^2 - \sum (X_i + g_i(X) - \mu_i)^2\}$$
$$= E_\mu\{-2\sum a(X_i)g_{ii}(X) - \sum g_i^2(X)\}$$
$$= E_\mu\left\{\frac{(p-3)^2}{S}\right\} \geqq 0.$$

It is simple to check that the conditions required for application of the identity are satisfied.

Other refinements which can be made to the James–Stein estimator have their analogue in the class $\mathscr{E}_0$ of distributions. The following theorem, for instance, is simple to obtain. The original proof for the normal distribution is due to Stein.

THEOREM 2. *Given* $X_1, \cdots, X_p$ *independently distributed as in Section* 3, *and squared error loss, set* $B_i = b(X_i)$, *with* $b(\cdot)$ *defined in* (3.1), *and arrange them in increasing order of absolute magnitude*

$$0 \leqq |B_{(1)}| \leqq |B_{(2)}| \leqq \cdots \leqq |B_{(p)}|.$$

For any $k \geqq 3$ *define* $S = \sum_1^k B_{(i)}^2 + (p-k)B_{(k)}^2$ *and* $g: R^p \to R$ *by*

$$g_i(X) = -\frac{(k-2)}{S}B_i \qquad \text{if } |B_i| \leqq |B_{(k)}|$$
$$= -\frac{(k-2)}{S}|B_{(k)}|\operatorname{sgn}(B_i) \qquad \text{if } |B_i| > |B_{(k)}|.$$

Then the estimator $X + g(X)$ dominates X and has risk

$$R(\mu, X) - E_\mu\left\{\frac{(k-2)^2}{S}\right\}.$$

5. Applications to discrete exponential families. The identity (2.10) for a discrete exponential family allows similar results to be obtained in multiparameter estimation of $\phi_1, \cdots, \phi_p$, where $X_1, \cdots, X_p$ are independent with distributions of the form (2.9), natural parameters $\theta_1, \cdots, \theta_p$, and $\phi_i = e^{\theta_i}$, $i = 1, \cdots, p$. The parameter ϕ is only of statistical interest for two distributions, however, these being the Poisson and negative binomial distributions.

If X is a Poisson variable with mean μ then the identity (2.10) is that

$$\mu Eg(X) = EXg(X-1) \tag{5.1}$$

provided $E|g(X)| < \infty$. This identity was first obtained by Stein. If X has the negative binomial distribution

$$f_\pi(x) = \binom{r+x-1}{x}\pi^x(1-\pi)^r \qquad x = 0, 1, 2, \cdots,$$

then the identity (2.10) provides that

$$\pi Eg(X) = E\left\{\frac{X}{r-1+X}\, g(X-1)\right\} \tag{5.2}$$

if $E|g(X)| < \infty$.

In multiparameter estimation of $\mu_1, \cdots, \mu_p$ and $\pi_1, \cdots, \pi_p$, respectively, from independent observations $X_1, \cdots, X_p$ having either Poisson, or negative binomial, distributions the MVUE can once again be bettered. The following argument is an extension of a use of the identity (5.1) in the Poisson case by Peng (1978) to show the inadmissibility of X under squared error loss if $p \geqq 3$.

Suppose $X_1, \cdots, X_p$ are independent observations satisfying the identity

$$\phi_i E\{g(X_i)\} = E\{t(X_i)g(X_i - 1)\},$$

where $t(X_i)$ is the MVUE of ϕ_i, provided $E|g(X_i)| < \infty$, $i = 1, \cdots, p$. Let $T = (t(X_1), \cdots, t(X_p))'$, $\phi = (\phi_1, \cdots, \phi_p)'$ and

$$L(\phi, \delta) = \textstyle\sum_{i=1}^p (\delta_i - \phi_i)^2. \tag{5.3}$$

Assume also that $E|g_i(X)| < \infty$, $i = 1, \cdots, p$. Then the improvement in risk of the estimator $T + g(X)$ on T is

$$\begin{aligned}\Delta R &= E\{\textstyle\sum_{i=1}^p (t(X_i) - \phi_i)^2 - \sum_{i=1}^p (t(X_i) + g_i(X) - \phi_i)^2\} \\ &= E\{-\textstyle\sum_{i=1}^p g_i(X)[2t(X_i) - 2\phi_i + g_i(X)]\} \\ &= E\{-2\textstyle\sum t(X_i)g_i(X) + 2\sum t(X_i)g_i(X - e_i) - \sum g_i^2(X)\} \\ &= E\{-2\textstyle\sum t(X_i)[g_i(X) - g_i(X - e_i)] - \sum g_i^2(X)\}\end{aligned}$$

where e_i is the unit ith coordinate vector $(0, 0, \cdots, 1, 0, \cdots, 0)'$. The identity (2.10) is used to obtain the third equality above, after conditioning on $\{X_j : j \neq i\}$.

Following Peng we shall treat the indices as interchangeable and, for notational convenience, define

$$\begin{aligned}
l &= \max\{X_i; i = 1, \cdots, p\}, \quad \text{the largest observation,} \\
N_j &= \#\{i: X_i = j\} \qquad j = 0, 1, \cdots, l, \\
N &= (N_0, N_1, \cdots, N_l)', \\
\phi_j(N) &= g_i(X) \quad \text{for those } i \text{ for which } X_i = j \\
&\qquad i = 1, \cdots, p; j = 1, \cdots, l.
\end{aligned}$$

Then, upon defining $t_j = t(j)$,

$$\Delta R = E\{-2 \textstyle\sum_{j=0}^{l} N_j t_j[\phi_j(N) - \phi_{j-1}(N - e_j + e_{j-1})] - \textstyle\sum_{j=0}^{l} N_j \phi_j^2(N)\}. \tag{5.4}$$

Finally, choose

$$\phi_j(N) = -\frac{p_0}{S} b_j \qquad j = 0, 1, \cdots, l \tag{5.5}$$

where $b_0 = 0$, $b_j = \sum_{k=1}^{j} t_k^{-1}$ if $j \geqq 1$, and $S = S(N) = \sum_{j=0}^{l} N_j b_j^2$. p_0 will be chosen later as either $(p - N_0 - 2)_+$ or $(p - N_0 - 3)_+$.

LEMMA. *With the above choice of $\phi_j(N)$, with $p_0 = (p - N_0 - 3)_+$,*

$$-2 \textstyle\sum_{j=0}^{l} N_j t_j[\phi_j(N) - \phi_{j-1}(N - 3_j + e_{j-1})] - \textstyle\sum_{j=0}^{l} N_j \phi_j^2(N) \geqq \frac{p_0^2}{S} \tag{5.6}$$

provided $t_0 = 0$ and $\{t_j\}$ is increasing in j. If $c_j = t_j^{-1}(b_j + b_{j-1})$, $j = 1, 2, \cdots$, and $\{c_j\}$ is decreasing for $j \geqq 2$, then, with the choice $p_0 = (p - N_0 - 2)_+$ in (5.5), *the inequality* (5.6) *remains true.*

PROOF. Note first that, if $N_j \geqq 1$,

$$S(N - e_j + e_{j-1}) = S(N) - \frac{1}{t_j}(b_j + b_{j-1}).$$

Thus

$$\begin{aligned}
&-2 \textstyle\sum_{j=0}^{l} N_j t_j[\phi_j(N) - \phi_{j-1}(N - e_j + e_{j-1})] \\
&\quad = 2N_1 t_1 \frac{p_0}{S}\frac{1}{t_1} + 2p_0 \textstyle\sum_{j=2}^{l} N_j t_j \left[\frac{b_j}{S} - \frac{b_{j-1}}{S - (1/t_j)(b_j + b_{j-1})}\right] \\
&\quad = 2N_1 \frac{p_0}{S} + \frac{2p_0}{S} \textstyle\sum_{j=2}^{l} N_j t_j \frac{(b_j - b_{j-1})S - (b_j/t_j)(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \\
&\quad = 2N_1 \frac{p_0}{S} + \frac{2p_0}{S} \textstyle\sum_{j=2}^{l} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})}.
\end{aligned}$$

Since $\sum_{j=0}^{l} N_j \phi_j^2(N) = p_0^2/S$ it is sufficient to show that

$$\begin{aligned}
&\textstyle\sum_{j=2}^{l} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \\
&\qquad \geqq p - N_0 - N_1 - 3, \qquad \text{if } p_0 = (p - N_0 - N_1 - 3)_+ \\
&\qquad \geqq p - N_0 - N_1 - 2, \qquad \text{if } p_0 = (p - N_0 - N_1 - 2)_+.
\end{aligned}$$

Let

$$\mathscr{C}_1 = \left\{j : 2 \leqq j < l, \frac{1}{t_j}(b_j + b_{j-1}) \geqq \frac{1}{t_l}(b_l + b_{l-1})\right\}$$

$$\mathscr{C}_2 = \left\{j : 2 \leqq j < l, \frac{1}{t_j}(b_j + b_{j-1}) \leqq \frac{1}{t_l}(b_l + b_{l-1})\right\}$$

and $n_1 = \sum_{j \in \mathscr{C}_1} N_j$, $n_2 = \sum_{j \in \mathscr{C}_2} N_j$. Note that $n_1 = p - N_0 - N_1 - N_2$ and $n_2 = 0$ when the sequence $t_j^{-1}(b_j + b_{j-1})$ is decreasing.

If $j \in \mathscr{C}_1$ and $N_j \geqq 1$

$$\frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \geqq \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_l)(b_l + b_{l-1})}$$

since $b_j(b_j + b_{j-1}) \leqq 2b_j^2 \leqq b_j^2 + b_l^2 \leqq S$. If $j \in \mathscr{C}_2$ and $N_j \geqq 1$

$$\frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} = 1 - \frac{b_{j-1}(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})}$$
$$\geqq 1 - \frac{b_{j-1}(b_j + b_{j-1})}{S - (1/t_l)(b_l + b_{l-1})} .$$

Thus

$$\begin{aligned}
&\sum_{j=2}^{l} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \\
&\quad \geqq \sum_{j \in \mathscr{C}_1} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_l)(b_l + b_{l-1})} \\
&\qquad + \sum_{j \in \mathscr{C}_2} N_j \left[1 - \frac{b_{j-1}(b_j + b_{j-1})}{S - (1/t_l)(b_l + b_{l-1})}\right] \\
&\qquad + N_l \frac{S - b_l(b_l + b_{l-1})}{S - (1/t_l)(b_l + b_{l-1})} \qquad (5.7) \\
&\quad = \frac{1}{S_l - (1/t_l)(b_l + b_{l-1})} \{(n_1 + N_l)S - \sum_{j \in \mathscr{C}_1 \cup \{l\}} N_j b_j (b_j + b_{j-1}) \\
&\qquad - \sum_{j \in \mathscr{C}_2} N_j b_{j-1}(b_j + b_{j-1})\} + n_2 \\
&\quad \geqq \frac{(n_1 + N_l - 2)S}{S - (1/t_l)(b_l + b_{l-1})} + n_2 \\
&\quad \geqq (n_1 + n_2 + N_l - 2) = p - N_0 - N_1 - 2 ,
\end{aligned}$$

where the last inequality is correct provided $n_1 + N_l \geqq 2$, which is certainly the case whenever $n_1 \geqq 1$ or $N_l \geqq 2$. If $n_1 = 0$ and $N_l = 1$, then (5.7) may be written

$$\sum_{j=2}^{l} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \geqq \sum_{j=2}^{l} N_j - \frac{\sum_{j=2}^{l} N_j b_{j-1}(b_j + b_{j-1})}{S - (1/t_l)(b_l + b_{l-1})} .$$

Then

$$\begin{aligned}
&\sum_{j=2}^{l} N_j b_{j-1}(b_j + b_{j-1}) \\
&\quad = 2 \sum_{j=2}^{l} N_j b_j^2 - \sum_{j=2}^{l} N_j (2b_j^2 - b_{j-1} b_j - b_{j-1}^2)
\end{aligned}$$

$$= 2S - \sum_{j=2}^{l} N_j \frac{1}{t_j}(2(b_j + b_{j-1}) - b_{j-1})$$

$$= 2S - 2\frac{1}{t_l}(b_l + b_{l-1}) + \frac{1}{t_l} b_{l-1} - \sum_{j=2}^{l-1} N_j \frac{1}{t_j}(2b_j + b_{j-1})$$

$$\leqq 2\left(S - \frac{1}{t_l}(b_l + b_{l-1})\right) + \frac{1}{t_l} b_{l-1}\,.$$

Therefore

$$\sum_{j=2}^{l} N_j \frac{S - b_j(b_j + b_{j-1})}{S - (1/t_j)(b_j + b_{j-1})} \geqq \sum_{j=2}^{l} N_j - 2 - \frac{(1/t_l)b_{l-1}}{S - (1/t_l)(b_l + b_{l-1})}$$

$$\geqq \sum_{j=2}^{l} N_j - 2 - \frac{1}{t_l b_{l-1}}$$

$$\geqq \sum_{j=2}^{l} N_j - 2 - \frac{t_{l-1}}{t_l} \geqq \sum_{j=2}^{l} N_j - 3$$

$$= p - N_0 - N_1 - 3$$

where the second inequality uses the fact that

$$S_l - \frac{1}{t_l}(b_l + b_{l-1}) = S(N - e_l + e_{l-1}) \geqq b_{l-1}^2$$

and the third inequality follows because $b_{l-1} \geqq 1/t_{l-1}$. Thus the proof is completed.

This lemma together with (5.4) shows the following result.

THEOREM. *Let $X_1, \cdots, X_p$ be independent, satisfying the identities $\phi_i E\{g(X_i)\} = E\{t(X_i)g(X_i - 1)\}$ whenever $E|g(X_i)| < \infty$, $i = 1, \cdots, p$. Define $g_i(X) = -(p_0/S)B_i$, $i = 1, \cdots, p$, where $b_j = \sum_1^j (1/t_k)$, $B_i = b_{X_i}$ and $S = \sum_{i=1}^p b_{X_i}^2$.*

1. *If $p \geqq 4$ and $p_0 = (p - N_0 - 3)_+$ the estimator $T + g(X)$ dominates $T = (t(X_1), \cdots, t(X_p))'$ under squared error loss* (5.3). *The improvement in risk exceeds $E_\phi\{p_0^2/S\}$.*
2. *If $p \geqq 3$, $p_0 = (p - N_0 - 2)_+$, and $(1/t_j)(b_j + b_{j-1}) \downarrow$ for $j \geqq 2$, then $T + g(X)$ dominates T under squared error loss* (5.3). *The improvement in risk exceeds $E_\phi\{p_0^2/S\}$.*

COROLLARY 1 (Peng (1978)). *If $X_1, \cdots, X_p$ are independent Poisson variables, with means $\mu_1, \cdots, \mu_p$, then, if $p \geqq 3$, the estimator $X + g(X)$ of μ dominates X under squared error loss, if $g_i(X) = -((p - N_0 - 2)_+/S) \sum_1^{X_i} (1/k)$, where $S = \sum_{i=1}^p (\sum_{k=1}^{X_i} (1/k))^2$.*

COROLLARY 2. *Let $X_1, \cdots, X_p$, $p \geqq 4$, be independent negative binomial variables measuring the number of successes before the* rth *failure, with the probabilities of success being $\pi_1, \cdots, \pi_p$ respectively. The* MVUE $T = (t(X_1), \cdots, t(X_p))'$, $t(x) = x/(r - 1 + x)$, *is dominated by $T + g(X)$, if $g_i(X) = -((p - N_0 - 3)_+/S)(X_i + (r - 1) \sum_1^{X_i} (1/k))$, where $S = \sum_{i=1}^p (X_i + (r - 1) \sum_1^{X_i} (1/k))^2$.*

6. Comments. The correspondence of the results obtained here from some exponential families of distributions with the James–Stein result for the normal distribution is striking. It would be surprising if dominating estimators could not be produced in many further cases.

In the class of distributions $\mathscr{E}_0$ the estimator $X - ((p-2)/S)B$, with $B_i = b(X_i)$, $b(x) = \int a(x)^{-1}\,dx$, was proposed. Recall that $E\{a(X_i)\} = \operatorname{Var} X_i$. For distributions outside the class $\mathscr{E}_0$ there may be similar unbiased estimators of the variance, and the above estimator would appear satisfactory. In cases where no unbiased estimator of the variance exists, the above results suggest the speculation that an estimator with reasonable properties could be obtained by defining

$$\alpha(\mu) = \operatorname{Var}_\mu(X_i)\,, \qquad \beta(x) = \int \alpha(x)^{-1}\,dx$$

and then proposing the estimator $X - ((p-2)/S)B$, where now $B_i = \beta(X_i)$, $S = \sum B_i^2$.

Further work is needed particularly in the discrete exponential family. The results of Sections 3 and 4 permitted a shift towards an arbitrary point in the range of a continuous variable but for the discrete distributions considered the results obtained only concern shifts towards the origin. The corresponding results for shifts towards nonzero origins are complicated by the discreteness of the distribution but do not appear intractable.

The identity is also useful in giving further insight into, and a simple proof of, the result of Clevenson and Zidek (1975) concerning minimax estimation of the means of Poisson random variables.

Acknowledgment. The foundations of this paper form part of the author's Stanford University Ph. D. dissertation. I am indebted to Professors Lincoln Moses and Charles Stein for many helpful discussions.

REFERENCES

[1] Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328.

[2] Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

[3] Brown, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *J. Amer. Statist. Assoc.* **70** 417–427.

[4] Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

[5] Hudson, H. M. (1974). Empirical Bayes estimation. Stanford Univ. Technical Report No. 58.

[6] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379, Univ. of California Press.

[7] Peng, J. C. (1978). Simultaneous estimation of the parameters of independent Poisson distribution. *Ann. Statist.* To appear.

[8] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. Stanford Univ. Technical Report No. 48.

School of Economic and Financial Studies
Macquarie University
North Ryde, New South Wales 2113
Australia