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

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

    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 $\mathcal{Y}$.

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

    1. $\mathrm{Cov}(z, y) = \mathrm{E}(zy^\top) - \mathrm{E}(z)\left(\mathrm{E}(y)\right)^\top$

    2. $\mathrm{Cov}(z, y) = \left(\mathrm{Cov}(y, z)\right)^\top$

    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 $\text{Cov}(z, y) = 0$

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

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)

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. $\varphi_{A+Bx}(t) = e^{it^T A} \varphi_x(B^T t)$ 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 $M_x(t) = \varphi_x(-it)$

Created on 2021/11/03 at 22:24:10 by Georgios Karagiannis

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

$$y \sim \chi_d^2$$

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

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

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

Created on 2021/11/03 at 22:24:10 by Georgios Karagiannis

Show that the marginal PDF of x is

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

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

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

<center>The Following exercise is part of Homework 1</center>

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

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

$$x|\xi \sim N_d(\mu, \Sigma\xi v)$$

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

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

$$x_1 \sim T_{d_1}(\mu_1, \Sigma_1, \nu)$$

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

2. Show that

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

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

   Hint: The PDF of $y \sim N_d(\mu, \Sigma)$ 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 $y \sim IG(a, b)$ is

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

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

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

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

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$\dot{\Sigma}_{2|1} = \frac{\nu + (\mu_1 - x_1)^\top \Sigma_1^{-1}(\mu_1 - x_1)}{\nu + d_1}\Sigma_{2|1}$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top$$

$$\nu_{2|1} = \nu + d_1$$

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 $y_1, \ldots, y_n$ 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 $z = y_{n+1}$.

The following exercise is about Bayesian treatment in problems with nuisance unknown parameters

          Created on 2021/11/03 at 22:24:10   by Georgios Karagiannis

- Assume observable quantities $y = (y_1, ..., y_n)$. Assume that the sampling distribution is $F(y|\theta)$ labeled by an unknown parameter $\theta \in \Theta$. Let $\theta = (\phi, \lambda)^\top$ with $\phi \in \Phi$ and $\lambda \in \Lambda$. Assume You are interested in learning parameter $\phi \in \Phi$, and You are not interested in learning the unknown parameter $\lambda \in \Lambda$; but both ϕ, λ are parts of the statistical model parameterisation. The unknown quantity $\lambda \in \Lambda$ is called <u>nuisance parameter</u>. We an call $\phi \in \Phi$ <u>parameter of interest</u>.

- In Bayesian Stats, learning (or quantifying uncertainty about) parameter of interest ϕ under the presence of a nuisance parameter $\lambda \in \Lambda$ is performed according to the Bayesian paradigm as usual: You specify a prior $\Pi(\phi, \lambda)$ with PDF/PMF $\pi(\phi, \lambda) = \pi(\phi|\lambda)\pi(\lambda)$ on the joint space of ALL Your unknown parameters $\theta = (\phi, \lambda)^\top$; you compute the joint posterior distribution $\mathrm{d}\Pi(\theta|y)$ of $\theta = (\phi, \lambda)^\top$ via the Bayesian theorem. Reasonably, Your posterior degree of believe about the parameter of interest ϕ given the data $y = (1_1, ..., y_n)$ is given through the marginal posterior distribution $\mathrm{d}\Pi(\phi|y)$.

- To summarize; Specify the Bayesian model as:

$$
\begin{cases}
y|\overbrace{\phi, \lambda}^{=\theta} \sim \mathrm{d}F(y|\overbrace{\phi, \lambda}^{=\theta}) & \text{, the statistical model} \\
\underbrace{(\phi, \lambda)}_{=\theta} \sim \mathrm{d}\Pi(\underbrace{\phi, \lambda}_{=\theta}) & \text{, the prior model}
\end{cases}
$$

The joint posterior of θ given y is $\mathrm{d}\Pi(\theta|y) = \mathrm{d}\Pi(\lambda|y, \phi)\mathrm{d}\Pi(\phi|y)$ is with PDF/PMF

-

$$
\pi(\overbrace{\phi, \lambda}^{=\theta}|y) = \frac{f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})}{f(y)} = \underbrace{\frac{f(y|\phi, \lambda)\pi(\lambda|\phi)}{f(y|\phi)}}_{=\pi(\lambda|y, \phi)} \underbrace{\frac{f(y|\phi)\pi(\phi)}{f(y)}}_{=\pi(\phi|y)} = \pi(\lambda|y, \phi)\pi(\phi|y)
$$

The (marginal) likelihood $f(y|\phi)$ of y given ϕ is

$$
f(y|\phi) = \underbrace{\int_\Lambda f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\lambda|\phi)}_{=\mathrm{E}_{\Pi(\lambda|\phi)}(f(y|\phi, \lambda)|\phi)} = \begin{cases} \int_\Lambda f(y|\phi, \lambda)\pi(\lambda|\phi)\mathrm{d}\lambda & \text{, if } \lambda \text{ cont} \\ \\ \sum_{\forall \lambda \in \Lambda} f(y|\phi, \lambda)\pi(\lambda|\phi) & \text{, if } \lambda \text{ discr} \end{cases}
$$

The PDF/PMF $\pi(\phi|y)$ of marginal posterior $\mathrm{d}\Pi(\phi|y)$ of ϕ is

$$
\pi(\phi|y) = \underbrace{\int_\Lambda \pi(\overbrace{\phi, \lambda}^{=\theta}|y)\mathrm{d}\lambda}_{=\mathrm{E}_{\Pi(\lambda|y)}(\pi(\phi|y, \lambda))} \qquad \text{or equivalently} \qquad \pi(\phi|y) = \frac{f(y|\phi)\pi(\phi)}{f(y)}
$$

The predictive distribution $G(z|y)$ of the next outcome $z = (y_{n+1}, ...y_{n+m})$ given y has pdf/pmf

$$
g(z|y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\overbrace{\phi, \lambda}^{=\theta}|y)
$$

and the marginal likelihood $f(y)$ is

$$
f(y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\phi\mathrm{d}\lambda
$$

Exercise 25. ($\star\star\star$)Assume observable quantities $y = (y_1, ..., y_n)$ forming the available data set of size n. Assume <-story that the observations are drawn i.i.d. from a sampling distribution which is judged to be in the Normal parametric

family of distributions $N(\mu, \sigma^2)$ with unknown mean μ and variance σ^2. We are interested in learning μ and the next outcome $z = y_{n+1}$. We do not care about σ^2.

Assume You specify a Bayesian model   <-set-up

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

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

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

$$\sigma^2|y \sim IG(a_n, k_n)$$

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

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

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

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

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

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

Hint-1: Consider that

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

where

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

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

The following is about the Normal linear model of regression.

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

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

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

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

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 $y_i|\theta \overset{\mathrm{iid}}{\sim} \mathrm{N}(\theta, \theta^2)$ 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.

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
$$0 \le f(x) \le g(x), \quad \text{for } x \ge a$$

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

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.

The Following exercise is part of Homework 1

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

Created on 2021/11/03 at 22:24:10   by Georgios Karagiannis

Part V

Exchangeability

We work on the proofs of the following theorems:

- Marginal distributions of finite exchangeable sequences $y_1, y_2, \ldots, y_k$ are invariant under permutations; i.e.:

$$\mathrm{d}F(y_{\mathfrak{p}(1)}, y_{\mathfrak{p}(2)}, \ldots, y_{\mathfrak{p}(k)}) = \mathrm{d}F(y_1, y_2, \ldots, y_k) \text{ for all } \mathfrak{p} \in \mathfrak{P}_n. \tag{3}$$

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 $y_1, y_2, \ldots, y_k$ are all identical:

$$\mathrm{E}(g(y_i)) = \mathrm{E}(g(y_1)) \text{ for all } i = 1, ..., k \text{ and all functions } g \colon \mathcal{Y} \to \mathbb{R} \tag{4}$$

- (Marginal) Variances of finite exchangeable sequences $y_1, y_2, \ldots, y_k$ are all identical:

$$\mathrm{Var}(y_i) = \mathrm{Var}(y_1). \tag{5}$$

- Covariances between elements of finite exchangeable sequences $y_1, y_2, \ldots, y_k$ are all identical:

$$\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(y_1, y_2) \text{ whenever } i \neq j. \tag{6}$$

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 $y_1, \ldots, y_n$. Let $g \colon \mathcal{Y}^k \to \mathbb{R}$ be any function of k of these, where $k \leq n$. Then, for any permutation $\pi \in \Pi_n$,*

$$E(g(Y_{\mathfrak{p}(1)}, Y_{\mathfrak{p}(2)}, \ldots, Y_{\mathfrak{p}(k)})) = E(g(Y_1, Y_2, \ldots, Y_k)) \tag{7}$$

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 $y_1, y_2, \ldots$ 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 $p(y_1, \ldots, y_n)$ for $y_1, \ldots, y_n$ has the form

$$p(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^n \underbrace{\theta^{y_i}(1-\theta)^{1-y_i}}_{f_{\mathrm{Br}(\theta)}(y_i|\theta)} \mathrm{d}\Pi(\theta)$$

where

$$\Pi(t) = \lim_{n \to \infty} \mathrm{Pr}(\frac{1}{n}\sum_{i=1}^n y_i \leq t) \quad \text{and} \quad \theta \stackrel{\mathrm{as}}{=} \lim_{n \to \infty} \frac{1}{n}\sum_{i=1}^n y_i$$

aka θ is the limiting relative frequency of 1s, by SLLN

Hint: (Helly's theorem [modified]) Given a sequence of distribution functions $\{F_1, F_2, \ldots\}$ 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 $\{F_{i_1}, F_{i_2}, \ldots\}$ such that $F_{i_j} \to F$.

Exercise 35. (⋆⋆)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.

Exercise 36. (⋆⋆)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:

$$\mathrm{E}(g(Y_i)) = \mathrm{E}(g(Y_1)) \text{ for all } i \in \{2, \ldots, n\} \tag{8}$$

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

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

$$\mathrm{E}(g(Y_i)) = \mathrm{E}(g(Y_1)) \text{ for all } i \in \{2, \ldots, n\} \tag{9}$$

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

$$\mathrm{Var}(Y_i) = \mathrm{Var}(Y_1) \text{ for all } i \in \{2, \ldots, n\} \tag{10}$$

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

$$cov(Y_i, Y_j) \geq -\frac{var(Y_1)}{n - 1} \tag{11}$$

Exercise 39. (⋆)What does

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

imply about the correlation of infinite exchangeable sequences?

Created on 2021/11/03 at 22:24:10 by Georgios Karagiannis

Part VI

Sufficiency

Exercise 40. (⋆⋆)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{iid}}{\sim} \text{Ex}(\theta), \ \forall i = 1, ..., n \\ \theta & \sim \text{Ga}(a, b) \end{cases}$$

Hint-1: The PDF of $x \sim \text{G}(a, b)$ is $\text{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0, +\infty)}(x)$

Hint-2: The PDF of $x \sim \text{Ex}(\theta)$ is $\text{Ex}(x|\theta) = \text{Ga}(x|1, \theta)$

1. Show that the parametric model is member of the Exponential family, and the sufficient statistic for a sample of observables $x = (x_1, ..., x_n)$.

2. Show that the posterior distribution θ given x is Gamma and compute its parameters.

3. Show that the predictive distribution $G(z|x)$ of a future z given $x = (x_1, ..., x_n)$, has PDF

$$g(z|x) = \frac{a^*(b^*)^{a^*}}{(z + b^*)^{a^*+1}} 1(x \geq 0)$$

Exercise 41. (⋆⋆⋆)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \text{Di}_k(a) \end{cases}$$

where $\theta \in \Theta$, with $\Theta = \{\theta \in (0,1)^k | \sum_{j=1}^k \theta_j = 1\}$ and $\mathcal{X}_k = \{x \in \{0, ..., n\}^k | \sum_{j=1}^k x_j = 1\}$.

Hint-1: Mu_k denotes the Multinomial probability distribution with PMF

$$\text{Mu}_k(x|\theta) = \begin{cases} \prod_{j=1}^k \theta_j^{x_j} & \text{, if } x \in \mathcal{X}_k \\ 0 & \text{, otherwise} \end{cases} \tag{12}$$

Hint-2: $\text{Di}_k(a)$ denotes the Dirichlet distribution with PDF

$$\text{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k \theta_j^{a_j - 1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

1. Show that the parametric model (12) is a member of the $k - 1$ exponential family.

2. Compute the likelihood $f(x_{1:n}|\theta)$, and find the sufficient statistic $t_n := t_n(x_{1:n})$.

3. Compute the posterior distribution. State the name of the distribution, and expresses its parameters with respect to the observations and the hyper-parameters of the prior. Justify your answer.

4. Compute the probability mass function of the predictive distribution for a future observation $y = x_{n+1}$ in closed form.

   Hint $\Gamma(x) = (x - 1)\Gamma(x - 1)$.

Exercise 42. ($\star\star$)Suppose that the vector $\mathbf{x} = (x, y, z)$ has a trinomial distribution depending on the index n and the parameter $\varpi = (\pi, \rho, \sigma)$ where $\pi + \rho + \sigma = 1$, that is

$$p(\mathbf{x}|\varpi) = \frac{n!}{x!\,y!\,z!}\pi^x \rho^y \sigma^z \qquad (x + y + z = n).$$

Show that this distribution is in the two-parameter exponential family.

Exercise 43. ($\star$)Establish the formula

$$(n_0^{-1} + n^{-1})^{-1}(\bar{x} - \theta_0)^2 = n\bar{x}^2 + n_0\theta_0^2 - n_1\theta_1^2$$

where $n_1 = n_0 + n$ and $\theta_1 = (n_0\theta_0 + n\bar{x})/n_1$.

This formula is often used to 'complete the square' in quadratiforms.

The following is a proof of a theorem

Exercise 44. ($\star\star\star$)Let $y_1, y_2...$ be an infinitely exchangeable sequence of random quantities. Let $t = t(y_1, ..., y_n)$ be a statistic for a finite $n \geq 1$. Then t is predictive sufficient if, and only if, it is parametric sufficient.

The following is a proof of a theorem

Exercise 45. ($\star\star\star$)(This is a theorem in Handout 5) Prove the following statement.

Let $t : \mathcal{Y} \to \mathcal{T}$ be a statistic. Then t is a parametric sufficient statistic for θ in the Bayesian sense if and only if the likelihood function $L(\cdot|\cdot)$ on $\mathcal{Y} \times \Theta$ can be factorized as the product of a kernel function k on $\mathcal{Y} \times \Theta$ and a residue function ρ on Θ as

$$L(y; \theta) = k(t(y)|\theta)\rho(y). \tag{13}$$

Created on 2021/11/03 at 22:24:10 by Georgios Karagiannis

Part VII

Priors

Exercise 46. ($\star\star$)Let $y = (y_1, ..., y_n)$ be observable quantities, generated from an exponential family of distributions as

$$y_i|\theta \overset{\text{iid}}{\sim} \text{Ef}(u, g, h, c, \phi, \theta, c), \ i = 1, ..., n$$

with density

$$\text{Ef}(y_i|u, g, h, c, \phi, \theta, c) = u(y_i)g(\theta)^n \exp(\sum_{j=1}^{k} c_j \phi_j(\theta)h_j(y_i))$$

and assume a conjugate prior $\Pi(\theta)$ with pdf/pmf

$$\pi(\theta) = \tilde{\pi}(\theta|\tau) \propto g(\theta)^{\tau_0} \exp(\sum_{j=1}^{k} c_j \phi_j(\theta)\tau_j)$$

1. Show that the posterior $\Pi(\theta|y)$ of θ has pdf/pmf $\pi(\theta|y) = \tilde{\pi}(\theta|\tau^*)$ with $\tau^* = (\tau_0^*, \tau_1^*, ..., \tau_k^*)$, $\tau_0^* = \tau_0 + n$, and $\tau_j^* = \sum_{i=1}^{n} h_j(x_i) + \tau_j$ for $j = 1, ..., k$, and pdf/pmf

$$\pi(\theta|y) = \pi(\theta|\tau^*) \propto g(\theta)^{\tau^*} \exp(\sum_{j=1}^{k} c_j \phi_j(\theta)\tau_j^*) \tag{14}$$

The operation $*$ here is addition $\tau * t(y) \longmapsto \tau + t(y) = \tau^*$

2. Show that the predictive distribution $G(z|y)$ for a new outcome $z = (y_{n+1}, ..., y_{n+m})$ has pdf/ pmf

$$g(z|y) = \prod_{i=1}^{m} u(z_i) \frac{K(\tau + t(y) + t(z))}{K(\tau + t_n(y))} \tag{15}$$

where $t(z) = (m, \sum_{i=1}^{m} h_1(z_i), ..., \sum_{i=1}^{m} h_k(z_i))$.

Exercise 47. ($\star\star$)

1. Show that the skew-logistic family of distributions, with

$$f(x|\theta) = \frac{\theta e^{-x}}{(1 + e^{-x})^{\theta+1}} \tag{16}$$

for $x \in \mathbb{R}$, labeled by $\theta > 0$, is a member of the exponential family and identify the factors u, g, h, ϕ, θ, c.

2. Show that the Gamma distribution

$$f(\theta|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0 \theta} \tag{17}$$

with hyperparameters $w_0 := (\alpha_0, \beta_0)$ (where $\alpha_0 > 1$ and $\beta_0 > 0$) is conjugate for i.i.d. sampling from for the skew-logistic distribution. Relate the hyperparameters (τ_0, τ_1) to the standard parameters (α_0, β_0) of the gamma distribution.

3. Given an i.i.d. sample $x_1, ..., x_n$ from the skew-logistic distribution, and assuming that the prior is Gamma(α_0, β_0), derive the posterior distribution of θ.

4. Given an i.i.d. sample $x_1, \ldots, x_n$ from the skew-logistic distribution, and assuming that the prior is Gamma(α_0, β_0), derive the predictive PDF for a future $y = x_{n+1}$ up to a normalising constant.

5. Give a minimal sufficient statistic for θ under i.i.d. sampling from the skew-logistic distribution. Would this statistic still be sufficient if we had chosen a prior for θ which was not a Gamma distribution?

The exercise below is theoretical, and very challenging.

Exercise 48. ($\star\star\star$)Prove the following statement.

Consider a PDF/PMF $f(x|\theta)$ with $x \in \mathcal{X}$ and $\theta \in \Theta$ where $\mathcal{X}$ does not depend on θ. If there exist a parametrised conjugate prior family $\mathcal{F} = (\pi(\theta|\tau), \tau \in T)$ with $\dim(\Lambda) < \infty$, then $f(x|\tau)$ is number of the exponential family.

Hint: Use the Pitman-Koopman Lemma:

> **Lemma.** *(Pitman-Koopman Lemma) If a family of distributions is such that for a large enough sample size there exist a sufficient statistic of constant dimension, then the family is an exponential family if the support does not depend on θ.*

PS: Please think about the importance of this result (!!!)

Exercise 49. ($\star\star$)Suppose that you have a prior distribution for the probability θ of success in a certain kind of gambling game which has mean 0.4, and that you regard your prior information as equivalent to 12 trials. You then play the game 25 times and win 12 times. What is your posterior distribution for θ?

Exercise 50. ($\star\star$)Find a (two-dimensional) sufficient statistic for (α, β) given an n-sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ from the two-parameter gamma distribution

$$p(x|\alpha, \beta) = \{\beta^\alpha \Gamma(\alpha)\}^{-1} x^{\alpha-1} \exp(-x/\beta) \qquad (0 < x < \infty)$$

where the parameters α and β can take any values in $0 < \alpha < \infty, 0 < \beta < \infty$.

Exercise 51. ($\star\star$)Suppose that your prior for θ with PDF

$$\pi(\theta) = \frac{2}{3}\mathrm{N}(\theta|0, 1) + \frac{1}{3}\mathrm{N}(\theta|1, 1)$$

that a single observation $x \sim \mathrm{N}(\theta, 1)$ turns out to equal 2. What is your posterior probability that $\theta > 1$?

Hint | We should use the following identity, discussed in the Lecture notes, :

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

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

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

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

The Exercise below has been set as Homework 2

Exercise 52. ($\star\star$)Let $x = (x_1, ..., x_n)$ be observables. Consider a Bayesian model such as

$$\begin{cases} x_i|\lambda & \overset{\text{iid}}{\sim} \text{Pn}(\lambda), \ \forall i = 1, ..., n \\ \lambda & \sim \Pi(\lambda) \end{cases}$$

Hint-1 Poisson distribution $x \sim \text{Pn}(\lambda)$ has PMF: $\text{Pn}(x|\lambda) = \frac{1}{x!}\lambda^x \exp(-\lambda)1_{\mathbb{N}}(x)$, where $\mathbb{N} = \{0, 1, 2, ...\}$ and $\lambda > 0$.

Hint-2 Gamma distribution $x \sim \text{Ga}(a, b)$ has PDF: $\text{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)1_{(0,\infty)}(x)$, with $a > 0$ and $b > 0$.

Hint-2 Negative Binomial distribution $x \sim \text{Nb}(r, \theta)$ has PMF: $\text{Nb}(x|r, \theta) = \binom{r+x-1}{r-1}\theta^r(1-\theta)^x 1_{\mathbb{N}}(x)$ with $\theta \in (0, 1)$, $r \in \mathbb{N} - \{0\}$, and $\mathbb{N} = \{0, 1, 2, ...\}$.

1. Compute the likelihood in the aforesaid Bayesian model.

2. Show that the sampling distribution is a member of the exponential family.

3. Specify the PDF of the conjugate prior distribution $\Pi(\lambda)$ of λ, and identify the parametric family of distributions as $\lambda \sim \text{Ga}(a, b)$, with $a > 0$, and $b > 0$. While you are deriving the conjugate prior distribution of λ, discuss which of the prior hyper-parameters can be considered as the 'strength of the prior information and which can be considered as summarizing the prior information.

4. Compute the PDF of the posterior distribution of λ, identify the posterior distribution as a Gamma distribution $\text{Ga}(\tilde{a}, \tilde{b})$, and compute the posterior hyper-parameters $\tilde{a}$, and $\tilde{b}$.

5. Compute the PMF of the predictive distribution of a future outcome $y = x_{n+1}$, identify the name of the resulting predictive distribution, and compute its parameters.

Exercise 53. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\theta \in \Theta$, with $\Theta = \{\theta \in (0, 1)^k | \sum_{j=1}^k \theta_j = 1\}$ and $\mathcal{X}_k = \{x \in \{0, ..., n\}^k | \sum_{j=1}^k x_j = 1\}$.

Hint-1: Mu_k denotes the Multinomial probability distribution with PMF

$$\text{Mu}_k(x|\theta) = \begin{cases} \prod_{j=1}^k \theta_j^{x_j} & \text{, if } x \in \mathcal{X}_k \\ 0 & \text{, otherwise} \end{cases}$$

Hint-2: $\text{Di}_k(a)$ denotes the Dirichlet distribution with PDF

$$\text{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k \theta_j^{a_j-1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

1. Derive the conjugate prior distribution for θ, and recognize that it is a Dirichlet distribution family of distributions.

2. Verify that the prior distribution you derived above is indeed conjugate by using the definition.

Exercise 54. (⋆⋆)Consider the Bayesian model

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

where $\mathrm{Ga}(a,\theta)$ is the Gamma distribution with expected value a/θ and density function

$$f(x|a,\theta) = \frac{\theta^a}{\Gamma(a)} x^{a-1} \exp(-\theta x) 1\,(x \geq 0)\,.$$

Specify a Jeffrey's prior density for θ.

The Exercise below has been set as Homework 2

Exercise 55. (⋆⋆)Assume observation x sampled from a Maxwell distribution with density

$$f(x|\theta) = \sqrt{\frac{2}{\pi}} \theta^{3/2} x^2 \exp(-\frac{1}{2}\theta x^2).$$

Find the Jeffreys prior density for the parameter θ.

Exercise 56. (⋆⋆)Consider the trinomial distribution

$$p(x,y|\pi,\rho) = \frac{n!}{x!\,y!\,z!} \pi^x \rho^y \sigma^z, \qquad (x+y+z=n)$$

$$\propto \pi^x \rho^y (1 - \pi - \rho)^{n-x-y}.$$

Specify a Jeffreys' prior for (π,ρ).

HINT: It is $\mathrm{E}(x) = n\pi$, $\mathrm{E}(y) = n\rho$.

Exercise 57. (⋆⋆)Suppose that x has a Pareto distribution $\mathrm{Pa}(\xi,\gamma)$ where ξ is known but γ is unknown, that is,

$$p(x|\gamma) = \gamma \xi^\gamma x^{-\gamma-1} I_{(\xi,\infty)}(x).$$

Use Jeffreys' rule to find a suitable reference prior for γ.

Exercise 58. (⋆⋆)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \stackrel{\mathrm{iid}}{\sim} \mathrm{Ga}(\alpha,\beta),\ \forall i = 1,...,n \\ (\alpha,\beta) & \sim \Pi(\alpha,\beta) \end{cases}$$

where $\mathrm{Ga}(a,\beta)$ is the Gamma distribution with expected value α/β. Specify a Jeffrey's prior for $\theta = (\alpha,\beta)$.

Hint-1: Gamma distr.: $x \sim \mathrm{Ga}(a,b)$ has pdf $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0,+\infty)}(x)$, and Expected value $\mathrm{E}_{\mathrm{Ga}}(x|a,b) = \frac{a}{b}$

Hint-2: You may also need that the second derivative of the logarithm of a Gamma function is the 'polygamma function of order 1'. Ie,

- $F^{(0)}(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha} \log(\Gamma(a))$
- $F^{(1)}(\alpha) = \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \log(\Gamma(a))$

Hint-3: You may leave your answer in terms of function $F^{(1)}(\alpha)$.

Exercise 59. (⋆⋆⋆)Consider the the model of Normal linear regression where the observables are pairs (ϕ_i, y_i) for $i = 1, ..., n$, assumed to be modeled according to the sampling distribution

$$y_i | \beta, \sigma^2 \sim \mathrm{N}(\phi_i^\top \beta, \sigma^2)$$

for $i = 1, ..., n$ with unknown $(\beta, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$. Namely,

$$y | \beta, \sigma^2 \sim \mathrm{N}_n(\Phi\beta, I\sigma^2)$$

where $y = (y_1, ...y_n)$, and Φ is the design matrix. Here β is d-dimensional. Find the Jeffreys' priors for (β, σ^2).

Hint: Recall your AMV: $\frac{\mathrm{d}}{\mathrm{d}x} x^\top A x = 2Ax$, $\frac{\mathrm{d}}{\mathrm{d}x}(c + Ax) = A$, and $\frac{\mathrm{d}}{\mathrm{d}x}(A(x))^\top = \left(\frac{\mathrm{d}}{\mathrm{d}x} A(x)\right)^\top$.

Hint: If $y|\beta, \sigma^2 \sim \mathrm{N}_n(\Phi\beta, I\sigma^2)$, then $\mathrm{E}_{y|\beta,\sigma^2 \sim \mathrm{N}_n(\Phi\beta, I\sigma^2)}\left((y - \Phi\beta)^\top (y - \Phi\beta)\right) = n\sigma^2$.

Exercise 60. (⋆⋆)Consider the Bayesian model

$$\begin{cases} x_i | \theta & \overset{\mathrm{iid}}{\sim} \mathrm{Pn}(\theta), \ \forall i = 1, ..., n \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\mathrm{Pn}(\theta)$ is the Poisson distribution with expected value θ. Specify a Jeffreys' prior for θ.

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

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

Exercise 61. (⋆⋆)Consider the Bayesian model

$$\begin{cases} x_i | \theta & \overset{\mathrm{iid}}{\sim} \mathrm{Pn}(\theta), \ \forall i = 1, ..., n \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\mathrm{Pn}(\theta)$ is the Poisson distribution with expected value θ. Specify a Maximum entropy prior under the constrain $\mathrm{E}(\theta) = 2$ and reference measure such as $\pi_0(\theta) = \frac{1}{\sqrt{\theta}}$. In particular, you also have to state the name of the derived Maximum entropy prior distribution and report the values of its parameters.

Hint-1: Poisson distribution: $x \sim \mathrm{Pn}(\theta)$ has PMF

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

Hint-2: Gamma distribution: $x \sim \mathrm{Ga}(a, b)$ has PDF

$$\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-\beta x) 1(x > 0)$$

Created on 2021/11/03 at 22:24:10   by Georgios Karagiannis