Ann. Inst. Statist. Math.
34 (1982), Part A, 433-439

EXAMPLES OF ESTIMATION PROBLEMS*

JIUNN TZON HWANG

(Received Oct. 15, 1981; revised Feb. 4, 1982)

Summary

Two examples of estimation problems are given. In the first example, X_1, X_2 and X_3 are independent random variables with X_1 having a Poisson distribution with mean θ_1, X_2 being $N(\theta_1, 1)$ and X_3/θ_3 having a chi-square distribution with n degrees of freedom. Based on these three observations, an estimator of $(\theta_1, \theta_2, \theta_3)$, strictly better than the standard one $(X_1, X_2, X_3/(n+2))$, is constructed by solving an inequality. In the second example, we establish a counter-example to the assertion that the lack of a nontrivial solution to a difference inequality (corresponding to the problem of improving upon an estimator δ through an identity of Hudson's (1974, *Technical Report* No. 58, Stanford University), and Stein's type (1973, *Proc. Prague Symp. Asymptotic Statist.*, 345–381)) implies the admissibility of δ. Implications of these two examples are discussed.

1. Introduction

The controversial results of Stein [10] stated that, based on three independent observations, X_i, $i=1, 2, 3$, having normal distributions with means θ_i and variances 1, the standard estimator (X_1, X_2, X_3) for $(\theta_1, \theta_2, \theta_3)$ is inadmissible under the sum of squared error loss, even though X_i is admissible for θ_i under the squared error loss function. The first reaction of many surprised statisticians was to think that the inadmissibility of the standard estimator is linked closely to the sum of squared error loss function. However, Brown's [2] results confirmed that Stein's example is in fact a phenomenon that occurs for general loss functions

* Research supported partly by AFOSR with Grant Number 77-3291 at Purdue University and partly by NSF with Grant Number MSC-8003568 at Cornell University. This manuscript is a revised version of part of the author's Ph.D. thesis submitted to Purdue University.

AMS 1970 subject classification: Primary 62C15, 62F10; Secondary 62H99, 39A10.

Key words and phrases: Admissibility, loss function, differential inequality, difference inequality, Poisson distribution, normal distribution and chi-square distribution.

and broad classes of distributions when the estimation problem is invariant under the location transformation.

Efforts have since been made to understand Stein's phenomenon. Using the empirical Bayes approach, Efron and Morris [4] derived the James-Stein estimator (an estimator better than the usual one), thus providing an appealing explanation of Stein's phenomenon. Their derivation, however, was based heavily on the symmetry assumption that the observations all have normal distributions with the same variance.

As mentioned above, most previous authors (see the reference in Hwang [7]) seemed to concentrate on the case in which the problem is either invariant or symmetric in the coordinates. From the practical point of view, these assumptions are of course reasonable, but it is also theoretically interesting to look at the more general case. In an asymmetric setting, where the observations have different distributions, Berger [1] also observed the Stein phenomenon, that is, combination of componentwise admissible estimators leads to an inadmissible estimator for the whole problem. No such results have been obtained, however, in the extreme case where some of the observations have discrete distributions and some continuous.

In Example 1 below, it is assumed that three independent observations have completely different distributions, namely Poisson, Normal, and Chi-square shifted according to a scalar parameter $\theta_3 > 0$, and yet an estimator that dominates the standard one can be constructed. This example, which is very hard to explain using the empirical Bayes method, shows that Stein's phenomenon appears to be a property more basic than invariance or coordinate symmetry. Additionally, the estimator constructed is interesting in its own right. For the first and second coordinates, this improved estimator corrects the standard one by shrinking toward zero and, for the third coordinate, by expanding toward infinity. This demonstrates a mixture of Stein's and Berger's phenomena [1].

The improved estimator is constructed by using a technique parallel to Stein's [11]. Specifically, let $\boldsymbol{\delta}^0$ be an estimator that one wants to improve under a loss function $L(\cdot, \cdot)$. Write a competitor as $\boldsymbol{\delta}^*(\boldsymbol{X}) = \boldsymbol{\delta}^0(\boldsymbol{X}) + \boldsymbol{\Phi}(\boldsymbol{X})$. It can be shown, (by integration by parts for continuous exponential families and by change of variables for discrete exponential families), that

$$R(\boldsymbol{\theta}, \boldsymbol{\delta}^*) - R(\boldsymbol{\theta}, \boldsymbol{\delta}^0) = \mathrm{E}_{\boldsymbol{\theta}}\, \mathcal{D}(\boldsymbol{X})$$

where $R(\boldsymbol{\theta}, \boldsymbol{\delta}) = \mathrm{E}_{\boldsymbol{\theta}}\, L(\boldsymbol{\theta}, \boldsymbol{\delta}(\boldsymbol{X}))$ is the risk (expected loss) of $\boldsymbol{\delta}$, and $\mathcal{D}(\boldsymbol{x})$ is an expression independent of $\boldsymbol{\theta}$. Typically, $\mathcal{D}(\boldsymbol{x})$ involves partial derivatives of $\boldsymbol{\Phi}$ for the continuous case and partial differences of $\boldsymbol{\Phi}$ for discrete case. Clearly, if one can find $\boldsymbol{\Phi}$ such that $\mathcal{D}(\boldsymbol{x}) < 0$ for all

$\boldsymbol{x}$, then the corresponding estimator $\boldsymbol{\delta}^*$ dominates $\boldsymbol{\delta}^0$.

Stein's technique, originally designed for normal families, has been generalized to many other situations. Many authors have since successfully solved a broad class of differential inequalities and difference inequalities. (See Hwang [7] and its references.) In Example 1, the inequality involved is a mixed type including both partial derivatives and partial differences. This is because we are dealing with a case in which two of the observations have continuous distributions and a third has a discrete one.

As mentioned above, the existence of a nontrivial solution to the inequality corresponding to an estimator $\boldsymbol{\delta}^0$, implies the inadmissibility of $\boldsymbol{\delta}^0$. Some statisticians therefore speculated that the lack of a nontrivial solution to the differential or difference inequality implies the admissibility of the corresponding estimator. Example 2 below shows that the statement is not true for the one dimensional discrete case and therefore is probably not true for higher dimensions. Another parallel example for the case when the observations are normally distributed was given in Brown [3]. While his inadmissible estimator, a generalized Bayes rule, is more appealing than ours, we note that our example is technically simpler, having arisen in a more natural context.

2. Two examples

Example 1. Assume that X_1, X_2 and X_3 are independent random variables: X_1 has a Poisson distribution with mean θ_1, X_2 has a normal distribution with mean θ_2 and variance 1, and X_3/θ_3 has a chi-square distribution with n degrees of freedom. We wish to estimate $(\theta_1, \theta_2, \theta_3)$ under the loss function

$$L_+(\boldsymbol{\theta}, \boldsymbol{a})=L_1(\theta_1, a_1)+L_2(\theta_2, a_2)+L_3(\theta_3, a_3)$$

where $L_1(\theta_1, a_1)=(\theta_1-a_1)^2$, $L_2(\theta_2, a_2)=(\theta_2-a_2)^2$ and $L_3(\theta_3, a_3)=\theta_3^{-1}(\theta_3-a_3)^2$. The standard estimator is

$$\delta^0(\boldsymbol{X})=(\delta_1^0(X_1), \delta_2^0(X_2), \delta_3^0(X_3))$$

where

$$\delta_i^0(X_i)=X_i\ , \qquad i=1, 2\ ,$$

and

$$\delta_3^0(X_3)=X_3/(n+2)\ .$$

It is known (see Hodges and Lehmann [5]) that $\delta_i^0(X_i)$ is an admissible estimator for θ_i under the loss function L_i. Note that $\delta_3^0(X_3)=X_3/(n+2)$ (rather than X_3/n) is also the best linear estimator under L_3 (or the squared error loss function) for estimating θ_3. However, a better estimator under L_+ can be constructed as shown in Theorem 1 below. Note that the loss function L_+ is chosen to simplify the following calculation. Results similar to what follows can be established for the sum of squared error loss function. In what follows, let

$$h_1(x_1)=\begin{cases}\sum_{k=1}^{x_1} k^{-1} & \text{if } x_1=1, 2, \cdots \\ 0 & \text{otherwise},\end{cases}$$

and for arbitrary real numbers x_2 and x_3,

$$h_2(x_2)=x_2\ , \qquad h_3(x_3)=-(n+2)^2/2x_3\ .$$

THEOREM 1. *Under L_+, $\boldsymbol{\delta}^0$ is dominated by an estimator $\boldsymbol{\delta}^*$ with components,*

$$\text{(1.1)} \qquad \begin{aligned} &\delta_i^*(\boldsymbol{X})=\delta_i^0(\boldsymbol{X})+\Phi_i(\boldsymbol{X}) \qquad i=1, 2\ , \\ &\delta_3^*(\boldsymbol{X})=\frac{X_3}{n+2}(1+\Phi_3(\boldsymbol{X})) \end{aligned}$$

where

$$\text{(1.2)} \qquad \begin{aligned} &\Phi_i(\boldsymbol{X})=-c(X_1)h_i(X_i)/D(\boldsymbol{X})\ , \\ &D(\boldsymbol{X})=h_1(X_1)h_1(X_1+1)+X_2^2+|h_3(X_3)|\ , \qquad \textit{and} \\ &c(X_1)=\begin{cases} 2/(n+2) & \text{if } x_1=1, 2, \cdots \\ 0 & \text{if } x_1=0\ . \end{cases} \end{aligned}$$

PROOF. By change of variables (as in equation (2.1) of Peng [9]) it can be shown that

$$\text{(1.3)} \qquad \mathrm{E}_{\boldsymbol{\theta}}\,[(\delta_1^*(\boldsymbol{X})-\theta_1)^2-(\delta_1^0(\boldsymbol{X})-\theta_1)^2]=\mathrm{E}_{\boldsymbol{\theta}}\,[2X_1\Delta_1\Phi_1(\boldsymbol{X})+\Phi_1^2(\boldsymbol{X})]$$

where for any function $F(\boldsymbol{x})$, $\Delta_1 F(\boldsymbol{x})=F(\boldsymbol{x})-F(\boldsymbol{x}-\boldsymbol{e}_1)$ and $\boldsymbol{e}_1=(1, 0, 0)$. Similarly, integration by parts can be used to show, as in equation (14) on page 6 of Stein [11], that

$$\text{(1.4)} \qquad \mathrm{E}_{\boldsymbol{\theta}}\,[(\delta_2^*(\boldsymbol{X})-\theta_2)^2-(\delta_2^0(\boldsymbol{X})-\theta_2)^2]=\mathrm{E}_{\boldsymbol{\theta}}\left[2\frac{\partial}{\partial x_2}\Phi_2(\boldsymbol{X})+\Phi_2^2(\boldsymbol{X})\right]\ ,$$

and, as in equation (2.2) of Berger [1], that

$$\text{(1.5)} \qquad \begin{aligned} &\frac{1}{\theta_3}\,\mathrm{E}_{\boldsymbol{\theta}}\,[(\delta_3^*(\boldsymbol{X})-\theta_3)^2-(\delta_3^0(\boldsymbol{X})-\theta_3)^2] \\ &\quad=\mathrm{E}_{\boldsymbol{\theta}}\left[\frac{4X_3^2}{(n+2)^2}\frac{\partial}{\partial X_3}\Phi_3(\boldsymbol{X})+\frac{X_3}{n+2}\Phi_3^2(\boldsymbol{X})\right. \end{aligned}$$

$$+\frac{4X_3^2}{(n+2)^2}\Phi_3(\boldsymbol{X})\frac{\partial}{\partial X_3}\Phi_3(\boldsymbol{X})\Big] .$$

Let R_+ denote the risk function with respect to L_+. Now by (1.3) through (1.5) and the fact $\Phi_3(\boldsymbol{x})(\partial\Phi_3(\boldsymbol{x})/\partial x_3)\leqq 0$, it is clear that

$$R_+(\boldsymbol{\theta}, \boldsymbol{\delta}^*)-R_+(\boldsymbol{\theta}, \boldsymbol{\delta}^0)\leqq 2\,\mathrm{E}_{\boldsymbol{\theta}}\,\mathscr{D}[\boldsymbol{\Phi}(\boldsymbol{X})] \tag{1.6}$$

where

$$\mathscr{D}[\boldsymbol{\Phi}(\boldsymbol{x})]=x_1\Delta_1\Phi_1(\boldsymbol{x})+\frac{\partial\Phi_2(\boldsymbol{x})}{\partial x_2}+\frac{2x_3^2}{(n+2)^2}\frac{\partial\Phi_3(\boldsymbol{x})}{\partial x_3} \tag{1.7}$$
$$+\frac{1}{2}\Phi_1^2(\boldsymbol{x})+\frac{1}{2}\Phi_2^2(\boldsymbol{x})+\frac{x_3}{2(n+2)}\Phi_3^2(\boldsymbol{x}) .$$

We next show that

$$\mathscr{D}[\boldsymbol{\Phi}(\boldsymbol{x})]\leqq 0 \qquad \text{and} \qquad \mathrm{E}_{\boldsymbol{\theta}}\,\mathscr{D}[\boldsymbol{\Phi}(\boldsymbol{X})]<0 , \tag{1.8}$$

which clearly completes the proof. By direct calculation,

$$\frac{\partial\Phi_2(\boldsymbol{x})}{\partial x_2}=c(x_1)\Big[\frac{-1}{D(\boldsymbol{x})}+\frac{2x_2^2}{D^2(\boldsymbol{x})}\Big] \tag{1.9}$$

$$\frac{2x_3^2}{(n+2)^2}\frac{\partial\Phi_3(\boldsymbol{x})}{\partial x_3}=c(x_1)\Big[\frac{-1}{D(\boldsymbol{x})}+\frac{|h_3(x_3)|}{D^2(\boldsymbol{x})}\Big]\leqq c(x_1)\Big[\frac{-1}{D(\boldsymbol{x})}+\frac{2|h_3(x_3)|}{D^2(\boldsymbol{x})}\Big] \tag{1.10}$$

and

$$x_1\Delta_1\Phi_1(\boldsymbol{x})\leqq x_1c(x_1)\Delta_1\Big[\frac{-h_1(x_1)}{D(\boldsymbol{x})}\Big] \tag{1.11}$$
$$=x_1c(x_1)\Big[\frac{-\Delta_1h_1(x_1)}{D(\boldsymbol{x})}+\frac{h_1(x_1-1)\Delta_1D(\boldsymbol{x})}{D(\boldsymbol{x})D(\boldsymbol{x}-\boldsymbol{e}_1)}\Big]$$
$$\leqq c(x_1)\Big[\frac{-1}{D(\boldsymbol{x})}+\frac{2h_1(x_1-1)h_1(x_1)}{D(\boldsymbol{x})D(\boldsymbol{x}-\boldsymbol{e}_1)}\Big] .$$

The last inequality is trivial if $x_1=0$, (for $c(x_1)$ is then zero). For $x_1>0$, it follows from the fact that $\Delta_1D(\boldsymbol{x})\leqq 2h_1(x_1)/x_1$. Hence (1.9), (1.10), and (1.11) together with the fact $D(\boldsymbol{x})\geqq D(\boldsymbol{x}-\boldsymbol{e}_1)$ imply

$$x_1\Delta_1\Phi_1(\boldsymbol{x})+\frac{\partial\Phi_2(\boldsymbol{x})}{\partial x_2}+\frac{2x_3^2}{(n+2)^2}\frac{\partial\Phi_3(\boldsymbol{x})}{\partial x_3}\leqq\frac{-c(x_1)}{D(\boldsymbol{x})} . \tag{1.12}$$

Meanwhile, it is clear that

$$\frac{1}{2}\Phi_1^2(\boldsymbol{x})+\frac{1}{2}\Phi_2^2(\boldsymbol{x})+\frac{x_3}{2(n+2)}\Phi_3^2(\boldsymbol{x})\leqq c^2(x_1)(n+2)/[4D(\boldsymbol{x})] . \tag{1.13}$$

Hence (1.7), (1.12), and (1.13) give

$$\mathscr{D}[\boldsymbol{\Phi}(\boldsymbol{x})] \leqq \frac{c(x_1)}{D(\boldsymbol{x})}\left(\frac{(n+2)c(x_1)}{4} - 1\right) = -c(x_1)/[2D(\boldsymbol{x})]$$

which establishes (1.8). Q.E.D.

Example 2. Let X be a one-dimensional random variable having logarithmic distribution, i.e.,

$$\mathrm{P}\,(X=x) = \frac{1}{-\log(1-\theta)}\,\frac{\theta^x}{x}\,, \qquad x=1, 2, \cdots,$$

where θ is some unknown parameter, $0<\theta<1$. It is clear that the unbiased estimator,

$$\delta^0(X) = \begin{cases} \dfrac{X}{X-1} & \text{if } X \geqq 2 \\ 0 & \text{if } X=1\,, \end{cases}$$

is inadmissible under the squared error loss function, since it estimates θ by some number greater than one, if $X \geqq 2$. Indeed $\delta^0(X)$ can certainly be improved if, for $X \geqq 2$, one estimates 1 (rather than $X/(X-1)$). However, we will consider the problem of improving upon δ^0 by solving a difference inequality derived by using Hudson's identity [6]. Again let $R(\theta, \delta)$ denote the risk function of δ with respect to the squared error loss function. One can derive (as in equation (2.9) of Hwang [7]) that

$$R(\theta, \delta^0+\Phi) - R(\theta, \delta^0) = \mathrm{E}_\theta\, \mathscr{D}[\Phi(X)]$$

$$\mathscr{D}[\Phi(x)] = 2v(x)[\Phi(x)-\Phi(x-1)] + \Phi^2(x)\,, \qquad \text{and} \tag{2.1}$$

$$v(x) = \begin{cases} x/(x-1) & x \geqq 2 \\ 0 & \text{otherwise}\,. \end{cases}$$

The following lemma will show that the only solution to $\mathscr{D}(\Phi) \leqq 0$ is $\Phi(x) \equiv 0$. Therefore the lack of a nontrivial solution to the difference inequality does not necessarily imply admissibility.

LEMMA 1. *If $\Phi(x)$ satisfies $\mathscr{D}[\Phi(x)] \leqq 0$, then $\Phi(x)=0$, $x=1, 2, \cdots$.*

PROOF. First, we will show that $\Phi(\cdot)$ is bounded. Clearly $\Phi(\cdot)$ is a nonincreasing function, since $\Phi(x)-\Phi(x-1)$ must be nonpositive. Also for $x=1$, $\mathscr{D}[\Phi(x)] \leqq 0$ implies $\Phi^2(1) \leqq 0$ or $\Phi(1)=0$. Thus $\Phi(x) \leqq 0$, for $x = 1, 2, \cdots$. Since $v(x) \leqq 2$ for all positive integer x, it follows that

$$0 \geqq \mathscr{D}[\Phi(x)] \geqq 4[\Phi(x)-\Phi(x-1)] + \Phi^2(x) \geqq 4\Phi(x) + \Phi^2(x) \tag{2.2}$$

which, since $\Phi(x) \leqq 0$, implies that

$$-4 \leqq \Phi(x) \leqq 0\,.$$

Next, let l be the limit of $\Phi(x)$ as $x \to \infty$. (This limit exists, since Φ is bounded and nonincreasing.) We then complete the proof by showing that $l=0$. Now clearly

$$-4 \leqq l \leqq 0 .$$

Suppose that $l<0$. Then there exists some $N>0$ such that $\Phi(x)<0$ for $x>N$. By (2.2),

$$\frac{4(\Phi(x)-\Phi(x-1))}{\Phi(x)}+\Phi(x) \geqq 0 , \tag{2.3}$$

whenever $x>N$. Letting x go to infinity, (2.3) implies that $l \geqq 0$, which is a contradiction. Q.E.D.

Acknowledgement

I would like to express my deepest thanks to Professor James O. Berger for his suggestions and helpful discussions in the preparation of this paper.

CORNELL UNIVERSITY

REFERENCES

[1] Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with application to simultaneous estimation of gamma scale parameters, *Ann. Statist.*, 8, 545-571.

[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. (1979). Counterexample—An inadmissible estimator which is generalized Bayes for a prior with "light" tails, *J. Multivariate Anal.*, **6**, 256-264.

[4] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach, *J. Amer. Statist. Ass.*, **68**, 117-130.

[5] Hodges, J. L., Jr. and Lehman, E. L. (1951). Some applications of the Cramer-Rao inequality, *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, 13-22.

[6] Hudson, M. (1974). Empirical Bayes estimation, *Technical Report* No. 58, Department of Statistics, Stanford University.

[7] Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and Negative Binomial cases, *Ann. Statist.*, **10**, 857-867.

[8] James, W. and Stein, C. (1960). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, **1**, Univ. of California Press, 361-379.

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

[10] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc. Third Berkeley Symp. Math. Statist. Prob.*, **1**, 197-206.

[11] Stein, C. (1973). Estimation of the mean of a multivariate distribution, *Proc. Prague Symp. Asymptotic Statist.*, 345-381.