POINT ESTIMATION OF MULTIVARIATE NORMAL MEAN USING t PRIORS

by

Mei-Mei Zen[1]
Purdue University

Technical Report #91-43

Department of Statistics
Purdue University

August 1991

[1]Research supported by NSF Grant 8923071–DMS.

Point Estimation of Multivariate Normal Mean Using t Priors

Abstract

Bayesian point estimation of a multivariate normal mean is considered under t priors. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$ and $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I)$. The posterior mode $\underset{\sim}{\nu}$ is employed as a point estimate of $\underset{\sim}{\theta}$. The risk behavior of the posterior mode is explored in this article. The unique unbiased estimator of the risk function of the posterior mode is derived in a closed form. It is proved that for any fixed m, τ^2 and σ^2, the posterior mode is minimax for moderate values of p.

1. Introduction

Consider the problem of estimating the mean vector $\underset{\sim}{\theta}$ of a $p \geq 3$ dimensional multivariate normal distribution, $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$, where σ^2 is assumed to be known. For the squared error loss function

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \frac{\| \underset{\sim}{\theta} - \underset{\sim}{a} \|^2}{\sigma^2},$$

the maximum likelihood estimator (MLE) $\hat{\underset{\sim}{\theta}}^0 = \underset{\sim}{X}$ has risk $R(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}^0) = p$. James and Stein [9] showed that the estimator $\hat{\underset{\sim}{\theta}}^1 = \left(1 - \frac{(p-2)\sigma^2}{\|\underset{\sim}{X}\|^2}\right)\underset{\sim}{X}$ does better than MLE $\hat{\underset{\sim}{\theta}}^0$. Its risk $R(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}^1)$, which is a function only of $\lambda = \frac{\|\underset{\sim}{\theta}\|^2}{2\sigma^2}$, increases from 2 at $\lambda = 0$ to the minimax value p as $\lambda \to \infty$. Baranchik [2] proved that, under squared error loss function, an estimator of the form

$$\hat{\underset{\sim}{\theta}} = \left(1 - \frac{\sigma^2 r(\frac{\|\underset{\sim}{X}\|^2}{\sigma^2})}{\| \underset{\sim}{X} \|^2}\right) \underset{\sim}{X} \tag{1}$$

is minimax under certain conditions. In this article, we consider Bayesian point estimation of a multivariate normal mean $\underset{\sim}{\theta}$ under t priors. DasGupta, Ghosh and Zen [4] found that if the posterior distribution of $\underset{\sim}{\theta}$ is starunimodal about the mode $\underset{\sim}{\nu}$, then $\underset{\sim}{\nu}$ can be expressed as $\underset{\sim}{\nu} = a\underset{\sim}{X}$, where '$a$' is the unique root of

$$h(a) = a^3 - a^2 + (\beta + \gamma)ay - \gamma y = 0, \tag{2}$$

$\alpha = \frac{\sigma^2}{m\tau^2}$, $\beta = m + p$, $\gamma = \frac{1}{\alpha}$ and $y = \frac{1}{\|\underset{\sim}{X}\|^2}$. Although the closed form of 'a' is messy, 'a' is a function of $\| \underset{\sim}{X} \|^2$ indeed; therefore, $\underset{\sim}{\nu} = (1 - (1 - a))\underset{\sim}{X}$, which is in the form of (1), is employed as a point estimator of $\underset{\sim}{\theta}$.

We first establish sufficient conditions for the minimaxity of the posterior mode in section 2. Also, the behavior of the function $r(\cdot)$ in (1) for the posterior mode is discussed in section 2. The results point out that for any given m, τ^2 and σ^2, $r(\cdot)$ is nondecreasing in $\| \underset{\sim}{X} \|^2$ if $p \leq m(\frac{\tau^2}{\sigma^2} - 1)$; furthermore, $r(\cdot)$ is bounded above by $2(p-2)$ for moderate p. Risk behavior of the posterior mode and comparison with the posterior mean $\hat{\underset{\sim}{\mu}}(\underset{\sim}{X})$ are both important issues. Notice that, unlike the posterior mode, the posterior mean doesn't have a closed form expression and needs to be approximated by numerical methods.

We denote the difference between $R(\underset{\sim}{\theta}, \underset{\sim}{\nu})$ and $R(\underset{\sim}{\theta}, \underset{\sim}{X})$ by $\Delta_{\underset{\sim}{\nu}}(\underset{\sim}{\theta})$. Then the unique unbiased estimator of $\Delta_{\underset{\sim}{\nu}}(\underset{\sim}{\theta})$, denoted by $u(\underset{\sim}{X})$, can be derived in a closed form. Therefore, $u(\underset{\sim}{X}) + p$ can serve as the unbiased estimator of the risk function of $\underset{\sim}{\nu}$. If the posterior is starunimodal, then 'a' is a function of $\underset{\sim}{X}$ through $\| \underset{\sim}{X} \|^2$ only. Thus, $u(\cdot)$ can be treated as a function of $\| \underset{\sim}{X} \|^2$. Bounding $u(\| \underset{\sim}{X} \|^2)$ above by 0, it is found that $\underset{\sim}{\nu}$ does better than MLE $\hat{\underset{\sim}{\theta}}^0$ under certain conditions; these results are given in section 3. Then in section 4, dealing with the posterior mean, we find that not only the posterior mean but also the posterior mode is tail minimax for all p. Also, the plots of posterior mean and posterior mode give a comparison between these two estimators.

2. Minimaxity of the Posterior Mode

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\theta}, \sigma^2 I)$ and $\underset{\sim}{\theta} \sim t(m, \underset{\sim}{\mu}, \tau^2 I)$. W.L.O.G., we take $\sigma^2 = 1$ and $\underset{\sim}{\mu} = 0$; if not, all assertions hold with $\frac{\underset{\sim}{X} - \underset{\sim}{\mu}}{\sigma}$ in place of $\underset{\sim}{X}$. Since the posterior mode $\underset{\sim}{\nu} = a\underset{\sim}{X}$ can be written as

$$\left[1 - \frac{r(\| \underset{\sim}{X} \|^2)}{\| \underset{\sim}{X} \|^2}\right] \underset{\sim}{X},$$

where $r(\| \underset{\sim}{X} \|^2) = (1-a) \| \underset{\sim}{X} \|^2$, it is of the form (1). Baranchik [2] proved that for $p \geq 3$ the conditions

$$0 \leq r(\| \underset{\sim}{X} \|^2) \leq 2(p-2) \tag{3}$$

$$\text{and} \qquad r(\| \underset{\sim}{X} \|^2) \ \text{ nondecreasing in } \ \| \underset{\sim}{X} \|^2 \tag{4}$$

are sufficient that $\underset{\sim}{\nu}$ be minimax.

We first need two Lemmas, and then give a sufficient condition for the minimaxity of the posterior mode. Recall that $\alpha = \frac{\sigma^2}{m\tau^2} = \frac{1}{m\tau^2}$, $\beta = m + p$, $\gamma = \frac{1}{\alpha}$ and $y = \frac{1}{\|\underset{\sim}{X}\|^2}$; for simplicity, sometimes we write 'a' instead of $a(y)$.

Lemma 1 *Let $r(\| X \|^2) = (1 - a(\| X \|^2)) \cdot \| X \|^2$, where $a(\cdot)$ is the unique root of (2). Then $r(\| X \|^2)$ is nondecreasing in $\| X \|^2$ if and only if $p \leq m(\tau^2 - 1)$.*

Proof: It is equivalent to proving that $r(y)$ is nonincreasing in y if and only if $\beta \leq \gamma$. This is because $\beta \leq \gamma \Longleftrightarrow p \leq m(\tau^2 - 1)$. It was proved that $a(\cdot)$ is nonincreasing in y. See Proposition 2.1 in DasGupta, Ghosh and Zen [4]. From (2), by implicit differentiation, we have

$$\frac{da}{dy} = a'(y) = \frac{\gamma - (\beta+\gamma)a}{3a^2 - 2a + (\beta+\gamma)y.} \tag{5}$$

Dividing both sides of (2) by a^2, it gives

$$1 - a = \frac{(\beta+\gamma)ay - \gamma y}{a^2};$$

therefore,

$$r(y) = \frac{1 - a(y)}{y} \tag{6}$$

$$= \frac{(\beta+\gamma)a - \gamma}{a^2}. \tag{7}$$

¿From (6), elementary calculations give that

$$\begin{aligned} 1 - a(y) &= y \cdot r(y) \\ \Longrightarrow \quad -a'(y) &= r(y) + yr'(y) \\ \Longrightarrow \quad r'(y) &= -\frac{a'(y) + r(y)}{y}. \end{aligned}$$

Since $y > 0, r(y)$ is nonincreasing in y is equivalent to $a'(y) + r(y) \geq 0, \quad \forall \ y > 0$. From (5) and (7), it reduces to

$$\frac{\gamma - (\beta+\gamma)a}{3a^2 - 2a + (\beta+\gamma)y} + \frac{(\beta+\gamma)a - \gamma}{a^2} \geq 0, \quad \forall \ y > 0.$$

$$\Longleftrightarrow \quad [(\beta+\gamma)a - \gamma]\left[\frac{1}{a^2} - \frac{1}{3a^2 - 2a + (\beta+\gamma)y}\right] \geq 0, \quad \forall \ y > 0. \tag{8}$$

Note that $(\beta+\gamma)a - \gamma \ > \ 0$ $\left(\text{or } a > \frac{\gamma}{\beta+\gamma}\right)$ for $\| X \|^2 > \ 0$ and $a \ = \ \frac{\gamma}{\beta+\gamma}$ for $\| X \|^2 = \ 0$. Further, from (5), we have that

$$3a^2 - 2a + (\beta+\gamma)y \ > \ 0, \ \ \forall \, y \ > \ 0.$$

Again, algebra reduces (8) to

$$2a^2 - 2a + (\beta+\gamma)y \geq 0. \tag{9}$$

Also, from (2), replacing $a^2 - a$ by $\frac{\gamma y}{a} - (\beta+\gamma)y$ in (9), straightforward algebra gives that

$$\frac{2\gamma y}{a} - (\beta+\gamma)y \geq 0, \quad \forall \ y > 0$$
$$\iff a \leq \frac{2\gamma}{\beta+\gamma}, \quad \forall \ \frac{\gamma}{\beta+\gamma} < a < 1 \tag{10}$$

If $r(y)$ is nonincreasing in y, (10) implies $\frac{2\gamma}{\beta+\gamma} > 1$; this proves the '*only if*' part. The '*if*' part is trivial from (10).

Lemma 2 *If* $p \leq m(\tau^2 - 1)$*, then (3) holds if and only if* $m + 4 \leq p$.

Proof: From (2), it is obvious that

$$\lim_{y\to 0} a(y) = 1 \tag{11}$$
$$\text{and} \quad \lim_{y\to\infty} a(y) = \frac{\gamma}{\beta+\gamma}.$$

¿From (5), (6) and (11), using l'Hôpital's rule, we have that

$$\begin{aligned}
\lim_{y\to 0} r(y) &\overset{(6)}{=} \lim_{y\to 0} \frac{1-a(y)}{y} \\
&= \lim_{y\to 0} -a'(y) \qquad\qquad \text{(l'Hôpital's rule)} \\
&\overset{(5)}{=} \lim_{y\to 0} -\frac{\gamma - (\beta+\gamma)a(y)}{3a^2(y) - 2a(y) + (\beta+\gamma)y} \\
&\overset{(11)}{=} -\frac{\gamma - (\beta+\gamma)}{3 - 2 + 0} \\
&= \beta.
\end{aligned}$$

Furthermore,

$$\lim_{y\to\infty} r(y) = \lim_{y\to\infty} \frac{1-a(y)}{y} = 0.$$

Since $p \leq m(\tau^2 - 1)$, it then follows from Lemma 1 that $r(y)$ is nonincreasing in y and $0 \leq r(y) \leq \beta$, $\forall \ y > 0$. By direct computation,

$$\beta \leq 2(p-2) \iff m \leq p - 4.$$

This proves the lemma.

We now state the following result without proof because it follows directly from Lemma 1 and Lemma 2.

Theorem 1 *For fixed m and τ^2, both (3) and (4) hold if and only if $m + 4 \leq p \leq m(\tau^2 - 1)$.*

Corollary 1 *For fixed m and τ^2, if $m + 4 \leq p \leq m(\tau^2 - 1)$, then the posterior mode $\underset{\sim}{\nu}$ is minimax.*

Remark: For fixed m and τ^2, the condition that $m(\tau^2 - 2) \geq 4$ is sufficient for the existence of a $p \geq 5$ which satisfies the inequalities in Theorem 1.

Example: If we take $m = 1$ and $\tau^2 = 10$, then, by Theorem 1, $5 \leq p \leq 9$ is a sufficient and necessary condition that both (3) and (4) hold.

Unlike (4), Alam [1] allowed $r(\| \underset{\sim}{X} \|^2)$ to decrease with increasing $(\| \underset{\sim}{X} \|^2)$, though not too quickly. See Efron and Morris [5]. Condition (4) is relaxed to the condition that

$$\frac{\| \underset{\sim}{X} \|^{p-2} r(\| \underset{\sim}{X} \|^2)}{2(p-2) - r(\| \underset{\sim}{X} \|^2)} \text{ is nondecreasing in } \| \underset{\sim}{X} \|^2 . \tag{12}$$

From (6), algebra reduces (12) to the condition that

$$\frac{1 - a(y)}{y^{\frac{p}{2}} \left[2(p-2) - \frac{1-a(y)}{y}\right]} \text{ is nonincreasing in } y \tag{13}$$

or

$$\frac{d}{dy} \cdot \frac{1 - a(y)}{y^{\frac{p}{2}} \left[2(p-2) - \frac{1-a(y)}{y}\right]} \leq 0, \ \forall \, y > 0. \tag{14}$$

Elementary calculations and substitutions yield that, under (3),

$$\frac{2a^2}{3a^2 - 2a + (\beta+\gamma)y} + \frac{(\beta+\gamma)a - \gamma}{2a^2} - p \leq 0, \ \forall \ \frac{\gamma}{\beta+\gamma} < a < 1. \tag{15}$$

Further simplification or characterization of (15) seems very difficult. Therefore, we directly proceed to an analysis of the "unbiased estimate of risk" itself in the next section. However, we will see subsequently that (15) has an interesting relation with the unbiased estimate of risk. In fact, from (7), if we treat $r(\cdot)$ as a function of a, $r(a) = \frac{(\beta+\gamma)a-\gamma}{a^2}$, then the following proposition gives the *l.u.b.* of $r(\cdot)$ directly.

Proposition 1 *If $\beta > \gamma$, then $r(a)$ is increasing on $(\frac{\gamma}{\beta+\gamma}, \frac{2\gamma}{\beta+\gamma})$ and decreasing on $(\frac{2\gamma}{\beta+\gamma}, 1)$; further, $\underset{\frac{\gamma}{\beta+\gamma}<a<1}{Max} r(a) = \frac{(\beta+\gamma)^2}{4\gamma}$.*

Proof: On direct computations, one has

$$\frac{dr(a)}{da} = \frac{2\gamma - (\beta+\gamma)a}{a^3}.$$

If $\beta > \gamma$, then $\frac{\gamma}{\beta+\gamma} < \frac{2\gamma}{\beta+\gamma} < 1$. The monotonicity of $r(a)$ follows directly from elementary calculations. And substituting $a = \frac{2\gamma}{\beta+\gamma}$ in (7), we have

$$\underset{\frac{\gamma}{\beta+\gamma}<a<1}{Max}\, r(a) = r(\frac{2\gamma}{\beta+\gamma}) = \frac{(\beta+\gamma)^2}{4\gamma},$$

which completes the proof.

Let $\Delta_{\underset{\sim}{\nu}}(\underset{\sim}{\theta}) = R(\underset{\sim}{\theta}, \underset{\sim}{\nu})$ - $R(\underset{\sim}{\theta}, \underset{\sim}{X})$ be the difference between the risks of the posterior mode $\underset{\sim}{\nu}$ and MLE $\underset{\sim}{X}$. In the following section, we will deal with the unique unbiased estimator of $\Delta_{\underset{\sim}{\nu}}(\underset{\sim}{\theta})$ first, denoted by $u(\underset{\sim}{X})$, and then derive a sufficient condition for $u(\underset{\sim}{X}) < 0,\ \forall\, \underset{\sim}{X}$.

3. Risk Function of the Posterior Mode and Unbiased Estimate of Risk

Recall that under squared error loss, the risk function of the posterior mode is given by

$$R(\underset{\sim}{\theta}, \underset{\sim}{\nu}) = E\left(\| \underset{\sim}{\theta} - \underset{\sim}{\nu} \|^2\right)$$

which depends on $\underset{\sim}{\theta}$ through $\lambda = \frac{\|\underset{\sim}{\theta}\|^2}{2}$. Stein [10] observed that for any absolutely continuous function $h(X_i)$ with Lebesgue measurable derivative $h'(X_i)$ satisfying $E|h'(X_i)| < \infty$,

$$E(X_i - \theta_i)h(X_i) = Eh'(X_i). \qquad \text{(Stein's Identity)}$$

If $\delta(\underset{\sim}{X}) = \underset{\sim}{X} + \underset{\sim}{h}(\underset{\sim}{X})$, to show that $\delta(\underset{\sim}{X})$ is better than $\underset{\sim}{X}$, $\forall\, \underset{\sim}{\theta}$, we need to show that

$$\Delta(\underset{\sim}{\theta}) = R(\underset{\sim}{\theta}, \delta(\underset{\sim}{X})) - R(\underset{\sim}{\theta}, \underset{\sim}{X}) \ < \ 0, \quad \forall\, \underset{\sim}{\theta}.$$

But

$$\begin{aligned}
\Delta(\underset{\sim}{\theta}) &= E(\| \underset{\sim}{\theta} - \delta(\underset{\sim}{X}) \|^2) - E(\| \underset{\sim}{\theta} - \underset{\sim}{X} \|^2) \\
&= E(\| \underset{\sim}{\theta} - \underset{\sim}{X} - \underset{\sim}{h}(\underset{\sim}{X}) \|^2) - \| \underset{\sim}{\theta} - \underset{\sim}{X} \|^2) \\
&= E\left(\| \underset{\sim}{h}(\underset{\sim}{X}) \|^2) + 2(\underset{\sim}{X} - \underset{\sim}{\theta})'\underset{\sim}{h}(\underset{\sim}{X})\right).
\end{aligned}$$

From Stein's Identity, we get that

$$\Delta(\underset{\sim}{\theta}) = E\left(\sum_{i=1}^{p} h_i^2(\underset{\sim}{X}) + 2\sum_{i=1}^{p} \frac{\partial h_i(\underset{\sim}{X})}{\partial X_i}\right),$$

if $\underset{\sim}{h}$ satisfies the assumptions in Stein's Identity.

Thus,

$$u(\underset{\sim}{X}) = \sum_{i=1}^{p} h_i^2(\underset{\sim}{X}) + 2\sum_{i=1}^{p} \frac{\partial h_i(\underset{\sim}{X})}{\partial X_i} \tag{16}$$

is an unbiased estimator of $\Delta(\underset{\sim}{\theta})$, and

$$u(\underset{\sim}{X}) \; < \; 0, \qquad\qquad \forall\, \underset{\sim}{X} \tag{17}$$

is a sufficient condition for $\Delta(\underset{\sim}{\theta}) \; < \; 0 \; \forall \, \underset{\sim}{\theta}$. See Stein [10], [11], Hwang [7], [8], Hudson [6] and Berger [3] etc.

If we take $\delta(\underset{\sim}{X}) = \underset{\sim}{v} = aX$ and $\underset{\sim}{h}(\underset{\sim}{X}) = (a-1)\underset{\sim}{X}$, where '$a$' is the unique root of (2), then $\underset{\sim}{v}$ is of the form $\underset{\sim}{X} + \underset{\sim}{h}(\underset{\sim}{X})$. Since '$a$' is a scalar function depending on $\underset{\sim}{X}$ through only $\| \underset{\sim}{X} \|^2$, we have

$$\sum_{i=1}^{p} h_i^2(\underset{\sim}{X}) = (a-1)^2 \sum_{i=1}^{p} X_i^2 = (a-1) \, \| \underset{\sim}{X} \|^2 . \tag{18}$$

Furthermore,

$$\begin{aligned} \frac{\partial h_i(\underset{\sim}{X})}{\partial X_i} &= (a-1) + X_i \frac{\partial a}{\partial X_i} \\ &= (a-1) + X_i \frac{da}{d \| \underset{\sim}{X} \|^2} \frac{\partial \| \underset{\sim}{X} \|^2}{\partial X_i} \\ &= (a-1) + 2X_i^2 \frac{da}{d \| \underset{\sim}{X} \|^2}. \end{aligned}$$

Thus,

$$\sum_{i=1}^{p} \frac{\partial h_i(\underset{\sim}{X})}{\partial X_i} = (a-1)p + 2 \, \| \underset{\sim}{X} \|^2 \, \frac{da}{d \| \underset{\sim}{X} \|^2}. \tag{19}$$

From (16), (18) and (19), one has

$$u(\underset{\sim}{X}) = (a-1)^2 \, \| \underset{\sim}{X} \|^2 + 2(a-1)p + 4 \, \| \underset{\sim}{X} \|^2 \, \frac{da}{d \| \underset{\sim}{X} \|^2}, \tag{20}$$

which is the unbiased estimator of

$$\begin{aligned} & \Delta_{\underset{\sim}{\nu}}(\underset{\sim}{\theta}) \\ = \; & E(\| \underset{\sim}{\theta} - \underset{\sim}{\nu} \|^2) - E(\| \underset{\sim}{\theta} - \underset{\sim}{X} \|^2) \\ = \; & E(\| \underset{\sim}{\theta} - \underset{\sim}{\nu} \|^2) - p, \end{aligned}$$

provided the expectation of each term in (20) exists.

Remark: Note that (20) shows that $u(\cdot)$ depends on $\underset{\sim}{X}$ through $\| \underset{\sim}{X} \|^2$ only. Also, $u(\underset{\sim}{X}) + p$ is the unique unbiased estimator of the risk of the posterior mode $\underset{\sim}{\nu}$. The uniqueness is due to that $\| \underset{\sim}{X} \|^2$ has a non-central chi-square distribution which is complete. It is obvious that (17) is a sufficient condition for the minimaxity of $\underset{\sim}{\nu}$. Therefore, the characterization of condition (17) is our next task. For convenience, we denote $z = \| \underset{\sim}{X} \|^2$; thus both $u(\cdot)$ and $a(\cdot)$ are functions of z.

From (2), by implicit differentiation, we have

$$\frac{da}{dz} = \frac{a^2(1-a)}{(3a^2 - 2a)z + (\beta + \gamma)}. \tag{21}$$

Also, (20) reduces to

$$u(\underset{\sim}{X}) = u(z) = (1-a)\left[(1-a)z + \frac{4z\frac{da}{dz}}{1-a} - 2p\right]. \tag{22}$$

Letting

$$g(z) = \frac{z\frac{da}{dz}}{1-a},$$

from (21), direct computation gives that

$$\begin{aligned} g(z) &= \frac{a^2 z}{(3a^2 - 2az) + (\beta + \gamma)} \\ &= \frac{a^2}{(3a^2 - 2a) + (\beta + \gamma)y}, \end{aligned} \tag{23}$$

where $y = \frac{1}{z}$. Note that $g(z) > 0, \; \forall \, z > 0$. Since $\frac{\gamma}{\beta+\gamma} < a < 1$, it follows, from (22) and (23), that (17) holds if and only if

$$(1-a)z + 4\frac{a^2}{3a^2 - 2a + (\beta+\gamma)y} - 2p < 0. \tag{24}$$

Recall from (7) that

$$(1-a)z = \frac{1-a}{y} = \frac{(\beta+\gamma)a-\gamma}{a^2}.$$

Multiplying (15) by 2, it is clear that (24) is equivalent to (15). That is, (12) subject to (3) is equivalent to $u(X) \le 0, \forall X$. Recall that, from (2),

$$z = \| X \|^2 = \frac{(\beta+\gamma)a-\gamma}{a^2(1-a)}, \tag{25}$$

which gives an explicit expression of z in terms of 'a'. For simplicity, we will sometimes treat $u(\cdot)$, $z(\cdot)$ as functions of 'a'. We are now in a position to prove the following results, which establishes a simple characterization of (24).

Theorem 2 *If $\beta \le \gamma$, then (17) holds if and only if $\beta \le 2(p-2)$.*

Proof: Recall that $r(z) = (1-a)z$ and $g(z) = \frac{z\frac{da}{dz}}{1-a}$. Then, from (22),

$$u(z) = (1-a)(r(z) + 4g(z) - 2p). \tag{26}$$

Since $0 < a < 1$, one has

$$\begin{aligned} & u(z) < 0, \quad \forall z > 0 \\ \iff & r(z) + 4g(z) < 2p, \quad \forall z > 0. \end{aligned} \tag{27}$$

It is mathematically difficult to obtain the *l.u.b.* of $r(z)+4g(z)$, but finding the *l.u.b.*'s of $r(z)$ and $g(z)$ separately is easier. If $\beta \le \gamma$, it follows directly from Lemma 1 that $r(z)$ is nondecreasing in z and $\lim_{z\to\infty} r(z) = \beta$, which is the *l.u.b.* of $r(z)$. Thus, on direct computations, one has

$$\frac{dr(z)}{dz} = (1-a) - z\frac{da}{dz} \ge 0, \quad \forall z > 0. \tag{28}$$

Moreover, dividing (28) by $1-a$, since $0 < a < 1$, we have that

$$\begin{aligned} \frac{\frac{dr(z)}{dz}}{1-a} &= 1 - \frac{z\frac{da}{dz}}{1-a} \\ &= 1 - g(z) \ge 0, \quad \forall z > 0 \end{aligned} \tag{29}$$

which implies that

$$g(z) \le 1, \quad \forall z > 0.$$

In fact, from (23), it is clear that

$$\lim_{z\to\infty} g(z) = 1,$$

which is the *l.u.b.* of $g(z)$. Indeed, $\beta+4$ is the *l.u.b.* of $r(z)+4g(z)$, because both *l.u.b.'s* of $r(z)$ and $g(z)$ are obtained $z\to\infty$. Thus, the restriction that $\beta+4 < 2p$ is equivalent to that $\beta < 2(p-2)$, which completes the proof.

Remark: The condition that both $\beta \le \gamma$ and $\beta \le 2(p-2)$ hold is equivalent to $m+4 \le p \le m(\tau^2-1)$ as in Theorem 1.

Corollary 2 *Under (4), (17) holds if and only if (3) holds.*

Theorem 3 *If $\gamma < \beta \le 8\gamma$, then (17) holds if $\frac{(\beta+\gamma)^2}{4\gamma} + \frac{8(\gamma-\beta)^2}{\beta(8\gamma-\beta)} < 2(p-2)$.*

Proof: Recall that $r(\cdot), g(\cdot)$ and $u(\cdot)$ are functions of 'a'. By Proposition 1, we know that $\frac{(\beta+\gamma)^2}{4\gamma}$ is the *l.u.b.* of $r(\cdot)$. Now, let us deal with $g(\cdot)$ first. Combining (23) and (25), on direct computations, it is obtained that

$$g(a) = \frac{[(\beta+\gamma)a-\gamma]\,a}{2(\beta+\gamma)a^2-(\beta+4\gamma)a+2\gamma}. \tag{30}$$

Let A, B denote the numerator and denominator of (30) separately. Note that B is positive, since both A and $g(\cdot)$ are positive. Then, on simplification, one has

$$B-A = (\beta+\gamma)a^2-(\beta+3\gamma)a+2\gamma. \tag{31}$$

Since the discriminant

$$\begin{aligned}\Delta &= (\beta+3\gamma)^2-8\gamma(\beta+\gamma)\\ &= (\beta+3\gamma)^2>0,\end{aligned}$$

we claim that

$$\begin{aligned} & B-A \ \ge\ 0, \qquad \text{if } a\le\frac{2\gamma}{\beta+\gamma} \\ \text{and}\qquad & B-A \ <\ 0, \qquad \text{if } a>\frac{2\gamma}{\beta+\gamma}. \end{aligned} \tag{32}$$

This is because, on direct computation,

$$B-A=0 \Longleftrightarrow a=\frac{2\gamma}{\beta+\gamma} \quad \text{or} \quad 1.$$

Furthermore, the minimum value of $B - A$, $\frac{-(\gamma-\beta)^2}{4(\beta+\gamma)}$, is obtained at $a = \frac{\beta+3\gamma}{2(\beta+\gamma)}$. That is,

$$\frac{-(\gamma-\beta)^2}{4(\beta+\gamma)} < B - A < 0, \qquad \text{if } \frac{2\gamma}{\beta+\gamma} < a < 1. \tag{33}$$

Dividing (32) and (33) by B, since B is positive, algebra implies that

$$\frac{A}{B} \leq 1, \qquad \text{if } a \leq \frac{2\gamma}{\beta+\gamma};$$

$$\text{and} \qquad 1 \leq \frac{A}{B} \leq 1 + \frac{(\gamma-\beta)^2}{4(\beta+\gamma)B}, \quad \text{if } a > \frac{2\gamma}{\beta+\gamma}. \tag{34}$$

Again, algebra results in that the minimum value B is attained at $a = \frac{\beta+4\gamma}{4(\beta+\gamma)}$,

$$\text{i.e.,} \qquad B \geq \frac{\beta(8\gamma-\beta)}{8(\beta+\gamma)}. \tag{35}$$

If $\beta \leq 8\gamma$, from (34) and (35), it then follows that

$$1 \leq \frac{A}{B} \leq 1 + \frac{2(\gamma-\beta)^2}{\beta(8\gamma-\beta)}, \quad \text{if } a > \frac{2\gamma}{\beta+\gamma}. \tag{36}$$

By Proposition 1 and (36), we have that

$$r(a) + 4g(a) < 2p, \quad \forall\, a$$

if

$$\frac{(\beta+\gamma)^2}{4\gamma} + \frac{8(\gamma-\beta)^2}{\beta(8\gamma-\beta)} < 2p - 4.$$

This proves the theorem.

Remark: In fact, if $\gamma < \beta < 4\gamma$, then $\frac{2\gamma}{\beta+\gamma} > \frac{\beta+4\gamma}{4(\beta+\gamma)}$; therefore, $B \geq \frac{2\gamma^2}{\beta+\gamma}$, which is attained at $a = \frac{2\gamma}{\beta+\gamma}$, if $a > \frac{2\gamma}{\beta+\gamma}$. Thus, (34) reduces to

$$1 \leq \frac{A}{B} \leq 1 + \frac{(\gamma-\beta)^2}{8\gamma^2}, \qquad \gamma < \beta < 4\gamma. \tag{37}$$

Note that the upper bound of (37) is sharper than that of (36); this is because

$$\frac{2(\gamma-\beta)^2}{\beta(8\gamma-\beta)} - \frac{(\gamma-\beta)^2}{8\gamma^2} = \frac{(\gamma-\beta)^2(4\gamma-\beta)^2}{8\gamma^2\beta(8\gamma-\beta)} > 0.$$

Corollary 3 *(3) holds if (17) holds.*

Proof: If $\beta \leq \gamma$, it is trivial from Corollary 2. If not, taking $a = \frac{2\gamma}{\beta+\gamma}$, one has that

$$\begin{aligned} & r(a) + 4g(a) \\ = \; & \frac{(\beta+\gamma)^2}{4\gamma} + 4. \end{aligned}$$

The result follows directly from (26) and Proposition 1.

Remark: This Corollary obeys Baranchik's theorem. Also, it points out that condition (3) is weaker than condition (17). That is, for fixed m and τ^2, the interval of p for which (3) holds is wider than the interval of p for which (17) holds. Recall that (12) subject to (3) is equivalent to (17). Thus Cor 3 explores a surprising result that (17) and Alam's conditions are equivalent.

Example: Let $m = 1$, $\tau^2 = 10$ and $\sigma^2 = 1$. Figures 1 and 2 illustrate the behavior of $u(z)$ for various values of p. Recall that $u(\cdot)$ is a function of $\underset{\sim}{X}$ through $\| \underset{\sim}{X} \|^2$ only, and $\| \underset{\sim}{X} \|^2$ has a noncentral chi-square distribution with parameter $\frac{\|\underset{\sim}{\theta}\|^2}{2}$. Thus, the expected value of $u(\| \underset{\sim}{X} \|^2)$ is a function of $\| \underset{\sim}{\theta} \|^2$ only, which can be numerically computed by using IMSL subroutines. Figures 3 and 4 give the plots of $E_{\underset{\sim}{\theta}}\left(u(\| \underset{\sim}{X} \|^2)\right)$ vs. $\| \underset{\sim}{\theta} \|^2$ for various values of p for the fixed m, τ^2 and σ^2. Since the risk function of the posterior mode is equal to $R(\underset{\sim}{\theta}, \underset{\sim}{\nu}) = E_{\underset{\sim}{\theta}}\left(u(\| \underset{\sim}{X} \|^2)\right) + p$, Figures 3 and 4 also give the illustration of risk functions. Note that for the minimaxity of the posterior mode, the condition $E_{\underset{\sim}{\theta}}\left(u(\| \underset{\sim}{X} \|^2)\right) < 0 \; \forall \, \underset{\sim}{\theta}$, is weaker than the condition $u(\| \underset{\sim}{X} \|^2) < 0 \; \forall \, \underset{\sim}{X}$.

4. Comparison With Posterior Mean

For $p > 1$, both posterior mean and posterior mode are of the form $\left[1 - \frac{r(\|\underset{\sim}{X}\|^2)}{\|\underset{\sim}{X}\|^2}\right] \underset{\sim}{X}$. As stated in the introduction, unlike the posterior mode, the posterior mean, $\hat{\underset{\sim}{\mu}}(\underset{\sim}{X})$, needs to be approximated by numerical methods. It seems necessary that a comparison between the posterior mean and the posterior mode be made. Figure 5-10 give the plots of the posterior mean and the posterior mode with $p = 1$ for some fixed m, τ^2 and σ^2. A plot of their risk functions is given to help make a comparison. Figure 10 describes the risk behavior for both the posterior mean and the posterior mode. The risk functions are computed by using IMSL subroutines. Figure 10 shows that both the posterior mean and the posterior mode are tail minimax when $p = 1$. Indeed, they both are tail minimax for all p. For higher p, the risk behavior of the posterior mode is described in Figure 4. It shows that if we take $m = 1$, $\tau^2 = 10$ and

$\sigma^2 = 1$, then the posterior mode $\underset{\sim}{v}$ is minimax for moderate values of p. Figure 10 indicates that the posterior mode is shrinking more than posterior mean; this can be a reason why the posterior mean is not minimax for any $p < \infty$.

References

[1] K. Alam. A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, 1:517–525, 1973.

[2] A. J. Baranchik. A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.*, 41:642–645, 1970.

[3] J. Berger. Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Stat.*, 8:545–571, 1980.

[4] A. Dasgupta and J. K. Ghosh and M. M. Zen. Bayesian analysis of a multivariate normal mean with flat tailed priors. Technical Report #90-22, 1990.

[5] B. Efron and C. Morris. Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Stat.*, 4(1):11–21, 1976.

[6] H. M. Hudson. A natural identity for exponential families with applications in multivariate estimation. *Ann. Statist.*, 6:473–484, 1978.

[7] J. T. Hwang. Certain bounds on the class of admissible estimators in continuous exponential families. *Statistical Decision Theory and Related Topics III, S. S. Gupta and J. Berger (eds.)*, 1982.

[8] J. T. Hwang. Improving upon standard estimators in discrete exponential families with applications to Possion and negative binomial cases. *Ann. Statist*, 10:857–867, 1982.

[9] W. James and C. Stein. Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1:361–379, 1961.

[10] C. Stein. Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statist.*, pages 345–381, 1973.

[11] C. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9:1135–1151, 1981.

Figure 1. The Curve of $u(z) : p = 1, 2, 3, 4, 5$

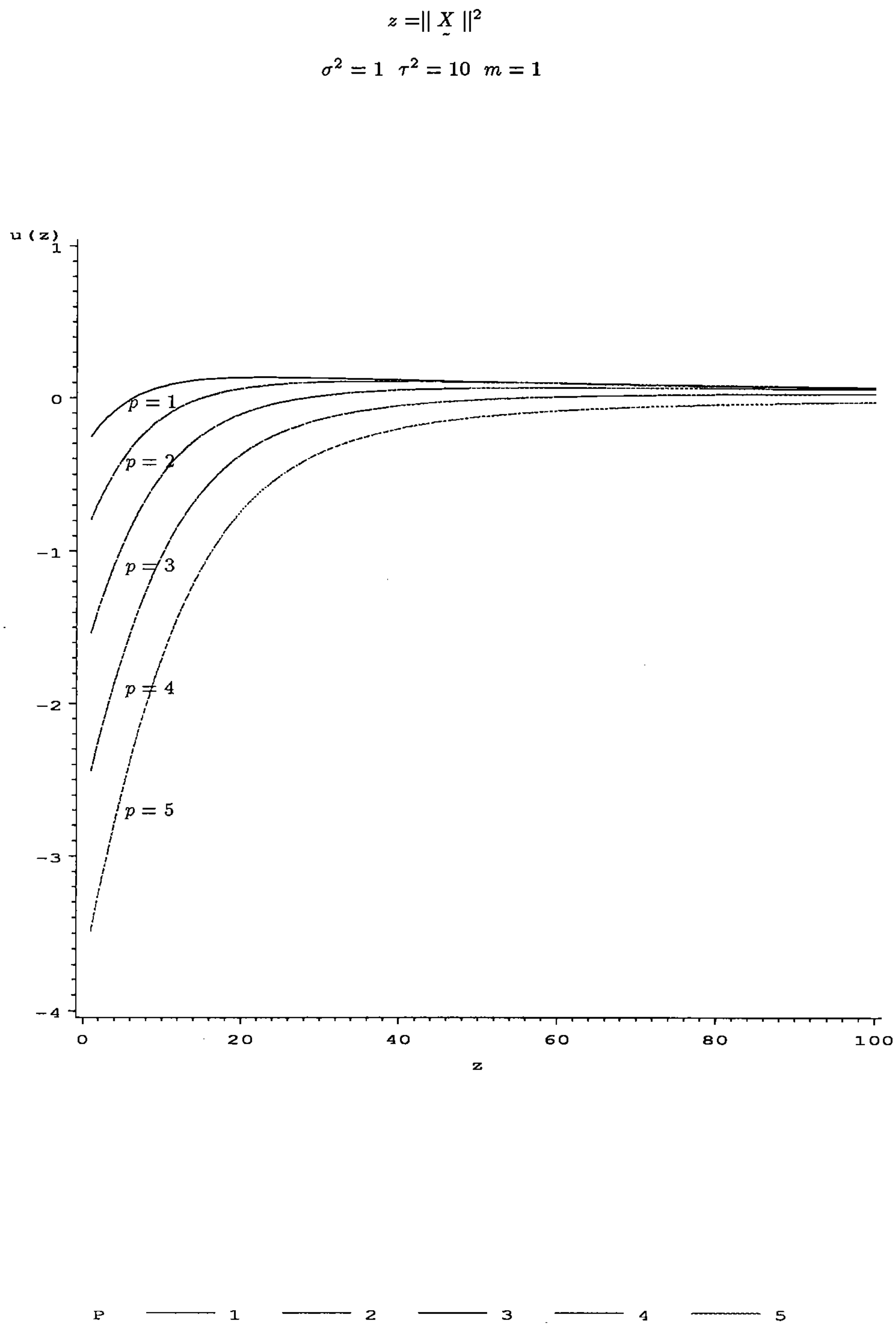


Figure 2. The Curve of $u(z) : p = 5, 25, 45, 65, 85$

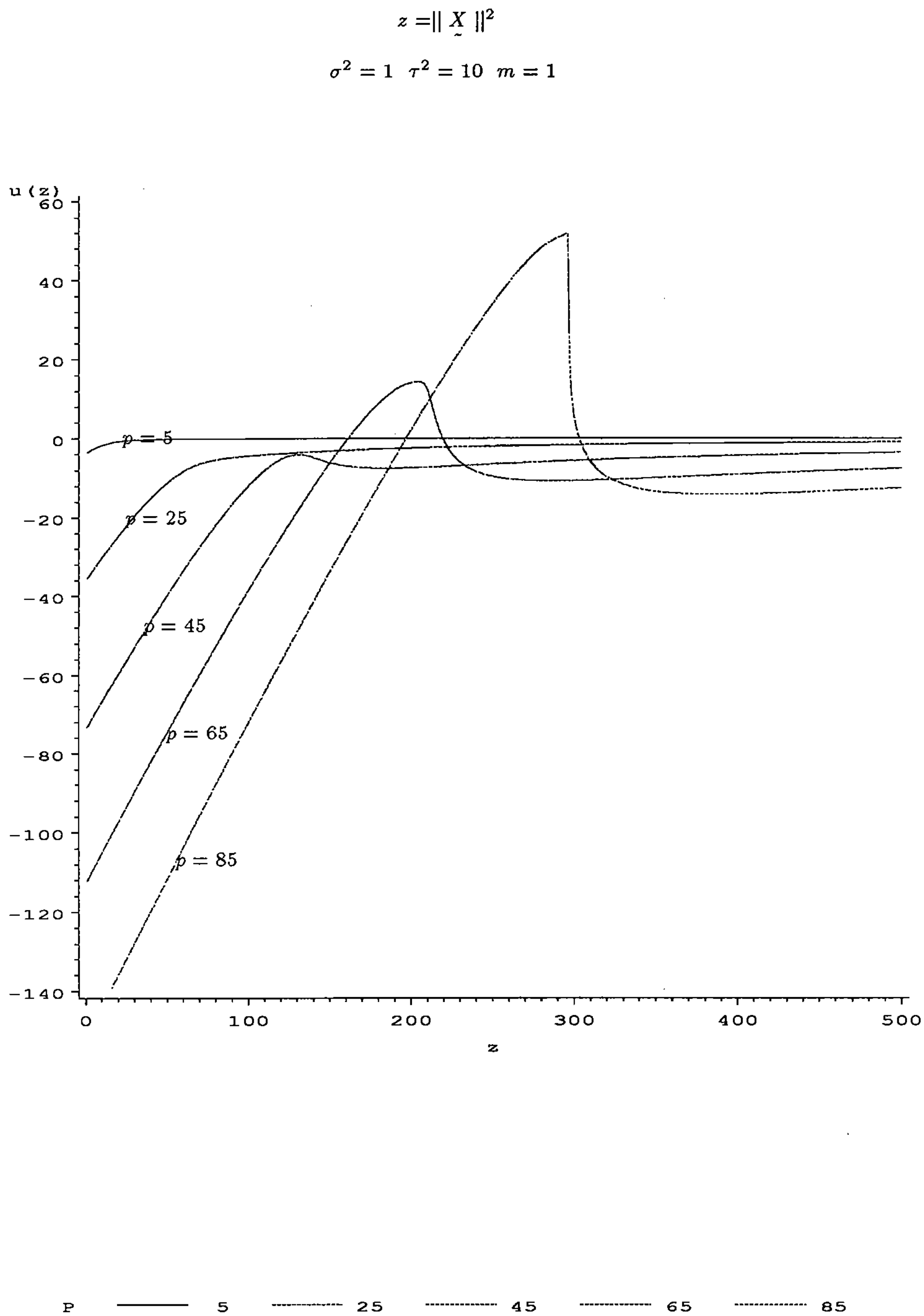


Figure 3. Curves of E(u(Z)) vs. $\lambda : p = 1, 2, 3, 4, 5$

$Z = \| \underset{\sim}{X} \|^2 \quad \lambda = \| \underset{\sim}{\theta} \|^2$
$\sigma^2 = 1 \quad m = 1 \quad \tau^2 = 10$

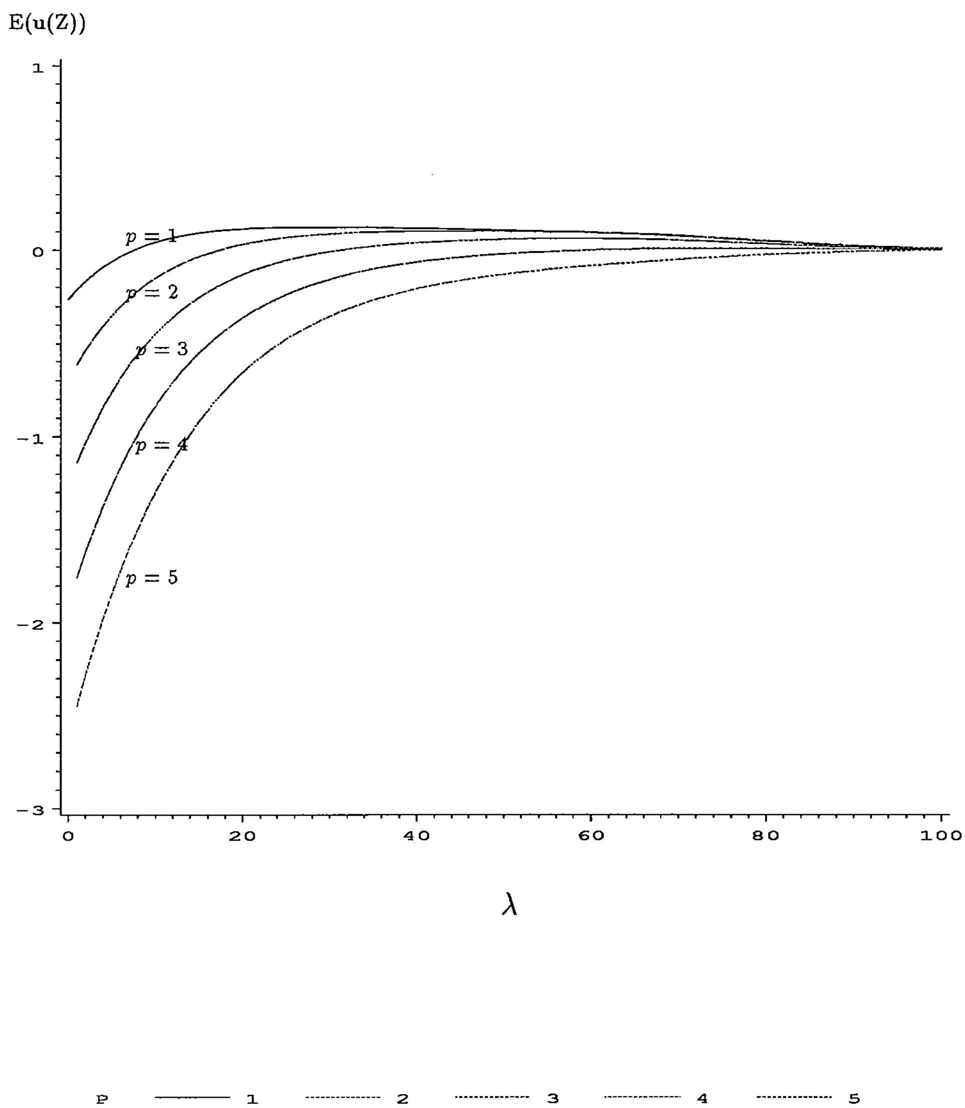


Figure 4. Curves of E(u(Z)) vs. $\lambda : p = 5, 25, 45, 65$

$$Z = \| \underset{\sim}{X} \|^2 \quad \lambda = \| \underset{\sim}{\theta} \|^2$$
$$\sigma^2 = 1 \quad m = 1 \quad \tau^2 = 10$$

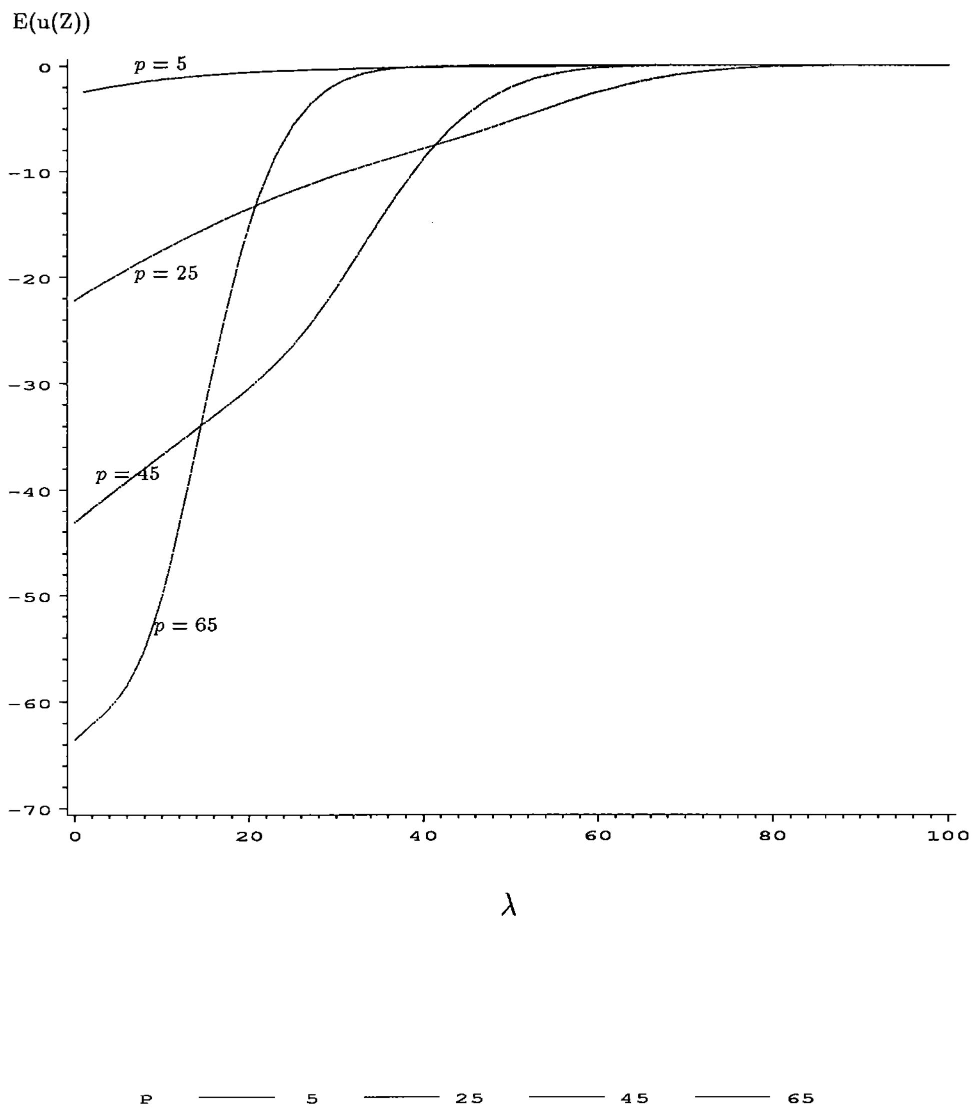


Figure 5. Plot of Posterior Mean and Mode (i)

$\sigma^2 = 1 \;\; \tau^2 = 0.5 \;\; m = 1 \;\; p = 1$

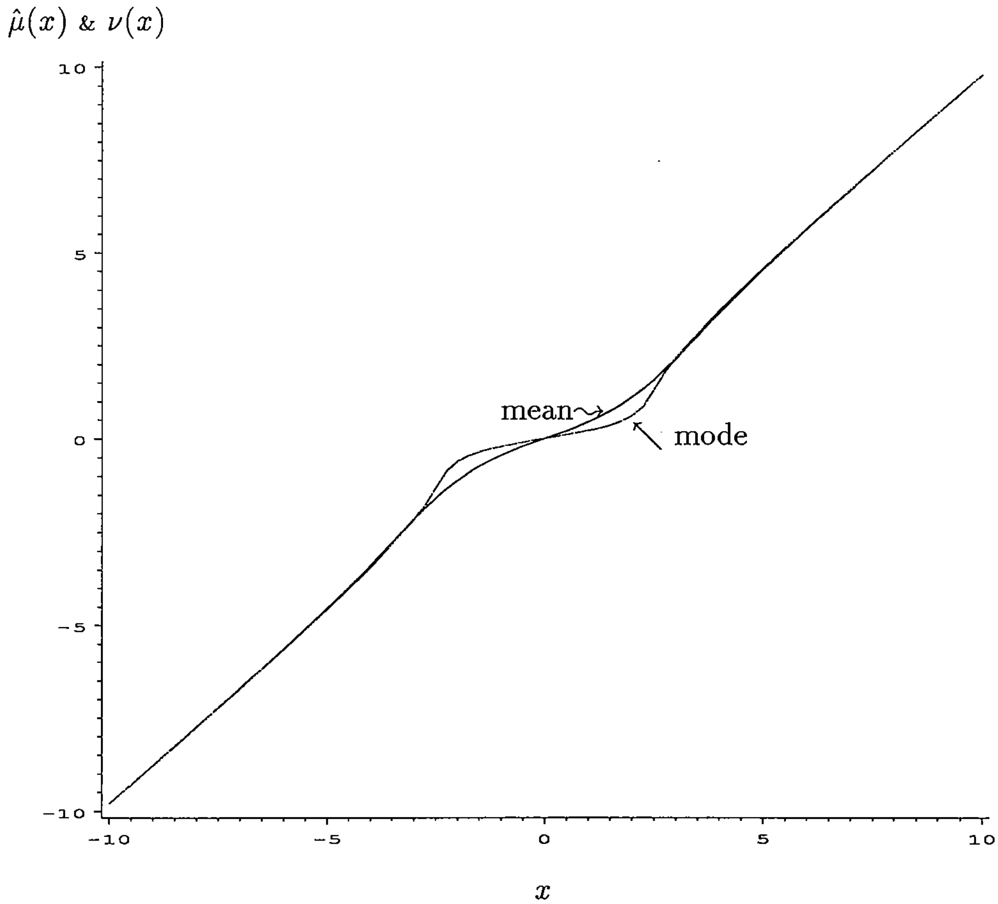


Figure 6. Plot of Posterior Mean and Mode (ii)

$\sigma^2 = 1 \;\; \tau^2 = 1 \;\; m = 1 \;\; p = 1$

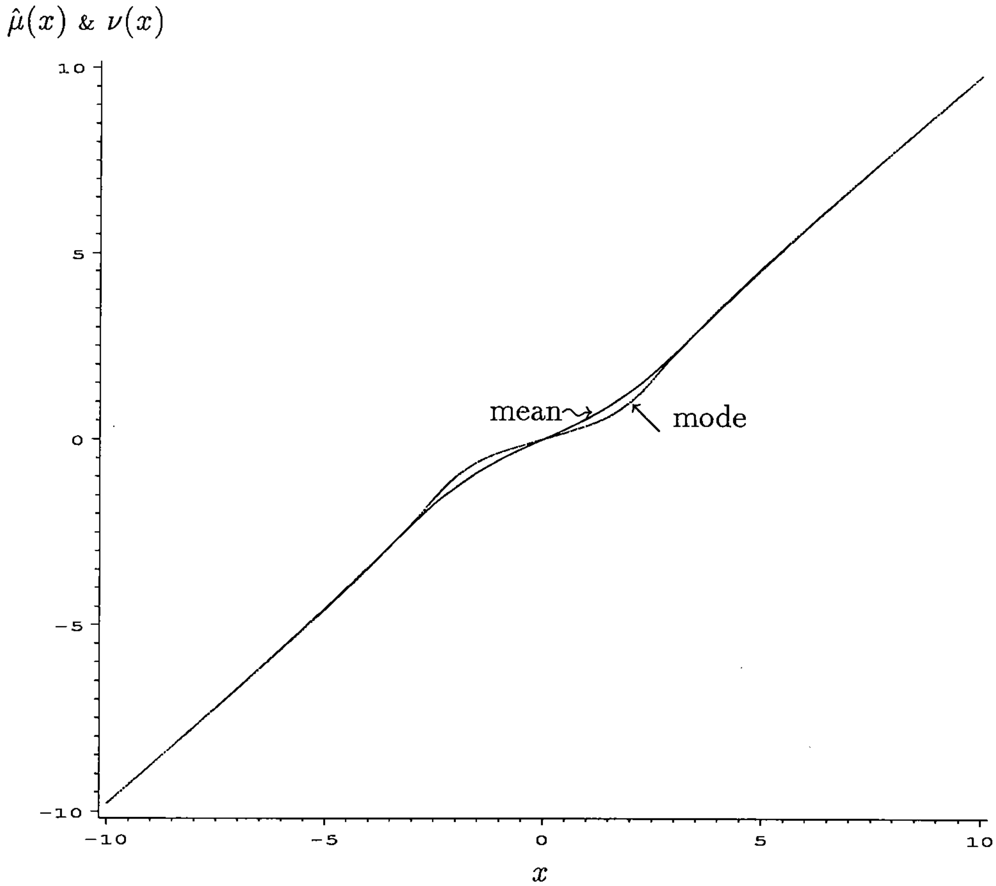


Figure 7. Plot of Posterior Mean and Mode (iii)

$\sigma^2 = 1 \;\; \tau^2 = 10 \;\; m = 1 \;\; p = 1$

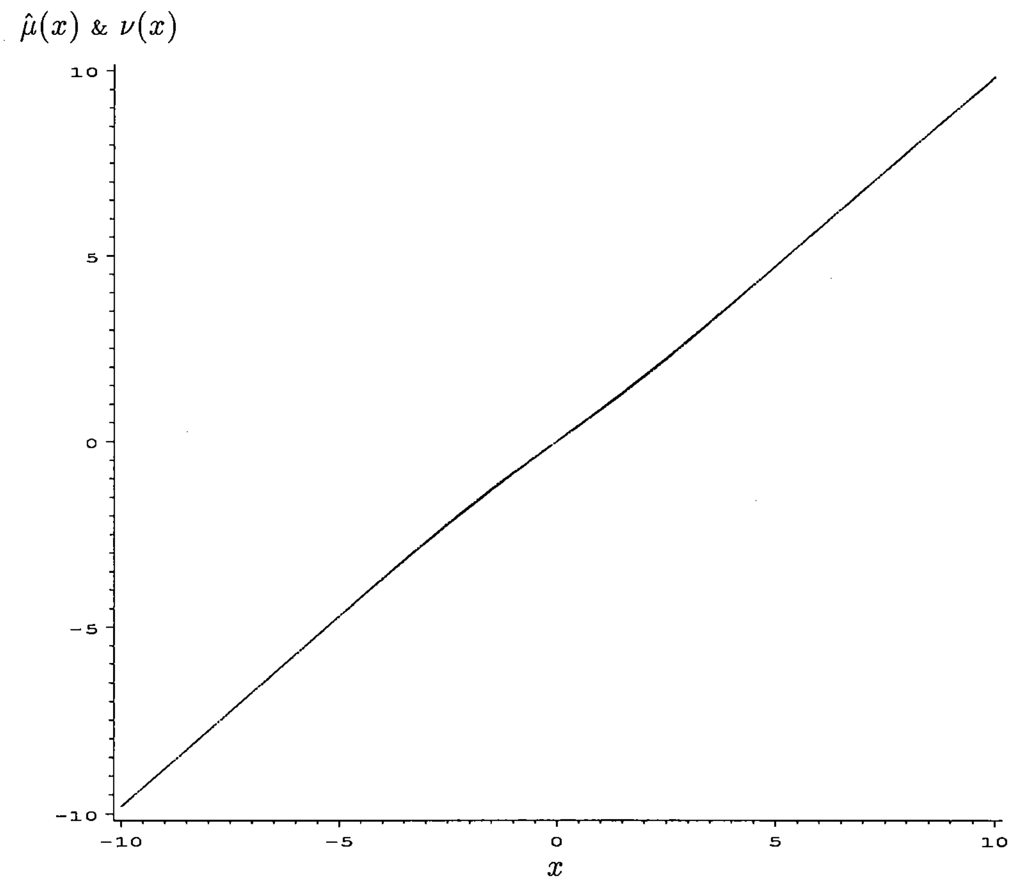


Figure 6. Plot of Posterior Mean and Mode (ii)

$\sigma^2 = 1 \;\; \tau^2 = 1 \;\; m = 1 \;\; p = 1$

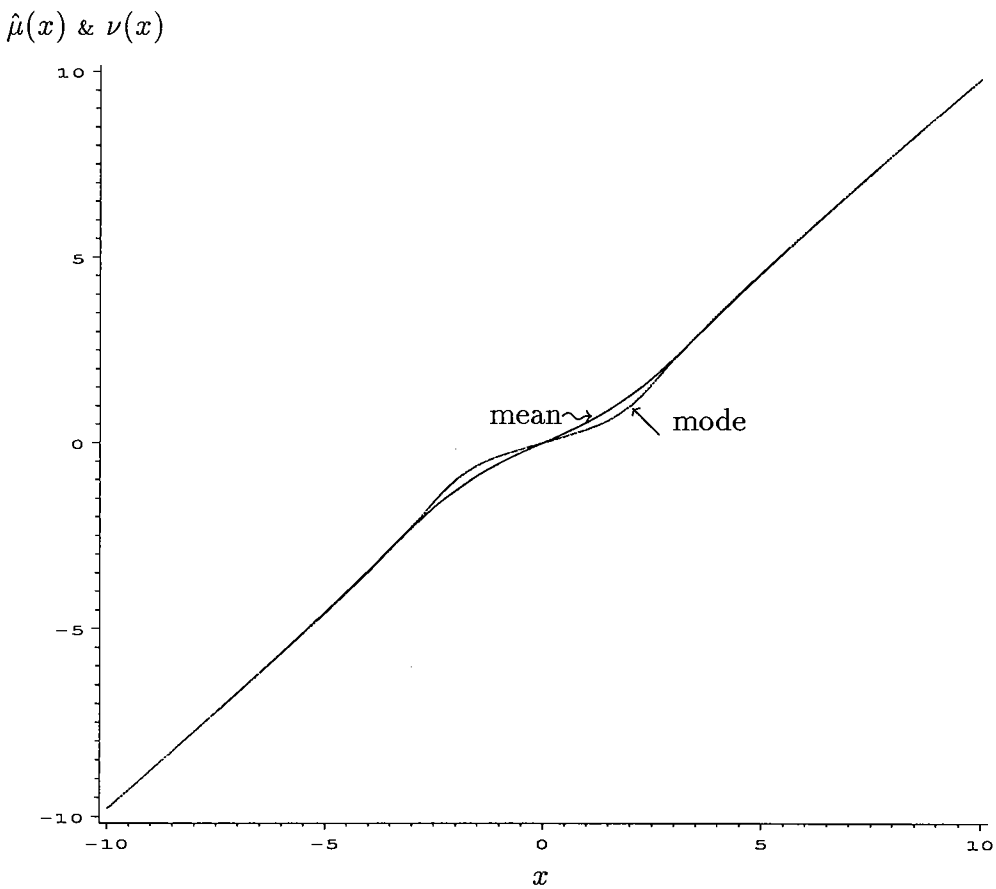
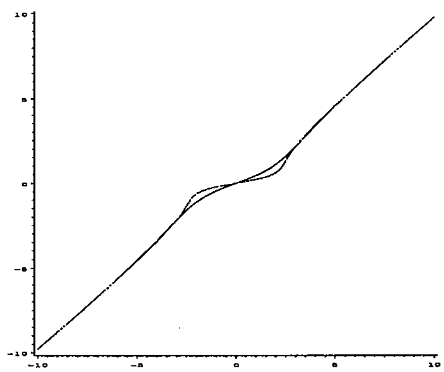


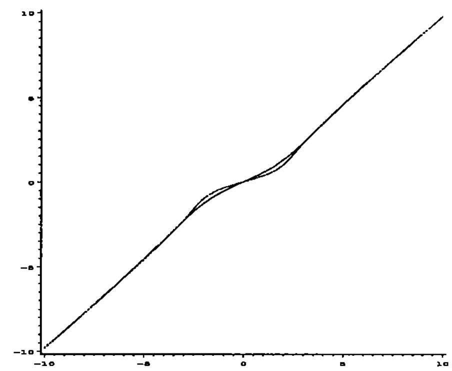
Figure 9. Plots of Posterior Mean and Mode

$\sigma^2 = 1 \quad p = 1$

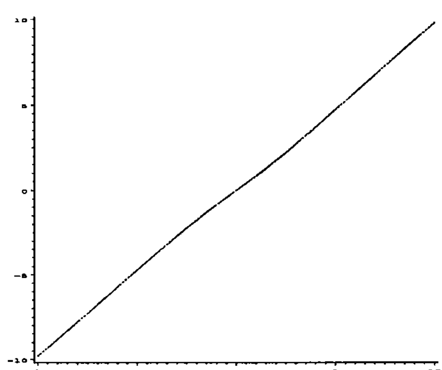
$\tau^2 = 0.5 \;\; m = 1$



$\tau^2 = 1 \;\; m = 1$



$\tau^2 = 10 \;\; m = 1$



$\tau^2 = 1 \;\; m = 3$

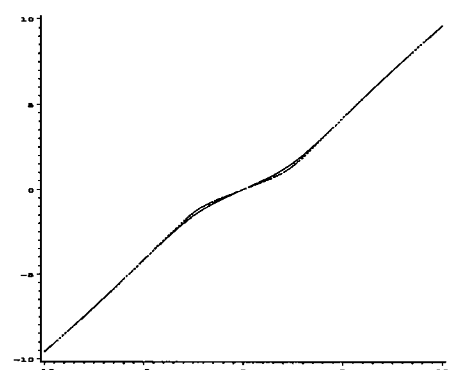


Figure 10. Risk Functions of Posterior Mean and Mode

$\sigma^2 = 1 \;\; \tau^2 = 1 \;\; m = 3 \;\; p = 1$

$\lambda = \theta^2$