

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. (★) 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. (★★)[Woodbury matrix identity] Verify that

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

if A and C are non-singular.

Solution.

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

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

So

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

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

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

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

Solution.

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

Exercise 4. (★★)[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, $B_{11} = [A^{-1}]_{11}$ 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

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

Solution. It is

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

So

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

So

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

Also

$$A_{11}B_{12} + A_{12}B_{22} = 0 \iff$$

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

$$A_{11}B_{12}B_{22}^{-1}B_{21} + A_{12}B_{21} = 0$$

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

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

$$A_{11}B_{11} + A_{12}B_{21} = I \iff$$

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

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

So

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

Part II

Random variables

Exercise 5. (*) Let $y \in \mathcal{Y} \subseteq \mathbb{R}$ be a univariate random variable with CDF $F_y(\cdot)$. Consider a bijective function $h : \mathcal{Y} \rightarrow \mathcal{Z}$ 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

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

For if $h \searrow$ it is

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

Exercise 6. (*) Let $y \in \mathcal{Y} \subseteq \mathbb{R}$ be a univariate random variable with PDF $f_y(\cdot)$. Consider a bijective function $h : \mathcal{Y} \rightarrow \mathcal{Z} \subseteq \mathbb{R}$ 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

$$f_z(z) = f_y(y) \left| \det\left(\frac{dy}{dz}\right) \right| = f_y(h^{-1}(z)) \left| \det\left(\frac{d}{dz} h^{-1}(z)\right) \right|$$

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

For if $h \nearrow$ it is

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

and

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

For if $h \searrow$ it is

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

and

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

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

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

Exercise 7. (*) 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

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

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

Exercise 8. (★) Prove the following properties

1. Let matrix $A \in \mathbb{R}^{q \times d}$, $c \in \mathbb{R}^q$, and $z = c + Ay$ then

$$\mathbb{E}(z) = \mathbb{E}(c + Ay) = c + A\mathbb{E}(y)$$

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

$$\mathbb{E}(\psi_1(z) + \psi_2(y)) = \mathbb{E}(\psi_1(z)) + \mathbb{E}(\psi_2(y))$$

3. If random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ are independent then

$$\mathbb{E}(\psi_1(z)\psi_2(y)) = \mathbb{E}(\psi_1(z))\mathbb{E}(\psi_2(y))$$

for any functions ψ_1 and ψ_2 defined on \mathcal{Z} and \mathcal{Y} .

Solution.

1. It is

$$\mathbb{E}(z) = \mathbb{E}(c + Ay) = \int (c + Ay) dF(y) = c + A \int y dF(y) = c + A\mathbb{E}(y)$$

2. It is

$$\begin{aligned} \mathbb{E}(\psi_1(z) + \psi_2(y)) &= \int (\psi_1(z) + \psi_2(y)) dF((z, y)) = \int \psi_1(z) dF((z, y)) + \int \psi_2(y) dF((z, y)) \\ &= \int \psi_1(z) dF(z) + \int \psi_2(y) dF(y) = \mathbb{E}(\psi_1(z)) + \mathbb{E}(\psi_2(y)) \end{aligned}$$

3. If random variables $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ then

$$dF(z, y) = dF(z)dF(y)$$

It is

$$\mathbb{E}(\psi_1(z)\psi_2(y)) = \int (\psi_1(z)\psi_2(y)) dF((z, y)) = \left(\int \psi_1(z) dF(z) \right) \left(\int \psi_2(y) dF(y) \right)$$

Exercise 9. (★) Prove the following properties of the covariance matrix

$$1. \text{Cov}(z, y) = \mathbb{E}(zy^\top) - \mathbb{E}(z)(\mathbb{E}(y))^\top$$

$$2. \text{Cov}(z, y) = (\text{Cov}(y, z))^\top$$

$$3. \text{Cov}_\pi(c_1 + A_1 z, c_2 + A_2 y) = A_1 \text{Cov}_\pi(z, y) A_2^\top, \text{ for fixed matrices } A_1, A_2, \text{ and vectors } c_1, c_2 \text{ with suitable dimensions.}$$

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

Solution.

1. It is

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

2. It is

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

3. It is

$$\begin{aligned}\text{Cov}(c_1 + A_1 z, c_2 + A_2 y) &= \mathbb{E}((c_1 + A_1 z)(c_2 + A_2 y)^\top) - \mathbb{E}(c_1 + A_1 z)(\mathbb{E}(c_2 + A_2 y))^\top \\ &= \dots = A_1 (\mathbb{E}(zy^\top) - \mathbb{E}(z)(\mathbb{E}(y))^\top) A_2^\top = A_1 \text{Cov}(z, y) A_2^\top\end{aligned}$$

4. Obviously since

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

Exercise 10. (★) 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

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

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

1. $\text{Var}(y) = \mathbb{E}(yy^\top) - \mathbb{E}(y)(\mathbb{E}(y))^\top$
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. $\text{Var}(y) = \text{Cov}(y, y) = \mathbb{E}(yy^\top) - \mathbb{E}(y)(\mathbb{E}(y))^\top$
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 $x \in \mathbb{R}^q$

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

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

Exercise 12. (★) Prove the following properties of characteristic functions

1. $\varphi_{A+Bx}(t) = e^{it^\top A} \varphi_x(B^\top t)$ 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 $M_x(t) = \mathbb{E}(e^{t^\top x})$ is the moment generating function, then $M_x(t) = \varphi_x(-it)$

Solution.

1. It is

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

2. straightforward

3. straightforward

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

Solution. It is

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

Exercise 14. (★)

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 $X_i \sim \text{Br}(p)$ $i = 1, \dots, n$ 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. (★★) Prove the following statement related to the Bayesian theorem:

Assume a probability space (Ω, \mathcal{F}, P) . Let a random variable $y : \Omega \rightarrow \mathcal{Y}$ 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

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)dF(\theta)} \quad (1)$$

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

$$\int_A f(x)g(x)dx = f(\xi) \int_A g(x)dx$$

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

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

By Bayes theorem it is

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

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

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

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

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

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

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

By mean value theorem¹ there exists $\zeta' \in B_r(y)$ such as

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

Then

$$P(\theta \in \Theta_0 | B_r(x)) = \frac{\int_{\Theta_0} [f(\zeta'|\theta) \|B_r(x)\|] dF(\theta)}{\int_{\Theta} [f(\zeta'|\theta) \|B_r(x)\|] dF(\theta)} \xrightarrow{r \rightarrow 0} \frac{\int_{\Theta_0} f(\zeta|\theta) dF(\theta)}{\int_{\Theta} f(\zeta|\theta) dF(\theta)}$$

Exercise 16. (★) Prove that:

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

Solution.

1. It is

$$\begin{aligned} \varphi_Z(t) &= E(\exp(it^T Z)) = E(\exp(i \sum_{j=1}^d (t_j Z_j))) = E(\prod_{j=1}^d \exp(it_j Z_j)) = \prod_{j=1}^d E(\exp(it_j Z_j)) \\ &= \prod_{j=1}^d \varphi_{Z_j}(t) = \prod_{j=1}^d \exp(-\frac{1}{2}t_j^2) = \exp(-\frac{1}{2} \sum_{j=1}^d t_j^2) = \exp(-\frac{1}{2}t^T t) \end{aligned}$$

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

$$\begin{aligned} \varphi_X(t) &= \varphi_{\mu + LZ}(t) = e^{it^T \mu} \varphi_Z(L^T t) = e^{it^T \mu} \exp(-\frac{1}{2}(L^T t)^T L^T t) \\ &= e^{it^T \mu} \exp(-\frac{1}{2}t^T L L^T t) = \exp(it^T \mu - \frac{1}{2}t^T \Sigma t) \end{aligned}$$

Exercise 17. (★) 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. $\varphi_{A+Bx}(t) = e^{it^T A} \varphi_x(B^T t)$ 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 prove the other way around)
4. if $M_x(t) = E(e^{t^T x})$ is the moment generating function, then $M_x(t) = \varphi_x(-it)$

Solution.

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

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

2. It is

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

¹ $\int_A f(x)g(x)dx = f(\xi) \int_A g(x)dx$ where $\xi \in A$ if A is connected, and $g(x) \geq 0$ for $x \in A$.

235 3. It is

236
$$\varphi_{x+y}(t) = \mathbb{E}(e^{it^T(x+y)}) = \mathbb{E}(e^{it^T x} e^{it^T y}) = \mathbb{E}(e^{it^T x}) \mathbb{E}(e^{it^T y}) = \varphi_x(t) \varphi_y(t)$$

Part III

Probability calculus

Exercise 18. (★) 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 $x = \frac{1}{c}y$

$$\begin{aligned} f(y) = f_{\text{IG}(a,b)}(x) \left| \frac{dx}{dy} \right| &\propto \left(\frac{1}{c}y \right)^{-a-1} \exp\left(-\frac{b}{\frac{1}{c}y}\right) 1_{(0,+\infty)}\left(\frac{1}{c}y\right) \frac{1}{c} \\ &\propto y^{-a-1} \exp\left(-\frac{cb}{y}\right) 1_{(0,+\infty)}(y) = f_{\text{IG}(a,cb)}(y) \end{aligned}$$

Exercise 19. (★★) 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 $F(x) \sim F(n, m)$, where $F(n, m)$ is F distribution with numerator and denominator degrees of freedom n and m , and PDF

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

2. Show that

$$E_{F(n,m)}(x) = \frac{m}{m-2}$$

3. Show that

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

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

Solution.

1. It is

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

So:

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

2. It is

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

3. It is

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

Exercise 20. (★★) Prove the following statement:

Let $x \sim \text{N}_d(\mu, \Sigma)$, $x \in \mathbb{R}^d$, and $y = (x - \mu)^\top \Sigma^{-1} (x - \mu)$. Then

$$y \sim \chi_d^2$$

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 $z \sim \text{N}_d(0, I)$. Because $z_i \sim \text{N}(0, 1)$, it is $\sum_{i=1}^d z_i^2 \sim \chi_d^2$ (from stats concepts 2).

Exercise 21. (★★) Let

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

with PDF

$$\begin{aligned} f_{\mathbf{N}_d(\mu, \Sigma\xi)}(x|\xi) &= (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \\ f_{\text{IG}(a, b)}(\xi) &= \frac{b^a}{\Gamma(a)} \xi^{-a-1} \exp\left(-\frac{b}{\xi}\right) \mathbf{1}_{(0, \infty)}(\xi) \end{aligned}$$

Show that the marginal PDF of x is

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

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

Solution. It is

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

Exercise 22. (★★★)

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

$$\begin{aligned} x|\xi &\sim \mathbf{N}_d(\mu, \Sigma\xi v) \\ \xi &\sim \text{IG}\left(\frac{v}{2}, \frac{1}{2}\right) \end{aligned}$$

Consider partition such that

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

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

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

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

2. Show that

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

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

Hint: The PDF of $y \sim \mathcal{N}_d(\mu, \Sigma)$ 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 $y \sim \text{IG}(a, b)$ is

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

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

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

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

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

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. (★★) Show that

1. If $x_i \sim \text{N}_d(\mu_i, \Sigma_i)$ for $i = 1, \dots, n$ and $y = c + \sum_{i=1}^n B_i x_i$, then

$$y \sim \text{N}_d\left(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top\right)$$

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

$$z \sim \text{T}_d\left(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top, v\right)$$

Solution.

1. For any $a \in \mathbb{R}^d$

$$a^\top y = a^\top \left(c + \sum_{i=1}^n B_i x_i \right) = a^\top c + \sum_{i=1}^n a^\top B_i x_i = a^\top c + \sum_{i=1}^n (B_i^\top a)^\top x_i$$

follows a univariate Normal distribution. So y follows a d -dimensional Normal by definition. Also

$$\text{E}(y) = \text{E}\left(c + \sum_{i=1}^n B_i x_i\right) = c + \sum_{i=1}^n \mu_i$$

and

$$\text{Var}(y) = \text{Var}\left(c + \sum_{i=1}^n B_i x_i\right) = \sum_{i=1}^n B_i \text{Var}(x_i) B_i^\top = \sum_{i=1}^n B_i \Sigma_i B_i^\top$$

So by definition $y \sim \text{N}_d\left(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top\right)$.

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 $y_i \sim \text{N}_d(0, \Sigma_i)$ and $\xi \sim \text{IG}\left(\frac{v}{2}, \frac{1}{2}\right)$, and hence

$$z = \left(c + \sum_{i=1}^n B_i \mu_i \right) + \tilde{y} \sqrt{v} \xi$$

where $\tilde{y} \sim \text{N}_d\left(0, \sum_{i=1}^n B_i \Sigma_i B_i^\top\right)$. Hence, $z \sim \text{T}_d\left(c + \sum_{i=1}^n \mu_i, \sum_{i=1}^n B_i \Sigma_i B_i^\top, v\right)$ by definition.

Part IV

Bayesian paradigm and calculations

Exercise 24. (★) Consider an i.i.d. sample y_1, \dots, y_n from the skew-logistic distribution with PDF

$$f(y_i|\theta) = \frac{\theta e^{-y_i}}{(1 + e^{-y_i})^{\theta+1}}$$

with parameter $\theta \in (0, \infty)$. To account for the uncertainty about θ we assign a Gamma prior distribution with PDF

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)),$$

and fixed hyper parameters a, b specified by the researcher's prior info.

1. Derive the posterior distribution of θ .
2. Derive the predictive PDF for a future $z = y_{n+1}$.

Solution. It is

$$f(y_i|\theta) = \frac{\theta e^{-y_i}}{(1 + e^{-y_i})^{\theta+1}} = \frac{\theta e^{-y_i}}{(1 + e^{-y_i})} \exp(-\theta \log(1 + e^{-y_i}))$$

1. By using the Bayes theorem

$$\begin{aligned} \pi(\theta|y) &\propto f(y|\theta) \pi(\theta) \propto \prod_{i=1}^n f(y_i|\theta) \pi(\theta) = \prod_{i=1}^n \frac{\theta e^{-y_i}}{(1 + e^{-y_i})^{\theta+1}} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)) \\ &\propto \prod_{i=1}^n \frac{e^{-y_i}}{(1 + e^{-y_i})} \theta^n \prod_{i=1}^n \exp(-\theta \log(1 + e^{-y_i})) \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)) \\ &\propto \theta^{n+a-1} \exp\left(-\theta \left[\sum_{i=1}^n \log(1 + e^{-y_i}) + b \right]\right) 1(\theta \in (0, \infty)) \propto \text{Ga}(\theta|a + n, b + \sum_{i=1}^n \log(1 + e^{-y_i})) \end{aligned}$$

So

$$\theta|y \sim \text{Ga}\left(\underbrace{a + n}_{=a^*}, \underbrace{b + \sum_{i=1}^n \log(1 + e^{-y_i})}_{=b^*}\right)$$

2. By using the definition for the predictive PDF, it is

$$\begin{aligned} f(z|y) &= \int_{\mathbb{R}} f(z|\theta) \pi(\theta|y) d\theta \\ &= \int_{\mathbb{R}_+} \frac{e^{-z}}{(1 + e^{-z})} \theta \exp(-\theta \log(1 + e^{-z})) \frac{(b^*)^{a^*}}{\Gamma(a^*)} \theta^{a^*-1} \exp(-\theta b^*) d\theta \\ &= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \frac{e^{-z}}{(1 + e^{-z})} \int_{\mathbb{R}_+} \theta^{a^*+1-1} \exp(-\theta(b^* + \log(1 + e^{-z}))) d\theta \\ &= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \frac{e^{-z}}{(1 + e^{-z})} \frac{\Gamma(a^* + 1)}{(b^* + \log(1 + e^{-z}))^{a^*+1}} = \frac{e^{-z}}{(1 + e^{-z})} \frac{(b^*)^{a^*}}{(b^* + \log(1 + e^{-z}))^{a^*+1}} a^* \end{aligned}$$

Exercise 25. (★★)(Nuisance parameters are involved)

<-story

Assume observable quantities $y = (y_1, \dots, y_n)$ forming the available data set of size n . Assume that the observations are drawn i.i.d. from a sampling distribution which is judged to be in the Normal parametric family of distributions $N(\mu, \sigma^2)$ with unknown mean μ and variance σ^2 . We are interested in learning μ and the next outcome $z = y_{n+1}$. We do not care about σ^2 .

Assume You specify a Bayesian model

<-set-up

$$\begin{cases} y_i | \mu, \sigma^2 \sim N(\mu, \sigma^2), \text{ for all } i = 1, \dots, n & , \text{Statistical model} \\ \mu | \sigma^2 \sim N(\mu_0, \sigma^2 \frac{1}{\tau_0}) & , \text{prior} \\ \sigma^2 \sim \text{IG}(a_0, k_0) & , \text{prior} \end{cases}$$

1. Show that

$$\sum_{i=1}^n (y_i - \theta)^2 = n(\bar{y} - \theta)^2 + ns^2,$$

$$\text{where } s^2 = \frac{1}{2} \sum_{i=1}^n (y_i - \bar{y})^2.$$

2. Show that the joint posterior distribution $\Pi(\mu, \sigma^2 | y)$ is such as

$$\begin{aligned} \mu | y, \sigma^2 &\sim N(\mu_n, \sigma^2 \frac{1}{\tau_n}) \\ \sigma^2 | y &\sim \text{IG}(a_n, k_n) \end{aligned}$$

with

$$\mu_n = \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0}; \quad \tau_n = n + \tau_0; \quad a_n = a_0 + n$$

$$k_n = k_0 + \frac{1}{2} ns_n^2 + \frac{1}{2} \frac{\tau_0 n (\mu_0 - \bar{y})^2}{n + \tau_0}$$

Hint: It is

$$-\frac{1}{2} \frac{(\mu - \mu_1)^2}{v_1} - \frac{1}{2} \frac{(\mu - \mu_2)^2}{v_2} \dots - \frac{1}{2} \frac{(\mu - \mu_n)^2}{v_n} = -\frac{1}{2} \frac{(\mu - \hat{\mu})^2}{\hat{v}} + C$$

where

$$\hat{v} = \left(\sum_{i=1}^n \frac{1}{v_i} \right)^{-1}; \quad \hat{\mu} = \hat{v} \left(\sum_{i=1}^n \frac{\mu_i}{v_i} \right); \quad C = \frac{1}{2} \frac{\hat{\mu}^2}{\hat{v}} - \frac{1}{2} \sum_{i=1}^n \frac{\mu_i^2}{v_i}$$

3. Show that the marginal posterior distribution $\Pi(\mu | y)$ is such as

$$\mu | y \sim T_1 \left(\mu_n, \frac{k_n}{a_n} \frac{1}{\tau_n}, 2a_n \right)$$

Hint-1: If $x \sim \text{IG}(a, b)$, $y = cx$, then $y \sim \text{IG}(a, cb)$.

Hint-2: The definition of Student T is considered as known

4. Show that the predictive distribution $\Pi(z | y)$ is Student T such as

$$z | y \sim T_1 \left(\mu_n, \frac{k_n}{a_n} \left(\frac{1}{\tau_n} + 1 \right), 2a_n \right)$$

Hint-1: Consider that

$$N(x | \mu_1, \sigma_1^2) N(x | \mu_2, \sigma_2^2) = N(x | m, v^2) N(\mu_1 | \mu_2, \sigma_1^2 + \sigma_2^2)$$

where

$$v^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}; \quad m = v^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right)$$

Hint-2: The definition of Student T is considered as known

Solution.

1. It is

$$\begin{aligned} \sum_{i=1}^n (y_i - \theta)^2 &= \sum_{i=1}^n [(y_i - \bar{y}) - (\theta - \bar{y})]^2 \\ &= \sum_{i=1}^n \left[(y_i - \bar{y})^2 + (\theta - \bar{y})^2 - 2(y_i - \bar{y})(\theta - \bar{y}) \right] \\ &= ns^2 + n(\bar{y} - \theta)^2, \text{ where } s^2 = \frac{1}{2} \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

2. I use the Bayes theorem

$$\begin{aligned} \pi(\mu, \sigma^2 | y) &\propto f(y | \mu, \sigma^2) \pi(\mu, \sigma^2) = \prod_{i=1}^n N(y_i | \mu, \sigma^2) N(\mu | \mu_0, \sigma^2 \frac{1}{\tau_0}) \text{IG}(\sigma^2 | a_0, k_0) \\ &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} \right) \times \left(\frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp \left(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma^2 / \tau_0} \right) \times \left(\frac{1}{\sigma^2} \right)^{a_0+1} \exp \left(-\frac{1}{\sigma^2} k_0 \right) \\ &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2} + \frac{1}{2} + a_0 + 1} \exp \left(\frac{1}{\sigma^2} \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{1} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{1/\tau_0} \right] - \frac{1}{\sigma^2} k_0 \right) \end{aligned}$$

It is

$$-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{1} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{1/\tau_0} = -\frac{1}{2} \frac{(\mu - \mu_n)^2}{\underbrace{v_n^2}_{=1/\tau_n}} + C_n$$

where

$$\begin{aligned} v_n &= \left(\sum_{i=1}^n \frac{1}{1} + \frac{1}{1/\tau_0} \right)^{-1} = \frac{1}{n + \tau_0} \implies \tau_n = n + \tau_0 \\ \mu_n &= v_n \left(\sum_{i=1}^n \frac{y_i}{1} + \frac{\mu_0}{1/\tau_0} \right) \implies \mu_n = \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0} \\ C_n &= \frac{1}{2} \frac{\mu_n^2}{v_n} - \frac{1}{2} \left(n \sum_{i=1}^n y_i^2 + \tau_0\mu_0^2 \right) = \frac{1}{2} \frac{(n\bar{y} + \tau_0\mu_0)^2}{n + \tau_0} - \frac{1}{2} \left(n \sum_{i=1}^n y_i^2 + \tau_0\mu_0^2 \right) \\ &= \dots \text{Quest. 1} \dots = -\frac{1}{2} ns_n^2 - \frac{1}{2} \frac{\tau_0 n (\mu_0 - \bar{y})^2}{n + \tau_0} \end{aligned}$$

So

$$\begin{aligned}\pi(\mu, \sigma^2|y) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2} + \frac{n}{2} + a_0 + 1} \exp\left(\frac{1}{\sigma^2} \left[-\frac{1}{2} \frac{(\mu - \mu_n)^2}{1/\tau_n} + C_n\right] - \frac{1}{\sigma^2} k_0\right) \\ &\propto \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{(\mu - \mu_n)^2}{\sigma^2/\tau_n}\right)}_{\propto N(\mu|\mu_n, \sigma^2/\tau_n)} \times \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2} + a_0 + 1} \exp\left(-\frac{1}{\sigma^2} \overbrace{(k_0 - C_n)}^{=k_n}\right)}_{\propto IG(\sigma^2|a_n, k_n)} \\ &\propto N(\mu|\mu_n, \sigma^2/\tau_n) IG(\sigma^2|a_n, k_n)\end{aligned}$$

where

$$\begin{aligned}\mu_n &= \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0}; & a_n &= \frac{n}{2} + a_0; \\ \tau_n &= n + \tau_0; & k_n &= k_0 + \frac{1}{2}ns_n^2 + \frac{1}{2}\frac{\tau_0n(\mu_0 - \bar{y})^2}{n + \tau_0}.\end{aligned}$$

3. It is

$$\pi(\mu|y) = \int \pi(\mu, \sigma^2|y) d\sigma^2 = \int N(\mu|\mu_n, \sigma^2/\tau_n) IG(\sigma^2|a_n, k_n) d\sigma^2$$

by change of variable $\xi = \sigma^2 \frac{1}{2k_n}$, it is

$$\begin{aligned}\pi(\mu|y) &= \int N(\mu|\mu_n, \xi 2k_n \frac{1}{\tau_n} \frac{2a_n}{2a_n}) IG(\xi|\frac{2a_n}{2}, \frac{1}{2}) d\xi = \int N(\mu|\mu_n, \xi \frac{1}{\tau_n} \frac{k_n}{a_n} 2a_n) IG(\xi|\frac{2a_n}{2}, \frac{1}{2}) d\xi \\ &= T_1(\mu|\mu_n, \frac{k_n}{a_n} \frac{1}{\tau_n}, 2a_n)\end{aligned}$$

4. It is

$$\begin{aligned}g(z|y) &= \int f(z|\mu, \sigma^2) \pi(\mu, \sigma^2|y) d\mu d\sigma^2 = \int N(z|\mu, \sigma^2) N(\mu|\mu_n, \sigma^2/\tau_n) IG(\sigma^2|a_n, k_n) d\mu d\sigma^2 \\ &= \int \left[\int N(z|\mu, \sigma^2) N(\mu|\mu_n, \sigma^2/\tau_n) d\mu \right] IG(\sigma^2|a_n, k_n) d\sigma^2\end{aligned}$$

Normal density is symmetric $N(z|\mu, \sigma^2) N(\mu|\mu_n, \sigma^2/\tau_n) = N(\mu|z, \sigma^2) N(\mu|\mu_n, \sigma^2/\tau_n)$, and by using the Hint

$$\int N(\mu|z, \sigma^2) N(\mu|\mu_n, \sigma^2/\tau_n) d\mu = \int N(\mu|\text{const.}, \text{const.}) N\left(z|\mu_n, \sigma^2 \left[\frac{1}{\tau_n} + 1\right]\right) d\mu = N\left(z|\mu_n, \sigma^2 \left[\frac{1}{\tau_n} + 1\right]\right)$$

So

$$g(z|y) = \int N\left(z|\mu_n, \sigma^2 \left[\frac{1}{\tau_n} + 1\right]\right) IG(\sigma^2|a_n, k_n) d\sigma^2$$

by change the variable $\xi = \sigma^2 \frac{1}{2k_n}$, it is

$$g(z|y) = \int N\left(z|\mu_n, \xi \left[\frac{1}{\tau_n} + 1\right] \frac{k_n}{a_n} 2a_n\right) IG(\xi|\frac{2a_n}{2}, \frac{1}{2}) d\xi = T_1\left(z|\mu_n, \left[\frac{1}{\tau_n} + 1\right] \frac{k_n}{a_n}, 2a_n\right)$$

The following is about the Normal linear model of regression. ~~The calculations are too challenging; (not anymore...)~~

Exercise 26. (★★)(Normal linear regression model with unknown error variance)

<-story

Consider we are interested in recovering the mapping

$$x \xrightarrow{\eta(x)} y$$

in the sense that y is the response (output quantity) that depends on x which is the independent variable (input quantity) in a procedure; E.g.:

- y : precipitation in log scale
- x = (longitude, latitude): geographical coordinates.

It is believed that the mapping $\eta(x)$ can be represented as an expansion of d known polynomial functions $\{\phi_j(x)\}_{j=0}^{d-1}$ such as

$$\eta(x) = \sum_{j=0}^{d-1} \phi_j(x) \beta_j = \Phi(x)^\top \beta; \quad \text{with } \Phi(x) = (\phi_0(x), \dots, \phi_{d-1}(x))^\top$$

where $\beta \in \mathbb{R}^d$ is unknown.

Assume observable quantities (data) in pairs (x_i, y_i) for $i = 1, \dots, n$; (E.g. from the i -th station at location x_i I got the reading y_i). Assume that the response observations $y = (y_1, \dots, y_n)$ may be contaminated by noise with unknown variance; such that

$$y_i = \eta(x_i) + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma^2)$ with unknown σ^2 .

You are interested in learning β , but you do not care about σ^2 . Also you want to learn the value of y_f at an untried x_f (i.e. the precipitation at any other location).

Consider the Bayesian model

<-set-up

$$y|\beta, \sigma^2 \sim N(\Phi\beta, I\sigma^2); \text{ the sampling distr}$$

$$\beta|\sigma^2 \sim N(\mu_0, V_0\sigma^2); \text{ prior distr}$$

$$\sigma^2 \sim \text{IG}(a_0, k_0) \text{ prior distr}$$

where Φ is the design matrix $[\Phi]_{i,j} = \Phi_j(x_i)$.

1. Show that the joint posterior distribution $d\Pi(\beta, \sigma^2|y)$ is such as

$$\beta|y, \sigma^2 \sim N(\mu_n, V_n\sigma^2); \quad \sigma^2|y \sim \text{IG}(a_n, k_n)$$

with

$$V_n^{-1} = \Phi^\top \Phi + V_0^{-1}; \quad \mu_n = V_n ((\Phi^\top \Phi)^{-1} \Phi^\top y + V_0^{-1} \mu_0); \quad a_n = \frac{n}{2} + a_0$$

$$k_n = \frac{1}{2} (y - \Phi \hat{\beta}_n)^\top (y - \Phi \hat{\beta}_n) - \frac{1}{2} \mu_n^\top V_n^{-1} \mu_n + \frac{1}{2} (\mu_0^\top V_0^{-1} \mu_0 + y^\top \Phi^\top (\Phi^\top \Phi)^{-1} \Phi y) + k_0$$

Hint-1:

$$(y - \Phi\beta)^\top (y - \Phi\beta) = (\beta - \hat{\beta}_n)^\top [\Phi^\top \Phi] (\beta - \hat{\beta}_n) + S_n; \quad S_n = (y - \Phi \hat{\beta}_n)^\top (y - \Phi \hat{\beta}_n); \quad \hat{\beta}_n = (\Phi^\top \Phi)^{-1} \Phi^\top y$$

Hint-2: If $\Sigma_1 > 0$ and $\Sigma_2 > 0$ symmetric

$$-\frac{1}{2} (x - \mu_1)^\top \Sigma_1^{-1} (x - \mu_1) - \frac{1}{2} (x - \mu_2)^\top \Sigma_2^{-1} (x - \mu_2) = -\frac{1}{2} (x - m)^\top V^{-1} (x - m) + C$$

where

$$V^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}; \quad m = V (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2); \quad C = \frac{1}{2} m^\top V^{-1} m - \frac{1}{2} (\mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_2^\top \Sigma_2^{-1} \mu_2)$$

2. Show that the marginal posterior of β given y is

$$\beta|y \sim T_d(\mu_n, V_n \frac{k_n}{a_n}, 2a_n)$$

3. Show that the predictive distribution of an outcome $y_f = \Phi_f \beta + \epsilon$ with $\Phi_f = (\phi_0(x_f), \dots, \phi_{d-1}(x_f))$ and $\epsilon \sim N(0, \sigma^2)$ at untried location x_f is

$$y_f|y \sim T_d(\mu_n, [\Phi^\top \Phi + 1] \frac{k_n}{a_n}, 2a_n)$$

Consider that

$$N(x|\mu_1, \sigma_1^2) N(x|\mu_2, \sigma_2^2) = N(x|m, v^2) N(\mu_1|\mu_2, \sigma_1^2 + \sigma_2^2)$$

where

$$v^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}; \quad m = v^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right)$$

Hint-2: The definition of Student T is considered as known

Solution.

1. I use the Bayes theorem

$$\begin{aligned} \pi(\mu, \sigma^2|y) &\propto f(y|\mu, \sigma^2) \pi(\mu, \sigma^2) = N(y|\Phi\beta, I\sigma^2) N(\beta|\mu_0, \sigma^2 V_0) \text{IG}(\sigma^2|a_0, k_0) \\ &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left(-\frac{1}{2} (y - \Phi\beta)^\top (I\sigma^2)^{-1} (y - \Phi\beta) \right) \times \left(\frac{1}{\sigma^2} \right)^{\frac{d}{2}} \exp \left(-\frac{1}{2} (\beta - \mu_0)^\top (V_0 \sigma^2)^{-1} (\beta - \mu_0) \right) \\ &\quad \times \left(\frac{1}{\sigma^2} \right)^{a_0+1} \exp \left(-\frac{1}{\sigma^2} k_0 \right) \end{aligned}$$

but

$$(y - \Phi\beta)^\top (y - \Phi\beta) = (\beta - \hat{\beta}_n)^\top [\Phi^\top \Phi] (\beta - \hat{\beta}_n) + S_n; \quad S_n = (y - \Phi\hat{\beta}_n)^\top (y - \Phi\hat{\beta}_n); \quad \hat{\beta}_n = (\Phi^\top \Phi)^{-1} \Phi^\top y$$

so

$$\begin{aligned} \pi(\mu, \sigma^2|y) &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left(-\frac{1}{2} \frac{1}{\sigma^2} (\beta - \hat{\beta}_n)^\top [\Phi^\top \Phi] (\beta - \hat{\beta}_n) - \frac{1}{2} \frac{1}{\sigma^2} S_n \right) \\ &\quad \times \left(\frac{1}{\sigma^2} \right)^{\frac{d}{2}} \exp \left(-\frac{1}{2} (\beta - \mu_0)^\top (V_0 \sigma^2)^{-1} (\beta - \mu_0) \right) \times \left(\frac{1}{\sigma^2} \right)^{a_0+1} \exp \left(-\frac{1}{\sigma^2} k_0 \right) \\ &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{d}{2}} \exp \left(-\frac{1}{2} \frac{1}{\sigma^2} (\beta - \hat{\beta}_n)^\top [\Phi^\top \Phi] (\beta - \hat{\beta}_n) - \frac{1}{2} \frac{1}{\sigma^2} (\beta - \mu_0)^\top V_0^{-1} (\beta - \mu_0) \right) \\ &\quad \times \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2} + a_0 + 1} \exp \left(-\frac{1}{2} \frac{1}{\sigma^2} S_n - \frac{1}{\sigma^2} k_0 \right) \end{aligned}$$

but

$$-\frac{1}{2} (\beta - \hat{\beta}_n)^\top [\Phi^\top \Phi] (\beta - \hat{\beta}_n) - \frac{1}{2} (\beta - \mu_0)^\top V_0^{-1} (\beta - \mu_0) = -\frac{1}{2} (\beta - \mu_n)^\top V_n^{-1} (\beta - \mu_n) + \frac{1}{2} C_n$$

$$V_n^{-1} = \Phi^\top \Phi + V_0^{-1}; \quad \mu_n = V_n \left(\Phi^\top \hat{\Phi} \hat{\beta}_n + V_0^{-1} \mu_0 \right) = V_n \left((\Phi^\top \Phi)^{-1} \Phi y + V_0^{-1} \mu_0 \right)$$

$$C_n = \frac{1}{2} \mu_n^\top V_n^{-1} \mu_n - \frac{1}{2} \left(\mu_0^\top V_0^{-1} \mu_0 + \hat{\beta}_n^\top [\Phi^\top \Phi] \hat{\beta}_n \right) = \frac{1}{2} \mu_n^\top V_n^{-1} \mu_n - \frac{1}{2} \left(\mu_0^\top V_0^{-1} \mu_0 + y^\top \Phi^\top (\Phi^\top \Phi)^{-1} \Phi y \right)$$

So

$$\pi(\mu, \sigma^2 | y) \propto \underbrace{\left(\frac{1}{|V_n \sigma^2|} \right)^{\frac{1}{2}} \exp \left(-\frac{1}{2} (\beta - \mu_n)^\top [V_n \sigma^2]^{-1} (\beta - \mu_n) \right)}_{\propto N_d(\beta | \mu_n, V_n \sigma^2)} \times \underbrace{\left(\frac{1}{\sigma^2} \right)^{\frac{n}{2} + a_0 + 1} \exp \left(-\frac{1}{\sigma^2} \left[\frac{1}{2} S_n - C_n + k_0 \right] \right)}_{\propto \text{IG}(\sigma^2 | a_n, k_n)}$$

So

$$\begin{cases} \mu | \sigma^2 \sim N(\mu_n, \sigma^2 V_n) \\ \sigma^2 \sim \text{IG}(a_n, k_n) \end{cases}$$

2. It is

$$\pi(\beta | y) = \int \pi(\beta, \sigma^2 | y) d\sigma^2 = \int N(\beta | \mu_n, V_n \sigma^2) \text{IG}(\sigma^2 | a_n, k_n) d\sigma^2$$

by change the variable $\xi = \sigma^2 \frac{1}{2k_n}$, it is

$$\begin{aligned} \pi(\beta | y) &= \int N(\beta | \mu_n, \xi 2k_n V_n \frac{2a_n}{2a_n}) \text{IG}(\xi | \frac{2a_n}{2}, \frac{1}{2}) d\xi = \int N(\beta | \mu_n, \xi V_n \frac{k_n}{a_n} 2a_n) \text{IG}(\xi | \frac{2a_n}{2}, \frac{1}{2}) d\xi \\ &= T_d(\beta | \mu_n, \frac{k_n}{a_n} V_n, 2a_n) \end{aligned}$$

3. It is

$$\begin{aligned} g(y_f | y) &= \int f(y_f | \Phi_f \beta, \sigma^2) \pi(\beta, \sigma^2 | y) d\beta d\sigma^2 = \int N(y_f | \Phi_f \beta, \sigma^2) N(\beta | \mu_n, V_n \sigma^2) \text{IG}(\sigma^2 | a_n, k_n) d\beta d\sigma^2 \\ &= \int \underbrace{\left[\int N(y_f | \Phi_f \beta, \sigma^2) N(\beta | \mu_n, V_n \sigma^2) d\beta \right]}_{=A} \text{IG}(\sigma^2 | a_n, k_n) d\sigma^2 \end{aligned}$$

by change of variable for $\xi' = \Phi_f \beta \sim N(\Phi_f \mu_n, \Phi_f^\top V_n \Phi_f \sigma^2)$

$$A = \int N(y_f | \xi', \sigma^2) N(\xi' | \Phi_f \mu_n, \Phi_f^\top V_n \Phi_f \sigma^2) d\xi'$$

because Normal is symmetric around the mean

$$A = \int N(\xi' | y_f, \sigma^2) N(\xi' | \Phi_f \mu_n, \Phi_f^\top V_n \Phi_f \sigma^2) d\xi'$$

by using the Hint

$$A = \int N(\xi' | \text{const.}, \text{const.}) N(y_f | \Phi_f \mu_n, \sigma^2 [\Phi_f^\top V_n \Phi_f + 1]) d\xi' = N(y_f | \Phi_f \mu_n, \sigma^2 [\Phi_f^\top V_n \Phi_f + 1])$$

So

$$g(y_f | y) = \int N(y_f | \Phi_f \mu_n, \sigma^2 [\Phi_f^\top V_n \Phi_f + 1]) \text{IG}(\sigma^2 | a_n, k_n) d\sigma^2$$

by change the variable $\xi = \sigma^2 \frac{1}{2k_n}$, it is

$$g(y_f|y) = \int \mathcal{N}\left(y_f|\Phi_f\mu_n, \xi [\Phi_f^\top V_n \Phi_f + 1] \frac{k_n}{a_n} 2a_n\right) \text{IG}\left(\xi|\frac{2a_n}{2}, \frac{1}{2}\right) d\xi = \mathcal{T}_1\left(y_f|\Phi_f\mu_n, [\Phi_f^\top V_n \Phi_f + 1] \frac{k_n}{a_n}, 2a_n\right)$$

So

$$y_f|y \sim \mathcal{T}_1\left(\Phi_f\mu_n, [\Phi_f^\top V_n \Phi_f + 1] \frac{k_n}{a_n}, 2a_n\right)$$

, or equiv.

$$y(x_f)|y \sim \mathcal{T}_1\left(\phi^\top(x_f)\mu_n, [\Phi_f^\top V_n \Phi_f + 1] \frac{k_n}{a_n}, 2a_n\right)$$

Exercise 27. (★★) Let $y = (y_1, \dots, y_n)$ be observables drawn iid from sampling distribution $y_i|\theta \stackrel{\text{iid}}{\sim} \mathcal{N}(\theta, \theta^2)$ for all $i = 1, \dots, n$, where $\theta \in \mathbb{R}$ is unknown. Specify a conjugate prior density for θ up to an unknown normalizing constant.

Solution. The sampling distribution is

$$f(y_i|\theta) = \mathcal{N}(y_i|\theta, \theta^2) \propto (\theta^2)^{-1/2} \exp\left(-\frac{1}{2} \frac{(y_i - \theta)^2}{\theta^2}\right) \propto |\theta|^{-1} \exp\left(-\frac{1}{2} y_i^2 \frac{1}{\theta^2} + y_i \frac{1}{\theta}\right)$$

and hence it belongs to the exponential family with $g(\theta) = |\theta|^{-1}$, $c_1 = -\frac{1}{2}$, $\phi_1(\theta) = \frac{1}{\theta^2}$, $h_1(y_i) = y_i^2$, $c_2 = 1$, $\phi_2(\theta) = \frac{1}{\theta}$, $h_2(y_i) = y_i$.

The corresponding conjugate prior has pdf such as

$$\pi(\theta) = \tilde{\pi}(\theta|\tau) \propto |\theta|^{-\tau_0} \exp\left(-\frac{1}{2} \frac{1}{\theta^2} \tau_1 + \frac{1}{\theta} \tau_2\right), \quad \text{where } \tau = (\tau_0, \tau_1, \tau_2).$$

I actually cannot recognize it as a standard distribution in this case. The posterior distribution has pdf such as

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta) = \prod_{i=1}^n \mathcal{N}(y_i|\theta, \theta^2)\pi(\theta) \propto |\theta|^{-(\tau_0+n)} \exp\left(-\frac{1}{2} \frac{1}{\theta^2} (\tau_1 + \sum_{i=1}^n y_i^2) + \frac{1}{\theta} (\tau_2 + \sum_{i=1}^n y_i)\right)$$

Namely, $\pi(\theta|y) = \tilde{\pi}(\theta|\tau^*)$, with $\tau^* = (\tau_0 + n, \tau_1 + \sum_{i=1}^n y_i^2, \tau_2 + \sum_{i=1}^n y_i)$; so it is conjugate.

Exercise 28. (★★) If the sampling distribution $F(\cdot|\theta)$ is discrete and the prior $\Pi(\theta)$ is proper, then the posterior $\Pi(\theta|y)$ is always proper.

Solution. It is

$$f(y) \leq \sum_{\forall y} f(y) = \sum_{\forall y} \overbrace{\int f(y|\theta) d\Pi(\theta)}^{f(y)=} \stackrel{\text{Fubini}}{=} \int \sum_{\forall y} f(y|\theta) d\Pi(\theta) = \int d\Pi(\theta) = 1$$

Exercise 29. (★★) If the sampling distribution $F(\cdot|\theta)$ is continuous and the prior $\Pi(\theta)$ is proper, then the posterior $\Pi(\theta|y)$ is almost always proper.

Solution. It is

$$\int f(y) dy = \int_{\forall y} \overbrace{\int_{\forall \theta} f(y|\theta) d\Pi(\theta)}^{f(y)=} \stackrel{\text{Fubini}}{=} \int_{\forall \theta} \int_{\forall y} f(y|\theta) dy d\Pi(\theta) = \int d\Pi(\theta) = 1$$

So it is $f(y) < \infty$ for every set of y (possibly) apart from a finite number of y 's with 'probability' zero.

The Limit Comparison Theorem for Improper Integrals

General: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$.

Let

$$0 \leq f(x) \leq g(x), \quad \text{for } x \geq a$$

Then

$$\begin{aligned} \int_a^\infty g(x) dx < \infty &\implies \int_a^\infty f(x) dx < \infty \\ \int_a^\infty f(x) dx = \infty &\implies \int_a^\infty g(x) dx = \infty \end{aligned}$$

Type I: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$, and let $g(x)$ be positive.

Let

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:

$$\int_a^\infty g(x) dx < \infty \iff \int_a^\infty f(x) dx < \infty$$

- If $c = 0$:

$$\int_a^\infty g(x) dx < \infty \implies \int_a^\infty f(x) dx < \infty$$

- If $c = \infty$:

$$\int_a^\infty f(x) dx = \infty \implies \int_a^\infty g(x) dx = \infty$$

Type II: Let integrable functions $f(x)$, and $g(x)$ for $a < x \leq b$, and let $g(x)$ be positive.

Let

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:

$$\int_a^\infty g(x) dx < \infty \iff \int_a^\infty f(x) dx < \infty$$

- If $c = 0$:

$$\int_a^\infty g(x) dx < \infty \implies \int_a^\infty f(x) dx < \infty$$

- If $c = \infty$:

$$\int_a^\infty f(x) dx = \infty \implies \int_a^\infty g(x) dx = \infty$$

Note: A useful test function is

$$\int_0^\infty \left(\frac{1}{x}\right)^p dx \begin{cases} < \infty & , \text{ when } p > 1 \\ = \infty & , \text{ when } p \leq 1 \end{cases}$$

Exercise 30. (★★) Consider the Bayesian model

$$\begin{cases} x|\sigma & \sim N(0, \sigma^2) \\ \sigma & \sim \text{Ex}(\lambda) \end{cases}$$

where $\text{Ex}(\lambda)$ is the exponential distribution with mean $1/\lambda$. Show that the posterior distribution is not defined always.

- HINT: Precisely, show that the posterior is not defined in the case that you collect only one observation $x = 0$.

Solution.

It is

$$\begin{aligned} f(x) &\propto \int_{\mathbb{R}_+} N(x|0, \sigma^2) \text{Ex}(\sigma|\lambda) d\sigma = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-0)^2\right) \lambda \exp(-\sigma\lambda) d\sigma \\ f(x=0) &\propto \int_0^\infty \frac{1}{\sigma} \exp(-\sigma\lambda) d\sigma \end{aligned}$$

We will use a convergence criteria in order to check if $\int_0^\infty \frac{1}{\sigma} \exp(-\sigma\lambda) d\sigma = \infty$.

I will use the Limit Comparison Test to check if $\int_0^\infty \frac{1}{\sigma} \exp(-\sigma\lambda) d\sigma = \infty$. Consider $h(\sigma) = \frac{1}{\sigma} \exp(-\sigma\lambda)$. The function $h(\sigma)$ has an improper behavior at 0, as it is not bounded there. Let $g(\sigma) = \frac{1}{\sigma}$. According to the Limit Comparison Test, it is

$$\lim_{\sigma \rightarrow 0^+} \frac{h(\sigma)}{g(\sigma)} = \lim_{\sigma \rightarrow 0^+} \frac{\frac{1}{\sigma} \exp(-\sigma\lambda)}{\frac{1}{\sigma}} = 1 \neq 0$$

and

$$\int_0^\infty g(\sigma) d\sigma = \int_0^\infty \frac{1}{\sigma} d\sigma = \infty.$$

Therefore, it will be

$$\underbrace{\int_0^\infty h(\sigma) d\sigma}_{=f(x=0)} = \infty$$

as well.

Exercise 31. (★★) Consider the Bayesian model

$$\begin{cases} x|\sigma & \sim N(0, \sigma^2) \\ \sigma & \sim \Pi(\sigma) \end{cases}$$

where $\Pi(\sigma)$ is an improper prior distribution with density such as $\pi(\sigma) \propto \sigma^{-1} \exp(-a\sigma^{-2})$ for $a > 0$. Show that we can use this prior on Bayesian inference.

Solution.

We will check the properness condition. It is

$$\begin{aligned} f(x) &= \int_{\mathbb{R}_+} N(x|0, \sigma^2) \text{Ex}(\sigma|\lambda) d\sigma \propto \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-0)^2\right) \sigma^{-1} \exp(-a\sigma^{-2}) d\sigma \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^2 + 2a)\right) d\sigma \\ &= \int_0^\infty \frac{1}{\sqrt{\xi}} \exp\left(-\frac{\xi}{2}(x^2 + 2a)\right) d\xi \end{aligned}$$

for $\xi = 1/\sigma^2$. It is

$$f(x) \propto \int_0^\infty \frac{1}{\sqrt{\xi}} \exp\left(-\underbrace{\frac{\xi}{2}(x^2 + 2a)}_{\substack{<0 \\ \in (0,1)}}\right) d\xi \leq \int_0^\infty \frac{1}{\sqrt{\xi}} d\xi < \infty$$

So the posterior is defined.

Exercise 32. (**) Let x be an observation. Consider the Bayesian model

$$\begin{cases} x|\theta & \sim \text{Pn}(\theta) \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\text{Pn}(\theta)$ is the Poisson distribution with expected value θ . Consider a prior $\Pi(\theta)$ with density such as $\pi(\theta) \propto \frac{1}{\theta}$. Show that the posterior distribution is not always defined.

Hint-1: It suffices to show that the posterior is not defined in the case that you collect only one observation $x = 0$.

Hint-2: Poisson distribution: $x \sim \text{Pn}(\theta)$ has PMF

$$\text{Pn}(x|\theta) = \frac{\theta^x \exp(-\theta)}{x!} 1(x \in \mathbb{N})$$

Solution.

The next exercise is about the Sequential processing of data via Bayes theorem

Exercise 33. (**) Assume that observable quantities x_1, x_2, \dots are generated i.i.d by a process that can be modeled as a sampling distribution $N(\mu, \sigma^2)$ with known σ^2 and unknown μ .

1. Assume that you have collected an observation x_1 . Specify a prior $\Pi(\mu)$ on μ as $\mu \sim N(\mu_0, \sigma_0^2)$ where μ_0, σ_0^2 are known.

- Derive the posterior $\Pi(\theta|x_1)$.

Next assume that you additionally observe an additional observation x_2 after collecting x_1 . Consider the posterior $\Pi(\mu|x_1)$ as the current state of your knowledge about θ .

- Derive the posterior $\Pi(\mu|x_1, x_2)$ in the light of the new additional observation x_2 .

2. Assume that you have collected two observations (x_1, x_2) . Specify a prior $\Pi(\mu)$ on μ as $\mu \sim N(\mu_0, \sigma_0^2)$ where μ_0, σ_0^2 are known.

- Derive the posterior $\Pi(\theta|x_1, x_2)$ in the light of the observations (x_1, x_2) .

3. What do you observe:

Hint: We considered the identity

$$-\frac{1}{2} \sum_{i=1}^n \frac{(y - \mu_i)^2}{\sigma_i^2} = -\frac{1}{2} \frac{(y - \hat{\mu})^2}{\hat{\sigma}^2} + c(\hat{\mu}, \hat{\sigma}^2),$$

$$c(\hat{\mu}, \hat{\sigma}^2) = -\frac{1}{2} \sum_{i=1}^n \frac{\mu_i^2}{\sigma_i^2} + \frac{1}{2} \left(\sum_{i=1}^n \frac{\mu_i}{\sigma_i^2} \right)^2 \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1}; \quad \hat{\sigma}^2 = \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1}; \quad \hat{\mu} = \hat{\sigma}^2 \left(\sum_{i=1}^n \frac{\mu_i}{\sigma_i^2} \right)$$

where $c(\hat{\mu}, \hat{\sigma}^2)$ is constant w.r.t. y .

Solution.

1. the posterior distribution $\Pi(\mu|x_1)$ has PDF

$$\begin{aligned}
 \pi(\mu|x_1) &\propto \overbrace{\text{N}(x_1|\mu, \sigma^2)}^{\text{likelihood}} \overbrace{\text{N}(\mu|\mu_0, \sigma_0^2)}^{\text{prior}} \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_1 - \mu)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_1 - \mu)^2}{\sigma^2} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \hat{\mu}_1)^2}{\hat{\sigma}_1^2}\right) \propto \text{N}(\mu|\hat{\mu}_1, \hat{\sigma}_1^2)
 \end{aligned} \tag{3}$$

where $\hat{\sigma}_1^2 = (\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2})^{-1}$, and $\hat{\mu}_1 = \hat{\sigma}_1^2(\frac{x_1}{\sigma^2} + \frac{\mu_0}{\sigma_0^2})$. In (3), we recognized the kernel of the Normal PDF. Hence, $\mu|x_1 \sim \text{N}(\hat{\mu}_1, \hat{\sigma}_1^2)$

Then the posterior distribution $\Pi(\mu|x_1, x_2)$ has PDF

$$\begin{aligned}
 \pi(\mu|x_1, x_2) &\propto \overbrace{\text{N}(x_2|\mu, \sigma^2)}^{\text{likelihood}} \overbrace{\text{N}(\mu|\hat{\mu}_1, \hat{\sigma}_1^2)}^{\text{prior}} \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_2 - \mu)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2} \frac{(\mu - \hat{\mu}_1)^2}{\hat{\sigma}_1^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_2 - \mu)^2}{\sigma^2} - \frac{1}{2} \frac{(\mu - \hat{\mu}_1)^2}{\hat{\sigma}_1^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \hat{\mu}_2)^2}{\hat{\sigma}_2^2}\right) \propto \text{N}(\mu|\hat{\mu}_2, \hat{\sigma}_2^2)
 \end{aligned} \tag{4}$$

where $\hat{\sigma}_2^2 = (\frac{1}{\sigma^2} + \frac{1}{\hat{\sigma}_1^2})^{-1} = (\frac{1}{\sigma^2} + \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2})^{-1}$, and $\hat{\mu}_2 = \hat{\sigma}_2^2(\frac{x_1}{\sigma^2} + \frac{\hat{\mu}_1}{\hat{\sigma}_1^2}) = \hat{\sigma}_2^2(\frac{x_1}{\sigma^2} + \frac{x_2}{\sigma^2} + \frac{\mu_0}{\sigma_0^2})$. In (3), we recognized the kernel of the Normal PDF. Hence, $\mu|x_1, x_2 \sim \text{N}(\hat{\mu}_2, \hat{\sigma}_2^2)$.

2. The posterior distribution $\Pi(\mu|x_1, x_2)$ has PDF

$$\begin{aligned}
 \pi(\mu|x_1, x_2) &\propto \overbrace{\text{N}(x_1|\mu, \sigma^2)\text{N}(x_2|\mu, \sigma^2)}^{\text{likelihood}} \overbrace{\text{N}(\mu|\mu_0, \sigma_0^2)}^{\text{prior}} \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_1 - \mu)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2} \frac{(x_2 - \mu)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(x_1 - \mu)^2}{\sigma^2} - \frac{1}{2} \frac{(x_2 - \mu)^2}{\sigma^2} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \\
 &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \hat{\mu})^2}{\hat{\sigma}^2}\right) \propto \text{N}(\mu|\hat{\mu}, \hat{\sigma}^2)
 \end{aligned} \tag{5}$$

where $\hat{\sigma}^2 = (\frac{1}{\sigma^2} + \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2})^{-1}$, and $\hat{\mu} = \hat{\sigma}^2(\frac{x_1}{\sigma^2} + \frac{x_2}{\sigma^2} + \frac{\mu_0}{\sigma_0^2})$. In (5), we recognized the kernel of the Normal PDF. Hence, $\mu|x_1, x_2 \sim \text{N}(\hat{\mu}, \hat{\sigma}^2)$

3. It is easy to see that $\hat{\mu}_2 = \hat{\mu}$, and $\hat{\sigma}_2^2 = \hat{\sigma}^2$, from (1) and (2). We observe the two Learning Scenarios are equivalent in the sense that they lead to the same posterior $d\Pi(\mu|x_1, x_2)$ at the end posterior $d\Pi(\mu|x_1, x_2)$ in a single application of Bayes theorem with the full data $x = (x_1, x_2)$.