Bayesian Statistics III/IV (MATH3361/4071)	Michaelmas term 2021

Exercise Sheet: Bayesian Statistics

Lecturer & author: Georgios P. Karagiannis — georgios.karagiannis@durham.ac.uk

Part I

Matrix & vector calculus

The exercises about Matrix & vector calculus are optional and can be skipped.

Exercise 1. ($\star$)Let A, B be $K \times K$ invertible matrices . Show that

$$(A+B)^{-1} = A^{-1}(A^{-1}+B^{-1})^{-1}B^{-1}$$

Solution. It is

$$\begin{aligned}(A+B)^{-1} &= A^{-1}(I+A^{-1}B)^{-1} \\ &= A^{-1}(A^{-1}+B^{-1})^{-1}B^{-1}\end{aligned}$$

Exercise 2. ($\star\star$)[Woodbury matrix identity] Verify that

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}$$

if A and C are non-singular.

Solution.

By checking that $(A+UCV)\,(A+UCV)^{-1} = I$

$$\begin{aligned}(A+UCV) \times &\left[A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}\right] \\ &= I + UCVA^{-1} - (U+UCVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\ &= I + UCVA^{-1} - UC(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\ &= I + UCVA^{-1} - UCVA^{-1} = I.\end{aligned}$$

So

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}$$

Exercise 3. ($\star\star$)[Sherman–Morrison formula] Let A be a $K \times K$ invertible matrix and u and v two $K \times 1$ column vectors. Verify that

$$\left(A+uv^{\top}\right)^{-1} = A^{-1} - \frac{1}{1+v^{\top}A^{-1}u}A^{-1}uv^{\top}A^{-1}$$

if $1+v^\top A^{-1}u \neq 0$, and if A is non-singular.

Solution.

$$\begin{aligned}
(A+uv^T)(A+uv^T)^{-1} &= (A+uv^T)\left(A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\right) \\
&= AA^{-1} + uv^TA^{-1} - \frac{AA^{-1}uv^TA^{-1} + uv^TA^{-1}uv^TA^{-1}}{1+v^TA^{-1}u} \\
&= I + uv^TA^{-1} - \frac{uv^TA^{-1} + uv^TA^{-1}uv^TA^{-1}}{1+v^TA^{-1}u} \\
&= I + uv^TA^{-1} - \frac{u\cancel{(1+v^TA^{-1}u)}v^TA^{-1}}{\cancel{1+v^TA^{-1}u}} \\
&= I + uv^TA^{-1} - uv^TA^{-1} \\
&= I
\end{aligned}$$

Exercise 4. ($\star\star\star$)[Block partition matrix inversion] Let A be $K \times K$ invertible matrix, and let $B = A^{-1}$ its inverse. Consider Partition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix};\ B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Namely, $B_{11} = \left[A^{-1}\right]_{11}$ is the upper corner of the A^{-1}, etc...
Show that

$$\begin{aligned}
A_{11}^{-1} = B_{11} &= B_{12}B_{22}^{-1}B_{21} \\
A_{11}^{-1}A_{12} &= -B_{12}B_{22}^{-1}
\end{aligned}$$

Hint: Start by noticing that

$$AB = I \Longleftrightarrow \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \Longleftrightarrow \begin{cases} A_{11}B_{11} + A_{12}B_{21} & = I \\ A_{11}B_{12} + A_{12}B_{22} & = 0 \end{cases}$$

Solution. It is

$$AB = I \Longleftrightarrow \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \Longleftrightarrow \begin{cases} A_{11}B_{11} + A_{12}B_{21} & = I \\ A_{11}B_{12} + A_{12}B_{22} & = 0 \end{cases}$$

So

$$\begin{aligned}
A_{11}B_{12} + A_{12}B_{22} = 0 &\Longleftrightarrow \\
A_{11}^{-1}\left(A_{11}B_{12} + A_{12}B_{22}\right)B_{22}^{-1} = 0 &\Longleftrightarrow \\
B_{12}B_{22}^{-1} + A_{11}^{-1}A_{12} = 0 &
\end{aligned}$$

So

$$A_{11}^{-1}A_{12} = -B_{12}B_{22}^{-1}$$

Also

$$
\begin{aligned}
A_{11}B_{12} + A_{12}B_{22} = 0 \iff \\
(A_{11}B_{12} + A_{12}B_{22})B_{22}^{-1}B_{21} = 0 \iff \\
A_{11}B_{12}B_{22}^{-1}B_{21} + A_{12}B_{21} = 0 \\
A_{12}B_{21} = -A_{11}B_{12}B_{22}^{-1}B_{21}
\end{aligned}
$$

Then, we plug in the above in $A_{11}B_{11} + A_{12}B_{21} = I$ we get

$$
\begin{aligned}
A_{11}B_{11} + A_{12}B_{21} = I \iff \\
A_{11}B_{11} - A_{11}B_{12}B_{22}^{-1}B_{21} = I \iff \\
B_{11} - B_{12}B_{22}^{-1}B_{21} = A_{11}^{-1}
\end{aligned}
$$

So

$$A_{11}^{-1} = B_{11} = B_{12}B_{22}^{-1}B_{21}$$

Part II

Random variables

Exercise 5. ($\star$)Let $y \in \mathcal{Y} \subseteq \mathbb{R}$ be a univariate random variable with CDF $F_y(\cdot)$. Consider a bijective function $h : \mathcal{Y} \to \mathcal{Z}$ with $z = h(y)$, and h^{-1} its inverse. The PDF of z is

$$F_z(z) = \begin{cases} F_Y(h^{-1}(z)) & \text{if } h \nearrow \\ \\ 1 - F_Y(h^{-1}(z)) & \text{if } h \searrow \end{cases}$$

Solution. It is $z = h(y) \Leftrightarrow y = h^{-1}(z)$

For if $h \nearrow$ it is

$$F_z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}(h^{-1}(Z) \leq h^{-1}(z)) = \mathrm{P}(Y \leq h^{-1}(z)) = F_Y(h^{-1}(z))$$

For if $h \searrow$ it is

$$F_z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}(h^{-1}(Z) \geq h^{-1}(z)) = \mathrm{P}(Y \geq h^{-1}(z)) = 1 - F_Y(h^{-1}(z))$$

Exercise 6. ($\star$)Let $y \in \mathcal{Y} \subseteq \mathbb{R}$ be a univariate random variable with PDF $f_y(\cdot)$. Consider a bijective function $h : \mathcal{Y} \to \mathcal{Z} \subseteq \mathbb{R}$ and let h^{-1} be the inverse function of h. Consider a univariate random variable such that $z = h(y)$. The PDF of z is

$$f_z(z) \ = f_y(y)|\det(\frac{\mathrm{d}y}{\mathrm{d}z})| \ = f_y(h^{-1}(z))|\det(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z))|$$

Solution. It is $z = h(y) \Leftrightarrow y = h^{-1}(z)$

For if $h \nearrow$ it is

$$F_z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}(h^{-1}(Z) \leq h^{-1}(z)) = \mathrm{P}(Y \leq h^{-1}(z)) = F_Y(h^{-1}(z))$$

and

$$f_z(z) = \frac{\mathrm{d}}{\mathrm{d}z}F_z(z) = \frac{\mathrm{d}}{\mathrm{d}z}F_Y(h^{-1}(z)) = \frac{\mathrm{d}}{\mathrm{d}h^{-1}}F_Y(h^{-1})\det(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z))$$

For if $h \searrow$ it is

$$F_z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}(h^{-1}(Z) \geq h^{-1}(z)) = \mathrm{P}(Y \geq h^{-1}(z)) = 1 - F_Y(h^{-1}(z))$$

and

$$f_z(z) = \frac{\mathrm{d}}{\mathrm{d}z}F_z(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left[1 - F_Y(h^{-1}(z))\right] = -\frac{\mathrm{d}}{\mathrm{d}h^{-1}}F_Y(h^{-1})\det(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z))$$

but $\det(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z)) < 0$ because $h \searrow$. So in both cases:

$$f_z(z) = f_y(h^{-1}(z))|\det(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z))|$$

Exercise 7. ($\star$)Let $y \sim \text{Ex}(\lambda)$ r.v. with Exponential distribution with rate parameter $\lambda > 0$, and $f_{\text{Ex}(\lambda)}(y) = \lambda \exp(-\lambda y)1(y \geq 0)$. Let $z = 1 - \exp(-\lambda y)$. Calculate the PDF of z, and recognize its distribution.

Solution. It is $z = 1 - \exp(-\lambda y) \iff y = -\frac{1}{\lambda}\log(1-z)$, and $z \in [0,1]$. So $h^{-1}(z) = -\frac{1}{\lambda}\log(1-z)$. Then

$$f_z(z) = f_{\text{Ex}(\lambda)}(h^{-1}(z)) \times \left|\det\left(\frac{\mathrm{d}}{\mathrm{d}z}h^{-1}(z)\right)\right| = f_{\text{Ex}(\lambda)}\left(-\frac{1}{\lambda}\log(1-z)\right) \times \left|\det\left(\frac{\mathrm{d}}{\mathrm{d}z}\frac{-1}{\lambda}\log(1-z)\right)\right|$$
$$= \exp\left(-\lambda\frac{-1}{\lambda}\log(1-z)\right) 1(-\frac{1}{\lambda}\log(1-z) \geq 0) \times \left|-\frac{1}{\lambda}\frac{1}{1-z}\right| = 1(z \in [0,1])$$

From the density, we recognize that $z \sim \mathrm{U}(0,1)$ follows a uniform distribution.

Exercise 8. ($\star$)Prove the following properties

1. Let matrix $A \in \mathbb{R}^{q \times d}$, $c \in \mathbb{R}^q$, and $z = c + Ay$ then
$$\mathrm{E}(z) = \mathrm{E}(c + Ay) = c + A\mathrm{E}(y)$$
2. Let random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$, and let functions ψ_1 and ψ_2 defined on $\mathcal{Z}$ and $\mathcal{Y}$, then
$$\mathrm{E}(\psi_1(z) + \psi_2(y)) = \mathrm{E}(\psi_1(z)) + \mathrm{E}(\psi_2(y))$$
3. If random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ are independent then
$$\mathrm{E}(\psi_1(z)\psi_2(y)) = \mathrm{E}(\psi_1(z))\mathrm{E}(\psi_2(y))$$
for any functions ψ_1 and ψ_2 defined on $\mathcal{Z}$ and $\mathcal{Y}$.

Solution.

1. It is
$$\mathrm{E}(z) = \mathrm{E}(c + Ay) = \int (c + Ay)\,\mathrm{d}F(y) = c + A\int y\mathrm{d}F(y) = c + A\mathrm{E}(y)$$
2. It is
$$\mathrm{E}(\psi_1(z) + \psi_2(y)) = \int (\psi_1(z) + \psi_2(y))\,\mathrm{d}F((z,y)) = \int \psi_1(z)\mathrm{d}F((z,y)) + \int \psi_1(z)\mathrm{d}F((z,y))$$
$$= \int \psi_1(z)\mathrm{d}F(z) + \int \psi_1(z)\mathrm{d}F(z) = \mathrm{E}(\psi_1(z)) + \mathrm{E}(\psi_2(y))$$
3. If random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ then
$$\mathrm{d}F(z,y) = \mathrm{d}F(z)\mathrm{d}F(y)$$
It is
$$\mathrm{E}(\psi_1(z)\psi_2(y)) = \int (\psi_1(z)\psi_2(y))\,\mathrm{d}F((z,y)) = \left(\int \psi_1(z)\mathrm{d}F(z)\right)\left(\int \psi_2(y)\mathrm{d}F(y)\right)$$

Exercise 9. ($\star$)Prove the following properties of the covariance matrix

1. $\mathrm{Cov}(z,y) = \mathrm{E}(zy^\top) - \mathrm{E}(z)\,(\mathrm{E}(y))^\top$
2. $\mathrm{Cov}(z,y) = (\mathrm{Cov}(y,z))^\top$
3. $\mathrm{Cov}_\pi(c_1 + A_1 z, c_2 + A_2 y) = A_1\mathrm{Cov}_\pi(x,y)A_2^\top$, for fixed matrices A_1, A_2, and vectors c_1, c_2 with suitable dimensions.

4. If z and y are independent random vectors then $\text{Cov}(z, y) = 0$

Solution.

1. It is

$$\begin{aligned}\text{Cov}(z,y) &= \text{E}\left((z-\text{E}(z))(y-\text{E}(y))^\top\right)\\ &\text{E}\left(zy^\top - z\text{E}(y)^\top - \text{E}(z)y^\top + \text{E}(z)\text{E}(y)^\top\right)\\ &= \text{E}(zy^\top) - \text{E}(z)\left(\text{E}(y)\right)^\top\end{aligned}$$

2. It is

$$\begin{aligned}(\text{Cov}(y,z))^\top &= \left(\text{E}\left((z-\text{E}(z))(y-\text{E}(y))^\top\right)\right)^\top = \text{E}\left(\left((z-\text{E}(z))(y-\text{E}(y))^\top\right)\right)^\top\\ &= \text{E}\left((y-\text{E}(y))(z-\text{E}(z))^\top\right) = \text{Cov}(y,z)\end{aligned}$$

3. It is

$$\begin{aligned}\text{Cov}(c_1+A_1z, c_2+A_2y) &= \text{E}\left((c_1+A_1z)(c_2+A_2y)^\top\right) - \text{E}(c_1+A_1z)\left(\text{E}(c_2+A_2y)\right)^\top\\ &= ... = A_1\left(\text{E}(zy^\top) - \text{E}(z)\left(\text{E}(y)\right)^\top\right)A_2^\top = A_1\text{Cov}(z,y)A_2^\top\end{aligned}$$

4. Obviously since

$$\text{Cov}(z,y) = 0 \Longleftrightarrow \text{Cov}(z_i, y_j) = \begin{cases} & i = j\\ & i \neq j\end{cases}$$

Exercise 10. ($\star$)Prove that the (i, j)-th element of the covariance matrix between vector z and y is the covariance between their elements z_i and y_j:

$$[\text{Cov}(z,y)]_{i,j} = \text{Cov}(z_i, y_j)$$

Solution.

It is

$$\begin{aligned}[\text{Cov}(z,y)]_{i,j} &= \left[\text{E}(zy^\top) - \text{E}(z)\left(\text{E}(y)\right)^\top\right]_{i,j} =\\ &= \left[\text{E}(zy^\top)\right]_{i,j} - \left[\text{E}(z)\left(\text{E}(y)\right)^\top\right]_{i,j}\\ &= \text{E}(z_iy_j^\top) - \text{E}(z_i)\left(\text{E}(y_j)\right)^\top = \text{Cov}(z_i, y_j)\end{aligned}$$

Exercise 11. ($\star$)Prove the following properties of $\text{Var}(Y)$ for a random vector $y \in \mathcal{Y} \subseteq \mathbb{R}^d$

1. $\text{Var}(y) = \text{E}(yy^\top) - \text{E}(y)\left(\text{E}(y)\right)^\top$
2. $\text{Var}(c + Ay) = A\text{Var}(y)A^\top$, for fixed matrix A, and vectors c with suitable dimensions.
3. $\text{Var}(y) \geq 0$; (semi-positive definite)

Solution.

1. $\text{Var}(y) = \text{Cov}(y, y) = \text{E}(yy^\top) - \text{E}(y)\left(\text{E}(y)\right)^\top$
2. $\text{Var}(c + Ay) = \text{Cov}(c + Ay, c + Ay) = A\text{Cov}(y, y)A^\top = A\text{Var}(y)A^\top$

3. For any vector $x \in \mathbb{R}^q$

$$
\begin{aligned}
t^\top \text{Var}(y) t &= t^\top \text{E}\left((y-\text{E}(y))(y-\text{E}(y))^\top\right) t \\
&= \text{E}\left(\left(t^\top(y-\text{E}(y))\right)\left(t^\top(y-\text{E}(y))\right)^\top\right) \\
&= \text{E}\left(zz^\top\right) = \text{E}\left(\sum_{j=1}^{d} z_j^2\right) \geq 0
\end{aligned}
$$

for $z = t^\top(y-\text{E}(y))$.

Exercise 12. ($\star$)Prove the following properties of characteristic functions

1. $\varphi_{A+Bx}(t) = e^{it^T A}\varphi_x(B^T t)$ if $A \in \mathbb{R}^d$ and $B \in \mathbb{R}^{k\times d}$ are constants
2. $\varphi_{x+y}(t) = \varphi_x(t)\varphi_y(t)$ if and only if x and y are independent
3. if $M_x(t) = \text{E}(e^{t^T x})$ is the moment generating function, then $M_x(t) = \varphi_x(-it)$

Solution.

1. It is

$$\varphi_{A+Bx}(t) = \text{E}(e^{it^T(A+Bx)}) = \text{E}(e^{A+it^T Bx}) = \text{E}(e^{it^T A}e^{iB^T tx}) = e^{it^T A}\text{E}(e^{i(B^T t)x}) = e^{it^T A}\varphi_x(B^T t)$$

2. straightforward
3. straightforward

Exercise 13. ($\star$)Show that if $X \sim \text{Ex}(\lambda)$ then $\varphi_X(t) = \frac{\lambda}{\lambda - it}$.

Solution. It is

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itX}\underbrace{\lambda e^{-\lambda x}1(X>0)}_{=f_{\text{Ex}}(x|\lambda)}\text{d}x = \lambda\int_{-\infty}^{\infty} e^{-x(\lambda - itX)}\text{d}x = \frac{\lambda}{\lambda - it}$$

Exercise 14. ($\star$)

1. Find $\varphi_X(t)$ if $X \sim \text{Br}(p)$.
2. Find $\varphi_Y(t)$ if $Y \sim \text{Bin}(n,p)$

Solution.

1. It is

$$\varphi_X(t) = \sum_{x=0,1} e^{itX}P(X=x) = e^{it0}(1-p) + e^{it1}p = (1-p) + pe^{it}$$

2. Because Binomial r.v. results as a summation of n IID Bernoulli r.v., it is $Y = \sum_{i=1}^n X_i$, where $X_i \sim \text{Br}(p)$ $i = 1, ..., n$ and IID. Then

$$\varphi_Y(t) = \varphi_{\sum X_i}(t) = \prod_{i=1}^{n}\varphi_{X_i}(t) = ((1-p)+pe^{it})^n$$

Exercise 15. ($\star\star\star$)Prove the following statement related to the Bayesian theorem: Assume a probability space $(\Omega, \mathscr{F}, P)$. Let a random variable $y : \Omega \to \mathcal{Y}$ with distribution $F(\cdot)$. Consider a partition $y = (x, \theta)$ with $x \in \mathcal{X}$ and $\theta \in \Theta$. Then the probability density function (PDF), or the probability mass function (PMF) of $\theta|x$ is

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)\mathrm{d}F(\theta)} \tag{1}$$

Hint Consider cases where x is discrete and continuous. In the later case use the mean value theorem :

$$\int_A f(x)g(x)\mathrm{d}x = f(\xi)\int_A g(x)\mathrm{d}x$$

where $\xi \in A$ if A is connected, and $g(x) \geq 0$ for $x \in A$.

Solution. We consider separately two cases.

x is discrete:

Let $\Theta_0 \subseteq \Theta$ be any sub-set of Θ; I need to show that

$$P(\theta \in \Theta_0|x) = \frac{\int_{\Theta_0} f(x|\theta)\mathrm{d}F(\theta)}{\int_{\Theta} f(x|\theta)\mathrm{d}F(\theta)} = \begin{cases} \int_{\Theta_0} \frac{f(x|\theta)f(\theta)}{\int_\Theta f(x|\theta)\mathrm{d}F(\theta)}\mathrm{d}\theta & ,\theta \text{ cont.} \\ \sum_{\theta\in\Theta_0} \frac{f(x|\theta)f(\theta)}{\int_\Theta f(x|\theta)\mathrm{d}F(\theta)} & ,\theta \text{ discr.} \end{cases}$$

By Bayes theorem it is

$$P(\theta \in \Theta_0|x) = \frac{P(\Theta_0, x)}{P(x)}$$

where $P(x) = \int_\Theta f(x|\theta)\mathrm{d}F(\theta)$ and $P(\Theta_0, x) = \int_{\Theta_0} f(x|\theta)\mathrm{d}F(\theta)$.

x is continuous:

Let $\Theta_0 \subseteq \Theta$ be any sub-set of Θ; because the probability $P(x) = 0$, I need to show that

$$\lim_{r\to 0} P(\theta \in \Theta_0|B_r(x)) = \frac{\int_{\Theta_0} f(x|\theta)\mathrm{d}F(\theta)}{\int_{\Theta} f(x|\theta)\mathrm{d}F(\theta)} = \begin{cases} \int_{\Theta_0} \frac{f(x|\theta)f(\theta)}{\int_\Theta f(x|\theta)\mathrm{d}F(\theta)}\mathrm{d}\theta & ,\theta \text{ cont.} \\ \sum_{\theta\in\Theta_0} \frac{f(x|\theta)f(\theta)}{\int_\Theta f(x|\theta)\mathrm{d}F(\theta)} & ,\theta \text{ discr.} \end{cases}$$

for an open ball $B_r(x) = \{x' \in \mathcal{X} : |x' - x| < r\}$. By Bayes theorem

$$P(\theta \in \Theta_0|B_r(x)) = \frac{P(\Theta_0, B_r(x))}{P(B_r(x))}$$

where

$$P(\Theta_0, B_r(x)) = \int_{\Theta_0} \left[\int_{B_r(x)} f(\zeta|\theta)\mathrm{d}\zeta \right] \mathrm{d}F(\theta)$$

$$P(B_r(x)) = \int_{\Theta} \left[\int_{B_r(x)} f(\zeta|\theta)\mathrm{d}\zeta \right] \mathrm{d}F(\theta)$$

By mean value theorem[1] there exists $\zeta' \in B_r(y)$ such as

$$\int_{B_r(x)} f(\zeta|\theta)\mathrm{d}\zeta = f(\zeta'|\theta)\int_{B_r(x)} \mathrm{d}\zeta = f(\zeta'|\theta)\ \|B_r(x)\|$$

Then

$$P(\theta\in\Theta_0|B_r(x)) = \frac{\int_{\Theta_0}\left[f(\zeta'|\theta)\ \|B_r(x)\|\right]\mathrm{d}F(\theta)}{\int_{\Theta}\left[f(\zeta'|\theta)\ \|B_r(x)\|\right]\mathrm{d}F(\theta)} \xrightarrow{r\to 0} \frac{\int_{\Theta_0} f(\zeta|\theta)\mathrm{d}F(\theta)}{\int_{\Theta} f(\zeta|\theta)\mathrm{d}F(\theta)}$$

Exercise 16. ($\star$)Prove that:

1. if $Z\sim \mathrm{N}(0,I)$ then $\varphi_Z(t)=\exp(-\frac{1}{2}t^Tt)$, where $Z\in\mathbb{R}^d$
2. if $X\sim \mathrm{N}(\mu,\Sigma)$ then $\varphi_X(t)=\exp(it^T\mu-\frac{1}{2}t^T\Sigma t)$, where $X\in\mathbb{R}^d$

Hint: Assume as known that if $Z\sim\mathrm{N}(0,1)$ then $\varphi_Z(t)=\exp(-\frac{1}{2}t^2)$, where $Z\in\mathbb{R}$

Solution.

1. It is

$$\varphi_Z(t) = \mathrm{E}(\exp(it^TZ)) = \mathrm{E}(\exp(i\sum_{j=1}^{d}(t_jZ_j))) = \mathrm{E}(\prod_{j=1}^{d}\exp(it_jZ_j)) = \prod_{j=1}^{d}\mathrm{E}(\exp(it_jZ_j))$$
$$= \prod_{j=1}^{d}\varphi_{Z_j}(t) = \prod_{j=1}^{d}\exp(-\frac{1}{2}t_j^2) = \exp(-\frac{1}{2}\sum_{j=1}^{d}t_j^2) = \exp(-\frac{1}{2}t^Tt)$$

2. Assume a matrix L such as $\Sigma=LL^T$. It is $X=\mu+LZ$. Then

$$\varphi_X(t) = \varphi_{\mu+LZ}(t) = e^{it^T\mu}\varphi_Z(L^Tt) = e^{it^T\mu}\exp(-\frac{1}{2}(L^Tt)^TL^Tt)$$
$$= e^{it^T\mu}\exp(-\frac{1}{2}t^TLL^Tt) = \exp(it^T\mu-\frac{1}{2}t^T\Sigma t)$$

Exercise 17. ($\star$)Show the following properties of the Characteristic Function

1. $\varphi_x(0)=1$ and $|\varphi_x(t)|\leq 1$ for all $t\in\mathbb{R}^d$
2. $\varphi_{A+Bx}(t)=e^{it^TA}\varphi_x(B^Tt)$ if $A\in\mathbb{R}^d$ and $B\in\mathbb{R}^{k\times d}$ are constants
3. x and y are independent then $\varphi_{x+y}(t)=\varphi_x(t)\varphi_y(t)$ (we do not proov the other way around)
4. if $M_x(t)=\mathrm{E}(e^{t^Tx})$ is the moment generating function, then $M_x(t)=\varphi_x(-it)$

Solution.

1. It is $\varphi_x(0)=\mathrm{E}(e^{i0^Tx})=\mathrm{E}(1)=1$. Also

$$|\varphi_x(t)| = \left|\mathrm{E}(e^{it^Tx})\right| = \left|\int\left(\cos(t^Tx)+i\sin(t^Tx)\right)\mathrm{d}F(x)\right| \leq \int\left|\cos(t^Tx)+i\sin(t^Tx)\right|\mathrm{d}F(x) \leq \int 1\mathrm{d}F(x)=1$$

2. It is

$$\varphi_{A+Bx}(t) = \mathrm{E}(e^{it^T(A+Bx)}) = \mathrm{E}(e^{it^TA+Bit^Tx}) = \mathrm{E}(e^{Ai}e^{i(B^Tt)^\top x}) = e^{it^TA}\varphi_x(B^Tt)$$

[1] $\int_A f(x)g(x)\mathrm{d}x = f(\xi)\int_A g(x)\mathrm{d}x$ where $\xi\in A$ if A is connected, and $g(x)\geq 0$ for $x\in A$.

3. It is

$$\varphi_{x+y}(t) = \mathrm{E}(e^{it^T(x+y)}) = \mathrm{E}(e^{it^Tx}e^{it^Ty}) = \mathrm{E}(e^{it^Tx})\mathrm{E}(e^{it^Ty}) = \varphi_x(t)\varphi_y(t)$$

Part III

Probability calculus

Exercise 18. ($\star$)Let a random variable $x \sim \text{IG}(a, b)$, a fixed value $c > 0$, and $y = cx$ then $y \sim \text{IG}(a, cb)$.

Solution. It is $y = cx$ and $x = \frac{1}{c}y$

$$f(y) = f_{\text{IG}(a,b)}(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right| \quad \propto (\frac{1}{c}y)^{-a-1} \exp(-\frac{b}{\frac{1}{c}y}) 1_{(0,+\infty)}(\frac{1}{c}y)\frac{1}{c}$$
$$\propto y^{-a-1} \exp(-\frac{cb}{y}) 1_{(0,+\infty)}(y) = f_{\text{IG}(a,cb)}(y)$$

Exercise 19. ($\star\star\star$)Consider that x given z is distributed according to $\text{Ga}(\frac{n}{2}, \frac{nz}{2})$, and that z is distributed according to $\text{Ga}(\frac{m}{2}, \frac{m}{2})$; i.e.

$$\begin{cases} x|z & \sim \text{Ga}(\frac{n}{2}, \frac{nz}{2}) \\ z & \sim \text{Ga}(\frac{m}{2}, \frac{m}{2}) \end{cases}$$

Here, $\text{Ga}(\alpha, \beta)$ is the Gamma distribution with shape and rate parameters α and β, and PDF

$$f_{\text{Ga}(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1(x > 0)$$

1. Show that the compound distribution of x is F $x \sim \text{F}(n, m)$, where $\text{F}(n, m)$ is F distribution with numerator and denumerator degrees of freedom n and m, and PDF

$$f_{\text{F}(n,m)}(x) = \frac{1}{x\,\text{B}\left(\frac{n}{2}, \frac{m}{2}\right)} \sqrt{\frac{(n\,x)^n\ m^m}{(n\,x + m)^{n+m}}} 1(x > 0)$$

2. Show that

$$\text{E}_{\text{F}(n,m)}(x) = \frac{m}{m-2}$$

3. Show that

$$\text{Var}_{\text{F}(n,m)}(x) = \frac{2m^2(n + m - 2)}{n(m - 2)^2(m - 4)}$$

Hint: If $\xi \sim \text{IG}(a, b)$ then $\text{E}_{\xi\sim\text{IG}(a,b)}(\xi) = \frac{b}{a-1}$, and $\text{Var}_{\xi\sim\text{IG}(a,b)}(\xi) = \frac{b^2}{(a-1)^2(a-2)}$

Solution.

1. It is

$$f_{\text{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z) = \frac{(\frac{nz}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{nz}{2}x} 1(x > 0) \ ; \qquad f_{\text{Ga}(\frac{m}{2},\frac{m}{2})}(z) = \frac{(\frac{m}{2})^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} z^{\frac{m}{2}-1} e^{-\frac{m}{2}z} 1(z > 0)$$

So:

$$
\begin{aligned}
f(x) &= \int f_{\mathrm{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z) f_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}(z)\mathrm{d}z \\
&= \int \underbrace{\frac{(\frac{nz}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{nz}{2}x} 1(x>0)}_{=f_{\mathrm{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z)} \underbrace{\frac{(\frac{m}{2})^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} z^{\frac{m}{2}-1} e^{-\frac{m}{2}z} 1(z>0)}_{=f_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}(z)}\mathrm{d}z \\
&= \frac{(\frac{n}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{(\frac{m}{2})^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} 1(x>0) x^{\frac{n}{2}-1} \int_0^\infty z^{\frac{n}{2}} e^{-\frac{nx}{2}z} z^{\frac{m}{2}-1} e^{-\frac{m}{2}z}\mathrm{d}z \\
&= \frac{(\frac{n}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{(\frac{m}{2})^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} 1(x>0) x^{\frac{n}{2}-1} \int_0^\infty z^{\frac{n}{2}+\frac{m}{2}-1} e^{-(\frac{m}{2}+\frac{nx}{2})z}\mathrm{d}z \\
&= \frac{(\frac{n}{2})^{\frac{n}{2}}(\frac{m}{2})^{\frac{m}{2}}}{\mathrm{B}(\frac{n}{2},\frac{m}{2})} 1(x>0) x^{\frac{n}{2}-1} \left(\frac{m}{2}+\frac{nx}{2}\right)^{-(\frac{n}{2}+\frac{m}{2})} \\
&= \frac{(n)^{\frac{n}{2}}(m)^{\frac{m}{2}}}{\mathrm{B}(\frac{n}{2},\frac{m}{2})} \frac{1}{x} \sqrt{\frac{x^n}{(m+nx)^{n+m}}} 1(x>0) \\
&= \frac{1}{x\,\mathrm{B}\left(\frac{n}{2},\frac{m}{2}\right)} \sqrt{\frac{(n\,x)^n\ m^m}{(n\,x+m)^{n+m}}} 1(x>0)
\end{aligned}
$$

2. It is

$$
\begin{aligned}
\mathrm{E}(x) \quad &= \mathrm{E}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\mathrm{E}_{\mathrm{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z)\right) \quad = \mathrm{E}_{z\sim\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\frac{1}{z}\right) \\
&= \mathrm{E}_{\xi\sim\mathrm{IG}(\frac{m}{2},\frac{m}{2})}(\xi) \quad = \frac{\frac{m}{2}}{\frac{m}{2}-1} = \frac{m}{m-2}
\end{aligned}
$$

3. It is

$$
\begin{aligned}
\mathrm{Var}(x) =& \mathrm{E}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\mathrm{Var}_{\mathrm{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z)\right) + \mathrm{Var}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\mathrm{E}_{\mathrm{Ga}(\frac{n}{2},\frac{nz}{2})}(x|z)\right) \\
=& \mathrm{E}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\frac{2}{nz^2}\right) + \mathrm{Var}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\frac{1}{z}\right) && = \frac{2}{n}\mathrm{E}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\frac{1}{z^2}\right) + \mathrm{Var}_{\mathrm{Ga}(\frac{m}{2},\frac{m}{2})}\left(\frac{1}{z}\right) \\
=& \frac{2}{n}\mathrm{E}_{\xi\sim\mathrm{IG}(\frac{m}{2},\frac{m}{2})}\left(\xi^2\right) + \mathrm{Var}_{\xi\sim\mathrm{IG}(\frac{m}{2},\frac{m}{2})}(\xi) \\
=& \frac{2}{n}\left(\frac{(\frac{m}{2})^2}{(\frac{m}{2}-1)(\frac{m}{2}-2)}\right) + \left(\frac{\frac{m}{2}}{\frac{m}{2}-1}\right) && = \ldots = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}
\end{aligned}
$$

Exercise 20. ($\star\star$)Prove the following statement:
Let $x \sim \mathrm{N}_d(\mu,\Sigma)$, $x \in \mathbb{R}^d$, and $y = (x-\mu)^\top \Sigma^{-1}(x-\mu)$. Then

$$y \sim \chi^2_d$$

Solution. It is

$$y = (x-\mu)^\top \Sigma^{-1}(x-\mu) = \left(\Sigma^{-1/2}(x-\mu)\right)^\top \left(\Sigma^{-1/2}(x-\mu)\right) = z^\top z = \sum_{i=1}^d z_i^2$$

where $z = \Sigma^{-1/2}(x-\mu)$, and $z \sim \mathrm{N}_d(0,I)$. Because $z_i \sim \mathrm{N}(0,1)$, it is $\sum_{i=1}^d z_i^2 \sim \chi^2_d$ (from stats concepts 2).

Exercise 21. ($\star\star$)Let

$$\begin{cases} x|\xi & \sim \mathrm{N}_d(\mu, \Sigma\xi) \\ \xi & \sim \mathrm{IG}(a, b) \end{cases}$$

with PDF

$$f_{\mathrm{N}_d(\mu,\Sigma\xi)}(x|\xi) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

$$f_{\mathrm{IG}(a,b)}(\xi) = \frac{b^a}{\Gamma(a)}\xi^{-a-1}\exp\left(-\frac{b}{\xi}\right)1_{(0,\infty)}(\xi)$$

Show that the marginal PDF of x is

$$\begin{aligned} f(x) &= \int f_{\mathrm{N}_d(\mu,\Sigma\xi)}(x|\xi) f_{\mathrm{IG}(a,b)}(\xi)\mathrm{d}\xi \\ &= \frac{2a^{-\frac{d}{2}}}{\pi^{\frac{n}{2}}\sqrt{\det(\frac{b}{a}\Sigma)}} \frac{\Gamma\left(a+\frac{d}{2}\right)}{\Gamma(a)} \left[1 + \frac{1}{2a}(x-\mu)^\top \left(\frac{b}{a}\Sigma\right)^{-1}(x-\mu)\right]^{-\frac{(2a+d)}{2}} \end{aligned} \tag{2}$$

FYI: For $a = b = \frac{v}{2}$, the marginal PDF is the PDF of the d-dimensional Student T distribution.

Solution. It is

$$\begin{aligned} &\int f_{\mathrm{N}_d(\mu,\Sigma\xi)}(x|\xi) f_{\mathrm{IG}(a,b)}(\xi)\mathrm{d}\xi = \\ &= \int \underbrace{\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sqrt{\det(\Sigma\xi)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \frac{\Sigma^{-1}}{\xi}(x-\mu)\right)}_{=\mathrm{N}_d(x|\mu,\Sigma\xi)} \underbrace{\frac{b^a}{\Gamma(a)}\xi^{-a-1}\exp\left(-\frac{b}{\xi}\right)1_{(0,\infty)}(\xi)}_{=\mathrm{IG}(\xi|a,b)}\mathrm{d}\xi \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sqrt{\det(\Sigma)}} \frac{b^a}{\Gamma(a)} \int \xi^{-a-1-\frac{d}{2}} \exp\left(-\frac{1}{\xi}\left[\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu) + b\right]\right)\mathrm{d}\xi \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sqrt{\det(\Sigma)}} \frac{b^a}{\Gamma(a)} \Gamma\left(a + \frac{d}{2}\right)\left[\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu) + b\right]^{-\left(a+\frac{d}{2}\right)} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sqrt{\det(\frac{b}{a}\Sigma)}} \frac{b^{-\frac{d}{2}}}{\Gamma(a)} \Gamma\left(a + \frac{d}{2}\right)\left[\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\frac{1}{b} + 1\right]^{-\frac{(2a+d)}{2}} \\ &= \frac{2a^{-\frac{d}{2}}}{\pi^{\frac{n}{2}}\sqrt{\det(\frac{b}{a}\Sigma)}} \frac{\Gamma\left(a+\frac{d}{2}\right)}{\Gamma(a)} \left[1 + \frac{1}{2a}(x-\mu)^\top \left(\frac{b}{a}\Sigma\right)^{-1}(x-\mu)\right]^{-\frac{(2a+d)}{2}} \end{aligned}$$

The Following one will be given as Homework

Exercise 22. ($\star\star\star$)

Let $x \sim \mathrm{T}_d(\mu, \Sigma, \nu)$. Recall that $x \sim \mathrm{T}_d(\mu, \Sigma, \nu)$ is the marginal distribution $f_x(x) = \int f_{x|\xi}(x|\xi) f_\xi(\xi)\mathrm{d}\xi$ of (x, ξ) where

$$x|\xi \sim \mathrm{N}_d(\mu, \Sigma\xi v)$$

$$\xi \sim \mathrm{IG}(\frac{v}{2}, \frac{1}{2})$$

Consider partition such that

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \qquad \Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix},$$

where $x_1 \in \mathbb{R}^{d_1}$ and $x_2 \in \mathbb{R}^{d_2}$.

Address the following:

1. Show that the marginal distribution of x_1 is such that

$$x_1 \sim \mathrm{T}_{d_1}(\mu_1, \Sigma_1, \nu)$$

Hint: Try to use the form $f_x(x) = \int f_{x|\xi}(x|\xi) f_\xi(\xi) \mathrm{d}\xi$.

2. Show that

$$\xi | x_1 \sim \mathrm{IG}(\frac{1}{2}(d_1 + v), \frac{1}{2}\frac{Q+v}{v})$$

where $Q = (\mu_1 - x_1)^\top \Sigma_1^{-1} (\mu_1 - x_1)$.

Hint: The PDF of $y \sim \mathrm{N}_d(\mu, \Sigma)$ is

$$f(y) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-\mu)^\top \Sigma^{-1} (y-\mu)\right)$$

Hint: The PDF of $y \sim \mathrm{IG}(a, b)$ is

$$f_{\mathrm{IG}(a,b)}(y) = \frac{b^a}{\Gamma(a)} y^{-a-1} \exp(-\frac{b}{y}) 1_{(0,+\infty)}(y)$$

3. Let $\xi' = \xi \frac{v}{Q+v}$, with $Q = (\mu_1 - x_1)^\top \Sigma_1^{-1} (\mu_1 - x_1)$, show that

$$\xi' | x_1 \sim \mathrm{IG}(\frac{v + d_1}{2}, \frac{1}{2})$$

4. Show that the conditional distribution of $x_2 | x_1$ is such that

$$x_2 | x_1 \sim \mathrm{T}_{d_2}(\mu_{2|1}, \dot{\Sigma}_{2|1}, \nu_{2|1})$$

where

$$\begin{aligned}
\mu_{2|1} =& \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \\
\dot{\Sigma}_{2|1} =& \frac{\nu + (\mu_1 - x_1)^\top \Sigma_1^{-1} (\mu_1 - x_1)}{\nu + d_1} \Sigma_{2|1} \\
\Sigma_{2|1} =& \Sigma_{22} - \Sigma_{21} \Sigma_1^{-1} \Sigma_{21}^\top \\
\nu_{2|1} =& \nu + d_1
\end{aligned}$$

Hint: You can use the Example [Marginalization & conditioning] from the Lecture Handout

Solution.

Exercise 23. ($\star\star\star$)Show that

1. If $x_i \sim \mathrm{N}_d(\mu_i, \Sigma_i)$ for $i = 1, .., n$ and $y = c + \sum_{i=1}^n B_i x_i$, then
$$y \sim \mathrm{N}_d(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top)$$
2. If $x_i \sim \mathrm{T}_d(\mu_i, \Sigma_i, v)$ for $i = 1, .., n$ and $z = c + \sum_{i=1}^n B_i x_i$, then
$$z \sim \mathrm{T}_d(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top, v)$$

Solution.

1. For any $a \in \mathbb{R}^d$
$$a^\top y = a^\top \left(c + \sum_{i=1}^n B_i x_i \right) = a^\top c + \sum_{i=1}^n a^\top B_i x_i = a^\top c + \sum_{i=1}^n \left(B_i^\top a \right)^\top x_i$$
follows a univariate Normal distribution. So y follows a d-dimensional Normal by definition. Also
$$\mathrm{E}(y) = \mathrm{E}(c + \sum_{i=1}^n B_i x_i) = c + \sum_{i=1}^n \mu_i$$
and
$$\mathrm{Var}(y) = \mathrm{Var}(c + \sum_{i=1}^n B_i x_i) = \sum_{i=1}^n B_i \mathrm{Var}(x_i) B_i^\top = \sum_{i=1}^n B_i \Sigma_i B_i^\top$$
So by definition $y \sim \mathrm{N}_d(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top)$.
2. It is
$$z = c + \sum_{i=1}^n B_i x_i = c + \sum_{i=1}^n B_i \left(\mu_i + y_i \sqrt{v\xi} \right) = \left(c + \sum_{i=1}^n B_i \mu_i \right) + \left(\sum_{i=1}^n B_i y_i \right) \sqrt{v\xi}$$
for $y_i \sim \mathrm{N}_d(0, \Sigma_i)$ and $\xi \sim \mathrm{IG}(\frac{v}{2}, \frac{1}{2})$, and hence
$$z = \left(c + \sum_{i=1}^n B_i \mu_i \right) + \tilde{y} \sqrt{v\xi}$$
where $\tilde{y} \sim \mathrm{N}_d(0, \sum_{i=1}^n B_i \Sigma_i B_i^\top)$. Hence, $z \sim \mathrm{T}_d(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top, v)$ by definition.
