STAT 501 Homework 2

Multinomial

February 10, 2018

1. (a) The conditional density of $\boldsymbol{X}_1|\boldsymbol{X}_2$ is $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so $A\boldsymbol{X}$ is also a Multinormal random variable. Then we set $A = \begin{bmatrix} \mathbf{1}_{1\times q} & \mathbf{0}_{1\times(p-q)} \end{bmatrix}$, we have $A\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{1\times q} & \mathbf{0}_{1\times(p-q)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} = \boldsymbol{X}_1$. So $\boldsymbol{X}_1 \sim N_q$, and

$$\mathrm{E}(\boldsymbol{X}_1) = \mathrm{E}(A\boldsymbol{X}) = A\,\mathrm{E}(\boldsymbol{X}) = A\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{1\times q} & \mathbf{0}_{1\times(p-q)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \boldsymbol{\mu}_1$$

and

$$\mathrm{Var}(\boldsymbol{X}_1) = \mathrm{Var}(A\boldsymbol{X}) = A\,\mathrm{Var}(\boldsymbol{X})A^T = A\boldsymbol{\Sigma}A^T = \begin{bmatrix} \mathbf{1}_{1\times q} & \mathbf{0}_{1\times(p-q)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{q\times 1} \\ \mathbf{0}_{(p-q)\times 1} \end{bmatrix} = \boldsymbol{\Sigma}_{11}$$

Hence $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$

(b) The density of $\boldsymbol{X}_1$ conditioning on $\boldsymbol{X}_2$ is

$$f(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)} = \frac{f(\boldsymbol{x})}{f(\boldsymbol{x}_2)}$$

We know

$$f(\boldsymbol{x}) = \left(\frac{1}{(2\pi)^p \det(\boldsymbol{\Sigma})}\right)^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

And from (a) we know

$$f(\boldsymbol{x}_2) = \left(\frac{1}{(2\pi)^p \det(\boldsymbol{\Sigma}_{22})}\right)^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\right)$$

So

$$\begin{aligned} f(\boldsymbol{x}_1|\boldsymbol{x}_2) &= \frac{f(\boldsymbol{x})}{f(\boldsymbol{x}_2)} \\ &= \left(\frac{1}{(2\pi)^p\{\det(\boldsymbol{\Sigma})/\det(\boldsymbol{\Sigma}_{22})\}}\right)^{1/2} \exp\left(-\frac{1}{2}\left\{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - (\boldsymbol{x}_2-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\right\}\right) \end{aligned}$$

Let $\boldsymbol{V} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, then

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{21} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V} & \boldsymbol{\Sigma}_{12} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Taking determinant on both sides, we have

$$\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{V})\det(\boldsymbol{\Sigma}_{22}) \Rightarrow \det(\boldsymbol{\Sigma})/\det(\boldsymbol{\Sigma}_{22}) = \det(\boldsymbol{V})$$

Also we have

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{V}^{-1} & -\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1} & \boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix}$$

So

$$
\begin{aligned}
(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= \begin{bmatrix}(x_1-\boldsymbol{\mu}_1)^T & (\boldsymbol{x}_2-\boldsymbol{\mu}_2)^T\end{bmatrix}\begin{bmatrix}\boldsymbol{V}^{-1} & -\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1} & \boldsymbol{\Sigma}_{22}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1 \\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{bmatrix} \\
&= (\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)+(\boldsymbol{x}_2-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2) \\
&\quad -(\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)-(\boldsymbol{x}_2-\boldsymbol{\mu}_2)^T\Sigma_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)
\end{aligned}
$$

As $\boldsymbol{\Sigma}$ is symmetric, $\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{21}^T$, and $\boldsymbol{\Sigma}_{22}^{-1}=(\boldsymbol{\Sigma}_{22}^{-1})^T$, so

$$
\begin{aligned}
\text{Above} &= (\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)+\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\right)^T\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2) \\
&\quad -(\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)-\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\right)^T\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1) \\
&= (\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2))^T\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2))
\end{aligned}
$$

Hence

$$
f(\boldsymbol{x}_1|\boldsymbol{x}_2)=\left(\frac{1}{(2\pi)^p\det(\boldsymbol{V})}\right)^{1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2))^T\boldsymbol{V}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2))\right)
$$

So the distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is multinormal. The expected value is $\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2)$ and the variance covariance matrix is $\boldsymbol{V}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, i.e.

$$
\boldsymbol{X}_1|\boldsymbol{X}_2\sim N_q(\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2),\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})
$$

(c) For all $\boldsymbol{t}\in\mathbb{R}^p$, $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})\Rightarrow\boldsymbol{t}^T\boldsymbol{X}\sim N_1(\boldsymbol{t}^T\boldsymbol{\mu},\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$. Thus

$$
M_{\boldsymbol{X}}(\boldsymbol{t})=\mathrm{E}[\exp(\boldsymbol{t}^T\boldsymbol{X})]=M_{\boldsymbol{t}^T\boldsymbol{X}}(1)=\exp(\boldsymbol{1}\cdot t^T\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\cdot 1^2)=\exp(\boldsymbol{t}^T\boldsymbol{\mu}-\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}/2)
$$

2. $\boldsymbol{Z}\sim N_p(\boldsymbol{0},\boldsymbol{I}_p)$, then $\boldsymbol{\Gamma}\boldsymbol{Z}\sim N_p$. And

$$
\mathrm{E}(\boldsymbol{\Gamma}\boldsymbol{Z})=\boldsymbol{\Gamma}\,\mathrm{E}(\boldsymbol{Z})=\boldsymbol{\Gamma}\boldsymbol{0}=\boldsymbol{0}
$$

$$
\mathrm{Var}(\boldsymbol{\Gamma}\boldsymbol{Z})=\boldsymbol{\Gamma}\,\mathrm{Var}(\boldsymbol{Z})\boldsymbol{\Gamma}^T=\boldsymbol{\Gamma}\boldsymbol{I}_p\boldsymbol{\Gamma}^T=\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T=\boldsymbol{I}_p
$$

Thus

$$
\boldsymbol{\Gamma}\boldsymbol{Z}\sim N_p(\boldsymbol{0},\boldsymbol{I}_p)
$$

3. (a) Sicne $\boldsymbol{B}$ is positive definite, we have $\boldsymbol{B}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$, where $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_i)_{i=1}^p$, $\lambda_i>0$, and $\boldsymbol{P}\boldsymbol{P}^T=\boldsymbol{I}_p$. Thus

$$
|\boldsymbol{B}|=|\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T|=|\boldsymbol{P}||\boldsymbol{\Lambda}||\boldsymbol{P}^T|=|\boldsymbol{\Lambda}||\boldsymbol{P}\boldsymbol{P}^T|=|\boldsymbol{\Lambda}|=\prod_{i=1}^p\lambda_i
$$

and

$$
\mathrm{trace}(\boldsymbol{B})=\sum_{i=1}^p\lambda_i
$$

Thus

$$
f(\boldsymbol{B})=n^{n/2}\left(\prod_{i=1}^p\lambda_i\right)^{n/2}\exp\left\{-\frac{n}{2}\left(\sum_{i=1}^p\lambda_i\right)\right\}
$$

Taking the log of f, we have

$$
\log(f(\boldsymbol{B}))=\frac{n}{2}\log n+\frac{n}{2}\sum_{i=1}^p\log(\lambda_i)-\frac{n}{2}\sum_{i=1}^p\lambda_i
$$

So

$$\frac{\partial f}{\partial \lambda_i} = \frac{n}{2}\frac{1}{\lambda_i} - \frac{n}{2}$$

Let $\frac{\partial f}{\partial \lambda_i} = 0$, we have $\lambda_i = 1$. And

$$\left.\frac{\partial^2 f}{\partial \lambda_i^2}\right|_{\lambda_i=1} = -\frac{n}{2}\frac{1}{\lambda_i^2}\Big|_{\lambda_i=1} = -\frac{n}{2} < 0,\ \frac{\partial^2 f}{\partial \lambda_i \partial \lambda_j} = 0$$

So the Hessian matrix

$$\left.\frac{\partial^2 f}{\partial \boldsymbol{\lambda}\boldsymbol{\lambda}^T}\right|_{\boldsymbol{\lambda}=\mathbf{1}} = -\frac{n}{2}\boldsymbol{I}_p$$

is negative definite. So $\lambda_i = 1$ gives the maximum of f. Thus

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T = \boldsymbol{P}\boldsymbol{I}_p\boldsymbol{P}^T = \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}_p$$

i. We know the log likelihood is

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \mu)$$

So

$$\frac{\partial \ell}{\partial \boldsymbol{\mu}} = -\frac{1}{2}\sum_{i=1}^{n}\left(-(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-T})(\boldsymbol{x}_i - \boldsymbol{\mu})\right) = \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}) = n\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$$

And

$$\frac{\partial \ell}{\partial \boldsymbol{\Sigma}} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}$$

Set $\frac{\partial \ell}{\partial \boldsymbol{\mu}} = \mathbf{0}$, since $\boldsymbol{\Sigma}^{-1}$ is full rank, then

$$\boldsymbol{\mu} = \bar{\boldsymbol{x}}$$

Then plug in $\boldsymbol{\mu} = \bar{\boldsymbol{x}}$ into $\frac{\partial \ell}{\partial \boldsymbol{\Sigma}}$ and set it to be $\mathbf{0}$. We have

$$\begin{aligned}
&\frac{\partial \ell}{\partial \boldsymbol{\Sigma}} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\boldsymbol{\Sigma}^{-1} = \mathbf{0} \\
\Rightarrow& - n + \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{x})(\boldsymbol{x}_i - \bar{x})^T\boldsymbol{\Sigma}^{-1} = \mathbf{0} \\
\Rightarrow& n\boldsymbol{\Sigma} = \sum_{i=1}^{p}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T \\
\Rightarrow& \boldsymbol{\Sigma} = \frac{\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T}{n}
\end{aligned}$$