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

Exercise Sheet: Bayesian Statistics

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

Part I

Matrix & vector calculus

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

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

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

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.

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.

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

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

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

Hint: Start by noticing that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1. $\text{Cov}(z, y) = \text{E}(zy^\top) - \text{E}(z)\,(\text{E}(y))^\top$
2. $\text{Cov}(z, y) = (\text{Cov}(y, z))^\top$
3. $\text{Cov}_\pi(c_1 + A_1 z, c_2 + A_2 y) = A_1 \text{Cov}_\pi(x, y) A_2^\top$, 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 $\text{Cov}(z, y) = 0$

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:

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

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

1. $\text{Var}(y) = \text{E}(yy^\top) - \text{E}(y)\,(\text{E}(y))^\top$
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)

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

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

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

Exercise 14. ($\star$)

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

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

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

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

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

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

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

1. if $Z \sim \text{N}(0, I)$ 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 $Z \in \mathbb{R}$

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

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

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 $y \sim \text{IG}(a, cb)$.

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 $\text{Var}_{\xi \sim \text{IG}(a,b)}(\xi) = \frac{b^2}{(a-1)^2(a-2)}$

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

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

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

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

with PDF

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

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

Show that the marginal PDF of x is

$$f(x) = \int f_{\mathrm{N}_d(\mu,\Sigma\xi)}(x|\xi) f_{\mathrm{IG}(a,b)}(\xi)\mathrm{d}\xi$$

$$= \frac{2a^{-\frac{d}{2}}}{\pi^{\frac{n}{2}}\sqrt{\det(\frac{b}{a}\Sigma)}} \frac{\Gamma\left(a+\frac{d}{2}\right)}{\Gamma(a)} \left[1 + \frac{1}{2a}(x-\mu)^\top \left(\frac{b}{a}\Sigma\right)^{-1}(x-\mu)\right]^{-\frac{(2a+d)}{2}} \tag{2}$$

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

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

---

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

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

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 $z = y_{n+1}$.

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

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

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

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

Exercise 26. ($\star\star\star$)(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

Exercise 27. ($\star\star$)Let $y = (y_1, ..., y_n)$ be observables drawn iid from sampling distribution $y_i|\theta \stackrel{\text{iid}}{\sim} \text{N}(\theta, \theta^2)$ 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.

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.

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.

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

- 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

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

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 $x = 0$.

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.

Exercise 32. ($\star\star$)Let x be an observation. Consider the Bayesian model

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

where $\mathrm{Pn}(\theta)$ 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: $x \sim \mathrm{Pn}(\theta)$ has PMF

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

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 $\mathrm{N}(\mu, \sigma^2)$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 $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$ 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.
