POINT ESTIMATION OF NORMAL MEAN USING T PRIORS

Mei-Mei Zen
Department of Statistics
National Cheng-kung University
Tainan 70101, Taiwan, ROC
email:mmzen@ibm.stat.ncku.edu.tw

1 Introduction

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

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

the maximum likelihood estimator (MLE) $\hat{\underline{\theta}}^0 = \underline{X}$ has risk $R(\underline{\theta}, \hat{\underline{\theta}}^0) = p$. James and Stein (1961) showed that the estimator $\hat{\underline{\theta}}^1 = \left(1 - \frac{(p-2)\sigma^2}{\|\underline{X}\|^2}\right)\underline{X}$ does better than MLE $\hat{\underline{\theta}}^0$. Its risk $R(\underline{\theta}, \hat{\underline{\theta}}^1)$, which is a function only of $\lambda = \frac{\|\underline{\theta}\|^2}{2\sigma^2}$, increases from 2 at $\lambda = 0$ to the minimax value p as $\lambda \to \infty$. Baranchik (1970) proved that, under squared error loss function, an estimator of the form

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

is minimax under certain conditions. We consider Bayesian point estimation of a multivariate normal mean $\underline{\theta}$ under t priors. DasGupta, Ghosh and Zen(1995) found that if the posterior distribution of $\underline{\theta}$ is starunimodal about the mode $\underline{\nu}$, then $\underline{\nu}$ can be expressed as $\underline{\nu} = a\underline{X}$, where '$a$' is the unique root of

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

Although the closed form of 'a' is messy, 'a' is a function of $\| \underline{X} \|^2$ indeed; therefore, $\underline{\nu} = (1 - (1-a))\underline{X}$, which is in the form of (1), is employed as an point estimator of $\underline{\theta}$ in this article.

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 $\| \underline{X} \|^2$ 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 $\hat{\underline{\mu}}(\underline{X})$ 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 $R(\underline{\theta}, \underline{\nu})$ and $R(\underline{\theta}, \underline{X})$ by $\Delta_{\underline{\nu}}(\underline{\theta})$. Then the unique unbiased estimator of $\Delta_{\underline{\nu}}(\underline{\theta})$, denoted by $u(\underline{X})$, can be derived in a closed form. Therefore, $u(\underline{X}) + p$ can serve as the unbiased estimator of the risk function of $\underline{\nu}$. If the posterior is starunimodal, then 'a' is a function of $\underline{X}$ through $\| \underline{X} \|^2$ only. Thus, $u(\cdot)$ can be treated as a function of $\| \underline{X} \|^2$. Bounding $u(\| \underline{X} \|^2)$ above by 0, it is found that $\underline{\nu}$ does better than MLE $\hat{\underline{\theta}}^0$ 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.

2 Minimaxity of the Posterior Mode

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

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

where $r(\| \underline{X} \|^2) = (1-a) \| \underline{X} \|^2$, it is of the form (1). Baranchik (1970) proved that for $p \geq 3$ the conditions

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

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

are sufficient that $\underline{\nu}$ be minimax.

Using two Lemmas, we obtain a sufficient condition for the minimaxity of the posterior mode.

<u>Lemma 1</u>

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

<u>Lemma 2</u>

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

<u>Theorem 1</u>

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

<u>Corollary 1</u>

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

Unlike (4), Alam (1973) allowed $r(\| \underline{X} \|^2)$ to decrease with increasing $(\| \underline{X} \|^2)$, though not too quickly. See Efron and Morris (1976). Condition (4) is relaxed to the condition that

$$\frac{\| \underline{X} \|^{p-2}\, r(\| \underline{X} \|^2)}{2(p-2) - r(\| \underline{X} \|^2)} \quad \text{is nondecreasing in} \quad \| \underline{X} \|^2 .$$

Further simplification or characterization of the above condition 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 (6) has an interesting relation with the unbiased estimate of risk. In fact, 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, $\underset{\frac{\gamma}{\beta+\gamma}<a<1}{Max}\; r(a) = \frac{(\beta+\gamma)^2}{4\gamma}$.

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

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

Under squared error loss, the risk function of the posterior mode is given by

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

which depends on $\underline{\theta}$ through $\lambda = \frac{\|\underline{\theta}\|^2}{2}$. Stein (1973) observed that for any absolutely continuous function $h(X_i)$ with Lebesgue measurable derivative $h^{'}(X_i)$ satisfying $E|h^{'}(X_i)| < \infty$,

$$E(X_i - \theta_i)h(X_i) = Eh^{'}(X_i). \tag{5}$$

If $\delta(\underline{X}) = \underline{X} + \underline{h}(\underline{X})$, to show that $\delta(\underline{X})$ is better than $\underline{X}$, $\forall\ \underline{\theta}$, we need to show that $\Delta(\underline{\theta}) = R(\underline{\theta}, \delta(\underline{X})) - R(\underline{\theta}, \underline{X}) < 0, \quad \forall\ \underline{\theta}$. From (5), we get that

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

if $\underline{h}$ satisfies the assumptions in Stein's Identity.
Thus,

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

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

$$u(\underline{X}) < 0,\ \forall\ \underline{X} \tag{6}$$

is a sufficient condition for $\Delta(\underline{\theta}) < 0\ \forall\ \underline{\theta}$. See Stein (1973)(1981), Hwang (1982a)(1982b), Hudson (1978) and Berger (1980) etc.

If we take $\delta(\underline{X}) = \underline{\nu} = aX$ and $\underline{h}(\underline{X}) = (a-1)\underline{X}$, then $\underline{\nu}$ is of the form $\underline{X} + \underline{h}(\underline{X})$. Since a is a scalar function depending on $\underline{X}$ through only $\| \underline{X} \|^2$, we have $\sum_{i=1}^{p} h_i^2(\underline{X}) = (a-1) \| \underline{X} \|^2$ and

$$u(\underline{X}) = (a-1)^2 \| \underline{X} \|^2 + 2(a-1)p + 4 \| \underline{X} \|^2 \frac{da}{d \| \underline{X} \|^2}, \tag{7}$$

which is the unbiased estimator of

$$\Delta_{\underline{\nu}}(\underline{\theta}) = E(\| \underline{\theta} - \underline{\nu} \|^2) - p,$$

provided the expectation of each term in (7) exists. Note that (7) shows that $u(\cdot)$ depends on $\underline{X}$ through $\| \underline{X} \|^2$ only. Also, $u(\underline{X}) + p$ is the unique unbiased estimator of the risk of the posterior mode $\underline{\nu}$. The uniqueness is due to that $\| \underline{X} \|^2$ has a non-central chi-square distribution which is complete. It is obvious that (6) is a sufficient condition for the minimaxity of $\underline{\nu}$. Therefore, the characterization of condition (6) is our next task.

Theorem 2

If $\beta \le \gamma$, then (6) holds if and only if $\beta \le 2(p-2)$.

Note that 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.

Theorem 3

If $\gamma < \beta \le 8\gamma$, then (6) holds if $\frac{(\beta+\gamma)^2}{4\gamma} + \frac{8(\gamma-\beta)^2}{\beta(8\gamma-\beta)} < 2(p-2)$.

4 Comparison With Posterior Mean

For $p > 1$, both posterior mean and posterior mode are of the form $\left[1 - \frac{r(\|\underline{X}\|^2)}{\|\underline{X}\|^2}\right]\underline{X}$. Unlike the posterior mode, the posterior mean, $\underline{\hat{\mu}}(\underline{X})$, needs to be approximated by numerical methods. It seems necessary that a comparison between the posterior mean and the posterior mode be made. The plots of the posterior mean and the posterior mode and a plot of their risk functions could be given to help make a comparison. The risk functions show 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 shows that if we take $m = 1$, $\tau^2 = 10$ and $\sigma^2 = 1$, then the posterior mode $\underline{\nu}$ is minimax for moderate values of p. Moreover, 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. afamily of a 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 familyes with applications to simultaneous estimation of gamma scale parameters. Ann. Stat., 8:545-571, 1980.

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

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

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

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

[8] A.Dasgupta, J.K. Ghosh and M.M. Zen. A new general method for constructing confidence sets in arbitrary dimensions: with applications. Ann. Statist., 23:1408-1432. 1995.

[9] C. Stain. Estimation of the mean of a multivariate normal distribution. Proc. Prague Symp. Asymptotic Statist., 345-381,1973