

# BIOS 767 Homework 2

Mingwei Fei

## 1 Problem 1- Derive fundamental properties of the multivariate normal distribution

Consider the  $n \times 1$  random vector  $Y$  satisfying  $Y \sim MVN(\mu, \Sigma)$ ,  $\Sigma$  is positive definite matrix.

- (a) Let  $Z = aY + c$ , where  $a$  is an  $r \times n$  matrix of constants having rank  $r$  ( $r \leq n$ ) and  $c$  is a  $r \times 1$  vector of constants. Show using the definition of moment generating function that  $Z \sim Normal(a\mu + c, a\Sigma a')$ .

As  $Y \sim MVN(\mu, \Sigma)$ ,  $\Sigma$  is positive definite matrix.

Then we can decompose  $\Sigma = BB^T$  for some nonsingular matrix  $B$  since  $\Sigma$  is positive definite.

$$X = B^{-1}(Y - \mu), \quad Y = \mu + BX$$

So we have  $X_1, \dots, X_n$  are independent standard normal,  $X = (X_1, \dots, X_n)^T \sim MVN(0, I_n)$ .

$$\begin{aligned} M_Y(t) &= E[e^{t^T Y}] = E[e^{t^T(\mu + BX)}] = e^{t^T \mu} E[e^{l^T X}], \quad l^T = t^T B \\ &= e^{t^T \mu} E[e^{\sum_{i=1}^n l_i Y_i}], \quad l = (l_1, \dots, l_n) \\ &= e^{t^T \mu} \prod_{i=1}^n E[e^{l_i Y_i}] \\ &= e^{t^T \mu} \prod_{i=1}^n e^{l_i^2/2} \\ &= \exp\left(\mu^T t + \frac{1}{2} l^T l\right) \\ &= \exp\left(\mu^T t + \frac{1}{2} t^T \Sigma t\right) \end{aligned}$$

While  $Z = aY + c$ , then we have the MGF for  $Z$

$$\begin{aligned} M_Z(t) &= E[e^{t^T Z}] = E[e^{t^T(aY + c)}] = e^{t^T c} E[e^{t^T aY}] \\ &= e^{t^T c} E[e^{(t^T a)Y}] \\ &= e^{t^T c} \exp\left(\mu^T (a^T t) + \frac{1}{2} (a^T t)^T \Sigma (a^T t)\right) \\ &= \exp\left(t^T (a\mu + c) + \frac{1}{2} t^T a \Sigma a^T t\right) \end{aligned}$$

So  $E(Z) = (a\mu + c)$ ,  $Var(Z) = a\Sigma a'$ , and  $Z \sim Normal(a\mu + c, a\Sigma a')$ .

(ii) Show that  $Y_1|Y_2$  is multivariate normal with mean and covariance given by

## 2 Problem 2 - Develop a theoretical understanding of REML estimation

$Y_i \sim MVN(X_i\beta, \Sigma)$

(a) Read "Bayesian inference for variance components using only error contrasts" by Harville(1974).

In the paper, an alternative method for deriving the likelihood function for error contrasts are provided using Bayesian method. I would like to walk through the derivation of the likelihood here in the homework. The error contrasts are the product of  $n - p$  independent normal distributions,

$$\begin{aligned} M &= X(X'X)^{-1}X', \quad \Sigma = \Sigma(\theta) \\ I - M &= AA', \quad A'A = I, \quad \text{Cholesky decomposition} \\ A'y &\sim MVN(A'X\hat{\beta}, A'\Sigma A), \quad Var(A'y) = A'Var(y)A = A'\Sigma A = \Sigma \\ f_w(A'y|\theta, \hat{\beta}) &= (2\pi)^{-\frac{n-p}{2}} \left(|\Sigma|\right)^{-1/2} \exp\left(-\frac{n-p}{2}(A'y - A'X\hat{\beta})'(\Sigma)^{-1}(A'y - A'X\hat{\beta})\right) \end{aligned}$$

where  $f_w(A'y|\theta, \hat{\beta})$  is the likelihood function of  $\theta, \hat{\beta}$ , and we could use Bayesian method to find the distribution of  $\hat{\beta}$  and further integrate out  $\hat{\beta}$  to get the distribution of  $f_w(A'y|\theta)$

Let  $\Sigma = QQ'$ , then we have the orthogonal projection matrix onto  $C(Q^{-1}X)$

$$\begin{aligned} P &= Q^{-1}y\left((Q^{-1}y)^T(Q^{-1}y)\right)^{-1}(Q^{-1}y)' = Q^{-1}y(y^TQ'^{-1}Q^{-1}y)^{-1}(Q^{-1}y)^T \\ &= \Sigma^{-1}y(y^T\Sigma^{-1}y)^{-1}y^T \end{aligned}$$

The estimate of  $\hat{\beta}$  could be denoted as

$$\begin{aligned} G &= \Sigma^{-1}y(y^T\Sigma^{-1}y)^{-1}, \quad G'Y = \hat{\beta} \\ G'y &\sim MVN(\hat{\beta}, G'\Sigma G), \quad Var(G'y) = G'Var(y)G = G'\Sigma G = (Y'\Sigma^{-1}Y)^{-1} \end{aligned}$$

We also have the relationship as below, it shows that the  $f_w(\beta, \theta)$  are normal distribution, and the quadratic form could be decomposed into  $f_w(\theta, \hat{\beta})$  and  $f_{\hat{\beta}}(\beta)$ .

$$(Y - X\beta)' \Sigma^{-1} (Y - X\beta) = (Y - X\hat{\beta})' \Sigma^{-1} (Y - X\hat{\beta}) + (\beta - \hat{\beta})' (X \Sigma^{-1} X) (\beta - \hat{\beta})$$

And the distribution of  $A'y$  is  $f_w(\theta, \hat{\beta})$  and distribution for  $G'y$  is  $f_{\hat{\beta}}(\beta)$ .

$$\begin{aligned}
f_w(A'y|\theta) &= \int f_w(A'y|\theta, \hat{\beta}) f_{\hat{\beta}}(G'y|\theta, \beta) d\beta \\
&= \left( \det(Y'Y) \right)^{1/2} \int f_y(y|\theta, \beta) d\beta \\
&= (2\pi)^{-\frac{1}{2}(n-p)} \left( \det(Y'Y) \right)^{1/2} \left( \det(\Sigma) \right)^{-1/2} \left( \det(Y'\Sigma Y) \right)^{-1/2} \\
&\quad \exp \left( -\frac{1}{2} (y - X\hat{\beta})^T \Sigma^{-1} (y - X\hat{\beta}) \right)
\end{aligned}$$

Overall, this paper provided a Bayesian method to get the marginal distribution of  $\Sigma(\theta)$ , so that we can use the maximum likelihood method to estimate.

- (b) Read the paper "A Conditional Derivation of Residual Maximum Likelihood" by Verbyla (1990).

This paper reviewed the Bayesian method in deriving likelihood function in Harville's paper and provided an alternative method in doing transformation without  $\Sigma$ . The approach is to construct the conditional distribution in matrix form. The estimate of  $\hat{\beta}$  is based on the conditional distribution  $Y_1|Y_2$  which based on the orthogonal projection matrix onto  $C(Q^{-1}X)$ .

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

And the REML log-likelihood for  $\sigma^2$  and  $\gamma$  is the marginal log-likelihood based on  $Y_2$ . The approach used the standard decomposition of the joint distribution of  $L'Y$  into conditional and marginal distributions. The idea of residual maximum likelihood as a marginal likelihood is actually the idea of the use of a conditional likelihood.

- (c) Derive the REML log-likelihood for  $\Sigma Y \sim X\beta + \epsilon$ , where  $\epsilon \sim N(0, \sigma^2\Omega(\gamma))$ . Let  $L_1(n \times p)$  and  $L_2(n \times (n-p))$  satisfying  $L_1'X = I_p$  and  $L_2'X = 0$ . Also

$$L'Y = [L_1', L_2']^T Y = [Y_1, Y_2]^T$$

Thus,

$$\begin{aligned}
Y &= [Y_1, Y_2]^T \sim N\{\mu, \Sigma\} \\
\mu &= \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \quad \Sigma = \sigma^2 \begin{pmatrix} L_1'\Omega L_1 & L_1'\Omega L_2 \\ L_2'\Omega L_1 & L_2'\Omega L_2 \end{pmatrix},
\end{aligned}$$

From the paper,  $Y_2 \sim N(0, \sigma^2 L_2'\Omega L_2)$ , and the REML log-likelihood for  $\sigma^2$  and  $\gamma$  is the marginal log-likelihood based on  $Y_2$  and is

$$\begin{aligned}
\text{Log}L(\sigma^2, \gamma, Y_2) &= -\frac{1}{2} \left( (n-p) \log \sigma^2 - \log \det(L_2'\Omega L_2) - \frac{Y_2'(L_2'\Omega L_2)^{-1}Y_2}{\sigma^2} \right) \\
&= -\frac{1}{2} \left( (n-p) \log \sigma^2 - \log \det(\Omega) - \log \det(X'\Omega^{-1}X) - \frac{Y_2'PY_2}{\sigma^2} \right) \\
P &= L_2(L_2'\Omega L_2)^{-1}L_2' = \Omega^{-1} - \Omega^{-1}X(X'\Omega X)^{-1}X'\Omega^{-1}
\end{aligned}$$

where we kept only the terms associated with  $\sigma^2$  and  $\gamma$ . We will use the following property in deriving the MLE

$$\begin{aligned}
tr[x'Ax] &= tr[xx'A], \\
\partial_A \log|A| &= A^{-1}, \\
\partial_A tr(BA) &= B, \text{ if } A \text{ is symmetric} \\
LogL(\sigma^2, \gamma, Y_2) &= -\frac{1}{2} \left( (n-p) \log \sigma^2 + \log|\Omega^{-1}| - \log|tr(XX'\Omega^{-1})| \right. \\
&\quad \left. - \frac{Y_2' \left[ \Omega^{-1} - \Omega^{-1}X(X'\Omega X)^{-1}X'\Omega^{-1} \right] Y_2}{\sigma^2} \right) \\
\partial_{\Omega^{-1}} Logf &= -\frac{1}{2}\Omega + \frac{1}{2}\Omega + \frac{1}{2\sigma^2}YY' - \frac{1}{2\sigma^2}YY'\Omega^{-1}X(X'\Omega X)^{-1} = 0
\end{aligned}$$