# Two Techniques of Integration By Parts and Some Applications

by

Christian Robert
Purdue University

Technical Report #87-52

Department of Statistics
Purdue University

February 1988
Revised Version July 1988
Second Revised Version November 1988

*

# Two Techniques of Integration By Parts and Some Applications

by

Christian Robert

Cornell University and Université de Rouen*

## Abstract

Integration by part is the main tool used to establish minimaxity in shrinkage estimation. In this paper, we extend it into two directions. First, we give a generalization of Stein's lemmae (1973) to non continuous functions; this result is obtained by the use of a very powerful analysis method, the *distribution theory.* We apply then the generalized lemmae to a class of discontinuous shrinkage estimators of a normal mean vector and derive a necessary condition of admissibility, generalizing Hwang's result (1982a) to the case when the variance is known up to a multiplicative factor.

Secondly, we propose a notion of integration by part with respect to the parameters in the general exponential family and we obtain some sufficient conditions of risk domination for estimators of the mean. These conditions can be expressed as partial differential inequalities; we give some applications in the gamma and normal cases.

**Key–words** Exponential family, shrinkage estimators, quadratic loss, distribution theory, admissibility, risk domination.

**AMS Classification (1980)** 62J07, 62F10, 62H25, 62C20, 62P20.

---

* Research supported by l'Institut National de la Statistique et des Etudes Economiques (Paris) and by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University. This work was performed while the author was visiting Purdue University.

**Two Techniques of Integration by Parts and Some Applications**

## 1. Introduction.

Integration by parts is the most widely used technique in shrinkage estimation. Introduced by Stein (1973), it has been generalized by Berger (1975) to spherically symmetric distributions and Stein (1981) to 'vectorial' shrinkage estimators. Concerning a very general class of distributions, Shinozaki's proof of minimaxity (1984) is based upon repeated integrations by parts (see also Akai (1986)). In the normal case, Gleser (1986) has also proposed some integrations by part techniques when the variance is unknown, generalizing Berger and Haff's results (1983). Integration by parts often leads to the notion of 'unbiased estimator of the risk' (see, e.g., Berger (1985), §5.4.2) which has revealed itself to be a very powerful tool to get sufficient conditions of minimaxity; for some distributions (Rukhin (1988)) or some classes of estimators (Brown (1987), Bock (1987)), it is yet useless.

One drawback of this technique is that the shrinkage function must be 'a.e. differentiable' (see Stein (1981)): it excludes discontinuous functions and therefore *pre-test* estimators (see Judge and Bock (1978)). In order to allow some discontinuities of the shrinkage functions, Cellier and Fourdrinier (1985) have used *Stieltjes* integration but their method does not produce compact expressions and, further, a Stieltjes integrability condition must be satisfied by the shrinkage function. The generalization we propose in Section 2 includes a wide class of shrinkage functions for which the usual notion of differentiability does not make sense. It has been first used by Hwang (1982a) in a special case to establish necessary conditions of admissibility.

In Section 3, we use this generalized integration by part to deduce necessary conditions of admissibility for shrinkage estimators of a normal mean vector when the variance matrix is known up to a multiplicative factor. The class of shrinkage estimators we consider is the class obtained in Fraisse, Robert and Roy (1987) and Proposition 3.5 is the generalization of Hwang (1982a) to our model. This condition allows to exclude wide classes of inadmissible estimators by defining a 'reasonable' amount of shrinkage. DasGupta (1984) gives a similar result for the gamma distribution.

Another drawback of the usual integration by parts techniques is that they work only for continuous families. Hwang (1982b) has developed an alternative method to deal with the estimation of natural parameters of discrete exponential families; Haff and Johnson (1986) have proposed, for some continuous exponential families, a generalization of Stein (1981). We propose in Section 4 a new way to get sufficient risk domination conditions for estimators of the *mean* for general exponential families (discrete and continuous); they are expressed through differential inequalities. Very few results have been obtained for this problem (see Brown (1985)).

## 2. Distribution theory.

This theory, which has no direct relationship with probability theory, has been introduced by Schwartz (1966) in order to formalize some tools used mainly by physicists; with respect to our problem, its main interest is that it generalizes the notion of differentiability to discontinuous functions.

If $(\mathcal{D})$ is the vector space of functions from $\mathbb{R}^n$ into $\mathbb{R}$ which are infinitely differentiable and have a compact support, a *distribution* $T$ is a real–valued linear functional on $(\mathcal{D})$ which is 'continuous' in the following way: *if* $(\varphi_j) \in (\mathcal{D})$ *is a sequence of functions whose supports are included into a common compact set and which uniformly converges to 0 in* $(\mathcal{D})$, *then* $T \cdot (\varphi_j)$ *is converging to 0.* Distributions are also called *'generalized functions'.*

If $f$ is a (Lebesgue) measurable function on $\mathbb{R}^n$, it is a particular distribution, defined by

$$f \cdot (\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)dx \qquad \forall \varphi \epsilon (\mathcal{D})$$

In the same way, a measure $\mu$ on $\mathbb{R}^n$ defines a distribution. But there exist also distributions which are not functions, nor even measures. One well-known example is the *Dirac derivative* defined by (for $n = 1$)

$$T \cdot (\varphi) = \varphi'(0) \qquad \text{for every } \varphi \epsilon (\mathcal{D}),$$

where $\varphi'$ is the derivative of $\varphi$.

Given a distribution $T$, the *derivative*, $\frac{\partial T}{\partial x_i}$, of $T$ is defined by

$$\left(\frac{\partial T}{\partial x_i}\right) \cdot (\varphi) = -T \cdot \left(\frac{\partial \varphi}{\partial x_i}\right) \text{ for every } \varphi \epsilon (\mathcal{D}) \tag{2.1}$$

Note that, if $f$ is a differentiable function, the derivative of the distribution defined by $f$ and the distribution defined by $\frac{\partial f}{\partial x_i}$ are the same:

$$\int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(x)\varphi(x)dx = -\int_{\mathbb{R}^n} f(x)\frac{\partial \varphi}{\partial x_i}(x)dx = -(f)\cdot\left(\frac{\partial \varphi}{\partial x_i}\right),$$

because $f$ and $\varphi$ are continuous and $\varphi$ is null outside a compact set. Therefore integration by part is the basis of this definition (Schwartz (1966, p. 35)).

Another useful notion introduced by Schwartz (1966, p. 116) is the *multiplication of two distributions*, which cannot be defined for two arbitrary distributions. For the product $ST$ to be defined, the more irregular $T$ is, the more regular $S$ has to be. If $S\epsilon(\mathcal{D})$, $ST$ is defined for every $T$ by

$$(ST)(\varphi) = T(S\varphi) \quad \textit{for every } \varphi\epsilon(\mathcal{D}).$$

Let $(\mathcal{D}^m)$ be the vector space of functions on $\mathbb{R}^n$ with compact support which have derivatives of order at least $m$. A distribution will be said of *order* $\leq m$ if it is defined on $(\mathcal{D}^m)$ and not on $(\mathcal{D}^{m-1})$. We have the following result (Schwartz (1966, p. 118)):

**lemma 2.0.** *If $T$ is a distribution of order $\leq m$, $S$ must be an $m$ continuously differentiable function in order for the distribution $TS$ to be defined. In particular, if $T$ is a measure $\mu$, $S$ has only to be a continuous function.*

When the product $ST$ is defined, we have the following rule of derivation (Schwartz (1966, p. 120)):

$$\frac{\partial}{\partial x_i}(ST) = \left(\frac{\partial}{\partial x_i}S\right)T + S\left(\frac{\partial}{\partial x_i}T\right) \tag{2.2}$$

The *Dirac distribution* in $a\,\epsilon\mathbb{R}^n$, $\delta_a$, is defined by $\delta_a(\varphi) = \varphi(a)$ for every function $\varphi$ defined in $a$. Then

**lemma 2.1.** *If $(a,b)\epsilon\mathbb{R}^2$ and $\mathbb{I}_{[a,b]}$ is the indicator function of $[a,b]$ (i.e. $\mathbb{I}_{[a,b]}(x) = 0$ if $x \notin [a,b]$, 1 if $x\epsilon[a,b]$), the derivative of $\mathbb{I}_{[a,b]}$ is $(\delta_a - \delta_b)$.*

In fact, we need more and less than what was established for the original theory of distributions. We will observe the image of the multivariate normal density by some distributions; this function is infinitely differentiable but does not have a compact support. On the other hand, we will only consider distributions deduced from functions. Let us denote $p_{\theta,\sigma}$ the density function of the normal distribution $N_n(\theta, \sigma^2 I_n)$ and $\mathbb{E}_{\theta,\sigma}$ the *expectation*:

$$\mathbb{E}_{\theta,\sigma}(g(x)) = \int_{\mathbb{R}^n} g(x) p_{\theta,\sigma}(x) dx$$

We will then consider $(\mathcal{E}^m)$ to be the set of functions on $\mathbb{R}^n$ such that

$$(g) \cdot (p_{\theta,\sigma}), (g) \cdot (\frac{\partial}{\partial x_i} p_{\theta,\sigma}), \dots, (g) \cdot \ (\frac{\partial^m}{\partial x_{i_1} \dots \partial x_{i_m}} p_{\theta,\sigma})$$

are defined. In particular, these functions must satisfy to:

$$\lim_{y_i \to \pm\infty} g(y_1, \dots, y_n) ||y_i||^m p_{\theta,\sigma}(y_1, \dots, y_n) = 0 \ for\ almost\ every\ (y_1, ..., y_{i-1}, y_{i+1}, \dots, y_n)$$

Usually we only have to work on $(\mathcal{E}^1)$.

With these restrictions, we can generalize the lemmas of Stein (1973) in the following way:

**lemma 2.2.** *(i) Let $g$ be a function in $(\mathcal{E}^1)$. Then, if $y \sim N_n(\theta, \sigma^2 I_n)$,*

$$\frac{1}{\sigma^2} \mathbb{E}_{\theta,\sigma}[(y_i - \theta_i) g(y)] = \left( \frac{\partial g}{\partial y_i} \right) \cdot (p_{\theta,\sigma}) \qquad (1 \leq i \leq n)$$

*(ii) Let $h$ be a real function and $K \sim \sigma^2 \chi_p^2$. Then*

$$\frac{1}{\sigma^2} \mathbb{E}^K[K h(K)] = p \mathbb{E}^K[h(K)] + 2(h') \cdot \left( K \phi_p \left( \frac{K}{\sigma^2} \right) \right)$$

*where $\phi_p$ is the density of the distribution $\chi_p^2$.*

**Proof.** (i) We have

$$\begin{aligned} \frac{1}{\sigma^2} \mathbb{E}_{\theta,\sigma}((y_i - \theta_i) g(y)) &= \int_{\mathbb{R}^n} g(y) \frac{y_i - \theta_i}{\sigma^2} p_{\theta,\sigma}(y) dy \\ &= \int_{\mathbb{R}^n} g(y) \left( - \frac{\partial}{\partial y_i} p_{\theta,\sigma}(y) \right) dy \\ &= \left( \frac{\partial g}{\partial y_i} \right) \cdot (p_{\theta,\sigma}) \qquad \text{by definition.} \end{aligned}$$

(ii) Given that $2(r\phi_p(\frac{r}{\sigma^2}))' = (-\frac{r}{\sigma^2} + p)\phi_p(\frac{r}{\sigma^2})$,

$$\begin{aligned} \frac{1}{\sigma^2} \mathbb{E}^K[K h(K)] &= p \mathbb{E}^K(h(K)) - 2 \int_0^\infty h(r) \left( r \phi_p \left( \frac{r}{\sigma^2} \right) \right)' dr \\ &= p \mathbb{E}^K(h(K)) + 2(h') \cdot \left( K \phi_p \left( \frac{K}{\sigma^2} \right) \right) \end{aligned}$$ □

We will now apply this result to discontinuous shrinkage functions. Hwang and Casella (1982) also present in their techniques of proof another application of the distribution theory.

## 3. A consequence.

**3.1. Model.** Let $y \in \mathbb{R}^k$ be an observation of a normal vector with unknown mean $\theta$ and covariance matrix $\sigma^2 I_k$, where $\sigma^2$ is unknown. We assume that an estimator of $\sigma^2, s^2$, is available; it is independent of $y$ and follows a $\sigma^2\chi^2_{n-k}$ distribution (note that the usual model of *linear regression* can always be written in this form).

Estimators of $\theta$ are compared with respect to the usual quadratic loss

$$L(\delta,\theta) = \sigma^{-2}\sum_{i=1}^{k}(\delta_i - \theta_i)^2$$

Since Stein (1956), who has established the inadmissibility of the usual least squares estimator, numerous papers have dealt with this model. We consider estimators of the form

$$\varphi(y,s^2) = y - h(y^tBy,s^2)By \tag{3.1}$$

where $B$ is a non-negative definite symmetric matrix and $h$ is a function from $\mathbb{R}_+ \times \mathbb{R}_+$ into $\mathbb{R}_+$. Fraisse *et al.* have established that, in the class of estimators

$$\varphi(y,s^2) = y - h(y^tBy,s^2)Cy,$$

where $C$ is a $k\times k$ matrix, admissible estimators are necessarily of the form (3.1). Further, Cellier, Fourdrinier and Robert (1987) have established the following sufficient condition of minimaxity.

**Proposition 3.1** *If*

(a) *there exists* $\lambda_1 \geq 1$ *such that* $t^{\lambda_1}h(t,u)$ *is non-decreasing for every* $u > 0$,

(b) *there exists* $\lambda_2 \geq 1$ *such that* $u^{\lambda_2}h(t,u)$ *is non-increasing for every* $t > 0$,

(c) *for every* $(t,u)$,

$$\frac{t}{u}\,h(t,u) \leq 2\,\frac{tr(B) - 2\lambda_1 ch_{max}(B)}{ch_{max}(B)}\;\frac{1}{n-k+2\lambda_2},$$

*where tr and $ch_{max}$ denote respectively the trace and the maximum eigenvalue of the considered matrix, $\varphi$ is uniformly (in $(\theta, \sigma)$) better than the least squares estimator.*

Note that, for this model, the least squares estimator is $\varphi^0(y, s^2) = y$.

**3.2. Admissibility conditions.** Within the extensive literature on shrinkage estimation, few papers have considered admissibility problems for this model because the estimated parameters are not *natural parameters.* However, a very general lemma of Hwang (1982a) also applies in our case:

**lemma 3.2.** *If $\varphi_1$ and $\varphi_2$ are two estimators of $\theta$ such that $\varphi_2$ dominates $\varphi_1$ for the usual quadratic risk and if*

$$d(y, s^2) = \varphi_2(y, s^2) - \varphi_1(y, s^2),$$

*any estimator $\delta$ satisfying*

$$d(y, s^2)^t \delta(y, s^2) \le d(y, s^2)^t \varphi_1(y, s^2) \tag{3.2}$$

*for every $(y, s^2)$ is inadmissible.*

A first application of this lemma is the following result.

**Proposition 3.3.** *If there exists $\alpha$ such that, for all $(t, u) \in \mathbb{R}_+ \times \mathbb{R}_+$,*

$$(a) \qquad \frac{t}{u}\, h(t, u) \le \alpha < \frac{k-2}{n-k+2}$$

*or*

$$(b) \qquad \frac{t}{u}\, h(t, u) \ge \alpha > \frac{k-2}{n-k+2},$$

*$\varphi$ is inadmissible in the class (3.1).*

**Proof.** Let $\varphi_1(y, s^2) = (1 - \frac{\alpha s^2}{||y||^2})y$ and $\varphi_2(y, s^2) = (1 - \frac{\alpha^* s^2}{||y||^2})y$, where $\alpha^* = \frac{k-2}{n-k+2}$. It is known that $\varphi_2$ dominates $\varphi_1$ for any $\alpha$ (James and Stein (1961)). Then

$$d(y, s^2) = (\alpha - \alpha^*) \frac{s^2}{||y||^2} y$$

and

$$\varphi(y, s^2)^t d(y, s^2) \le \varphi_1(y, s^2)^t d(y, s^2)$$

if

$$(\alpha-\alpha^*)h(y^tBy,s^2)\frac{s^2}{||y||^2}y^tBy \geq (\alpha-\alpha^*)\alpha\,\frac{(s^2)^2}{||y||^2}\ ;$$

(a) if $\alpha < \alpha^*$, the condition becomes $\frac{y^tBy}{s^2}h(y^tBy,s^2) \leq \alpha\ < \frac{k-2}{n-k+2}$;

(b) if $\alpha > \alpha^*$, we get $\frac{y^tBy}{s^2}h(y^tBy,s^2) \geq \alpha > \frac{k-2}{n-k+2}$ .

A dominating estimator in the class (3.1) can be deduced from Hwang (1982a). □

We deduce immediately from Proposition 3.3 (a).

**Corollary 3.4.** *A necessary condition of admissibility for estimators satisfying the conditions of Proposition 3.1 is*

$$tr(B) > \frac{k+2}{2}\,ch_{max}(B).$$

This result is interesting because it is stronger than the usual necessary condition of minimaxity (Brown (1975))

$$tr(B) \geq 2ch_{max}(B)$$

and thus allows to eliminate a larger class of matrices $B$ if one is interested in estimators which perform well on the average (*minimaxity*) and at a particular spot (*admissibility*). In particular, estimators which shrink towards projections over subspaces of too large dimension are excluded by this condition.

Using lemma 2.2., we generalize now the results of Brown (1971) and Hwang (1982a) and obtain a sufficient condition of inadmissibility weaker than in Proposition 3.3.

**Proposition 3.5.** *If there exist* $\alpha$, $M_1 > 0$ *and* $M_2 > 0$ *such that*

*(a) for all* $(t,u) \in [M_1,+\infty) \times (0,M_2]$,

$$\frac{t}{u}\ h(t,u) \leq \alpha < \frac{k-2}{n-k+2}$$

*or*

*(b) for all* $(t,u) \in [0,M_1] \times [M_2,+\infty)$,

$$\frac{t}{u}\ h(t,u) \geq \alpha > \frac{k-2}{n-k+2},$$

$\varphi$ *is not admissible in the class (3.1).*

**Proof.** The proof is a generalization of the proof of Hwang (1982,a): we first establish the domination of a truncated James-Stein estimator over a given class and then apply lemma 3.2. The dominating estimator in the class (3.1) can again be deduced from Hwang(1982a). The estimators we consider are

$$\varphi_c(y,s^2) = y - \frac{cs^2}{||y||^2}\, \mathbb{I}_A(y,s^2)y \qquad (c \geq 0) \tag{3.3}$$

where $A$ is $[K_1,+\infty)^k \times [0,M_2]$ or $[0,K_1]^k \times [M_2,+\infty)$ with $K_1 > 0$. We will prove that, if

$$c^* = \frac{k-2}{n-k+2},$$

$\varphi_{c^*}$ is 'optimal' in a way defined below.

The difference between the risk of the least squares estimator and the risk of $\varphi_c$ is

$$\begin{aligned}\Delta_c &= \frac{1}{\sigma^2}\, \mathbb{E}_{\theta,\sigma}(||y-\theta||^2) - \frac{1}{\sigma^2}\, \mathbb{E}_{\theta,\sigma}(||\varphi_c(y,s^2)-\theta||^2) \\ &= \frac{2}{\sigma^2}\, \mathbb{E}_{\theta,\sigma}[(y-\theta)^t y\, \frac{cs^2}{||y||^2}\, \mathbb{I}_A(y,s^2)] - \frac{1}{\sigma^2}\, \mathbb{E}_{\theta,\sigma}[||y||^2\, \frac{c^2s^4}{||y||^4}\, \mathbb{I}_A(y,s^2)]\end{aligned} \tag{3.4}$$

From lemma 2.2 (i), we deduce

$$\begin{aligned}\mathbb{E}_{\theta,\sigma}\, \left[(\frac{y-\theta}{\sigma^2})^t\, y\, \frac{cs^2}{||y||^2}\, \mathbb{I}_A\, (y,s^2)\right] &= \sum_{i=1}^k\, \mathbb{E}_{\theta,\sigma}\left(\frac{(y_i-\theta_i)}{\sigma^2}\, y_i \frac{cs^2}{||y||^2}\, \mathbb{I}_A\, (y,s^2)\right) \\ &= \sum_{i=1}^k\, \left(\frac{\partial}{\partial y_i}\, [y_i\, \frac{cs^2}{||y||^2}\, \mathbb{I}_A\, (y,s^2)]\right)\cdot(p_{\theta,\sigma})\end{aligned}$$

The function $g_i(y,s^2) = y_i \frac{s^2}{||y||^2}$ being smooth enough, we can deduce from (2.2) that

$$\frac{1}{\sigma^2}\, \mathbb{E}_{\theta,\sigma}\, \left[(y-\theta)^t y\, \frac{cs^2}{||y||^2}\, \mathbb{I}_A(y,s^2)\right] =$$
$$c\, \left(\sum_{i=1}^k \left(\frac{\partial}{\partial y_i} g_i\right)\cdot(\mathbb{I}_A p_{\theta,\sigma}) + \sum_{i=1}^k \left(\frac{\partial}{\partial y_i}\mathbb{I}_A\right)\cdot(g_i p_{\theta,\sigma})\right)$$

As $g_i$ is differentiable, we have

$$\left(\frac{\partial g_i}{\partial y_i}\right)\cdot(\mathbb{I}_A p_{\theta,\sigma}) = \int_{\mathbb{R}^k\times\mathbb{R}_+} \frac{\partial g_i}{\partial y_i}(y,s^2)\mathbb{I}_A(y,s^2)p_{\theta,\sigma}(y,s^2)dyds^2$$

Therefore

$$c\sum_{i=1}^{k}\left(\frac{\partial g_i}{\partial y_i}\right)\cdot(\mathbb{I}_A p_{\theta,\sigma}) = \mathbb{E}_{\theta,\sigma}\left[\sum_{i=1}^{k}\left(1-\frac{2y_i^2}{||y||^2}\right)\frac{cs^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right]$$
$$= \mathbb{E}_{\theta,\sigma}\left[(k-2)\,\frac{cs^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right],$$

which is a term we usually obtain in the decomposition of the risk.

Consider now the second term in (3.4): lemma 2.2 (ii) implies

$$\mathbb{E}_{\theta,\sigma}\left[\frac{s^2}{\sigma^2}\,\frac{c^2s^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right] = (n-k)\,\mathbb{E}_{\theta,\sigma}\left[\frac{c^2s^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right]$$
$$+2\left(\frac{\partial}{\partial s^2}\left[\frac{c^2s^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right]\right)\cdot(s^2p_{\theta,\sigma}(y,s^2))$$
$$= (n-k+2)\,\mathbb{E}_{\theta,\sigma}\left[\frac{c^2s^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right]+2\left(\frac{\partial}{\partial s^2}\mathbb{I}_A\right)\cdot\left(\frac{c^2s^4}{||y||^2}\,p_{\theta,\sigma}(y,s^2)\right).$$

Therefore the difference of the risks can be written

$$\Delta_c = \mathbb{E}_{\theta,\sigma}\left[\frac{s^2}{||y||^2}\,\mathbb{I}_A(y,s^2)\right](2(k-2)-c(n-k+2))c$$
$$+2c\sum_{i=1}^{k}\left(\frac{\partial}{\partial y_i}\mathbb{I}_A\right)\cdot(g_ip_{\theta,\sigma})-2c^2\left(\frac{\partial}{\partial s^2}\mathbb{I}_A\right)\cdot(g_{k+1}p_{\theta,\sigma}),$$

where $g_{k+1}(y,s^2)=\frac{s^4}{||y||^2}$. The first part of this sum is maximized for $c=c^*$. Therefore, to establish optimality of $\varphi_c^*$, we have only to show

$$\text{(3.5)}\qquad (c-c^*)\left\{\sum_{i=1}^{k}\left(\frac{\partial}{\partial y_i}\mathbb{I}_A\right)\cdot(g_ip_{\theta,\sigma})-\left(\frac{\partial}{\partial s^2}\mathbb{I}_A\right)\cdot(g_{k+1}p_{\theta,\sigma})(c+c^*)\right\}\le 0$$

(a) *If* $c\le c^*$ *and* $A=[K_1,+\infty)^k\times[0,M_2]$, we have

$$\left(\frac{\partial}{\partial y_i}\mathbb{I}_A\right)\cdot(g_ip_{\theta,\sigma})=$$
$$-\int_{[K_1,+\infty)^{k-1}\times[0,M_2]}\left(-y_i\frac{s^2}{||y||^2}p_{\theta,\sigma}(y,s^2)\right)|_{y_i=K_1}dy_1\ldots dy_{i-1}dy_{i+1}\ldots dy_kds^2=$$
$$K_1\int_{[K_1,+\infty)^{k-1}\times[0,M_2]}\left(\frac{s^2}{||y||^2}p_{\theta,\sigma}(y,s^2)\right)|_{y_i=K_1}dy_1\ldots dy_{i-1}dy_{i+1}\ldots dy_kds^2\ge 0,$$

following from lemma 2.1. And

$$\left(\frac{\partial}{\partial s^2}\,\mathbb{I}_A\right)\cdot(g_{k+1}p_{\theta,\sigma}) = -\int_{[K_1,+\infty)^k}\left(\frac{s^4}{||y||^2}\,p_{\theta,\sigma}(y,s^2)\right)|_{s^2=M_2}dy_1\dots dy_k$$
$$= -M_2^2\int_{[K_1,+\infty)^k}\frac{1}{||y||^2}\,p_{\theta,\sigma}(y,M_2)dy_1\dots dy_k \le 0.$$

Therefore (3.5) is satisfied. For this choice of $A$, $\varphi_c^*$ dominates $\varphi_c$ for every $c < c^*$.

If we apply now lemma 3.2., we have

$$d(y,s^2) = \varphi_c^*(y,s^2) - \varphi_c(y,s^2)$$
$$= (c-c^*)\frac{s^2}{||y||^2}\mathbb{I}_A(y,s^2)y$$

and, for $\varphi$ defined by (3.1),

$$d(y,s^2)^t\varphi(y,s^2) = (c-c^*)\frac{s^2}{||y||^2}\mathbb{I}_A(y,s^2)(||y||^2 - h(y^tBy,s^2)y^tBy)$$

with

$$d(y,s^2)^t\varphi_c(y,s^2) = (c-c^*)\frac{s^2}{||y||^2}\mathbb{I}_A(y,s^2)(||y||^2 - \frac{cs^2}{||y||^2}\mathbb{I}_A(y,s^2)||y||^2]$$

As $c < c^*$, the inequality (3.2) is equivalent to

$$\frac{y^tBy}{s^2}h(y^tBy,s^2) \le c \qquad \text{for every } (y,s^2)\in A. \tag{3.6}$$

The part (a) of Proposition 3.5. follows from the fact that, if $y \in [K_1,+\infty)^k$,

$$y^tBy \ge ch_{min}(B)kK_1^2 = M_1.$$

Therefore, if $\frac{t}{u}h(t,u) \le c$ for every $(t,u) \in [M_1,+\infty)\times[0,M_2]$, (3.6) is satisfied for $A = [K_1,+\infty)^k \times [0,M_2]$.

(b) *If $c > c^*$ and $A = [0,K_1]^k \times [M_2,+\infty)$*, the 'opposite' choice of $A$ implies that (3.5) is satisfied. And a proof similar to the previous one gives the part (b) of Proposition 3.5 as $y \in [0,K_1)^k$ implies

$$y^tBy \le ch_{max}(B)kK_1^2 = M_1.$$

□

**Remark 1.** We can see from this result that the original James-Stein estimator gives a 'minimal amount of shrinkage' for large values of $\frac{t}{u}$ and a 'maximal amount' for small values of $\frac{t}{u}$. As DasGupta (1984) points out, this estimator *"stands as a dividing line between admissible and inadmissible estimators"*. However, it is not admissible (because it is dominated by its truncated counterpart).

**Remark 2.** (3.3) *is not* the class of positive part James-Stein estimators. However, for $A = [M_1, +\infty)^k \times [0, M_2)$, $\varphi_c^*$ still dominates the least squares estimators ($c = 0$). This class appears as a direct generalization of the estimators introduced in Example 3 of Baranchik (1970).

**Remark 3.** The conditions of Proposition 3.5 may appear rather counter–intuitive because, in the class of positive–part James–Stein estimators, for small values of $||\theta||$, the estimator associated with $2(k-2)$ dominates all the other ones and, for large values of $||\theta||$, the estimator associated with $(k-2)$ is the best one. However those estimators are not admissible (see e.g. Brown (1985)).

**Remark 4.** The second part of the result can be deduced more directly from Judge and Bock (1978). In fact, if $h$ satisfies to this inequality, the shrinkage factor will become negative for $\frac{t}{u}$ small enough and the associated estimator will be dominated by its positive–part.

## 4. A general differential inequality for the exponential family.

**4.1.** As it is pointed out in the introduction, usual integration by parts techniques only apply to continuous distributions. We propose here a new type of integration by part which covers discrete and continuous exponential families of distributions. It gives a sufficient condition of domination of an 'usual' estimator; this condition is expressed as a partial differential inequality (see (4.2) or (4.5)).

Let $y$ be a random vector in $\mathbb{R}^k$ with density $f(y|\theta) = \psi(\theta)e^{\theta^t y}$ w.r.t. a $\sigma$-finite measure $\nu$; we assume that $\theta$ belongs to the *natural parameter set*, i.e.

$$C = \left\{ \int e^{\theta^t y} d\nu(y) < +\infty \right\} \subset \mathbb{R}^k.$$

This class of distributions is called the *exponential family* (the measure $\nu$ allows to include both discrete and continuous distributions in our model) and is rather exhaustively studied in Brown (1985). A lot of papers deal with the problem of the estimation of the natural parameter $\theta$, in continuous (Berger (1980), Chen (1983), Ghosh, Hwang and Tsui (1983), Berger and DasGupta (1986)) and discrete (Hwang (1982b)) cases. We are mainly concerned in this paper with the estimation of the mean of the observation and the comparison with the estimator $\varphi_0(y) = y$. A few papers actually consider the estimation of the mean (Brown and Hwang (1982), Brown and Farrell (1985), DasGupta and Sinha (1986)) for other distributions than the normal distribution; their main concern is admissibility, not risk domination.

**4.2.** A first class of estimators we consider is

$$\varphi_c(y) = (1 - ch(y))y \tag{4.1}$$

where $c \in \mathbb{R}$ and $0 \le h(\cdot) \le 1$. It contains both shrinkage estimators ($c \ge 0$) which perform well in the normal case and 'expanders' ($c < 0$) which can be of interest for the gamma distribution (see Berger (1980), DasGupta (1984)).

For a function $h$, defined on $\mathbb{R}^k$, we define the *generalized Laplace transform of h w.r.t. $\nu$* to be

$$\tilde{h}(\theta) = \int_{\mathbb{R}^k} e^{\theta^t x} h(x) d\nu(x)$$

(see Brown (1971) and also Berger and Srinivasan (1978), Brown (1985)). We have then $\phi(\theta) = \mathbb{E}_\theta(h(x)) = \psi(\theta)\tilde{h}(\theta)$ .

For a function $h$ from $\mathbb{R}^k$ into $\mathbb{R}$, we use the notations

$$\nabla h(y) = \left(\frac{\partial h}{\partial y_1}(y), \ldots, \frac{\partial h}{\partial y_k}(y)\right)^t \qquad \text{and} \qquad \Delta h(y) = \sum_{i=1}^k \frac{\partial^2}{\partial y_i^2} h(y).$$

Thus

**Proposition 4.1.** *If, for every $\theta \in C$,*

$$(2-c)c(\Delta\phi \cdot \psi - \phi \cdot \Delta\psi) + 2c(1-c)\left(\frac{\nabla\psi}{\psi}\right)^t (\nabla\psi \cdot \phi - \psi\nabla\phi) \ge 0, \tag{4.2}$$

*$\varphi_c$ dominates $\varphi_0$ (for the usual quadratic loss).*

Before establishing this result, let us recall the following well-known identities (see e.g. Brown (1985)).

**lemma 4.2.** *If $x$ has density $\psi(\theta)e^{\theta\cdot x}$ w.r.t. $\nu$,*

$$(i)\quad \mathbb{E}_\theta(x) = -\nabla Ln\psi$$

$$(ii)\quad \mathbb{E}_\theta(||x+\nabla Ln\psi||^2) = -\Delta Ln\psi.$$

**Proof.** The difference between the risk of $\varphi_0$ and the risk of $\varphi_c$ is

$$2c\mathbb{E}_\theta[h(y)y^t(y+\nabla Ln\psi(\theta))] - c^2\mathbb{E}_\theta[h^2(y)||y||^2]$$

We have

$$\begin{aligned}
\mathbb{E}_\theta[h(y)y_i\left(y_i+\frac{\partial}{\partial\theta_i}Ln\psi(\theta)\right)] &= \int_{\mathbb{R}^k} h(y)y_i\left(y_i+\frac{\partial}{\partial\theta_i}Ln\psi(\theta)\right)e^{\theta^t y+Ln\psi(\theta)}d\nu(y)\\
&= \frac{\partial}{\partial\theta_i}\int_{\mathbb{R}^k} h(y)y_i e^{\theta^t y+Ln\psi(\theta)}d\nu(y)\\
&= \frac{\partial}{\partial\theta_i}\mathbb{E}_\theta(h(y)y_i)
\end{aligned}$$

In the same way,

$$\mathbb{E}_\theta(h(y)y_i) = \frac{\partial}{\partial\theta_i}\,\mathbb{E}_\theta(h(y)) - \frac{\partial}{\partial\theta_i}(Ln\ \psi\ (\theta))\ \mathbb{E}_\theta(h(y))$$

Furthermore,

$$\mathbb{E}_\theta(h^2(y)||y||^2) \le \mathbb{E}_\theta(h(y)||y||^2)$$

and

$$\mathbb{E}_\theta(h(y)y_i^2) = \frac{\partial}{\partial\theta_i}(\mathbb{E}_\theta(h(y)y_i)) - \frac{\partial}{\partial\theta_i}(Ln\psi(\theta))\mathbb{E}_\theta(h(y)y_i)$$

Therefore

$$\mathbb{E}_\theta(||y+\nabla Ln\psi(\theta)||^2) - \mathbb{E}_\theta(||\varphi_c(y)+\nabla Ln\psi(\theta)||^2)$$

$$\ge 2c\sum_{i=1}^k\left\{\frac{\partial}{\partial\theta_i}\mathbb{E}_\theta(h(y)y_i)\right\} - c^2\sum_{i=1}^k\left\{\frac{\partial}{\partial\theta_i}(\mathbb{E}_\theta(h(y)y_i)) - \frac{\partial}{\partial\theta_i}(Ln\psi(\theta))\mathbb{E}_\theta(h(y)y_i)\right\}$$

$$= (2-c)c\sum_{i=1}^k\frac{\partial}{\partial\theta_i}\left\{\frac{\partial}{\partial\theta_i}\mathbb{E}_\theta(h(y)) - \frac{\partial}{\partial\theta_i}(Ln\psi(\theta))\mathbb{E}_\theta(h(y))\right\} +$$

$$c^2\sum_{i=1}^k\frac{\partial}{\partial\theta_i}(Ln\psi(\theta))\left\{\frac{\partial}{\partial\theta_i}\mathbb{E}_\theta(h(y)) - \frac{\partial}{\partial\theta_i}(Ln\psi(\theta))\mathbb{E}_\theta(h(y))\right\}$$

$$= (2-c)c\Delta\phi(\theta) - (2-c)c\Delta Ln\psi(\theta)\cdot\phi(\theta) - (2-c)c(\nabla Ln\psi(\theta))^t\nabla\phi(\theta) +$$

$$c^2(\nabla Ln\psi(\theta))^t\nabla\phi(\theta) - c^2||\nabla Ln\psi(\theta)||^2\phi(\theta)$$

$$= (2-c)c[\Delta\phi(\theta) - \frac{\Delta\psi(\theta)}{\psi(\theta)}\phi(\theta)] + 2c(1-c)(||\nabla Ln\psi(\theta)||^2\phi(\theta) - (\nabla Ln\psi(\theta))^t\nabla\phi(\theta))$$

as $\Delta Ln\psi = \frac{\Delta\psi}{\psi} - ||\nabla Ln\psi||^2$. The last term will be non-negative if and only if

$$(2-c)c(\Delta\phi\cdot\psi - \Delta\psi\cdot\phi) + 2c(1-c)\left(\frac{\nabla\psi}{\psi}\right)^t(\nabla\psi\cdot\phi - \nabla\phi\cdot\psi) \geq 0 \qquad \square$$

By elementary manipulations, we obtain

**Corollary 4.3.** *A sufficient condition of domination of* $\varphi_0$ *by* $\varphi_c$ *is*

$$(i)\quad \text{if } c > 0, (\nabla\psi^2)^t(\nabla\tilde{h}) + (2-c)\psi^2\Delta\tilde{h} \geq 0,$$
$$(ii)\quad \text{if } c < 0, (\nabla\psi^2)^t(\nabla\tilde{h}) + (2-c)\psi^2\Delta\tilde{h} \leq 0.$$

**Example.** Consider $k$ independent *gamma* r.v.'s $y_1, y_2, \ldots, y_k$ with respective densities

$$f_{\theta_i}(y_i) = \frac{1}{\Gamma(\alpha)}e^{\theta_i y_i}(-\theta_i)^\alpha y_i^{\alpha-1} \quad \theta_i < 0,\ i = 1,\ldots,k,$$

where $\alpha > 0$ is known. With the previous notations, we have

$$\psi(\theta) = \prod_{i=1}^k(-\theta_i)^\alpha,\ \ \nabla\psi(\theta) = \alpha\psi(\theta)\left(\frac{1}{\theta_1},\frac{1}{\theta_2},\ldots,\frac{1}{\theta_k}\right)^t \ \text{ and } \ \Delta\psi(\theta) = \alpha(\alpha-1)\psi(\theta)\sum_{i=1}^k\frac{1}{\theta_i^2}.$$

Therefore, if we consider the estimator $\varphi(y) = (1-c)y$, we get $\phi(\theta) = 1$ and (4.2) implies

$$-(2-c)c\alpha(\alpha-1)\sum_{i=1}^k\frac{1}{\theta_i^2} + 2c(1-c)\alpha^2\sum_{i=1}^k\frac{1}{\theta_i^2} \geq 0\ ,$$

i.e. $c \leq \frac{2}{\alpha+1}$. Note further that the constant maximizing the right hand side of (4.2) is $c = \frac{1}{\alpha+1}$ . This constant gives the *natural* estimator of the mean,

$$\varphi(y) = \frac{\alpha}{\alpha+1}\,y\ . \tag{4.3}$$

Berger (1980) and DasGupta (1984) give estimators which improve over the estimator (4.3), but they are not of the form (4.1), the shrinkage factor being different for each component. However, there exists a similar differential inequality which gives a sufficient condition for the domination of (4.3).

The class (4.1) contains in particular the *positive-part shrinkage estimators* which are of main importance for normal distributions (see e.g. Judge and Bock (1978)). In the normal case, the minimaxity conditions usually consider $th(t)$ bounded rather than $h$ bounded (see Proposition 3.1). The next paragraph gives domination conditions corresponding to this class of estimators.

**4.3.** Consider now the class

$$\varphi_c(x) = \left(1 - c\,\frac{r(x)}{||x||^2}\right)\, x, \tag{4.4}$$

where $c \in \mathbb{R}$ and $0 \leq r(x) \leq 1$. We note again $\frac{r(x)}{||x||^2} = h(x)$. Thus

**Proposition 4.4.** *If*

$$2c\Delta\phi \cdot \psi - 2c\nabla\psi \cdot \nabla\phi + c(2||\nabla\psi||^2 - c\psi^2 - 2\Delta\psi \cdot \psi)\frac{\phi}{\psi} \geq 0, \tag{4.5}$$

*$\varphi_c$ dominates $\varphi_0$. If $r$ is constant and equal to 1, this condition is also necessary.*

**Proof.** As $0 \leq r(\cdot) \leq 1$, we have

$$\begin{aligned} &2\mathbb{E}_\theta[ch(x)x^t(x + \nabla Ln\psi(\theta))] - c^2\mathbb{E}_\theta[h^2(x)||x||^2] \\ &\geq 2\mathbb{E}_\theta[ch(x)x^t(x + \nabla Ln\psi(\theta))) - c^2\mathbb{E}_\theta[h(x)] \end{aligned}$$

This inequality is an equality if $r$ is equal to 1. From the proof of Proposition 4.1, we deduce

$$\begin{aligned} 2c\mathbb{E}_\theta[h(x)x^t(x + \nabla Ln\psi(\theta))] &= 2c\{\Delta\phi - \Delta Ln\psi \cdot \phi - (\nabla Ln\psi)^t\nabla\phi\} \\ &= \frac{2c}{\psi}\{\Delta\phi \cdot \psi - \Delta\psi \cdot \phi + (\nabla Ln\psi)^t(\phi\nabla\psi - \psi\nabla\phi)\} \end{aligned}$$

A sufficient condition is then

$$2c\left\{\Delta\phi \cdot \psi - (\nabla\psi)^t(\nabla\phi) + \left(\frac{||\nabla\psi||^2}{\psi} - \Delta\psi\right)\phi\right\} - c^2\phi\psi \geq 0 \qquad \square$$

Note that the class (4.4) contains the estimator introduced by James and Stein (1961) to dominate $\varphi_0$ . We will show in §4.4 that it is a solution of (4.5).

Once again, if we express the condition (4.5) with respect to $\tilde{h}$, we get simpler conditions.

**Corollary 4.5.** *$\varphi_c$ will dominate $\varphi_0$ if*

$$c\left\{2\Delta\tilde{h}\psi^2 + (\nabla\psi^2)^t\nabla\tilde{h} - c\psi^2\tilde{h}\right\} \geq 0$$

**4.4. The normal case.** The normal distribution allows some simplifications in the previous conditions. In fact, except for the uniform distributions on spheres, it is the only *usual* spherically symmetric distribution which belongs to the exponential family.

We suppose that $h$ is only a function of $||x||^2$. Then $\tilde{h}$ is only a function of $||\theta||^2$ and can be written $\tilde{h}(\theta) = g(||\theta||^2)$. With this assumption, condition (i) of corollary 4.3 becomes

$$0 \leq (-2\theta e^{-||\theta||^2})^t(2\theta g'(||\theta||^2)) + (2-c)e^{-||\theta||^2}(2kg'(||\theta||^2) + 4||\theta||^2 g''(||\theta||^2))$$

which is a unidimensional differential inequality

$$-2tg'(t) + (2-c)(kg'(t) + 2tg''(t)) \geq 0 \tag{4.6}$$

In the same way, when $c > 0$, corollary 4.5 can be written

$$4tg''(t) + 2(k-t)g'(t) - \frac{c}{2}g(t) \geq 0 \tag{4.7}$$

Among the solutions of (4.6) (or (4.7)), one has then to choose the solutions $g$ such that

$$g(||\theta||^2) = \int_{\mathbb{R}^k} h(||x||^2)e^{\theta^t x}e^{-||x||^2/2}dx$$

and $0 \leq h(t) \leq 1$ (or $0 \leq th(t) \leq 1$).

One of these solutions is the original James–Stein estimator. In fact, in this case, we have $h(x) = \frac{1}{||x||^2}$. Then

$$\begin{aligned}\tilde{h}(\theta) &= e^{||\theta||^2/2}\mathbb{E}_\theta(||x||^{-2}) \\ &= \sum_{i=0}^{\infty}\frac{(||\theta||^2)^i}{2^i\, i!}\,\mathbb{E}\left[1/\chi^2_{k+2i}\right] = \sum_{i=0}^{\infty}\frac{(||\theta||^2)^i}{2^i\, i!}\frac{1}{k+2i+2},\end{aligned}$$

due to the Poisson decomposition of the non-central $\chi_k^2$ distribution. Therefore (4.7) becomes

$$\sum_{i=0}^{\infty} \left\{ \frac{k+2+2i}{k+4+2i} (2i+k) - (2i + \frac{c}{2}) \right\} \frac{t^i}{2^i \; i!(k+2i+2)} \geq 0.$$

And

$$(k+2+2i)(2i+k) - (2i+\frac{c}{2})(k+4+2i) = i\ (2(k-2)-c) + (k+2)\ k - c\ \frac{k+4}{2}$$

which is positive for every $i$ if and only if $c \leq 2(k-2)$. We have then obtained by this method the classical necessary and sufficient condition of James and Stein (1961).

Another application of these inequalities can be found in *truncated parameters estimation*. For simplicity's sake, we consider only the normal case. Let us suppose that $\theta$ belongs to the ball of radius $\rho$,

$$\Theta = \left\{ ||\theta||^2 \leq \rho \right\} \ .$$

We are looking for sufficient conditions of domination of the restricted mle,

$$\varphi_0(x) = \begin{cases} x & \text{if } x \in \Theta, \\ \rho \frac{x}{||x||} & \text{otherwise,} \end{cases}$$

by shrinkage estimators in the class

$$\varphi_c(x) = \begin{cases} (1-ch(x))x & \text{if } x \in \Theta, \\ \varphi_0(x) & \text{otherwise.} \end{cases} \tag{4.8}$$

where $0 \leq h(x) \leq 1$ and $c \in \mathbb{R}$. It is then straighforward to prove that a sufficient condition for this domination is

$$c(2-c)\Delta\phi(\theta) + 2c(1-c)\theta^t \nabla\phi(\theta) + c(p(2-c) - c||\theta||^2)\phi(\theta) \geq 0 \ , \tag{4.9}$$

where $\phi(\theta) = \mathbb{E}_\theta \left( h(x) \mathbb{I}_\Theta(x) \right) = e^{-||\theta||^2/2} \tilde{h}(\theta)$. In the particular case where $\tilde{h}(\theta) = g(||\theta||^2)$, (4.9) becomes

$$c\left[ 2(2-c)tg''(t) + (p(2-c)-2t))g'(t) \right] \geq 0. \tag{4.10}$$

For instance, for $c > 0$, a necessary condition for (4.10) to be satisfied is $c < 2$. And, if $\rho < p$, it can be shown that (4.10) is satisfied for any function $h$ (bounded by 1) if $c \leq 2(p-\rho)/p$.

**Acknowledgements**

The author is grateful to Jim Berger, Larry Brown, Bernhard Flury and Jiunn Hwang for helpful comments on an earlier version of this paper.

## 5. References

Akai, T. (1986). Simultaneous estimation of location parameters of the distributions with finite support. *Ann. Inst. Stat. Math.* **38(A)**, 85–99.

Baranchik, A. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Stat.* **41**, 642–645.

Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Stat.* **3(6)**, 1318–28.

Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of Gamma scale parameters. *Ann. Stat.* **8**, 545–571.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*; Springer (2nd edition).

Berger, J., Das Gupta, A. (1986). Estimation of multiple gamma scale-parameters: Bayes estimation subject to uniform domination. *Comm. Stat. A* **15**, 2065–2086.

Berger, J., Haff L. (1983). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Stat. Dec.* **1**, 1–27.

Berger, J., Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems, *Ann. Stat.* **6(4)**, 783–801.

Bock, M. E. (1987) Shrinkage estimators: pseudo-Bayes rules for normal mean vectors, Technical Report #87-18, Purdue University.

Brown, L. (1971). Admissible estimators, recurrent diffusions and insoluble boundary problems, *Ann. Math. Stat.* **42(3)**, 855–903.

Brown, L. (1975). Estimation with incompletely specified loss functions, *JASA* **70**, 417–26.

Brown, L. (1985). *Foundations of Exponential Families.* IMS Monograph Series, Hayward, California.

Brown, L. (1987) The differential inequality of a statistical estimation problem. *Proc. Fourth Purdue Symp. Stat. Dec. Theo. Rel. Topics* (S.S. Gupta and J. Berger, eds.) Springer Verlag, New York, Vol. 1, 299–324.

Brown, L., Farrell, R. (1985). Complete class theorems for estimation of multivariate Poisson means and related problems, *Ann. Stat.* **13(2)**, 706–726.

Brown, L., Hwang, J. (1982). A unified admissibility proof. In *Stat. Dec. Theo. and Rel. Topics III* (S. S. Gupta and J. Berger, eds.) **(1)**, 205–230, Academic Press.

Cellier, D., Fourdrinier, D. (1985). Estimateurs à rétrécisseur de la moyenne d'une loi normale multidimensionnelle, pour un coût quadratique général *Statistique et Analyse des Données* **10(3)**, 26–41.

Cellier, D., Fourdrinier, D., Robert C. (1987). Estimateurs à rétrécisseur du paramètre de position d'une loi à symétrie sphérique. *Note aux CRAS* Paris t. **304(14)**, Série I, 439–442.

Chen, S. (1983). Improving upon some standard estimators in continuous exponential families under arbitrary quadratic loss, Technical Report #83-2, Purdue University.

Das Gupta, A. (1984). Admissibility in the gamma distributions: two examples, *Sankhyā A* **40(3)** 395–407.

Das Gupta, A., Sinha B. (1986). Estimation in the multiparameter exponential family: admissibility and inadmissibility results. *Stat. Dec.* **4**, 101–130.

Fraisse, A.M., Robert, C., Roy, M. (1987). Une condition nécessaire d'admissibilité pour les familles exponentielles et ses conséquences sur les estimateurs à rétrécisseur matriciel, *Note aux CRAS* Paris t. **305**, Série I, 889–891.

Gleser, L. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix, *Ann. Stat.* **14(4)** 1625–1633.

Ghosh, M., Hwang, J., Tsui, K. (1983). Construction of improved estimators in multiparameter estimation for continuous exponential families. *J. Mult. Anal.* **14**, 212–220.

Haff, L., Johnson, R. (1986). The superharmonic condition for simultaneous estimation of means in exponential families, *Can. J. Stat.* **14(1)**, 43–54.

Hwang, J. (1982a). Certain bounds on the class of admissible estimators in continuous exponential families, in *Stat. Dec. Theo. and Rel. Topics III* (S. S. Gupta and J. Berger eds.), **(2)**, 15–30, Academic Press.

Hwang, J. (1982b). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Stat.* **11(2)**, 15–30.

Hwang, J., Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution, *Ann. Stat.* **10(3)**, 868–881.

James, W., Stein, C. (1961) Estimation with quadratic loss, *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, **1**, 361–379.

Judge, G., Bock, M.E. (1978) *The statistical implications of pre-test and Stein-rule estimators in Econometrics.* North-Holland.

Rukhin, A. (1988). Improved estimation in lognormal regression models, *J. Stat. Plann. Inf.* **18**(3), 291–273 .

Schwartz, L. (1966). *Théorie des Distributions*, Hermann Paris.

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss, *Ann. Stat.* **12(1)**, 322–335.

Stein, C. (1973). Estimation of the mean of a multivariate distribution. *Proc. Prague Symp. Asymp. Stat.*, 345–381.

Stein, C. (1981). Estimation of the mean of a multivariate distribution, *Ann. Stat.* **9(6)** 1136–1151.