Winter 22 – STAT206B Homework 1 Solution

1. Let $X \sim \text{Exp}(\lambda)$, where $\text{E}(X) = 1/\lambda$. What is the pmf (probability mass function) of $Y = \lfloor X \rfloor$ (the floor of X)? Do you recognize it as a distribution that you have studied in the past?

   We first find $\Pr(Y = y) = \Pr(y \le X < y - 1) = \int_y^{y+1} \lambda e^{-\lambda x} dx = e^{-\lambda y}(1 - e^{-\lambda})$, $y = 0, 1, 2, \ldots$. We observe it is the probability function for the geometric distribution with parameter $e^{-\lambda}$, i.e., $Y \sim \text{Geometric}(e^{-\lambda})$.

2. Let X_1 and X_2 be two independent random variables such that $X_i \sim \text{Ga}(a_i, b)$ for any $a_1, a_2, b > 0$. Define $Y = X_1/(X_1 + X_2)$ and $Z = (X_1 + X_2)$.

   (a) Find the joint pdf for Y and Z and show that these two random variables are independent.

   $Y = X_1/(X_1 + X_2)$ and $Z = (X_1 + X_2)$, $X_1, X_2 \in \mathbb{R}^+$ are differentiable one-to-one transformations. We observe $X_1 = YZ$ and $X_2 = Z(1 - Y)$. We use a change of variables. We have $0 < Y < 1$ and $Z \in \mathbb{R}^+$, and find

   $$\begin{aligned} g(y,z) &= f(yz, z(1-y))|J| \\ &= \frac{b^{a_1}}{\Gamma(a_1)}(yz)^{a_1-1}e^{-byz}\frac{b^{a_2}}{\Gamma(a_2)}((1-y)z)^{a_2-1}e^{-b(1-y)z}z \\ &\propto \underbrace{y^{a_1-1}(1-y)^{a_2-1}}_{\text{function of } y \text{ only}}\underbrace{z^{a_1+a_2-1}e^{-bz}}_{\text{function of } z \text{ only}}, y \in (0,1) \text{ and } 0 < z. \end{aligned}$$

   Since the joint pdf of Y and Z can be expressed as a product of a function of Y only and a function of Z only, Y and Z are independent.

   (b) Find the marginal pdf of Z. Do you recognize this pdf as belonging to some family that you know?

   From part (a), we have $f(z) \propto z^{a_1+a_2-1}e^{-bz}$, $z \in \mathbb{R}^+$ and recognize it is proportional to the pdf of $\text{Gamma}(a_1 + a_2, b)$. Thus, $Z \sim \text{Gamma}(a_1 + a_2, b)$.

   (c) Find the marginal pdf of Y. Do you recognize this pdf as belonging to some family that you know?

   From part (a), we have $f(y) \propto y^{a_1-1}(1-y)^{a_2-1}$, $0 < y < 1$ and recognize it is proportional to the pdf of $\text{Be}(a_1, a_2)$. Thus, $Y \sim \text{Be}(a_1, a_2)$.

   (d) Compute $\text{E}(Y^k)$ for any $k > 0$.

   We have

   $$\text{E}(Y^k) = \int_0^\infty y^k \frac{1}{B(a_1, a_2)} y^{a_1-1}(1-y)^{a_2-1} dy = \frac{B(a_1 + k, a_2)}{B(a_1, a_2)}.$$

(e) What does this result imply if $a_i = b = 1$?

We observe $Y \sim \text{Be}(1,1) = \text{Unif}(1,1)$. For $Z = X_1 + \ldots, X_n$, $Z \sim \text{Gamma}(n, 1)$, that is, Erlang distribution.

3. Consider three independent random variables X_1, X_2 and X_3 such that $X_i \overset{indep}{\sim} \text{Gamma}(a_i, b)$, $i = 1, 2, 3$. Let

$$\boldsymbol{Y} = (Y_1, Y_2, Y_3) = \left(\frac{X_1}{X_1 + X_2 + X_3}, \frac{X_2}{X_1 + X_2 + X_3}, \frac{X_3}{X_1 + X_2 + X_3} \right).$$

(a) Show that $\boldsymbol{Y} \sim \text{Dirichlet}(a_1, a_2, a_3)$, a Dirichlet distribution.

First, we consider the joint density of the three independent Gamma-distributed RVs:

$$p(x_1, x_2, x_3) = \prod_{i=1}^{3} p(x_i) = \prod_{i=1}^{3} \frac{x_i^{a_i - 1} e^{-x_i/b}}{\Gamma(a_i) b^{a_i}} = \frac{e^{\sum_{i=1}^{3} x_i/b} \prod_{i=1}^{3} x_i^{a_i - 1}}{b^{\sum_{i=1}^{3} a_i} \prod_{i=1}^{3} \Gamma(a_i)}.$$

Since $Y_i = X_i / \sum_{i=1}^{3} X_i$, we find

$$\begin{aligned} X_1 &= Y_1 Z \\ X_2 &= Y_2 Z \\ X_3 &= Y_3 Z = (1 - Y_1 - Y_2) Z, \end{aligned}$$

where $Z = \sum_{i=1}^{3} X_i$. To obtain the joint distribution of (Y_1, Y_2, Z), we find the Jacobian for this change of variables. The matrix is

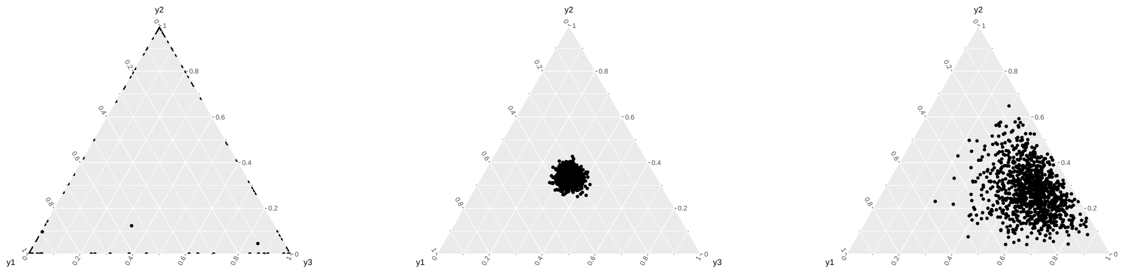
$$J = \begin{bmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} & \frac{dx_1}{d_z} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} & \frac{dx_2}{d_z} \\ \frac{dx_3}{dy_1} & \frac{dx_3}{dy_2} & \frac{dx_3}{d_z} \end{bmatrix} = \begin{pmatrix} Z & 0 & Y_1 \\ 0 & Z & Y_2 \\ -Z & -Z & (1 - Y_1 - Y_2) \end{pmatrix}$$

So, the Jacobian, $|J|$ is Z^2.

$$p(Y_1, Y_2, Z) = \frac{(y_1 z)^{a_1 - 1} e^{-y_1 z/b} (y_2 z)^{a_2 - 1} e^{-y_2 z/b} \{(1 - y_1 - y_2) z\}^{a_3 - 1} e^{-(1 - y_1 - y_2) z/b}}{b^{\sum_{i=1}^{3} a_i} \prod_{i=1}^{3} \Gamma(a_i)} z^2,$$

where $0 < y_1, y_2 < 1$, $y_1 + y_2 < 1$ and $0 < z$.
By letting $y_3 = 1 - y_1 - y_2$,

$$p(Y_1, Y_2, Z) = \frac{y_1^{a_1 - 1} y_2^{a_2 - 1} y_3^{a_3 - 1} z^{\sum_{i=1}^{3} a_i - 1} e^{z \overbrace{\sum_{i=1}^{3} y_i}^{=1} / b}}{b^{\sum_{i=1}^{3} a_i} \prod_{i=1}^{3} \Gamma(a_i)}.$$



(a) $\boldsymbol{a} = (0.01, 0.01, 0.01)$ (b) $\boldsymbol{a} = (100, 100, 100)$ (c) $\boldsymbol{a} = (3, 5, 10)$

Figure 1: $\boldsymbol{y} = (y_1, y_2, y_3)$ simulated from Dirichlet distribution $\text{Dir}(\boldsymbol{a})$.

We now integrate out z to obtain $p(y_1, y_2, y_3)$.

$$\begin{aligned} p(y_1, y_2, y_3) &= \int_{\mathbb{R}} \frac{y_1^{a_1-1} y_2^{a_2-1} y_3^{a_3-1} \qquad \overbrace{z^{\sum_{i=1}^3 a_i - 1} e^{-z/b}}^{\text{kernel for Gamma}(\sum_{i=1}^3 a_i, b)}}{b^{\sum_{i=1}^3 a_i} \prod_{i=1}^3 \Gamma(a_i)} dz \\ &= \frac{\Gamma(\sum_{i=1}^3 a_i)}{\prod_{i=1}^3 \Gamma(a_i)} y_1^{a_1-1} y_2^{a_2-1} y_3^{a_3-1}. \end{aligned}$$

Thus, $\boldsymbol{Y} = (Y_1, Y_2, Y_3) \sim \text{Dir}(a_1, a_2, a_3)$.

(b) How can this result be used to generate random variables according to a Dirichlet distribution? Write a simple function in `R` or `Matlab` (your choice) that takes as inputs n, the number of trivariate vectors to be generated, and $\boldsymbol{a} = (a_1, a_2, a_3)$ and generates a matrix of size $n \times 3$ whose rows correspond to independent samples from a Dirichlet distribution with parameter (a_1, a_2, a_3).

Use each of $\boldsymbol{a} = (0.01, 0.01, 0.01)$, $(100, 100, 100)$, and $(3, 5, 10)$ and comment how the density of $\boldsymbol{Y}$ changes over $\boldsymbol{a}$.

```
#a is a vector of length p; a=(a_1, a_2, ..., a_p)
#n is the sample size

> dirichlet <- function(a, n){
        p <- length(a)
        y <- array(NA, dim=c(n, p)) #Each row of y is iid sample from Dir(a)

        for (i in 1:n) {
              tmp <- rgamma(p, a, 1)
              y[i, ] <- tmp / sum(tmp)
        }
```

```
    return(y)
}
```

Fig 1 illustrates trinary plots of 1000 simulated $\boldsymbol{y}$ with different $\boldsymbol{a}$. For $a < 1$, $\boldsymbol{y}$ are at the vertices and edges, meaning either one or two of y_1, y_2 or y_3 are close to zero. For $a = 100$, all y_i are close to 1/3 (i.e., small variance) as shown in Fig 1(b). For $\boldsymbol{a} = (3, 5, 10)$, $\boldsymbol{y}$ are around the mean $(3/18, 5/18, 10/18)$ and its variance is larger than that with $a = 100$.

4. Y follows an inverse Gamma distribution with shape parameter a and scale parameter b ($Y \sim \mathrm{IG}(a, b)$) if $Y = 1/X$ with $X \sim \mathrm{Gamma}(a, b)$ (assume the Gamma distribution is parameterized so that $\mathrm{E}(X) = ab$.

   (a) Find the density of Y.

   Since $X \sim \mathrm{Gamma}(a, b)$,

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \text{for } x > 0.$$

   Let $y = 1/x$. Then $x = 1/y$ and $\frac{dx}{dy} = -1/y^2$. Therefore,

$$p(y) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{a-1} \exp(-b/y) \left|-\frac{1}{y^2}\right| = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp(-b/y),\ y > 0,$$

   which is an inverse Gamma with shape a and rate b, $y \sim \mathrm{IG}(a, b)$.

   (b) Compute $\mathrm{E}(Y^k)$. Do you need to impose any constrain on the problem for this expectation to exists?

   We have

$$\mathrm{E}(Y^k) = \int_{\mathbb{R}+} y^k \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp(-b/y) dy = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-k)}{b^{a-k}} = \frac{b^k \Gamma(a-k)}{\Gamma(a)}.$$

   That is, $\mathrm{E}(Y^k) = b^k \Gamma(a-k)/\Gamma(a)$ for $a - k > 0$.

   (c) Compare $\mathrm{E}(Y^k)$ to $1/\mathrm{E}(X^k)$ (hint: look at the ratio of the two quantities).

   We first find

$$\mathrm{E}(X^k) = \frac{b^a \Gamma(a+k)}{\Gamma(a) b^{a+k}} = \frac{\Gamma(a+k)}{\Gamma(a) b^k}.$$

   Then,

$$\zeta \equiv \frac{1/\mathrm{E}(X^k)}{\mathrm{E}(Y^k)} = \frac{\Gamma(a)\Gamma(a)}{\Gamma(a+k)\Gamma(a-k)}$$

   which implies that $\zeta = 1 \Leftrightarrow k = 0$. That is, none of the moments of X is invariant to the reciprocal transformation.

5. Y follows a log normal distribution with parameters μ and σ^2 (denotes as $Y \sim$ Log-N(μ, σ^2) if $Y = \exp(X)$ where $X \sim \mathrm{N}(\mu, \sigma^2)$).

   (a) Find the density of Y.

   We have $Y = \exp(X) \in \mathbb{R}^+$. We use a change of a variable and find

$$g(y) = f(\log(y)) \left|\frac{1}{y}\right| = \frac{1}{\sqrt{2\pi\sigma^2}y} \exp\left\{-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right\}, \quad y \in \mathbb{R}^+.$$

   (b) Compute the mean and the variance of Y.

   Observe the $k-$th moment of Y, $\mathrm{E}^Y(Y^k) = \mathrm{E}^X(e^{kX})$ and find

$$\begin{aligned} \mathrm{E}^X(e^{kX}) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2} + kx\right\} dx \\ &= \exp\left\{-\frac{\mu^2}{2\sigma^2} + \frac{(\mu + k\sigma^2)^2}{2\sigma^2}\right\} \\ &= \exp\left(k\mu + k^2\sigma^2/2\right). \end{aligned}$$

   Thus, $\mathrm{E}(Y) = \exp\left(\mu + \sigma^2/2\right)$ and $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \mathrm{E}(Y)^2 = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$.

6. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ with $X \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ and set $\boldsymbol{Z}_1 = (X_1, \ldots, X_q)$ and $\boldsymbol{Z}_2 = (X_{q+1}, \ldots, , X_p)$ with $1 < q < p$. Show that

$$\boldsymbol{Z}_1 \mid \boldsymbol{Z}_2 \sim \mathrm{N}_q\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{Z}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right),$$

   where $\boldsymbol{\mu}_k$ and $\Sigma_{k\ell}$ denote the blocks of $\boldsymbol{\mu}$ and Σ where the rows correspond to the variables in $\boldsymbol{Z}_k$ and the columns to the variables in $\boldsymbol{Z}_\ell$.

   Using some results of the inverse of a partitioned matrix, we find for symmetric and $\Sigma > 0$,

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1} & \Sigma_{22.1}^{-1} \end{bmatrix} = \begin{bmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1} \\ -\Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22.1}^{-1} \end{bmatrix},$$

   where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $\Sigma_{22.2} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. We have

$$\begin{aligned} g(\boldsymbol{z}_1 \mid \boldsymbol{z}_2) &\propto f(\boldsymbol{z}_1, \boldsymbol{z}_2) \\ &\propto \exp\left\{-\frac{1}{2}\begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{z}_2 - \boldsymbol{\mu}_2 \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{z}_2 - \boldsymbol{\mu}_2 \end{bmatrix}\right\} \\ &\propto \exp\Big[-\frac{1}{2}\Big\{(\boldsymbol{z}_1 - \boldsymbol{\mu}_1)'\Sigma_{11.2}^{-1}(\boldsymbol{z}_1 - \boldsymbol{\mu}_1) - (\boldsymbol{z}_1 - \boldsymbol{\mu}_1)'\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2) \\ &\qquad -(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)'\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}(\boldsymbol{z}_1 - \boldsymbol{\mu}_1)\Big\}\Big] \\ &\propto \exp\Big[-\frac{1}{2}\Big\{\boldsymbol{z}_1'\Sigma_{11.2}^{-1}\boldsymbol{z}_1 - \boldsymbol{z}_1'\left(\Sigma_{11.2}^{-1}\boldsymbol{\mu}_1 + \Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)\right) \\ &\qquad -(\boldsymbol{\mu}_1'\Sigma_{11.2}^{-1} + (\boldsymbol{z}_2 - \boldsymbol{\mu}_2)'\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1})\boldsymbol{z}_1\Big\}\Big]. \end{aligned}$$

We recognize the kernel for $\mathrm{N}_q(\Sigma_{11.2}\left(\Sigma_{11.2}^{-1}\boldsymbol{\mu}_1 + \Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)\right), \Sigma_{11.2})$.
That is, $\boldsymbol{Z}_1 \mid \boldsymbol{Z}_2 \sim \mathrm{N}_q\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{Z}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$.

7. Show that if $X \sim \mathrm{Exp}(\beta)$, then

   (a) $Y = X^{1/\gamma}$ has a Weibull distribution with parameters γ and β with $\gamma > 0$ a constant.

   We have $Y = X^{1/\gamma} \in \mathbb{R}^+$. We use a change of a variable and find
   $$g(y) = f(y^\gamma)|\gamma y^{\gamma-1}| = \beta\gamma y^{\gamma-1}\exp\left(-\beta y^\gamma\right), \quad y \in \mathbb{R}^+.$$
   We also observe $\mathrm{E}^Y(Y^k) = \mathrm{E}^X(X^{k/\gamma}) = \frac{\Gamma(1+k/\gamma)}{\beta^{k/\gamma}}$ (use the gamma kernel). We find $\mathrm{E}^Y(Y) = \frac{\Gamma(1+1/\gamma)}{\beta^{1/\gamma}}$ and $\mathrm{Var}(Y) = \mathrm{E}^X(X^{2/\gamma}) - \mathrm{E}^X(X^{1/\gamma})^2 = \left\{\Gamma(1+2/\gamma) - (\Gamma(1+1/\gamma))^2\right\}/\beta^{2/\gamma}$.

   (b) $Y = (2X/\beta)^{1/2}$ has the Rayleigh distribution.

   We have $Y = (2X/\beta)^{1/2} \in \mathbb{R}^+$. We use a change of a variable and find
   $$g(y) = f(\beta y^2/2)|\beta y| = \beta^2 y \exp\left(-\beta^2 y^2/2\right), \quad y \in \mathbb{R}^+.$$
   We also observe $\mathrm{E}^Y(Y^k) = \mathrm{E}^X((2X/\beta)^{k/2}) = \frac{2^{k/2}\Gamma(1+k/2)}{\beta^k}$ (use the gamma kernel). We find $\mathrm{E}^Y(Y) = \sqrt{2}\Gamma(1.5)/\beta$ and $\mathrm{Var}(Y) = \mathrm{E}^X(2X/\beta) - \mathrm{E}^X((2X/\beta)^{1/2})^2 = 2(1 - \Gamma(1.5)^2)/\beta^2$.

   For both parts, derive the form of the pdf, verify that is a pdf, and calculate the mean and the variance.

8. Let $Y \mid X \sim \mathrm{Poisson}(X)$ and let $X \sim \mathrm{Exp}(\lambda)$. What is the marginal distribution of Y?

   We find
   $$m(y) = \int_{\mathbb{R}^+} \frac{e^{-x}x^y}{y!}\lambda e^{-\lambda x}dx = \frac{\Gamma(y+1)\lambda}{y!(1+\lambda)^{y+1}} = \frac{\lambda}{(1+\lambda)^{y+1}}, y = 0, 1, 2, \ldots.$$

9. (Robert) If $y \sim \mathrm{Binomial}(n, \theta)$ and $x \sim \mathrm{Binomial}(m, \theta)$, and $\theta \sim \mathrm{Beta}(\alpha, \beta)$. Find the predictive distribution of y given x.

   Assume conditional independence of X and Y given θ. We know that $\theta \mid x \sim \mathrm{Be}(\alpha + x, \beta + m - x)$ and then have for $y = 0, \ldots, n$,
   $$\begin{aligned} f(y \mid x) &= \int_0^1 f(y \mid \theta)\pi(\theta \mid x)d\theta \\ &= \int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{\theta^{\alpha+x-1}(1-\theta)^{\beta+m-x-1}}{B(\alpha+x, \beta+m-x)}d\theta \\ &= \frac{\binom{n}{y}B(\alpha+x+y, \beta+m-x+n-y)}{B(\alpha+x, \beta+m-x)}, \quad y = 0, 1, 2, \ldots \end{aligned}$$

that is, $y \mid x \sim \text{Beta-Binomial}(n, \alpha + x, \beta + m - x)$.

10. (Robert) Give the posterior and the marginal distributions in the following cases:

    (a) $x \mid \sigma^2 \sim \mathrm{N}(0, \sigma^2)$ and $1/\sigma^2 \sim \text{Gamma}(1, 2)$.

    We know $1/\sigma^2 \sim \text{Gamma}(1, 2) \;\Rightarrow \sigma^2 \sim \text{IG}(1, 2)$.

$$\begin{aligned} m(x) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} (\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{2}{\Gamma(1)} (\sigma^2)^{-2} \exp\left(-\frac{2}{\sigma^2}\right) d\sigma^2 \\ &= \frac{2\Gamma(3/2)}{\sqrt{2\pi}(2 + x^2/2)^{3/2}}, \quad x \in \mathbb{R}. \end{aligned}$$

    We have

$$\pi(\sigma^2 \mid x) \propto (\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right) (\sigma^2)^{-2} \exp\left(-\frac{2}{\sigma^2}\right), 0 < \sigma^2$$

    that is, $\sigma^2 \mid x \sim \text{IG}(3/2, 2 + x^2/2)$ ($\Rightarrow \quad 1/\sigma^2 \mid x \sim \text{Gamma}(3/2, 2 + x^2/2)$).

    (b) $x \mid p \sim \text{Negative-Binomial}(10, p)$ and $p \sim \text{Beta}(1/2, 1/2)$.

    Following the parameterization of CR p522, we have

$$\begin{aligned} m(x) &= \int_0^\infty \binom{n + x + 1}{x} p^n (1-p)^x \frac{1}{B(1/2, 1/2)} p^{-1/2} (1-p)^{-1/2} dp \\ &= \frac{\binom{n+x+1}{x} B(n + 1/2, x + 1/2)}{B(1/2, 1/2)}, x = 0, 1, \ldots. \end{aligned}$$

    We have

$$\pi(p \mid x) \propto p^n (1-p)^x p^{-1/2} (1-p)^{-1/2}, 0 < p < 1,$$

    that is, $p \mid x \sim \text{Be}(n + 1/2, x + 1/2)$.

11. (a) The prior $\tilde{\pi}(\theta, \sigma^2) = 1/\sigma^2$ is improper, and we need to check if the marginal distribution

$m(\boldsymbol{x})$ is finite (so the posterior distribution of θ and σ^2 is a legitimate distribution).

$$\begin{aligned}
m(\boldsymbol{x}) &= \int_0^\infty \int_{-\infty}^\infty f(\boldsymbol{x} \mid \theta, \sigma^2)\tilde{\pi}(\theta, \sigma^2) d\theta d\sigma^2 \\
&= \int_0^\infty \int_{-\infty}^\infty (2\pi)^{-n/2}(\sigma^2)^{-n/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\} d\theta d\sigma^2 \\
&= \int_0^\infty \frac{(2\pi)^{-(n-1)/2}(\sigma^2)^{-(n-1)/2-1}}{\sqrt{n}} \exp\left\{-\frac{s^2}{2\sigma^2}\right\} \int_{-\infty}^\infty \underbrace{\frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}}_{\text{the pdf of } N(\bar{x}, \sigma^2/n)} d\theta d\sigma^2 \\
&= \int_0^\infty \underbrace{\frac{(2\pi)^{-(n-1)/2}(\sigma^2)^{-(n-1)/2-1}}{\sqrt{n}} \exp\left\{-\frac{s^2}{2\sigma^2}\right\}}_{\text{a kernel of } IG((n-1)/2, s^2/2)} d\sigma^2 \\
&= \frac{(2\pi)^{-(n-1)/2}}{\sqrt{n}} \frac{\Gamma\left(\frac{(n-1)}{2}\right)}{(s^2/2)^{(n-1)/2}} < \infty.
\end{aligned}$$

This ensures that the posterior distribution of θ and σ^2 is a legitimate distribution. We write down the posterior up to proportionality as below.

$$\begin{aligned}
\pi(\theta, \sigma^2 \mid \boldsymbol{x}) &= \frac{f(\boldsymbol{x} \mid \theta, \sigma^2)\tilde{\pi}(\theta, \sigma^2)}{m(\boldsymbol{x})} \\
&\propto f(\boldsymbol{x} \mid \theta, \sigma^2)\tilde{\pi}(\theta, \sigma^2) \\
&\propto (\sigma^2)^{-n/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}.
\end{aligned}$$

(b)

$$\begin{aligned}
\pi(\theta \mid \boldsymbol{x}, \sigma^2) &= \frac{\pi(\theta, \sigma^2 \mid \boldsymbol{x})}{\pi(\sigma^2 \mid \boldsymbol{x})} \\
&\propto \pi(\theta, \sigma^2 \mid \boldsymbol{x}) \\
&\propto (\sigma^2)^{-n/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}.
\end{aligned}$$

We recognize it is a kernel of $N(\bar{x}, \sigma^2/n)$. So $\pi(\theta \mid \boldsymbol{x}, \sigma^2)$ is $N(\bar{x}, \sigma^2/n)$.

$$\begin{aligned}
\pi(\sigma^2 \mid \boldsymbol{x}) &= \int_{-\infty}^{\infty} \pi(\theta, \sigma^2 \mid \boldsymbol{x}) d\theta \\
&\propto \int_{-\infty}^{\infty} f(\boldsymbol{x} \mid \theta, \sigma^2) \tilde{\pi}(\theta, \sigma^2) d\theta \\
&\propto \int_{-\infty}^{\infty} (\sigma^2)^{-n/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\} d\theta \\
&\propto (\sigma^2)^{-(n-1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}}_{\text{the pdf of } N(\bar{x}, \sigma^2/n)} d\theta \\
&= (\sigma^2)^{-(n-1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2}\right\}.
\end{aligned}$$

We recognize it is a kernel of $\mathrm{IG}((n-1)/2, s^2/2)$. So $\pi(\sigma^2 \mid \boldsymbol{x})$ is $\mathrm{IG}((n-1)/2, s^2/2)$.

(c)

$$\begin{aligned}
\pi(\theta \mid \boldsymbol{x}) &= \int_{0}^{\infty} \pi(\theta, \sigma^2 \mid \boldsymbol{x}) d\sigma^2 \\
&\propto \int_{0}^{\infty} f(\boldsymbol{x} \mid \theta, \sigma^2) \tilde{\pi}(\theta, \sigma^2) d\sigma^2 \\
&= \int_{0}^{\infty} (\sigma^2)^{-n/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\} d\sigma^2 \\
&= \left\{\frac{s^2}{2} + \frac{n(\bar{x}-\theta)^2}{2}\right\}^{-n/2} \\
&= \left\{1 + \frac{n(\bar{x}-\theta)^2}{s^2}\right\}^{-n/2}.
\end{aligned}$$

We recognize it is a kernel of a t-distribution, $t(n-1, \bar{x}, s^2/(n(n-1)))$.

11-1. Assume that observations, $x_1, \ldots, x_n$ are iid from $\mathrm{N}(\theta, \sigma^2)$, where θ and σ^2 are unknown. Consider

$$\pi(\theta, \sigma^2) = \pi_1(\theta \mid \sigma^2)\pi_2(\sigma^2),$$

where π_1 is a normal distribution $\mathrm{N}(\mu, \sigma^2/n_0)$ and π_2 is a inverse gamma distribution $\mathrm{IG}(v/2, s_0^2/2)$.

(a) Find the joint posterior distribution $\pi(\theta, \sigma^2 \mid \boldsymbol{x})$.

Since $\pi(\theta, \sigma^2)$ is proper, $\pi(\theta, \sigma^2 \mid \boldsymbol{x})$ is also proper. We use the proportionality argument and find the posterior distribution. From lecture, we have

$$p(\boldsymbol{x} \mid \theta, \sigma) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2}\right\},$$

where $\bar{x} = \sum_{i=1}^n x_i/n$ and $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$. So,

$$
\begin{aligned}
\pi(\theta, \sigma^2 \mid \boldsymbol{x}) &\propto p(\boldsymbol{x} \mid \theta, \sigma)\pi(\theta, \sigma^2) \\
&\propto \underbrace{(\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2}\right\}}_{\text{from } p(\boldsymbol{x} \mid \theta, \sigma)} \underbrace{(\sigma^2)^{-1/2} \exp\left\{-\frac{n_0(\theta-\mu)^2}{2\sigma^2}\right\}}_{\text{from } \pi_1(\theta \mid \sigma^2)} \\
&\quad \times \underbrace{(\sigma^2)^{-v/2-1} \exp\left\{-\frac{s_0^2}{2\sigma^2}\right\}}_{\text{from } \pi_2(\sigma^2)} \\
&\propto (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2} - \frac{n_0(\theta-\mu)^2}{2\sigma^2} - \frac{s_0^2}{2\sigma^2}\right\}.
\end{aligned}
$$

(b) Find the posterior distributions $\pi_1(\theta \mid \boldsymbol{x}, \sigma^2)$ and $\pi_2(\sigma^2 \mid \boldsymbol{x})$.

$$
\begin{aligned}
\pi_1(\theta \mid \boldsymbol{x}, \sigma^2) &\propto \pi(\theta, \sigma^2 \mid \boldsymbol{x}) \\
&\propto (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2} - \frac{n_0(\theta-\mu)^2}{2\sigma^2} - \frac{s_0^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{n(\theta-\bar{x})^2}{2\sigma^2} - \frac{n_0(\theta-\mu)^2}{2\sigma^2}\right\} \\
&\propto \exp\left[-\frac{(n+n_0)}{2\sigma^2}\left\{\theta^2 - 2\left(\frac{n_0\mu + n\bar{x}}{n_0+n}\right)\theta\right\}\right].
\end{aligned}
$$

We recognize a kernel for a normal distribution and identify the parameters of the normal distribution. We find $\pi_1(\theta \mid \boldsymbol{x}, \sigma^2)$ is $\mathrm{N}\left(\theta_1, \sigma^2/n_1\right)$, where $n_1 = (n + n_0)$ and $\theta_1 = (n_0\mu + n\bar{x})/n_1$. We next find $\pi_2(\sigma^2 \mid \boldsymbol{x})$. We have

$$
\begin{aligned}
\pi_2(\sigma^2 \mid \boldsymbol{x}) &\propto \int_{\mathbb{R}} \pi(\theta, \sigma^2 \mid \boldsymbol{x}) d\theta \\
&\propto \int_{\mathbb{R}} (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2} - \frac{n_0(\theta-\mu)^2}{2\sigma^2} - \frac{s_0^2}{2\sigma^2}\right\} d\theta \\
&\propto (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{1}{2\sigma^2}\left(s^2 + s_0^2 + n\bar{x}^2 + n_0\mu^2\right)\right\} \\
&\quad \times \int_{\mathbb{R}} \frac{\sqrt{2\pi\sigma^2/n_1}}{\sqrt{2\pi\sigma^2/n_1}} \exp\left\{-\frac{n_1}{2\sigma^2}\left(\theta^2 - 2\theta_1\theta \pm \theta_1^2\right)\right\} d\theta \\
&\propto (\sigma^2)^{-(v+n)/2-1} \exp\left\{-\frac{1}{2\sigma^2}\left(s^2 + s_0^2 + n\bar{x}^2 + n_0\mu^2 + n_1\theta_1^2\right)\right\}
\end{aligned}
$$

We recognize a kernel for IG and find their parameters. We find $\pi_2(\sigma^2 \mid \boldsymbol{x})$ is IG$((v + n)/2, s_1^2/2)$, where $s_1^2 = s^2 + s_0^2 + n\bar{x}^2 + n_0\mu^2 + n_1\theta_1^2 = s_0^2 + s^2 + n_0 n(\mu - \theta)^2/(n_0 + n)$.

(c) Find the marginal posterior distribution of θ, $\pi(\theta \mid \boldsymbol{x})$.

$$\begin{aligned}
\pi_2(\theta \mid \boldsymbol{x}) &\propto \int_{\mathbb{R}^+} \pi(\theta, \sigma^2 \mid \boldsymbol{x}) d\sigma^2 \\
&\propto \int_{\mathbb{R}} (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2} - \frac{n_0(\theta-\mu)^2}{2\sigma^2} - \frac{s_0^2}{2\sigma^2}\right\} d\sigma^2 \\
&\propto \int_{\mathbb{R}} (\sigma^2)^{-(v+n+1)/2-1} \exp\left\{-\frac{1}{2\sigma^2}\left(s_1^2 + n_1(\theta-\theta_1)^2\right)\right\} d\sigma^2 \\
&\propto \left\{s_1^2 + n_1(\theta-\theta_1)^2\right\}^{-(v+n+1)/2} \\
&\propto \left\{1 + \frac{(\theta-\theta_1)^2}{s_1^2/n_1}\right\}^{-(v+n+1)/2}.
\end{aligned}$$

We recognize a kernel for a t-distribution and find their parameters. We find $\pi(\theta \mid \boldsymbol{x})$ is $t(v+n, \theta_1, s_1^2/n_1/(v+n))$.

12. (a) First, similar to the univariate case, we write down the likelihood using the sufficient statistics,

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}, \Sigma) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} tr(\Sigma^{-1} S) - \frac{n}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{x}})^t \Sigma^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{x}})\right\}$$

where $\bar{\boldsymbol{x}}$ is the sample mean vector (p-dim) and $S = \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^t (p \times p \text{ dim})$. Note that since the prior is proper, the posterior is proper and we express the joint posterior distribution up to proportionality;

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \Sigma^{-1} \mid \boldsymbol{x}) &\propto \pi(\boldsymbol{\theta}, \Sigma) \prod_{i=1}^n f(\boldsymbol{x}_i \mid \boldsymbol{\theta}, \Sigma) \\
&\propto |\Sigma|^{-(\alpha+p+n)/2-1} \exp\left\{-\frac{n+n_0}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_1)^t \Sigma^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_1) - \frac{1}{2} tr(W_1^{-1} \Sigma^{-1})\right\}
\end{aligned}$$

where

$$\boldsymbol{\mu}_1 = \frac{n_0 \boldsymbol{\mu} + n \bar{\boldsymbol{x}}}{n_0 + n}, \text{ and } W_1^{-1} = \left(W^{-1} + S + \frac{n_0 n (\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'}{n_0 + n}\right)$$

$$\begin{aligned}
\pi_1(\boldsymbol{\theta} \mid \boldsymbol{x}, \Sigma^{-1}) &= \frac{\pi(\boldsymbol{\theta}, \Sigma^2 \mid \boldsymbol{x})}{\pi(\Sigma \mid \boldsymbol{x})} \\
&\propto \pi(\boldsymbol{\theta}, \Sigma^2 \mid \boldsymbol{x}) \\
&\propto |\Sigma|^{-(\alpha+p+n)/2-1} \exp\left\{-\frac{n+n_0}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_1)^t \Sigma^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_1) - \frac{1}{2} tr(W_1^{-1} \Sigma^{-1})\right\} \\
&\propto \exp\left\{-\frac{n+n_0}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_1)^t \Sigma^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_1)\right\}.
\end{aligned}$$

We recognize it is a kernel of $N_p(\boldsymbol{\mu}_1, \Sigma/(n_0+n))$. So $\pi_1(\boldsymbol{\theta} \mid \boldsymbol{x}, \Sigma)$ is $N_p(\boldsymbol{\mu}_1, \Sigma/(n_0+n))$.

$$\begin{aligned}
\pi_2(\Sigma^{-1} \mid \boldsymbol{x}) &= \int \pi(\boldsymbol{\theta}, \Sigma \mid \boldsymbol{x}) d\boldsymbol{\theta} \\
&\propto \int f(\boldsymbol{x} \mid \boldsymbol{\theta}, \Sigma)\pi(\boldsymbol{\theta}, \Sigma) d\boldsymbol{\theta} \\
&= \int |\Sigma|^{-(\alpha+p+n)/2-1} \exp\left\{-\frac{n+n_0}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_1)^t \Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_1) - \frac{1}{2} tr(W_1^{-1}\Sigma^{-1})\right\} d\boldsymbol{\theta} \\
&= |\Sigma|^{-(\alpha+p+n+1)/2} \exp\left\{-\frac{1}{2} tr(W_1^{-1}\Sigma^{-1})\right\}.
\end{aligned}$$

We recognize it is a kernel of Wishart$(\alpha+n, W_1)$. So $\pi_2(\Sigma^{-1} \mid \boldsymbol{x})$ is Wishart$(\alpha+n, W_1)$.

(b) Yes, it is conjugate since $\pi_1(\boldsymbol{\theta} \mid \boldsymbol{x}, \Sigma)$ and $\pi_2(\Sigma^{-1} \mid \boldsymbol{x})$ have the same forms as $\pi_1(\boldsymbol{\theta} \mid \Sigma)$ and $\pi_2(\Sigma^{-1})$.