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.

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

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

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

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

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

So

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

Also

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

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

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

For if $h \searrow$ it is

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

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

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

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

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

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 $y \sim \mathrm{Ex}(\lambda)$ r.v. with Exponential distribution with rate parameter $\lambda > 0$, and $f_{\mathrm{Ex}(\lambda)}(y) = \lambda\exp(-\lambda y)1(y \ge 0)$. 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

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

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
$$\begin{aligned} \mathrm{E}(\psi_1(z) + \psi_2(y)) =& \int (\psi_1(z) + \psi_2(y))\,\mathrm{d}F((z,y)) = \int \psi_1(z)\mathrm{d}F((z,y)) + \int \psi_1(z)\mathrm{d}F((z,y)) \\ =& \int \psi_1(z)\mathrm{d}F(z) + \int \psi_1(z)\mathrm{d}F(z) = \mathrm{E}(\psi_1(z)) + \mathrm{E}(\psi_2(y)) \end{aligned}$$
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 $x \in \mathbb{R}^q$

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

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

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

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

Solution.

1. It is
$$\varphi_{A+Bx}(t) = \text{E}(e^{it^T(A+Bx)}) = \text{E}(e^{A+it^TBx}) = \text{E}(e^{it^TA}e^{iB^Ttx}) = e^{it^TA}\text{E}(e^{i(B^Tt)x}) = e^{it^TA}\varphi_x(B^Tt)$$
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

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

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

1. if $Z \sim \mathrm{N}(0, I)$ then $\varphi_Z(t) = \exp(-\frac{1}{2}t^T t)$, 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

$$\varphi_Z(t) = \mathrm{E}(\exp(it^T Z)) = \mathrm{E}(\exp(i\sum_{j=1}^{d}(t_j Z_j))) = \mathrm{E}(\prod_{j=1}^{d} \exp(it_j Z_j)) = \prod_{j=1}^{d} \mathrm{E}(\exp(it_j Z_j))$$

$$= \prod_{j=1}^{d} \varphi_{Z_j}(t) = \prod_{j=1}^{d} \exp(-\frac{1}{2}t_j^2) = \exp(-\frac{1}{2}\sum_{j=1}^{d} t_j^2) = \exp(-\frac{1}{2}t^T t)$$

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

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

$$= e^{it^T\mu}\exp(-\frac{1}{2}t^T LL^T t) = \exp(it^T\mu - \frac{1}{2}t^T\Sigma t)$$

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

Solution.

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

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

2. It is

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

[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

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

So:

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

2. It is

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

3. It is

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

Exercise 20. ($\star\star$)Prove the following statement:
Let $x \sim \text{N}_d(\mu, \Sigma)$, $x \in \mathbb{R}^d$, 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 $z \sim \text{N}_d(0, I)$. Because $z_i \sim \text{N}(0,1)$, 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

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

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

Solution. It is

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

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.

Part IV

Bayesian paradigm and calculations

Exercise 24. ($\star$)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

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

1. By using the Bayes theorem

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

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

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

Exercise 25. ($\star\star\star$)(Nuisance parameters are involved) <-story

Assume observable quantities $y=(y_1,...,y_n)$ 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 $\mathrm{N}(\mu,\sigma^2)$ 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

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

1. Show that

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

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

$$\mu|y,\sigma^2 \sim \mathrm{N}(\mu_n,\sigma^2\frac{1}{\tau_n})$$
$$\sigma^2|y \sim \mathrm{IG}(a_n,k_n)$$

with

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

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

Hint: It is

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

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

$$\mu|y \sim \mathrm{T}_1\left(\mu_n,\frac{k_n}{a_n}\frac{1}{\tau_n},2a_n\right)$$

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

$$z|y \sim \mathrm{T}_1\left(\mu_n,\frac{k_n}{a_n}(\frac{1}{\tau_n}+1),2a_n\right)$$

Hint-1: Consider that

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

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

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

2. I use the Bayes theorem

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

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

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

So

$$\pi(\mu,\sigma^2|y) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}+\frac{n}{2}+a_0+1} \exp\left(\frac{1}{\sigma^2}\left[-\frac{1}{2}\frac{(\mu-\mu_n)^2}{1/\tau_n}+C_n\right]-\frac{1}{\sigma^2}k_0\right)$$

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

$$\propto \mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{IG}(\sigma^2|a_n,k_n)$$

where

$$\mu_n = \frac{n\bar{y}+\tau_0\mu_0}{n+\tau_0}\,; \qquad a_n = \frac{n}{2}+a_0\,;$$

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

3. It is

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

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

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

$$= \mathrm{T}_1(\mu|\mu_n,\frac{k_n}{a_n}\frac{1}{\tau_n},2a_n)$$

4. It is

$$g(z|y) = \int f(z|\mu,\sigma^2)\pi(\mu,\sigma^2|y)\mathrm{d}\mu\mathrm{d}\sigma^2 = \int \mathrm{N}(z|\mu,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\mu\mathrm{d}\sigma^2$$

$$= \int \left[\int \mathrm{N}(z|\mu,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{d}\mu\right]\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\sigma^2$$

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

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

So

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

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

$$g(z|y) = \int \mathrm{N}\left(z|\mu_n,\xi\left[\frac{1}{\tau_n}+1\right]\frac{k_n}{a_n}2a_n\right)\mathrm{IG}(\xi|\frac{2a_n}{2},\frac{1}{2})\mathrm{d}\xi = \mathrm{T}_1\left(z|\mu_n,\left[\frac{1}{\tau_n}+1\right]\frac{k_n}{a_n},2a_n\right)$$

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

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

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
- $x = (\text{longitude}, \text{latitude})$: 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

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

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

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

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

where $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$ 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

$$\begin{aligned} y|\beta, \sigma^2 &\sim \mathrm{N}(\Phi\beta, I\sigma^2); \text{the sampling distr} \\ \beta|\sigma^2 &\sim \mathrm{N}(\mu_0, V_0\sigma^2); \text{prior distr} \\ \sigma^2 &\sim \mathrm{IG}(a_0, k_0) \text{prior distr} \end{aligned}$$

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

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

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

with

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

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

Hint-1:

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

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

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

where

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

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

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

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

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

Consider that

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

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

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

but

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

so

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

but

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

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

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

So

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

So

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

2. It is

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

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

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

3. It is

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

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

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

because Normal is symmetric around the mean

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

by using the Hint

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

So

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

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

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

So

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

, or equiv.

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

Exercise 27. ($\star\star$)Let $y=(y_1,...,y_n)$ be observables drawn iid from sampling distribution $y_i|\theta\overset{\text{iid}}{\sim}\mathrm{N}(\theta,\theta^2)$ for all $i=1,...,n$, 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

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

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

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

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

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

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. ($\star\star$)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

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

Exercise 29. ($\star\star$)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

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

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

$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

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

Let

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

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

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

Let

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

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Note: A useful test function is

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

Exercise 30. ($\star\star$)Consider the Bayesian model

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

where $\mathrm{Ex}(\lambda)$ 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

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

$$f(x=0) \propto \int_0^\infty \frac{1}{\sigma}\exp(-\sigma\lambda)\mathrm{d}\sigma$$

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

I will use the Limit Comparison Test to check if $\int_0^\infty \frac{1}{\sigma}\exp(-\sigma\lambda)\mathrm{d}\sigma = \infty$. 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

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

and

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

Therefore, it will be

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

as well.

Exercise 31. ($\star\star$)Consider the Bayesian model

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

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

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

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

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

So the posterior is defined.

Exercise 32. ($\star\star$)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. ($\star\star$)Assume that observable quantities $x_1, x_2, ...$ are generated i.i.d by a process that can be modeled as a sampling distribution $\text{N}(\mu, \sigma^2)$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 $\mu \sim \text{N}(\mu_0, \sigma_0^2)$ where μ_0, σ_0^2 are known.

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

    Next assume that you additionally aobserve 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 $\mu \sim \text{N}(\mu_0, \sigma_0^2)$ 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

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

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

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

Solution.

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

$$\pi(\mu|x_1) \propto \overbrace{\mathrm{N}(x_1|\mu,\sigma^2)}^{\text{likelihood}}\overbrace{\mathrm{N}(\mu|\mu_0,\sigma_0^2)}^{\text{prior}}$$

$$\propto \exp(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2})\exp(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2}-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(\mu-\hat{\mu}_1)^2}{\hat{\sigma}_1^2}) \propto \mathrm{N}(\mu|\hat{\mu}_1,\hat{\sigma}_1^2) \tag{3}$$

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, $\mu|x_1 \sim \mathrm{N}(\hat{\mu}_1,\hat{\sigma}_1^2)$

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

$$\pi(\mu|x_1,x_2) \propto \overbrace{(x_2|\mu,\sigma^2)}^{\text{likelihood}}\overbrace{\mathrm{N}(\mu|\hat{\mu}_1,\hat{\sigma}_1^2)}^{\text{prior}}$$

$$\propto \exp(-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2})\exp(-\frac{1}{2}\frac{(\mu-\hat{\mu}_1)^2}{\hat{\sigma}_1^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2}-\frac{1}{2}\frac{(\mu-\hat{\mu}_1)^2}{\hat{\sigma}_1^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(\mu-\hat{\mu}_2)^2}{\hat{\sigma}_2^2}) \propto \mathrm{N}(\mu|\hat{\mu}_2,\hat{\sigma}_2^2) \tag{4}$$

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

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

$$\pi(\mu|x_1,x_2) \propto \overbrace{\mathrm{N}(x_1|\mu,\sigma^2)\mathrm{N}(x_2|\mu,\sigma^2)}^{\text{likelihood}}\overbrace{\mathrm{N}(\mu|\mu_0,\sigma_0^2)}^{\text{prior}}$$

$$\propto \exp(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2})\exp(-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2})\exp(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2}-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2}-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2})$$

$$\propto \exp(-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\hat{\sigma}^2}) \propto \mathrm{N}(\mu|\hat{\mu},\hat{\sigma}^2) \tag{5}$$

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, $\mu|x_1,x_2 \sim \mathrm{N}(\hat{\mu},\hat{\sigma}^2)$

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 $\mathrm{d}\Pi(\mu|x_1,x_2)$ at the end posterior $\mathrm{d}\Pi(\mu|x_1,x_2)$ in a single application of Bayes theorem with the full data $x=(x_1,x_2)$.

Part V

Exchangeability

We work on the proofs of the following theorems:

- Marginal distributions of finite exchangeable sequences $y_1, y_2, \dots, y_k$ are invariant under permutations; i.e.:
$$\mathrm{d}F(y_{\mathfrak{p}(1)}, y_{\mathfrak{p}(2)}, \dots, y_{\mathfrak{p}(k)}) = \mathrm{d}F(y_1, y_2, \dots, y_k) \text{ for all } \mathfrak{p} \in \mathfrak{P}_n. \tag{6}$$
In particular, for $k = 1$, it follows that all y_i are identically distributed (but not necessarily independently, as stated in the Lecture notes)
- (Marginal) Expectations of finite exchangeable sequences $y_1, y_2, \dots, y_k$ are all identical:
$$\mathrm{E}(g(y_i)) = \mathrm{E}(g(y_1)) \text{ for all } i = 1, ..., k \text{ and all functions } g\colon \mathcal{Y} \to \mathbb{R} \tag{7}$$
- (Marginal) Variances of finite exchangeable sequences $y_1, y_2, \dots, y_k$ are all identical:
$$\mathrm{Var}(y_i) = \mathrm{Var}(y_1). \tag{8}$$
- Covariances between elements of finite exchangeable sequences $y_1, y_2, \dots, y_k$ are all identical:
$$\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(y_1, y_2) \text{ whenever } i \neq j. \tag{9}$$

Just for your information The properties above are implied by the following general theorem. However, you should not use this theorem, directly, to solve the exercises below...

Theorem. *Consider an exchangeable sequence $y_1, \dots, y_n$. Let $g\colon \mathcal{Y}^k \to \mathbb{R}$ be any function of k of these, where $k \leq n$. Then, for any permutation $\pi \in \Pi_n$,*
$$E(g(Y_{\mathfrak{p}(1)}, Y_{\mathfrak{p}(2)}, \dots, Y_{\mathfrak{p}(k)})) = E(g(Y_1, Y_2, \dots, Y_k)) \tag{10}$$

This is not an exercise to solve. Feel free to read the solution of this exercise, as it may help you understand the the Interpretation of the 'representation Theorem with $0-1$ quantities'.

Exercise 34. ($\star\star\star\star\star$)(Representation Theorem with $0-1$ quantities). If $y_1, y_2, ...$ is an infinitely exchangeable sequence of $0-1$ random quantities with probability measure P, there exists a distribution function Π such that the joint mass function $p(y_1, ..., y_n)$ for $y_1, ..., y_n$ has the form
$$p(x_1, ..., x_n) = \int_0^1 \prod_{i=1}^n \underbrace{\theta^{y_i}(1-\theta)^{1-y_i}}_{f_{\mathrm{Br}(\theta)}(y_i|\theta)} \mathrm{d}\Pi(\theta)$$
where
$$\Pi(t) = \lim_{n\to\infty} \Pr(\frac{1}{n}\sum_{i=1}^n y_i \leq t) \quad \text{and} \quad \theta \overset{\text{as}}{=} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n y_i$$
aka θ is the limiting relative frequency of 1s, by SLLN

Hint: (Helly's theorem [modified]) Given a sequence of distribution functions $\{F_1, F_2, ...\}$ that satisfy the tightness condition; [for each $\epsilon > 0$ there is a such that for all sufficiency large i it is $F_i(a) - F_i(-a) > 1 - \epsilon$], there exists a distribution F and a sub-sequence $\{F_{i_1}, F_{i_2}, ...\}$ such that $F_{i_j} \rightarrow F$.

Solution. Let the sum of random quantities be $S_n = \sum_{i=1}^{n} y_i$, and assume that the sum S_n is equal to value s_n; i.e. $S_n = t_n$. By exchangeability, for $0 \leq t_n < n$, it is

$$p(S_n = t_n) = \binom{n}{t_n} p(y_{\mathfrak{p}(1)}, ..., y_{\mathfrak{p}(n)})$$

for any permutation operator $\mathfrak{p}$. For finite N, let $N \geq n \geq t_n \geq 0$,

$$\begin{aligned}
p(S_n = t_n) &= \sum_{t_N=0}^{N} p(S_n = t_n | S_N = t_N) p(S_N = t_N) \\
&= \underbrace{\sum_{t_N=0}^{t_n-1} p(S_n = t_n | S_N = t_N) p(S_N = t_N)}_{=0} \qquad (11) \\
&\qquad + \sum_{y_N=y_n}^{N-(n-y_n)} p(S_n = t_n | S_N = t_N) p(S_N = t_N) \\
&\qquad\quad + \underbrace{\sum_{t_N=N-(n-t_n)+1}^{N} p(S_n = t_n | S_N = t_N) p(S_N = t_N)}_{=0} \qquad (12) \\
&= \sum_{y_N=y_n}^{N-(n-y_n)} p(S_n = t_n | S_N = t_N) p(S_N = t_N)
\end{aligned}$$

The terms in (11, 12) are zero because $p(S_n = t_n | S_N = t_N) = 0$ for $t_N < t_n$ and $t_N > N - (n - t_n)$ because we contrition on $S_N = t_N$.

We work out on $p(S_n = t_n | S_N = t_N)$ which is the conditional probability for S_n given $S_N = t_N$. We observe that the random variable $S_n | S_N = t_N$ follows a Hypergeometric distribution $S_n | S_N = t_N \sim \text{Hy}(t_N, N - t_N, n)$. This is because it describes a Hypergeometric experiment[2]. i.e., $S_n = t_n$ is the number of successes (random draws for which the object drawn has a specified feature) in n random draws without replacement, from a finite population of size N that contains exactly $S_N = t_N$ objects of that feature, wherein each draw is either a success or a failure (aka $x_i = 0$ or 1). Hence, $p(S_n = t_n | S_N = t_N)$ is a Hypergeometric PMF, namely

$$p(S_n = t_n | S_N = t_N) = \text{Hy}(S_n = t_n | t_N, N - t_N, n) = \frac{\binom{t_N}{t_n}\binom{N-t_N}{n-t_n}}{\binom{N}{n}}, \ 0 \leq t_n \leq n$$

Rewriting the binomial coefficients by rearranging the terms in the product, we get

$$\begin{aligned}
p(S_n = t_n) &= \sum \binom{N}{n}^{-1} \binom{t_N}{t_n} \binom{N - t_N}{n - t_n} p(S_N = t_N) \\
&= \binom{n}{t_n} \sum \frac{(t_N)_{t_n} (N - t_N)_{n-t_n}}{(N)_n} p(S_N = t_N)
\end{aligned}$$

where $(y)_r = y(y-1)...(y-r+1)$.

[2] `https://en.wikipedia.org/wiki/Hypergeometric_distribution`

Now, define a function $\Pi_N(\theta)$ on $\mathbb{R}$ as the step function which is zero for $\theta < 0$, and has steps of size $p(S_N = t_N)$ at $\theta = t_N/N$ for $t_N = 0, 1, 2, ..., N$. Then, by changing variable we get,

$$p(S_n = t_n) = \binom{n}{t_n} \int_0^1 \frac{(\theta N)((1-\theta)N)_{n-t_n}}{(N)_n} \mathrm{d}\Pi_N(\theta).$$

This result holds for any finite N. Now we need to consider $N \to \infty$. In the limit, we get

$$\lim_{N\to\infty} \frac{(\theta N)((1-\theta)N)_{n-t_n}}{(N)_n} = \theta^{t_n}(1-\theta)^{n-t_n} = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} \tag{13}$$

Note that function $\Pi_N(t)$ is a step function, starting at zero and ending at one with N steps of varying sizes at particular values of t. By Helly's theorem, there exists a subsequence $\{\Pi_{N_1}, \Pi_{N_2}, ...\}$ such that

$$\lim_{N_j\to\infty} \Pi_{N_j} = \Pi$$

where Π is a distribution function.

Exercise 35. ($\star\star$)Clearly a set of independent and identically distributed random variables form an exchangeable sequence. Thus sampling with replacement generates an exchangeable sequence. What about sampling without replacement? Prove that sampling n items from N distinct objects without replacement (where $n \leq N$) is exchangeable.

Solution. Sampling without replacement is clearly not iid. However, it is exchangeable. Assume that we sample n items from N distinct objects without replacement, we have that:

$$f(y_1, ..., y_n) = \frac{1}{N^{\underline{n}}} = \frac{(N-n)!}{N!} \tag{14}$$

Clearly, the probability mass function does not depend on the ordering of the sequence. Therefore the sequence is exchangeable.

Exercise 36. ($\star\star$)Let $Y_1, \ldots, Y_n$ be an exchangeable sequence, and let g be any function on $\mathcal{Y}$. Show, directly from the definition of exchangeability in the summary notes) that $\mathrm{E}(g(Y_i))$ does not depend on i:

$$\mathrm{E}(g(Y_i)) = \mathrm{E}(g(Y_1)) \text{ for all } i \in \{2, \ldots, n\} \tag{15}$$

For ease of exposition, you may restrict your proof to the case $i = 2$.

Solution. For ease of exposition, we show that $\mathrm{E}(g(Y_1)) = \mathrm{E}(g(Y_2))$. The general case follows similarly.

$$\mathrm{E}(g(Y_1)) = \sum_{(y_1,y_2,y_3,\ldots,y_n)\in\mathcal{Y}^n} g(y_1) f(y_1, y_2, y_3, \ldots, y_n) \tag{16}$$

and by exchangeability, we can swap the indices 1 and 2 in the probability mass function, so

$$= \sum_{(y_1,y_2,y_3,\ldots,y_n)\in\mathcal{Y}^n} g(y_1) f(y_2, y_1, y_3, \ldots, y_n) \tag{17}$$

and swapping y_1 and y_2 (we can always do this, exchangeability is not used here),

$$= \sum_{(y_2,y_1,y_3,\ldots,y_n)\in\mathcal{Y}^n} g(y_2) f(y_1, y_2, y_3, \ldots, y_n) = \mathrm{E}(g(Y_2)) \tag{18}$$

Exercise 37. ($\star\star$)Let $Y_1, \ldots, Y_n$ be an exchangeable sequence. Use

$$\mathrm{E}(g(Y_i)) = \mathrm{E}(g(Y_1)) \text{ for all } i \in \{2, \ldots, n\} \tag{19}$$

to show that $\mathrm{Var}(Y_i)$ does not depend on i:

$$\mathrm{Var}(Y_i) = \mathrm{Var}(Y_1) \text{ for all } i \in \{2, \ldots, n\} \tag{20}$$

Solution. By the usual properties of variance,

$$\mathrm{Var}(Y_i) = \mathrm{E}(Y_i^2) - \mathrm{E}(Y_i)^2 \tag{21}$$

and now applying19 twice

$$\mathrm{Var}(Y_i) = \mathrm{E}(Y_1^2) - \mathrm{E}(Y_1)^2 = \mathrm{Var}(Y_1)$$

Exercise 38. ($\star\star$)Let $Y_1, \ldots, Y_n$ be an exchangeable sequence. By expanding $var(\sum_{k=1}^{n} Y_k)$, show that when $i \neq j$,

$$cov(Y_i, Y_j) \geq -\frac{var(Y_1)}{n-1} \tag{22}$$

Solution. It is

$$0 \leq var\left(\sum_{k=1}^{n} Y_k\right) = \sum_{k=1}^{n} var(Y_k) + 2\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n} cov\left(Y_k, Y_\ell\right) \tag{23}$$

and because, by exchangeability, var $(Y_k) = var\ (Y_1)$ and cov $(Y_k, Y_\ell) =$ cov (Y_i, Y_j) for all $k \neq \ell$,

$$= n \,\mathrm{var}\,(Y_1) + (n^2 - n)\mathrm{cov}\,(Y_i, Y_j) \tag{24}$$

where the $n^2 - n$ factor can be derived as follows: note that the pairs of indices (k, ℓ) appearing in the sum can be put into a matrix—the sum does not include the diagonal of this matrix (n pairs), but otherwise covers precisely half of it, and the full matrix has n^2 pairs, so there are $(n^2 - n)/2$ terms in the sum.

Consequently,

$$\mathrm{Cov}\,(Y_i, Y_j) \geq -\frac{n\mathrm{var}\,(Y_1)}{n^2 - n} = -\frac{\mathrm{var}\,(Y_1)}{n-1} \tag{25}$$

Exercise 39. ($\star$)What does

$$cov(Y_i, Y_j) \geq -\frac{var(Y_1)}{n-1}$$

imply about the correlation of infinite exchangeable sequences?

Solution. The correlation must be non-negative: because, as $n \to \infty$, $cov(Y_i, Y_j) \geq 0$ for all $i \neq j$.
