Homework 2

20213073 Donggyu Kim

April 11, 2021

1. (a) *Proof.*
Whenever $A \subset B$, we get

$$B = A \cup (B \setminus A)$$

By the definition of probability measures,

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A \cup (B \setminus A)) \\ &= \mathbb{P}(A) + \mathbb{P}(B \setminus A) && (\because A, B \setminus A : \text{disjoint}) \\ &\geq \mathbb{P}(A) && (\because 0 \leq \mathbb{P}(B \setminus A) \leq 1) \end{aligned}$$

$\square$

(b) *Proof.*
For $n = 1$, the equality holds.
Assume that the inequality holds for $n \leq k$.
Let $A = \bigcup_{i=1}^{k} A_i$. Then, for $n = k + 1$,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{k+1} A_i\right) &= \mathbb{P}\left(A \cup (A_{k+1} \setminus A)\right) \\ &= \mathbb{P}\left(A\right) + \mathbb{P}(A_{k+1} \setminus A)) && (\because A, A_{k+1} \setminus A : \text{disjoint}) \\ &\leq \sum_{i=1}^{k} \mathbb{P}(A_i) + \mathbb{P}(A_{k+1} \setminus A) && (\because \text{Inductive hypothesis}) \\ &\leq \sum_{i=1}^{k} \mathbb{P}(A_i) + \mathbb{P}(A_{k+1}) && (\because A_{k+1} \setminus A \subset A_{k+1} \Rightarrow \text{by (a)}) \\ &= \sum_{i=1}^{k+1} \mathbb{P}(A_i) \end{aligned}$$

Thus, the inequality is proved by induction. $\square$

2. The conjugate prior of a binomial likelihood is a beta distribution.

$$p(\pi) = \mathrm{Beta}(\pi | a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}$$

Applying Bayes' rule with the prior, we have

$$\begin{aligned} p(\pi|x=k,m) &\propto p(x=k|m,\pi)p(\pi) \\ &\propto \pi^k(1-\pi)^{m-k}\pi^{a-1}(1-\pi)^{b-1} \\ &= \pi^{k+a-1}(1-\pi)^{m-k+b-1} \\ &\propto \mathrm{Beta}(a+k, b+m-k) \end{aligned}$$

The posterior is again a beta distribution.

3. (a) $\underbrace{0.7 \times \frac{4}{4+2}}_{\text{bag 1}} + \underbrace{0.3 \times \frac{4}{4+4}}_{\text{bag 2}} = \frac{37}{60}$

   (b) $\frac{0.3 \times \frac{4}{4+4}}{\frac{37}{60}} = \frac{9}{37}$

   (c) There are 8 cases to pick three apples.

Coin	Bag	Probability
HHH	1,1,1	$(0.7)^3 \times \left(\frac{1}{3}\right)^3 = \frac{343}{27} \times 10^{-3}$
HHT	1,1,2	$(0.7)^2(0.3)^1 \times \left(\frac{1}{3}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{49}{6} \times 10^{-3}$
HTH	1,2,2	$(0.7)^2(0.3)^1 \times \left(\frac{1}{3}\right)^1 \left(\frac{1}{2}\right)^2 = \frac{49}{4} \times 10^{-3}$
HTT	1,2,1	$(0.7)^1(0.3)^2 \times \left(\frac{1}{3}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{7}{2} \times 10^{-3}$
THH	2,2,2	$(0.7)^2(0.3)^1 \times \left(\frac{1}{2}\right)^3 = \frac{147}{8} \times 10^{-3}$
THT	2,2,1	$(0.7)^1(0.3)^2 \times \left(\frac{1}{3}\right)^1 \left(\frac{1}{2}\right)^2 = \frac{21}{4} \times 10^{-3}$
TTH	2,1,1	$(0.7)^1(0.3)^2 \times \left(\frac{1}{3}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{7}{2} \times 10^{-3}$
TTT	2,1,2	$(0.3)^3 \times \left(\frac{1}{3}\right)^1 \left(\frac{1}{2}\right)^2 = \frac{9}{4} \times 10^{-3}$

   So the probability that those apples were only picked from bag 1 is

$$\frac{\frac{343}{27}}{\frac{343}{27} + \frac{49}{6} + \frac{49}{4} + \frac{7}{2} + \frac{147}{8} + \frac{21}{4} + \frac{7}{2} + \frac{9}{4}} = \frac{2744}{14255}$$

4. (a) $\boldsymbol{b}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$ are independent.

$$p\left(\boldsymbol{b} \,\middle|\, \begin{bmatrix} x \\ y \end{bmatrix}\right) = p(\boldsymbol{b}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$$

   Since x and y are given for the posterior probability, $\boldsymbol{A}\begin{bmatrix} x \\ y \end{bmatrix}$ is fixed. Thus,

$$\begin{aligned} p(z,w|x,y) &= p\left(\boldsymbol{A}\begin{bmatrix} x \\ y \end{bmatrix} + \boldsymbol{b} \,\middle|\, \begin{bmatrix} x \\ y \end{bmatrix}\right) \\ &= \mathcal{N}\left(\boldsymbol{A}\begin{bmatrix} x \\ y \end{bmatrix}, \boldsymbol{Q}\right) \end{aligned}$$

(b) It is known that any linear/affine transformation of a Gaussian random variable also follows a Gaussian distribution.

$$p\left(\boldsymbol{A}\begin{bmatrix}x\\y\end{bmatrix}\right) = \mathcal{N}\left(\boldsymbol{A}\begin{bmatrix}\mu_x\\\mu_y\end{bmatrix}, \boldsymbol{A}\begin{bmatrix}\sigma_{xx}^2 & \sigma_{xy}^2\\\sigma_{yx}^2 & \sigma_{yy}^2\end{bmatrix}\boldsymbol{A}^\top\right)$$

Again, $\boldsymbol{b}$ and $\begin{bmatrix}x\\y\end{bmatrix}$ are independent. Therefore,

$$\begin{aligned}p(z,w) &= p\left(\boldsymbol{A}\begin{bmatrix}x\\y\end{bmatrix} + \boldsymbol{b}\right)\\&= \mathcal{N}\left(\boldsymbol{A}\begin{bmatrix}\mu_x\\\mu_y\end{bmatrix}, \boldsymbol{A}\begin{bmatrix}\sigma_{xx}^2 & \sigma_{xy}^2\\\sigma_{yx}^2 & \sigma_{yy}^2\end{bmatrix}\boldsymbol{A}^\top + \boldsymbol{Q}\right)\end{aligned}$$

5.

$$\begin{aligned}h[\mathbf{x}] &= -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx\\&= -\int_{-\infty}^{\infty} p(x)\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\right]dx\\&= \int_{-\infty}^{\infty} p(x)\left[\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]dx + \int_{-\infty}^{\infty} p(x)\left[\frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\right]dx\\&= \frac{1}{2}\mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\\&= \frac{1}{2}\mathbb{E}\left[\mathrm{tr}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\Big)\right] + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\\&= \frac{1}{2}\,\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\right]\right) + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\\&= \frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\right]\right) + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\\&= \frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right) + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\}\\&= \frac{n}{2} + \frac{1}{2}\log\{(2\pi)^n|\boldsymbol{\Sigma}|\} = \frac{1}{2}\log\{(2\pi e)^n|\Sigma|\}\end{aligned}$$

6. *Proof.*
Let $\mathbf{y} = (\mathrm{y}_1, \cdots, \mathrm{y}_\mathrm{n})$, $\mathbf{z} = (\mathrm{y}_1, \cdots, \mathrm{y}_{\mathrm{n}-1}, \mathrm{z})$ where

$$\mathrm{z} := \sum_{i=1}^{n} \mathrm{x}_\mathrm{i},\ \ \mathrm{y}_\mathrm{i} := \frac{\mathrm{x}_\mathrm{i}}{\mathrm{z}} \text{ for } i = 1, \cdots, n$$

To perform a change of variables $(\mathrm{x}_1, \cdots, \mathrm{x}_\mathrm{n}) \to (\mathrm{y}_1, \cdots, \mathrm{y}_{\mathrm{n}-1}, \mathrm{z})$, we need the transformation Jacobian $\boldsymbol{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$. Since $\mathrm{x}_i = \mathrm{y}_i\mathrm{z}$ for $i = 1, \cdots n-1$ and $\mathrm{x}_n = \mathrm{z} - \sum_{i=1}^{n-1}\mathrm{x}_\mathrm{i} = \mathrm{z}\left(1 - \sum_{i=1}^{n-1}\mathrm{y}_\mathrm{i}\right)$,

$$\boldsymbol{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \begin{bmatrix}\mathrm{z} & 0 & \cdots & \mathrm{y}_1\\0 & \mathrm{z} & \cdots & \mathrm{y}_2\\\vdots & \vdots & \ddots & \vdots\\-\mathrm{z} & -\mathrm{z} & \cdots & 1 - \sum_{i=1}^{n-1}\mathrm{y}_\mathrm{i}\end{bmatrix}$$

By adding the first $n-1$ rows to the n-th row, we can easily compute $|\boldsymbol{J}| = \mathrm{z}^{n-1}$ Then, we can compute $p(\boldsymbol{z}) = p(\boldsymbol{x}) \left|\frac{\partial \mathbf{x}}{\partial \mathbf{z}}\right|$.

$$\begin{aligned}
p(\boldsymbol{z}) &= p(\boldsymbol{x}) \left|\frac{\partial \mathbf{x}}{\partial \mathbf{z}}\right| \\
&= \left(\prod_{i=1}^{n} \frac{x_i^{\alpha_i - 1} e^{-x_i}}{\Gamma(\alpha_i)}\right) z^{n-1} \\
&= \left(\prod_{i=1}^{n} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right) z^{n-1} e^{-\sum_{i=1}^{n} x_i} \\
&= \frac{\prod_{i=1}^{n} (y_i z)^{\alpha_i - 1}}{\prod_{i=1}^{n} \Gamma(\alpha_i)} z^{n-1} e^{-z} \\
&= \frac{\prod_{i=1}^{n} y_i^{\alpha_i - 1}}{\prod_{i=1}^{n} \Gamma(\alpha_i)} z^{\sum_{i=1}^{n} \alpha_i - 1} e^{-z}
\end{aligned}$$

By using this, we can get the PDF of $\mathbf{y}$

$$\begin{aligned}
p(\boldsymbol{y}) &= p(y_1, \cdots, y_n) \\
&= p(y_1, \cdots, y_{n-1}) \\
&= \int_0^\infty p(y_1, \cdots, y_{n-1}, z)\, dz \\
&= \int_0^\infty \frac{\prod_{i=1}^{n} y_i^{\alpha_i - 1}}{\prod_{i=1}^{n} \Gamma(\alpha_i)} z^{\sum_{i=1}^{n} \alpha_i - 1} e^{-z} dz \\
&= \frac{\prod_{i=1}^{n} y_i^{\alpha_i - 1}}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \int_0^\infty z^{\sum_{i=1}^{n} \alpha_i - 1} e^{-z} dz \\
&= \frac{\Gamma\left(\sum_{i=1}^{n} \alpha_i\right)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n} y_i^{\alpha_i - 1} \\
&= \mathrm{Dir}(\boldsymbol{y} | \boldsymbol{\alpha})
\end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$.
To sum up, $\mathbf{y} \in [0,1]^n$ with $\sum_{i=1}^{n} y_i = 1$, and has PDF $p(\boldsymbol{y}) = \frac{\Gamma\left(\sum_{i=1}^{n} \alpha_i\right)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n} y_i^{\alpha_i - 1}$.

$$\therefore \mathbf{y} = \left(\frac{\mathrm{x}_1}{\sum_{i=1}^{n} \mathrm{x}_i}, \cdots, \frac{\mathrm{x}_\mathrm{n}}{\sum_{i=1}^{n} \mathrm{x}_i}\right) \sim \mathrm{Dir}\left(\boldsymbol{\alpha}\right)$$

$\square$