An Explicit Formula for the Risk of the Positive-Part James-Stein Estimator

by

Christian Robert[1]
Purdue University

Technical Report #87-44

Department of Statistics
Purdue University

March 1988

[1] Research partially supported by l'Institut National de la Statistique et des Etudes Economiques.

An Explicit Formula for the Risk of the Positive-Part James-Stein Estimator

by

Christian Robert[1]
Purdue University and Université de Rouen

Abstract

When estimating a normal mean vector with variance known up to a multiplicative factor, it is well known that the positive-part James-Stein estimator is not admissible but, until now, no one has been able to exhibit a uniformly better estimator. We propose here an explicit formula for the risk of the positive-part James-Stein estimator.

Key Words Normal mean, positive-part James-Stein estimator, quadratic risk

AMS (1985) Subject Classification: 62C20, 62J07

[1] Research partially supported by l'Institut National de la Statistique et des Etudes Economiques.

1. Introduction.

Since Stein (1956) has shown the inadmissibility of the least squares estimator of the mean of a random vector $y \sim N_p(\theta, \sigma^2 I_p)$ when $p \geq 3$, a lot of shrinkage estimators dominating the least squares estimator have been proposed, following the primitive James-Stein estimator (1961). In particular, Baranchick (1964) has shown that the positive-part James-Stein estimator dominates the James-Stein estimator. Even though this estimator is not admissible, no estimator uniformly better than it is known. It seems very difficult to find a dominating estimator; Bock (1987) and Brown (1988) have shown that the usual technique of unbiased estimation of the risk is of no use in that case.

Because of the importance of this estimator, we give, in this paper, an exact formula for the risk of the positive-part James-Stein estimator as Egerton and Laycock (1982) have done for the James-Stein estimator. We consider the problem of the estimation of a normal mean vector when the variance is known up to a multiplicative factor.

2. Gain of the positive-part James-Stein estimator.

This estimator can be written

$$\phi(y, s^2) = \left(1 - \frac{(k-2)(n-k)}{n-k+2} \frac{s^2}{||y||^2}\right)^+ \quad y$$

where s^2 is an estimator of σ^2 independent of y, $(n-k)s^2 \sim \sigma^2 \chi^2_{n-k}$ and $(x)^+ = \max(0, x)$. We can also write

$$\phi(y, s^2) = (1 - h(y, s^2))y$$

with $h(y, s^2) = \alpha \frac{s^2}{||y||^2} \mathbf{I}(\frac{\alpha s^2}{||y||^2} < 1) + \mathbf{I}(\frac{\alpha s^2}{||y||^2} > 1)$,

$$\alpha = \frac{(k-2)(n-k)}{n-k+2},$$

$\mathbf{I}\left(\frac{\alpha s^2}{||y||^2} < 1\right) = 1$ if $\left(\frac{\alpha s^2}{||y||^2}\right) < 1$, 0 otherwise and $\mathbf{I}\left(\frac{\alpha s^2}{||y||^2} > 1\right) = 1 - \mathbf{I}\left(\frac{\alpha s^2}{||y||^2} < 1\right)$.

Note that the function h is continuous and almost everywhere differentiable. We can then apply usual integration by parts lemmas. The increase in risk of the usual estimator, y, over $\phi(y, s^2)$, is given by

$$\frac{1}{\sigma^2}\mathbb{E}_{\theta,\sigma}[||y-\theta||^2] - \frac{1}{\sigma^2}\mathbb{E}_{\theta,\sigma}[||\phi(y,s^2)-\theta||^2] =$$
$$\frac{2}{\sigma^2}\mathbb{E}_{\theta,\sigma}[h(y,s^2)y^t(y-\theta)] - \frac{1}{\sigma^2}\mathbb{E}_{\theta,\sigma}[h^2(y,s^2)y^t y],$$

which is equal to

$$2\left\{\mathbb{E}_{\theta,\sigma}\left[(k-2)\alpha\,\frac{s^2}{y^t y}\,\mathbf{I}\left(\frac{\alpha s^2}{y^t y} < 1\right)\right] + k\ \mathbb{E}_{\theta,\sigma}\left[\mathbf{I}\left(\frac{\alpha s^2}{y^t y} > 1\right)\right]\right\}$$
$$-\left\{\mathbb{E}_{\theta,\sigma}\left[\alpha^2 \frac{s^2}{y^t y}\ \frac{n-k+2}{n-k}\ \mathbf{I}\left(\frac{\alpha s^2}{y^t y} < 1\right)\right] + \frac{n-k-2}{n-k}\ \mathbb{E}_{\theta,\sigma}\left[\frac{y^t y}{s^2}\,\mathbf{I}\left(\frac{\alpha s^2}{y^t y} > 1\right)\right]\right\}$$

(see e.g. Judge and Bock (1978)).

This gain can also be written as

$$\text{(2.1)} \qquad k - \lambda + \mathbb{E}_{\theta,\sigma}\left[\left\{\frac{(k-2)^2(n-k)}{n-k+2}\frac{s^2}{y^ty} - 2k + \frac{n-k-2}{n-k}\ \frac{y^ty}{s^2}\right\}\mathbf{I}\left(\alpha\frac{s^2}{y^ty} < 1\right)\right]$$

where $\lambda = \theta^t\theta/\sigma^2$. Note that $y^ty \sim \chi_k^2(\lambda/2)$.

We will suppose that $(n-k)$ is even and greater than 2. Note that this constraint can be satisfied in practical cases by adding or subtracting an observation and is far less restrictive than a constraint about the number of parameters.

3. An exact formula of the gain.

As in Egerton and Laycock (1982), the formulas differ according to the parity of k. There also appears the *Dawson-integral*

$$D(\lambda) = e^{-\lambda^2}\int_0^\lambda e^{t^2}\,dt,$$

which is tabulated in Abramowitz-Stegun (1964, p. 319). It is worthwhile to remark that the approximation of e^{t^2} by $\sum_{i=0}^{100}\frac{t^{2i}}{i!}$ in the above integral gives exactly the same results as the Abramowitz-Stegun table.

Proposition

(a) If k *is even*, the gain is equal to

$$\frac{(k-2)(n-k)}{n-k+2}\left[\sum_{j=1}^{\frac{k-2}{2}}(-1)^{j+1}\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}-j)}(\frac{2}{\lambda})^j\left\{1-\left(\frac{k-2}{n}\right)^{-j+\frac{k-2}{2}}e^{-\frac{\lambda}{2}\frac{n-k+2}{n}}\right\}\right]$$

$$-\,e^{-\frac{\lambda}{2}\frac{n-k+2}{n}}\left\{\sum_{i=1}^{\frac{n-k}{2}}\left[(n-k)(n-k+2)-4\frac{k(n-k+2)}{k-2}i+4i(i-1)\right]\left(\frac{n-k+2}{n}\right)^{i-2}\times\right.$$

$$\text{(3.a)} \qquad \left.\frac{\Gamma(\frac{k-2}{2}+i)}{\Gamma(\frac{k}{2})}\frac{1}{2i!}\sum_{j=0}^{i-1}\binom{i-1}{j}\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}+j\right)}\left(\frac{k-2}{n}\right)^j\left(\frac{\lambda}{2}\right)^j\right\}(\frac{k-2}{n})^{\frac{k+2}{2}}$$

(b) If k *is odd*, the gain is equal to

$$\frac{(k-2)(n-k)}{n-k+2}\Big[\sum_{j=1}^{\frac{k-3}{2}}(-1)^{j+1}\ \frac{(k-2)\dots(k-2j)}{\lambda^j}\Big\{1-\left(\frac{k-2}{n}\right)^{\frac{k-2}{2}-j}e^{-\frac{\lambda}{2}\frac{n-k+2}{n}}\Big\}$$

$$+\frac{(-1)^{\frac{k-3}{2}}(k-2)!}{(2\lambda)^{\frac{k-3}{2}}(\frac{k-3}{2})!\sqrt{\lambda/2}}\Big\{D(\sqrt{\lambda/2})-e^{-\frac{\lambda}{2}\frac{n-k+2}{n}}D((\frac{(k-2)\lambda}{2n})^{1/2})\Big\}\Big]$$

$$-e^{-\frac{\lambda}{2}\frac{n-k+2}{n}}\Big\{\sum_{i=1}^{\frac{n-k}{2}}[(n-k)(n-k+2)-4\ \frac{k(n-k+2)}{k-2}\ i+4i(i-1)]\left(\frac{n-k+2}{n}\right)^{i-2}\times$$

$$\frac{\Gamma(\frac{k-2}{2}+i)}{\Gamma(\frac{k}{2})}\frac{1}{2i!}\sum_{j=0}^{i-1}\binom{i-1}{j}\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+j)}\ \left(\frac{k-2}{n}\right)^j\left(\frac{\lambda}{2}\right)^j\Big\}\left(\frac{k-2}{n}\right)^{\frac{k+2}{2}} \qquad (3.b)$$

Note that the ratios $\frac{\Gamma(x\pm i)}{\Gamma(x)}$ which appear in these formulas can be very simply calculated through the recursion formula $\Gamma(x+1)=x\Gamma(x)$. With the previous partial series approximation, these formulas can then be incorporated into a package without any need for numerical tables. The graphs in Section 5 give the computations of this gain for several values of k and n.

Remark 1. If, for the positive-part James-Stein estimator ϕ, we replace $\frac{(k-2)(n-k)}{n-k+2}$ with α, where $\frac{(k-2)(n-k)}{n-k+2}\leq\alpha\leq\frac{2(k-2)(n-k)}{n-k+2}$, we get a class of shrinkage estimators which are not comparable (see Figure 1). It is straightforward to generalize the formulas (3.a) and (3.b) for an arbitrary α by using the following proof.

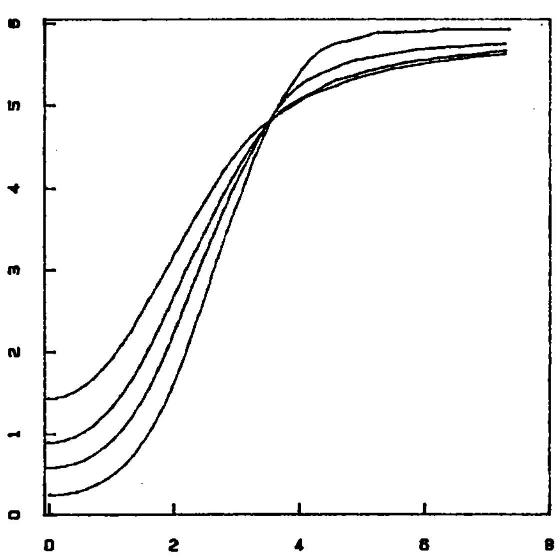


Figure 1. Risk of several positive-part James-Stein estimators ($n=6, k\ \sigma^2$ known).

Remark 2. We can also deduce from these formulas the exact risk of the positive-part James-Stein estimator when the variance is totally known by taking the limit of (3.a) and (3.b) as n goes to infinity.

4. Proof.

We have

$$\mathbb{E}_{\theta,\sigma}\left[\left\{\frac{(k-2)^2}{n-k+2}\frac{(n-k)s^2}{y^ty}-2k+(n-k-2)\frac{y^ty}{(n-k)s^2}\right\}\mathbf{I}\left[y^ty>\alpha s^2\right]\right]$$
$$=\int_0^\infty\int_0^\infty\left\{\frac{(k-2)^2}{n-k+2}\;w-\left(2k-(n-k-2)\;\frac{t+\phi(w)}{w}\right)(t+\phi(w))\right\}\times$$
$$\times\sum_{j=0}^\infty\frac{e^{-\lambda/2}(\lambda/2)^j}{j!}\;\frac{(t+\phi(w))^{\frac{k}{2}+j-2}}{2^{\frac{k}{2}+j}}\frac{e^{-(t+\phi(w))/2}}{\Gamma(\frac{k}{2}+j)}\,dt\;\frac{w^{\frac{n-k}{2}-1}e^{-w/2}}{2^{\frac{n-k}{2}}\Gamma(\frac{n-k}{2})}\;dw$$

where $\phi(w)=\frac{k-2}{n-k+2}w$.

For $\nu\in\mathbb{R}_+$, the modified Bessel function I_ν can be written (see Abramowitz-Stegun (1964, p. 375))

$$I_\nu(z)=(\frac{1}{2}z)^\nu\sum_{j=0}^\infty\frac{(\frac{1}{2}z)^{2j}}{j!\Gamma(\nu+j+1)}.$$

The above integral is equal to

$$\frac{e^{-\lambda/2}}{2\lambda^{\frac{k-2}{4}}}\int_0^\infty\int_0^\infty\left\{\frac{(k-2)^2}{n-k+2}w-\left(2k-(n-k-2)\frac{t+\phi(w)}{w}\right)(t+\phi(w)\right\}\times$$
$$\times I_{\frac{k-2}{2}}\left(\lambda^{1/2}(t+\phi(w))^{1/2}\right)(t+\phi(w))^{\frac{k-6}{4}}e^{-(t+\phi(w))/2}\,dt\;\frac{w^{\frac{n-k}{2}-1}\;e^{-w/2}}{2^{\frac{n-k}{2}}\;\Gamma(\frac{n-k}{2})}dw$$
$$=\frac{e^{-\lambda/2}}{\lambda^{\frac{k-2}{2}}}\int_0^\infty\int_0^{\delta(u)}\left\{\frac{(k-2)^2}{n-k+2}w-(2k-(n-k-2)\frac{u^2}{\lambda w})\frac{u^2}{\lambda}\right\}\frac{w^{\frac{n-k}{2}-1}\;e^{-w/2}}{2^{\frac{n-k}{2}}\;\Gamma(\frac{n-k}{2})}\;dw\;\times$$
$$\times\;I_{\frac{k-2}{2}}(u)u^{\frac{k-4}{2}}e^{-u^2/2\lambda}\;du$$

where $\delta(u)=\frac{n-k+2}{k-2}\frac{u^2}{\lambda}$.

We have, for every $p\in\mathbb{N}$

$$\int_0^{\delta(u)}w^pe^{-w/2}dw=p!\;\;2^{p+1}\left(1-e^{-\delta(u)/2}\sum_{i=0}^p\frac{(\delta(u)/2)^i}{i!}\right).$$

Thus, if $(n-k) \geq 4$, the expectation we are considering can be written

$$\frac{e^{-\lambda/2}}{\lambda^{\frac{k-2}{2}}} \int_0^\infty \Big\{ \frac{(k-2)^2(n-k)}{n-k+2} \left(1 - e^{-\delta(u)/2} \sum_{i=0}^{\frac{n-k}{2}} \frac{(\delta(u)/2)^i}{i!} \right) -$$

$$2k\frac{u^2}{\lambda} \left(1 - e^{-\delta(u)/2} \sum_{i=0}^{\frac{n-k}{2}-1} \frac{(\delta(u)/2)^i}{i!} \right)$$

$$+ (\frac{u^2}{\lambda})^2 \left(1 - e^{-\delta(u)/2} \sum_{i=0}^{\frac{n-k}{2}-2} \frac{(\delta(u)/2)^i}{i!} \right) \Big\} I_{\frac{k-2}{2}}(u) u^{\frac{k-4}{2}} e^{-u^2/2\lambda}\, du$$

From Gradshteyn-Ryzhik (1980, **6.631**), we deduce that, for $\nu + \mu > 0$,

$$\int_0^{+\infty} I_\nu(t) t^{\mu-1} e^{-p^2t^2} dt = \frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2}\mu)}{2^{\nu+1} p^{\mu+\nu} \Gamma(\nu+1)}\ {}_1F_1(\frac{1}{2}\nu + \frac{1}{2}\mu; \nu+1; \frac{1}{4p^2}),$$

where ${}_1F_1(\alpha;\beta;z)$ is the confluent hypergeometric function (see Abramowitz-Stegun, p. 503). This equality allows us to compute the above integral, which is equal to

$$\frac{e^{-\lambda/2}}{\lambda^{\frac{k-2}{2}}} \Big\{ \Big[\frac{(k-2)(n-k)}{n-k+2}\ {}_1F_1\left(\frac{k-2}{2}; \frac{k}{2}; \frac{\lambda}{2}\right) - 2k\ {}_1F_1\left(\frac{k}{2}; \frac{k}{2}; \frac{\lambda}{2}\right) +$$

$$k\ {}_1F_1\left(\frac{k+2}{2}; \frac{k}{2}; \frac{\lambda}{2}\right) \Big] \lambda^{\frac{k-2}{2}} - \sum_{i=0}^{\frac{n-k}{2}} (\frac{n-k+2}{k-2})^{i-2} [(n-k)(n-k+2) - 4i\frac{k(n-k+2)}{k-2}$$

$$+ 4i(i-1)] \frac{\lambda^{\frac{k-2}{2}}}{2i!} \frac{\Gamma(\frac{k-2}{2}+i)}{\Gamma(\frac{k}{2})} \left(\frac{k-2}{n}\right)^{\frac{k-2}{2}+i}\ {}_1F_1\left(\frac{k-2}{2}+i; \frac{k}{2}; \frac{\lambda(k-2)}{2n}\right) \Big\}$$

$$= -e^{-\lambda/2} \Big\{ \sum_{i=1}^{\frac{n-k}{2}} \Big[\Big[(n-k)(n-k+2) - 4i\ \frac{k(n-k+2)}{k-2} + 4i\ (i-1) \Big] \left(\frac{n-k+2}{n}\right)^{i-2} \times$$

$$\times (\frac{k-2}{n})^{\frac{k+2}{2}} \frac{\Gamma\left(\frac{k-2}{2}+i\right)}{2i!\Gamma(\frac{k}{2})}\ {}_1F_1\left(\frac{k-2}{2}+i;\ \frac{k}{2}; \frac{\lambda(k-2)}{2n}\right) \Big]$$

$$+ \frac{(n-k)(k-2)}{n-k+2}\ {}_1F_1\left(\frac{k-2}{2}; \frac{k}{2}; \frac{\lambda}{2}\right) \Big\} - k + \lambda.$$

Using the properties of the confluent hypergeometric functions (see lemmas), we get then the formulas (3.a) and (3.b). □

5. Examples

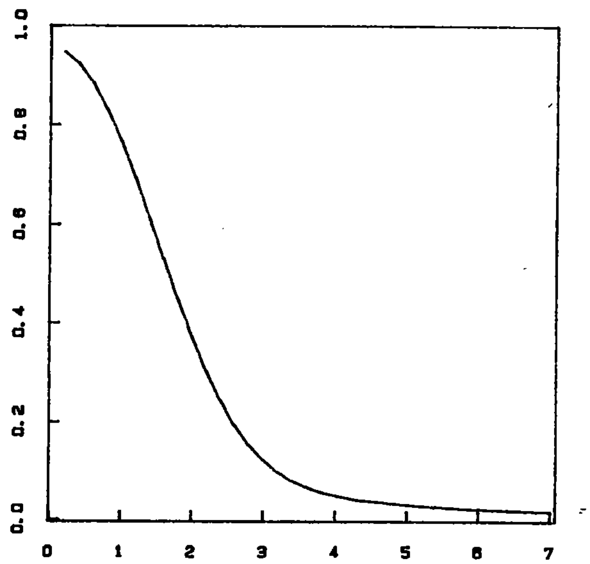


Figure 2 – Gain of the positive-part James-Stein estimator ($k = 3, n = 7$).

Even if the formulas (3.a) and (3.b) seem rather unwieldy, they can be reduced a great deal in particular cases. For example, when $k = 3$ and $n = 7$, we deduce from (3.b) that the gain is

$$\frac{2}{3}\frac{D\left(\sqrt{\lambda/2}\right) - e^{-\frac{3}{7}\lambda} D\left(\sqrt{\lambda/14}\right)}{\sqrt{\lambda/2}} + \frac{2e^{-\frac{3}{7}\lambda}}{7\sqrt{7}}\{5 + \frac{\lambda}{7}\}$$

For $\lambda = 0$, the value of the gain is $\frac{2}{3} + \frac{16}{21\sqrt{7}}$ (as $\lim_{\lambda \to 0} \frac{D(\lambda)}{\lambda} = 1$). One can compare this value with the value of the risk of the James-Stein estimator, $\frac{2}{3}$ (computed through Egerton and Laycock's formula). And when $k = 4$ and $n = 8$, we get (using (3.a)) the following gain:

$$\frac{8}{3\lambda}(1 - e^{-\frac{3\lambda}{8}}) + \frac{e^{\frac{-3\lambda}{8}}}{4}(3 + \frac{\lambda}{8})$$

When $\lambda = 0$, the gain is $\frac{7}{4}$, compared with $\frac{4}{3}$ for the James-Stein estimator.

As k and n increase, the number of terms in the formulas become larger. For k even, we have $\frac{n+k-4}{2}$ terms and for k odd, $\frac{n-k-2}{2}$ terms (including the Dawson integrals).

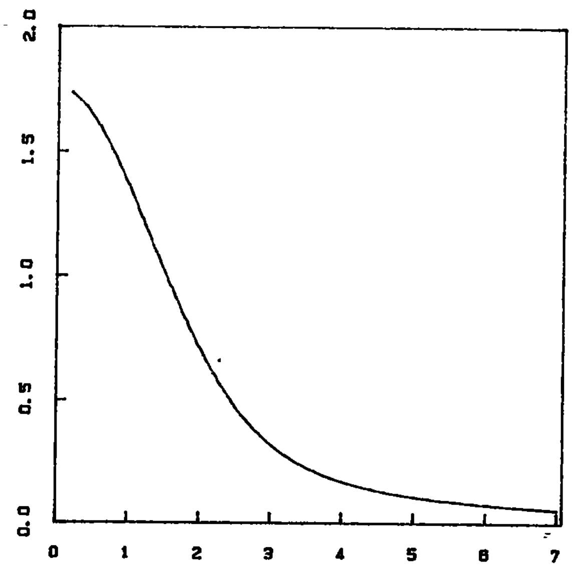


Figure 3 – Gain of the positive-part James-Stein estimator ($k = 4, n = 8$).

Appendix

The following lemmas can be proved by using the recursion formula (13.4.4) of Abramowitz and Stegun

$$ {}_1F_1(a;b;z) = {}_1F_1(a-1;b;z) + \frac{z}{b}\, {}_1F_1(a;b+1;z) $$

Some of them have been proved in Bock, Judge and Yancey (1984, p. 223)

Lemma 1 $\forall\, a, \forall\, j \geq 0$,

$$ {}_1F_1(a+j;a;z) = e^z \sum_{l=1}^{j} \binom{j}{l} \frac{z^l}{a(a+1)\dots(a+l)}. $$

Lemma 2 If $p \in \mathbb{N}^*$, if $z > 0$,

$$ {}_1F_1(p;p+1;z) = e^z \sum_{i=1}^{p} (-1)^{i+1} \frac{p\dots(p-i+1)}{z^i} + (-1)^p \frac{p!}{z^p}. $$

Lemma 3 If $p \in \mathbb{N}^*$, if $z > 0$,

$$ {}_1F_1(p+\frac{1}{2};p+\frac{3}{2};z) = e^z \{ \sum_{i=1}^{p} (-1)^i \frac{(2p+1)\dots(2p+3-2i)}{(2z)^i} + (-1)^p \frac{(2p+1)!}{p!(4z)^p} \frac{D(\sqrt{z})}{\sqrt{z}} \}. $$

Acknowledgement

I would like to thank Mary Ellen Bock for suggesting the use of modified Bessel functions and for constant help during the preparation of this paper.

References

ABRAMOWITZ, M., STEGUN I. (1964) Handbook of mathematical functions and formulas, graphs and mathematical tables (U.S. Department of Commerce, Washington, D.C.).

BARANCHICK A. (1964) Multiple regression and estimation of the mean of a multivariate normal distribution, Techn. Rep. No. 51, Department of Statistics, Stanford University.

BOCK, M. E. (1987) Shrinkage estimators: pseudo-Bayes rules for normal mean vectors, Techn. Rep. #87-18, Purdue University.

BOCK, M. E., JUDGE, G., YANCEY, T. (1986) A simple form for the inverse moments of non-central χ^2 and $\mathcal{F}$ random variables and certain confluent hypergeometric functions, *J. Econom.* 25(1), 217–234.

BROWN, L. (1988) 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, 299–324.

EGERTON, M., LAYCOCK, P. (1982) An explicit formula for the risk of James-Stein estimators, *Can. J. Stat.* 10(3), 199–205.

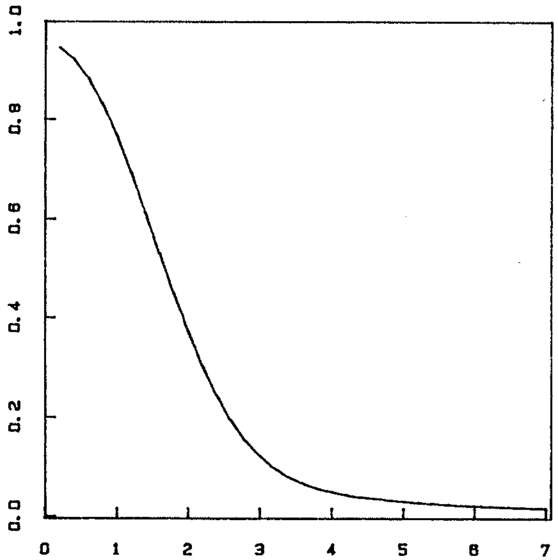
GRADSHTEYN, I., RYZHIK, I. (1980) Table of Integrals, Series and Products. Academic Press.

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

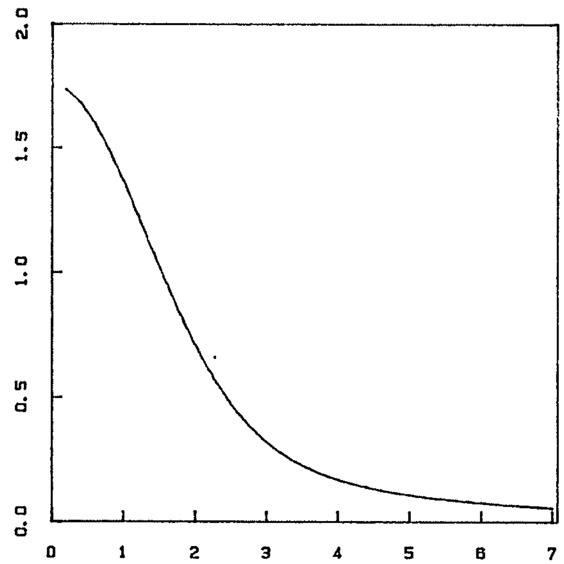
JUDGE, G., BOCK, M.E. (1978) The statistical implications of pre-test and Stein-rule estimators in econometrics. North-Holland.

STEIN, C. (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. 3rd Berkeley Symp. Math. Stat. Prob., 1, 197–206.

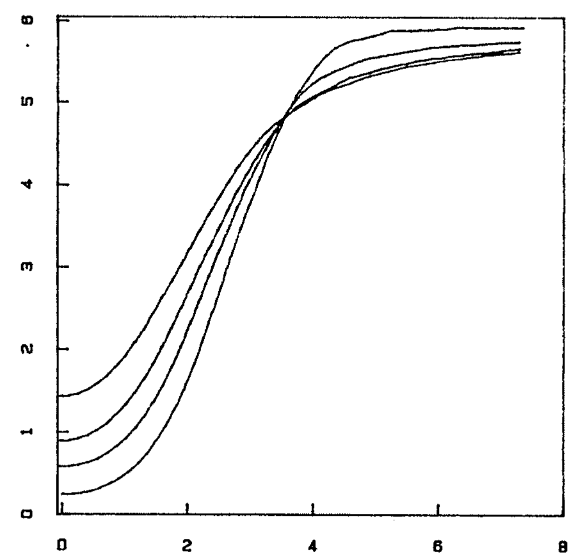
(2)

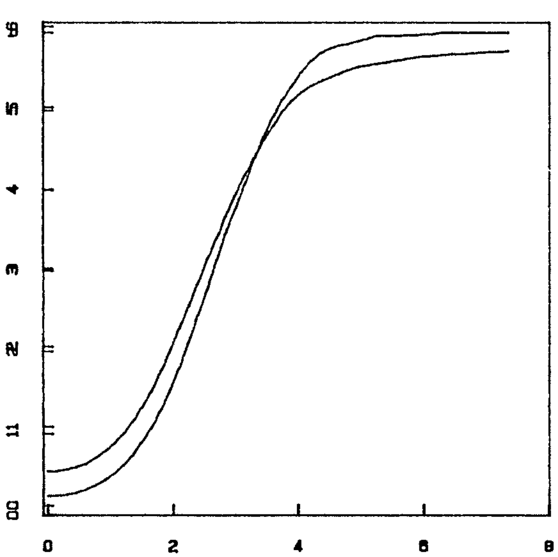


(3)



(1)





<- Don't use

Orig of figures if changes are made 3/21/88