Estimation of a Mean Vector in Two Sample Problems

by

François Perron
Université de Montréal and Purdue University

Technical Report # 90–34C

Department of Statistics
Purdue University

July, 1990

△

ESTIMATION OF A MEAN VECTOR IN TWO SAMPLE PROBLEMS

by

François Perron
Université de Montréal and Purdue University

ABSTRACT

Abstract. We consider the problem of estimating a p-dimensional vector μ_1 based on independent variables X_1, X_2 and U where X_1 is $N_p(\mu_1, \sigma^2\Sigma_1), X_2$ is $N_p(\mu_2, \sigma^2\Sigma_2)$ and U is $\chi_n^2(\Sigma_1$ and Σ_2 are known). A family of minimax estimators is proposed. Bayesian estimators are also developed. These estimators are expressed as the sum of two shrinkage estimators. Finally, the idea of shrinking towards an independent estimators is explored. This idea is generalized to an abstract problem.

Key words and phrases: Bayes risk, group, invariance, minimaxity, right hand modulus, shrinkage estimator.

AMS 1980 Subject classifications: Primary 62F10, 62H12; Secondary 62C12.

1. Introduction

Imagine a situation in which two related experiments are conducted independently. In each experiment p-dimensional valued measurements are recorded. The data is summarized into one mean vector per experiment (X_1 and X_2 say) plus an univariate variable U. Let us denote by μ_1 and μ_2 the expectation of X_1 and X_2 respectively. Since the two experiments are related μ_2 is in the vicinity of μ_1. Thus, not only X_1 and U but also X_2 can be used for estimating μ_1.

Assume from now on that X_1 and X_2 are normally distributed with mean vectors μ_1 and μ_2 and covariance matrices $\sigma^2\Sigma_1$, and $\sigma^2\Sigma_2(X_1 \sim N_p(\mu_1, \sigma^2\Sigma_1), X_2 \sim N_p(\mu_2, \sigma^2\Sigma_2))$. Assume also that U/σ^2 has a chi-square distribution ($U \sim \sigma^2\chi_n^2$). The matrices Σ_1 and Σ_2 are positive definite and are supposed known. Let $A = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}, Y_1 = A(X_1 - X_2)$ and $Y_2 = X_1 - Y_1$. When μ_2 is equal to μ_1 the statistic (U, Y_2) is sufficient. However, if μ_2 is far apart from μ_1, X_2 is of little use in the inference. Intuitively, depending on whether $\mu_2 \approx \mu_1$ (or equivalently $\theta_1 = E[Y_1] = A(\mu_1 - \mu_2) \approx 0$) one should use a different estimator. A first way of approaching the estimation of μ_1 consists in using a preliminary test estimator (PTE). Namely, one has to test $H_0 : \theta_1 = 0$ against $H_1 : \theta_1 \neq 0$. This test might depend on (Y_1, U). If the test is rejected then the inference is based on (X_1, U). Otherwise, the inference is based on (Y_2, U).

The lack of continuity of the PTE implies its inadmissibility. More smoothness is needed. A second approach to the problem of estimating μ_1 is the following. Rather than basing the inference on (X_1, U) or (Y_2, U), according to the value of (Y_1, U), we can generalize the PTE approach by considering a convex combination of X_1 and Y_2 depending on the value of (Y_1, U). In other words, $\hat{\mu}_1 = (1 - \beta(Y_1, U))X_1 + \beta(Y_1, U)Y_2 = X_1 - \beta(Y_1, U)(X_1 - Y_2)$ with β taking values in [0, 1]. The last expression in the equalities is the one proposed by Ghosh and Sinha (1988, page 213, expression 3.2). Let us call this estimator μ_1^*. The two authors give an unbiased estimator of the risk function, derive minimax estimators using both the frequentist and bayesian arguments and propose estimators which dominate the PTE.

In this article, we introduce a third approach. The basic idea of this new approach will consist of adapting the technique developed for shrinkage estimators to our problem. Actually, μ_1^* can be rewritten as $\mu_1^* = (1 - \beta(Y_1, U))Y_1 + Y_2$. This procedure shrinks Y_1 toward zero and does not affect Y_2. We might think of shrinking Y_2 as well and considering $\hat{\mu}_1 = \Sigma_{i=1}^2(1 - h_i(Y_i, U))Y_i$. Conditions on h_1 and h_2 under which this estimator is minimax are given in Section 2 along with an unbiased estimator of the risk functions. In the remaining part of the paper σ^2 is assumed to be known and we do not make use of the variable U. Bayesian aspects of the problem are taken in consideration in Section 3. Admissible results are also developed. In Section 4, the original problem is formulated in terms of an abstract one. Minimax and generalized Bayes estimators are given. These estimators generalized μ_1^*.

2. Minimax results

Let X_1, X_2 and U be independent, $X_1 \sim N_p(\mu_1, \sigma^2\Sigma_1), X_2 \sim N_p(\mu_2, \sigma^2\Sigma_2)$ and $U \sim \sigma^2\chi_n^2$. Consider the loss function given by $L(\mu_1, \mu_2, \sigma^2; \hat{\mu}_1) = (\hat{\mu}_1 - \mu_1)'Q(\hat{\mu}_1 - \mu_1)/\sigma^2$. It is well known that X_1 is a minimax estimator of μ_1 with constant risk tr $(Q\Sigma_1)$. In this section we propose a family of minimax estimators of μ_1. This family generalizes the one proposed by Ghosh and Sinha (1988). Minimaxity is obtained by comparing an unbiased estimate of the risk to tr $(Q\Sigma_1)$. The estimate is derived by using the following identities;

a) $E[(Z-\theta)h(Z)] = \sigma^2 E[h'(Z)]$ if $Z \sim N(\theta, \sigma^2)$,

b) $E[(V-nb)h(V)] = 2bE[Vh'(V)]$ if $V \sim b\chi^2_n$ (2.1)

where h satisfies the conditions for integrating by parts (cf. Efron and Morris 1976).

Let us first transform (X_1, X_2) into (Y_1, Y_2) as we did in Section 1. Thus, $Y_1 = A(X_1 - X_2)$ and $Y_2 = X_1 - Y_1$ with $A = \Sigma_1(\Sigma_1+\Sigma_2)^{-1}$. Similarly, let $\theta_1 = A(\mu_1 - \mu_2), \theta_2 = \mu_1 - \theta_1$ and set $\Lambda_i = A\Sigma_i, i = 1, 2$. The problem is now reformulated in terms of estimating $\mu_1 = \theta_1 + \theta_2$ based on Y_1, Y_2 and U where Y_1, Y_2 and U are independent, $Y_i \sim N_p(\theta_i, \Lambda_i), i = 1, 2$ and $U \sim \sigma^2\chi^2_n$. Think of $\hat{\mu}_1$ as $\hat{\mu}_1 = \hat{\theta}_1 + \hat{\theta}_2$ where $\hat{\theta}_i$ is an estimate of θ_i based on (Y_i, U) and let $\hat{\theta}_i = (1 - g_i(F_i)/F_i)Y_i$ where $F_i = (n+2)Y_i'\Lambda_i^{-1}Y_i/U, i = 1, 2$. Define also T_i and Δ as $T_i = Y_i'QY_i/Y_i'\Lambda_i^{-1}Y_i, i = 1, 2$ and $\Delta = \Sigma_{i=1}^2\{2[tr(Q\Lambda_i) - 2T_i]g_i(F_i)/F_i + 4g_i'(F_i)T_i - g_i^2(F_i)T_i/F_i + 4g_i(F_i)g_i'(F_i)T_i/(n+2)\} + \{4g_1'(F_1)g_2(F_2)/F_2 + 4g_1(F_1)g_2'(F_2)/F_1 - 2(n+2)g_1(F_1)g_2(F_2)/F_1F_2\}Y_1'QY_2$.

Theorem 2.1. If the functions g_i satisfy the regularity conditions for integrating by parts and Δ is positive then $\hat{\mu}_1$ is minimax.

Proof. Let $R^* = tr(Q\Sigma_1) - R$. Showing minimaxity is equivalent of showing that R^* is positive. Developing R^* we get

$$R^* = tr(Q\Sigma_1) - E(X_1 - \mu_1)'Q(X_1 - \mu_1)/\sigma^2$$

$$+ 2E\{((g_1(F_1)/F_1)Y_1 + (g_2(F_2)/F_2)Y_2)'Q(Y_1 - \theta_1 + Y_2 - \theta_2)\}/\sigma^2$$

$$- E\{((g_1(F_1)/F_1)Y_1 + (g_2(F_2)/F_2)Y_2)'Q((g_1(F_1)/F_1)Y_1 + (g_2(F_2)/F_2)Y_2)\}/\sigma^2$$

$$= \sum_{i=1}^{2} E\{2(g_i(F_i)/F_i)Y_i'Q(Y_i - \theta_i) - (g_i(F_i)/F_i)^2 Y_i'QY_i\}/\sigma^2$$

$$- 2E\{g_1(F_1)g_2(F_2)Y_1'QY_2/F_1F_2\}/\sigma^2$$

$$= E\Delta$$

and the last equality comes from the identities given in expression (2.1). ■

Actually, Δ is an unbiased estimator of R^*. Characterizing all the functions g_1 and g_2 such that Δ is positive is not an easy task. In the particular case where g_2 is null Efron and Morris (1976) give the solution for $\Lambda_1 Q = I$. This result can be easily generalized for $\Lambda_1 Q \neq I$. It gives a family of estimators slightly larger than the one proposed by Ghosh and Sinha (1988). If we want g_1 and g_2 both positive the next theorem might be useful. The idea of the proof consists in making positive an estimator of R^* who underestimate R^*.

Before giving the theorem let us introduce some notations. If A is a $p \times p$ matrix define $ch_{\max}(A)$ as the largest eigenvalue of A. Let $\gamma_i(T_i) = tr(Q\Lambda_i)/T_i - 2, \overline{\gamma}_i = tr(Q\Lambda_i)/ch_{\max}(Q\Lambda_i) - 2$ and $\delta_i(T_i) = 4/(n+2) + 2/\gamma_i(T_i)$.

Theorem 2.2. If the regularity conditions for integrating by parts g_1 and g_2 are satisfied and

a) $\overline{\gamma}_i > 0 \qquad i = 1, 2$

b) $0 \leq g_i \leq \overline{\gamma}_i$

c) $xg_i(x)^{2/\gamma_i(x)}/(\gamma_i(x) - g_i(x))^{\delta_i(x)}$ is nondecreasing in $x, i = 1, 2$, then $\hat{\mu}_1$ is minimax.

Proof. We have

$$
\begin{aligned}
R^* &= 2\sum_{i=1}^{2} E\{(g_i(F_i)/F_i)Y_i'Q(Y_i-\theta_i)\}/\sigma^2 \\
&\quad - E\{((g_1(F_1)/F_1)Y_1+(g_2(F_2)/F_2)Y_2)'Q((g_1(F_1)/F_1)Y_1+(g_2(F_2)/F_2)Y_2\}/\sigma^2 \\
&\geq 2\sum_{i=1}^{2} E\{(g_i(F_i)/F_i)Y_i'Q(Y_i-\theta_i)-(g_i(F_i)/F_i)^2Y_i'QY_i\}/\sigma^2 \\
&= 2\sum_{i=1}^{2} E\{(tr(\Lambda_iQ)-2T_i)(g_i(F_i)/F_i)-T_ig_i^2(F_i)/F_i \\
&\qquad + 2T_ig_i'(F_i)+4T_ig_i(F_i)g_i'(F_i)/(n+2)\} \\
&= 2\sum_{i=1}^{2} E\left\{\frac{T_i(\gamma_i(T_i)-g_i(F_i))^{\delta_i(T_i)+1}}{F_ig_i(F_i)^{(2/\gamma_i(T_i))-1}}\frac{\partial}{\partial F_i}\frac{F_ig_i(F_i)^{2/\gamma_i(T_i)}}{(\gamma_i(T_i)-g_i(F_i))^{\delta_i(T_i)}}\right\} \\
&\geq 0 \qquad \blacksquare
\end{aligned}
$$

Corollary. If the condition c) in Theorem 2.2 is replaced by saying that g_i is increasing, $i=1,2$, then the result holds.

3. Bayesian results

Let Y_1 and Y_2 be independent, $Y_1 \sim N_p(\theta_1, \sigma^2\Lambda_1)$ and $Y_2 \sim N_p(\theta_2, \sigma^2\Lambda_2)$. We want to estimate $\mu_1 = \theta_1 + \theta_2$ and the loss function is quadratic. (See Section 2 for the connection between this problem and the original one.) For a quadratic loss the Bayes estimator is simply the posterior mean of μ_1, or, equivalently, the sum of the posterior means of θ_1 and θ_2. Taking independent priors on θ_1 and θ_2 leads to an estimator of the form $\hat{\mu}_1 = E[\theta_1|Y_1] + E[\theta_2|Y_2] = \hat{\theta}_1 + \hat{\theta}_2$ (say). Thus, our problem is reduced to the more familiar one of estimating the mean of a multivariate normal distribution in one sample problem. Many results exists on this subject (cf. Berger and Robert 1990 for a good discussion). Depending on the priors we can recapture some of the estimators introduced in Section 2 (cf. Berger 1980).

In the following we shall develop a new estimator for estimating the mean of a multivariate normal distribution in one sample problem. Let Y be distributed as a $N_p(\theta, \sigma^2\Lambda)$ and suppose that we want to develop a prior for θ when σ^2 and Λ are known. Assume that $\Lambda = cI$. In many cases it is reasonable to believe that the components of θ are exchangeable. Thus, one can determine a hierarchical prior the following way. On the first stage θ is $N_p(\mu\mathbf{1}, A)$ where $\mathbf{1} = (1, \ldots, 1)'$ and $A = \sigma_\pi^2(\rho\mathbf{1}\mathbf{1}' + (1-\rho)I), \sigma_\pi^2 \geq 0, -1/(p-1) \leq \rho \leq 1$. On the second stage a joint distribution on σ_π^2 and ρ is chosen.

We expressed A in terms of σ_π^2 and ρ for two reasons. Firstly, these parameters represent directly the variances and the correlations. Secondly, for any vector having exchangeable components with second moments there exists $\sigma_\pi^2 \geq 0$ and $-1/(p-1) \leq \rho \leq 1$ such that the covariance matrix is given by $\sigma_\pi^2(\rho\mathbf{1}\mathbf{1}'+(1-\rho)I)$. Although this representation has its advantages we would like to switch to another one by setting $\alpha = \sigma_\pi^2\{1+(p-1)\rho\}$ and $\beta = \sigma_\pi^2(1-\rho)$. In this new representation $A = \alpha P + \beta(I-P)$ where $P = (1/p)\mathbf{1}\mathbf{1}'$. This representation makes computations simple and we can interpret it easily. Actually, α represents the variance of $\mathbf{1}'\theta/\sqrt{p}$ and β is the variance of $\eta'\theta$ for any vector η satisfying $\eta'\eta = 1$ and $\eta'\mathbf{1} = 0$.

Let ω be the joint density of (α, β). Given (α, β) we have

$$\begin{pmatrix} X \\ \theta \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} c\sigma^2 I + A & A \\ A & A \end{pmatrix}\right)$$

so the Bayes estimator becomes

$$\begin{aligned}
\hat{\theta} = E[\theta|x] &= x - c\sigma^2 E[(c\sigma^2 I + A)^{-1}|x](x - \mu\mathbf{1}) \\
&= x - E[(1+\alpha/c\sigma^2)^{-1}|x]P(x-\mu\mathbf{1}) - E[(1+\beta/c\sigma^2)^{-1}|x](I-P)(x-\mu\mathbf{1}) \\
&= x - E[(1+\alpha/c\sigma^2)^{-1}|x](\overline{x}-\mu)\mathbf{1} - E[(1+\beta/c\sigma^2)^{-1}|x](x-\overline{x}\mathbf{1})
\end{aligned}$$

and the conditional density of (α, β) is given by

$$\omega(\alpha, \beta | x) \;\propto\; (1+\alpha/c\sigma^2)^{-1/2}(1+\beta/c\sigma^2)^{-(p-1)/2} \exp -\frac{1}{2}\Big\{ p(\overline{x}-\mu)^2(1+\alpha/c\sigma^2)^{-1} + s^2(1+\beta/c\sigma^2)^{-1} \Big\} \omega(\alpha, \beta)$$

with $\overline{x} = \mathbf{1}'x/p$ and $s^2 = ||x - \overline{x}\mathbf{1}||^2$.

Depending on the assumptions on g some simplifications are possible. Assuming independence between α and β ($\omega(\alpha, \beta) = \omega_1(\alpha)\omega_2(\beta)$ say) we get

$$\hat{\theta} = x - h_1(p(\overline{x}-\mu)^2/2c\sigma^2)(\overline{x}-\mu)\mathbf{1} - h_2(s^2/2c\sigma^2)(x - \overline{x}\mathbf{1})$$

where h_1 and h_2 are both increasing functions,

$$h_i = \frac{\int_0^\infty (1+t/c\sigma^2)^{-(p_i/2+1)} \exp -\frac{1}{2}\{a_i(1+t/c\sigma^2)^{-1}\}\omega_i(t)dt}{\int_0^\infty (1+t/c\sigma^2)^{-p_i/2} \exp -\frac{1}{2}\{a_i(1+t/c\sigma^2)^{-1}\}\omega_i(t)dt} \quad i = 1, 2$$

with $p_1 = 1, p_2 = p-1, a_1 = p(\overline{x}-\mu)^2/2c\sigma^2, a_2 = s^2/2c\sigma^2$. In practice, h_1 and h_2 can be evaluated numerically.

We can even obtain closed expressions for $h_i, i = 1, 2$, if we make the prior dependent of the parameters c and σ^2 through $c\sigma^2$. In fact, if we set $u_1 = (1+\alpha/c\sigma^2)^{-1}, u_2 = (1+\beta/c\sigma^2)^{-1}$ and give density f_i to $u_i, i = 1, 2$ where

$$f_i(t) \;\propto\; t^{(m_i-p_i)/2-1}e^{-b_i t/2}, t \in (0, 1)$$

then we get

$$h_i(x) = m_i(x+b_i)\left(1 - \left[\frac{m_i}{2}\int_0^{\prime} \lambda^{m_i/2-1} \exp -\{(\lambda-1)(x+b_i)/2\} d\lambda\right]^{-1}\right)$$

and $\frac{m}{2}\int_0^{\prime} \lambda^{m/2-1} \exp -\{(\lambda-1)b\} d\lambda$

$$= \begin{cases} \frac{b^{m/2}}{(m/2)![\exp(b)-\Sigma_{i=0}^{m/2-1} b^i/i!]} & \text{if } m \text{ is even} \\ \frac{b^{m/2}}{\Gamma(m/2+1)[\exp(b)P[|N(0,1)|<\sqrt{2b}]-\Sigma_{i=0}^{(m-3)/2} b^{(i+1/2)}/\Gamma(i+3/2)]} & \text{if } m \text{ is odd} \end{cases}$$

(cf. Berger 1980).

The last result is interesting because it leads to an admissible estimator easy to compute. However, in a practical situation, c will depend on the sample size so the prior will change according to the sample size which is troublesome from a bayesian point of view.

4. Abstract results

In this section we shall extend the original problem to a more abstract one. The key element is going to be the invariance. Let us start with two examples.

In the first example X_1 is $N_p(\mu, \Sigma), \Sigma$ is known, and X_2 is independent of X_1. We desire to estimate μ. No further assumptions are made on the distribution of X_2. We assume that we have two estimators of μ (a minimax estimator $\hat{\mu}_1$ based on X_1 and another estimator $\hat{\mu}_2$ based on X_2). We can combine these two estimators and obtain a minimax estimator $\hat{\mu}$ by setting $\hat{\mu}(X_1, X_2) = \hat{\mu}_2(X_2) + \hat{\mu}_1(X_1 - \hat{\mu}_2(X_2))$. For instance, if A is positive definite, $\hat{\mu}_1(X_1) = X_1 - \frac{\Phi(X_1'(\Sigma+A)^{-1}X_1)}{X_1'(\Sigma+A)^{-1}X_1}\Sigma(\Sigma+A)^{-1}X_1$ and $\hat{\mu}_2(X_2) = X_2$ then, subject to some regularity conditions (cf. Berger 1980), $\hat{\mu}_1$ is generalized Bayes and minimax. Moreover, if we take $\Sigma = \sigma^2\Sigma_1$ and $A = \sigma^2\Sigma_2$ then we obtain μ_1^*. Intuitively, if the risk function of $\hat{\mu}_1$ is minimum at the origin and $\hat{\mu}_2$ is an accurate estimate of μ then the risk function of $\hat{\mu}$ will be smaller than the one of $\hat{\mu}_1$ for large values of μ.

In our second example S_1 is Wishart $W_p(\Sigma, n)$, S_2 is independent of S_1 and we desire to estimate Σ. A minimax estimator of Σ is given by $\hat{\Sigma}_1$ where $\hat{\Sigma}_1$ depends only on S_1.

Another estimator of Σ based on S_2 is given by $\hat{\Sigma}_2$. This second estimator is always positive definite. Write $\hat{\Sigma}_2$ as $\hat{\Sigma}_2 = TT'$ where T is lower triangular (the Cholesky decomposition). Define $\hat{\Sigma}$ as $\hat{\Sigma}(S_1, S_2) = T\hat{\Sigma}_1(T^{-1}S_1T^{-1'})T'$. This new estimator is minimax again. If the risk function of $\hat{\Sigma}_1$ is minimum at $\Sigma = I$ and $\hat{\Sigma}_2$ is accurate then the risk function of $\hat{\Sigma}$ will be smaller than the one of $\hat{\Sigma}_1$ as $(\Sigma - I)$ becomes "large".

Consider now an invariant problem with a group G, a sample space $\mathcal{X}$, a parameter space Θ, a decision space $\mathcal{D}$ and a loss function L. Let R be the risk function. Corresponding to $(g, x, \theta, d) \in G \times \mathcal{X} \times \Theta \times \mathcal{D}$ denote by $(\tilde{g}x, \overline{g}\theta, g^*d)$ the outcome resulting from the transformations, on $\mathcal{X} \times \Theta \times \mathcal{D}$, induced by $g, g \in G$. Assume that G is transitive on Θ (that is, for any $\theta, \theta_0 \in \Theta$ there is a $g \in G$ such that $\theta = \overline{g}\theta_0$). Given $g \in G$ define δ^g as $\delta^g(x) = g^*\delta(\tilde{g}^{-1}x)$ for all $x \in \mathcal{X}$. Suppose that δ is minimax and let $m = \inf\{R(\theta, \delta) : \theta \in \Theta\}$ and $\Omega = \{\theta \in \Theta : R(\theta, \delta) = m\}$. Since the problem is invariant we have $R(\theta, \delta^g) = R(\overline{g}^{-1}\theta, \delta)$ which implies that δ^g is also minimax. If we happen to know a $g \in G$ such that $\overline{g}^{-1}\theta \in \Omega$ we can replace δ by δ^g. Usually we do not know which g satisfies the relation $\overline{g}^{-1}\theta \in \Omega$ but we can always guess one and replace δ by δ^g keeping in mind that δ^g still minimax (cf. Stein 1962).

The basic idea in this section consists in replacing what is considered as an arbitrary guess for g by an estimate coming from another independent experiment. Independence will preserve minimaxity. In fact, if M denotes the minimax risk then

$$R(\theta, \delta^g) = E_\theta[R(\theta, \delta^g)|g] \leq E_\theta[M|g] = M$$

no matter what is the distribution of g. In the first example we took $g = \hat{\mu}_2(X_2)$ and in the second example we had $g = T$. Suppose now that δ is Bayes. Is δ^g Bayes also? We shall answer this question by considering the case where g is fixed and then the one where g is estimated.

Let $g \in G$ and $\theta_0 \in \Theta$ be fixed. Since G is transitive on Θ assume that the prior distribution on Θ is induced by a probability measure on $G(P(\{\theta : \theta \in A\}) = P(\{g \in G : \overline{g}\theta_0 \in A\}))$ having density π with respect to a left invariant measure λ on G. Denote the corresponding Bayes risk by $R(\pi, \delta)$ where δ represents the estimator. Finally, define π^g as $\pi^g(h) = \pi(g^{-1}h), g, h \in G$, and denote the density of X, with respect to a measure μ, by $f(x;\theta)$.

Theorem 4.1. $R(\pi, \delta) = R(\pi^g, \delta^g)$.

Proof.

$$\begin{aligned}
R(\pi, \delta) &= \int\int L(\overline{h}\theta_0, \delta(x)) f(x; \overline{h}\theta_0)\pi(h)\lambda(dh)\mu(dx) \\
&= \int\int L(\overline{g}^{-1}\overline{h}\theta_0, \delta(x)) f(x; \overline{g}^{-1}\overline{h}\theta_0)\pi^g(h)\lambda(dh)\mu(dx) \\
&= \int\int L(\overline{h}\theta_0, g^*\delta(x)) f(x; \overline{g}^{-1}\overline{h}\theta_0)\pi^g(h)\lambda(dh)\mu(dx) \\
&= \int\int L(\overline{h}\theta_0, \delta^g(x)) f(x; \overline{h}\theta_0)\pi^g(h)\lambda(dh)\mu(dx) \\
&= R(\pi^g, \delta^g)
\end{aligned}$$

■

Therefore, if δ is a Bayes estimator with respect to π and g is fixed then δ^g is a Bayes estimator with respect to π^g.

Let us now treat the case where g is random. Let λ be a left invariant measure on G and take this measure as the reference measure for all of the densities define on G. As before, decompose θ as $\theta = \overline{h}\theta_0, \theta \in \Theta, h \in G$. Suppose that the density of g is r^β where β is an unknown parameter, $\beta \in G$, and consider a joint distribution p on (h, β) given in the following way. Conditionally on $\beta, \beta \in G, h$ has density u^β and marginally, β as (possibly improper) density $\Delta(\beta^{-1})$ where Δ is the right hand modulus of λ. Define also π^g on G as $\pi^g(h) = \int r^\beta(g)u^\beta(h)\Delta(\beta^{-1}g)\lambda(d\beta)$. Notice that π satisfies the following property:

Property 4.1 There exists two independent random variables H_1 and H_2 on G having densities r^β and u^β respectively such that $H_1^{-1}H_2$ has density π.

Theorem 4.2. $R(p, \delta) = \int R(\pi^g, \delta)\Delta(g^{-1})\lambda(dg)$

Proof:

$$\begin{aligned} R(p,\delta) &= \int\int\int\int L(\overline{h}\theta_0, \delta(x))r^\beta(g)u^\beta(h)\Delta(\beta^{-1})\lambda(d\beta)\lambda(dh)\mu(dx)\lambda(dg) \\ &= \int\int\int L(\overline{h}\theta_0, \delta(x))\pi^g(h)\lambda(dh)\mu(dx)\Delta(g^{-1})\lambda(dg) \\ &= \int R(\pi^g, \delta(x))\Delta(g^{-1})\lambda(dg) \end{aligned} \qquad \blacksquare$$

Therefore, if δ is a Bayes estimator with respect to π and there is a density r such that property 4.1 holds then δ^g is a generalized Bayes estimator with respect to p. Let us conclude this article by applying our last result to the original problem. If we set $G = \mathbb{R}^p, \theta_0 = 0, g = X_2, h = \mu_1, \beta = \mu_2$ and suppose that, conditionally on β, h is $N(\beta, \frac{(1-\lambda)}{\lambda}\sigma^2(\Sigma_1 + \Sigma_2))$, λ having a certain distribution on [0, 1], then μ_1^* turns out to be a generalized Bayes estimator.

Acknowledgement. I am very grateful to Professor James O. Berger for many discussions and advice.

References

Berger, J.O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean, *Ann. Statist.*, **8**, 716–761.

Berger, J.O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.*, **18**, 617–651.

Ghosh, M. and Sinha, B.K. (1988). Empirical and hierarchical Bayes competitors of preliminary test estimators in two sample problems. *Journal of Multivariate Analysis*, **27**, 206–227.

Stein, C.M. (1962). Confidence sets for the mean of a multivariate normal distribution. *Journal of the Royal Statistical Society, Series B*, **24**, 265–296.