Coordinator
Chong Liu

Probability Theory

Solution sheet 8

Solution 8.1 Let Q_1 be the distribution of $(X_1, \ldots, X_n)$ and Q_2 the distribution of $(X_{\pi(1)}, \ldots, X_{\pi(n)})$.

(a) We claim that $Q_1 = Q_2$ over $\mathcal{B}(\mathbb{R}^n)$. To show this, by a consequence of Dynkin's lemma (see (1.3.11), p. 18 in lecture notes), it suffices to prove that

$$Q_1[A_1 \times \ldots \times A_n] = Q_2[A_1 \times \ldots \times A_n] \quad \text{with } A_j \in \mathcal{B}(\mathbb{R}).$$

By the definion of X_i,

$$\begin{aligned} P[X_1 \in A_1, \ldots, X_n \in A_n] &\overset{\text{i.i.d.}}{=} \prod_{j=1}^{n} P[X_j \in A_j] \\ &= \prod_{j=1}^{n} P[X_{\pi(j)} \in A_j] \\ &= P[X_{\pi(1)} \in A_1, \ldots, X_{\pi(n)} \in A_n]. \end{aligned}$$

Hence $Q_1 = Q_2$. Therefore,

$$E_P[g(X_1, \ldots, X_n)] = E_{Q_1}[g] = E_{Q_2}[g] = E_P[g(X_{\pi(1)}, \ldots, X_{\pi(n)})].$$

(b) Let π be a permutation such that $\pi(1) = 2,\ \pi(2) = 1,\ \pi(j) = j,\ \forall j \geq 3$. By definition of $\sigma(S)$,

$$\forall A \in \sigma(S),\ \exists B \in \mathcal{B}(\mathbb{R}) \text{ with } A = S^{-1}(B).$$

For $A \in \sigma(S)$,

$$E[X_1 \cdot 1_A] = E[X_1(1_B \circ S)] = E[X_2(1_B \circ S)] = E[X_2 \cdot 1_A],$$

where the second equality follows from part (a) with $g(x_1, \ldots, x_n) = x_1 1_B \circ \left(\sum_{j=1}^{n} x_j\right)$.

Therefore, it holds that

$$E[X_1|S] = E[X_2|S] \quad P\text{-a.s.}$$

Simlarly, $E[X_j|S] = E[X_1|S]$ for $j = 1, \ldots, n$, whence

$$S = E[S|S] = E\left[\sum_{j=1}^{n} X_j \middle| S\right] = \sum_{j=1}^{n} E\left[X_j|S\right] = nE[X_1|S] \quad P\text{-a.s.}$$

This implies that

$$E[X_1|S] = \frac{1}{n} S \quad P\text{-a.s.}$$

Solution 8.2 The total number of balls after the n-th iteration is given by $K(n) = s + w + n(t-1)$. For $n \geq 1$, let A_n be the event that the n-th ball to be drawn is black. Then the conditional probability of A_{n+1} given $Y_1, \ldots, Y_n$ equals Y_n, that is, for $n \geq 1$,

$$P[A_{n+1}|\sigma(Y_1, \ldots, Y_n)] = Y_n. \tag{1}$$

Note that we have, for $n \geq 1$,

$$Y_{n+1}(\omega) = \begin{cases} \frac{Y_n K(n) + (t-1)}{K(n+1)}, & \text{if } \omega \in A_{n+1}, \\ \frac{Y_n K(n)}{K(n+1)}, & \text{if } \omega \in A_{n+1}^c. \end{cases} \tag{2}$$

Thus we get, setting $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$,

$$\begin{aligned} E[Y_{n+1}|\mathcal{F}_n] &= E[Y_{n+1}\mathbf{1}_{A_{n+1}} + Y_{n+1}\mathbf{1}_{A_{n+1}^c}|\mathcal{F}_n] \\ &\overset{(2)}{=} E\left[\frac{Y_n K(n) + (t-1)}{K(n+1)}\mathbf{1}_{A_{n+1}} + \frac{Y_n K(n)}{K(n+1)}\mathbf{1}_{A_{n+1}^c}\middle|\mathcal{F}_n\right] \\ &= \frac{Y_n K(n) + (t-1)}{K(n+1)} P[A_{n+1}|\mathcal{F}_n] + \frac{Y_n K(n)}{K(n+1)} P[A_{n+1}^c|\mathcal{F}_n] \\ &\overset{(1)}{=} \frac{Y_n K(n) + (t-1)}{K(n+1)} Y_n + \frac{Y_n K(n)}{K(n+1)}(1 - Y_n) = Y_n, \end{aligned}$$

since $K(n) + (t-1) = K(n+1)$.

Solution 8.3
Alternative 1. Write

$$\Sigma = A^{-1} = \left(\begin{array}{c|c} \sigma_{1,1} & d^T \\ \hline d & \Sigma' \end{array}\right).$$

Note that $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,n}$ is the covariance matrix of X so that $\sigma_{i,j} = E[X_i X_j]$ for all $i, j = 1, \ldots, n$. First let us observe the following two important properties of n-dimensional normal distribution (with mean zero):
Claim 1: If $X = (X_1, \ldots, X_n)$ is an n-dim. normal random vector (with mean zero and covariance matrix Σ), then any linear combination of $X_1, \ldots, X_n$ is a real-valued normal random variable with mean zero.
Indeed, for X being an n-dim. normal random vector with mean 0 and covariance matrix Σ, we can compute that its characteristic function φ_X satisfies that for all $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$, $t \in \mathbb{R}$,

$$\varphi_X(t\xi) = \varphi_{\xi \cdot X}(t) = E[\exp\{it\sum_{j=1}^n \xi_j X_j\}] = \exp\{-\frac{1}{2}t^2 \xi^\top \Sigma \xi\}.$$

Of course this means that the linear combination $\sum_{j=1}^n \xi_j X_j$ has the normal distribution $\mathcal{N}(0, \xi^\top \Sigma \xi)$.
Warning: We have shown that if $X = (X_1, \ldots, X_n)$ is an n-dim. normal random vector, then each X_i is a real-valued normal random variable for $i = 1, \ldots, n$. However, the converse fails in general. For example, let X be a real valued random variable with standard normal istribution, Y is a random variable independent of X with distribution $P[Y = 1] = p$ and $P[Y = -1] = 1 - p$ for some $p \in (0, 1)$. Then one can check that $Z = XY$ also has standard normal distribution but the random vector (X, Z) does not have 2-dim. normal distribution.
Claim 2: If Z is an n-dim. normal random vector (with mean zero and covariance matrix Σ) and Z_1 is orthogonal to each $Z_2, \ldots, Z_n$ in the L^2-sense, i.e., $E[Z_1 Z_j] = 0$ holds for all $j = 2, \ldots, n$, then Z_1 and $(Z_2, \ldots, Z_n)$ are independent.
Indeed, in this case, the covariance matrix Σ of X has the following form:

$$\Sigma = A^{-1} = \left(\begin{array}{c|c} \sigma_{1,1} & 0 \\ \hline 0 & \Sigma' \end{array}\right),$$

where 0 is the zero vector in $\mathbb{R}^{n-1}$ and Σ' is obviously the covariance matrix of the $(n-1)$-dim. normal random vector $Z' := (Z_2, \dots, Z_n)$. Then from the special form of characteristic function of multi-dimensional normal distribution given above, we can easily check that for all $a \in \mathbb{R}$, $b = (b_1, \dots, b_{n-1}) \in \mathbb{R}^{n-1}$,

$$\varphi_{(Z_1,Z')}(a,b) = \exp\{-\frac{1}{2}(a,b)^\top \Sigma(a,b)\} = \exp\{-\frac{1}{2}a^2\sigma_{1,1}\}\exp\{-\frac{1}{2}b^\top \Sigma' b\} = \varphi_{Z_1}(a)\varphi_{Z'}(b).$$

Then by using the result obtained from *Exercise 6.2* we can conclude that Z_1 and Z' are independent.

Thanks to the above two observations, we may use a geometric approach to compute $E[X_1 \mid \mathcal{F}] = E[X_1 \mid Y]$ with $Y = (X_2, \dots, X_n)$. The strategy can be described as follows: we consider $X_1, \dots, X_n$ as elements in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ with inner product $\langle U, V\rangle = E[UV]$ for $U, V \in L^2(\Omega, \mathcal{A}, P)$. Let the *orthogonal projection* of X_1 onto the linear subspace spanned by $X_2, \dots, X_n$ in $L^2(\Omega, \mathcal{A}, P)$ be denoted by $\pi(X_1)$. It can be written as a linear combination of $X_2, \dots, X_n$ such that $\pi(X_1) = \sum_{k=2}^n \alpha_k X_k$ and satisfies that $\langle X_1 - \pi(X_1), X_k\rangle = 0$ for all $k = 2, \dots, n$. So, if we can show that the random vector $Z = (Z_1, \dots, Z_n)$ with $Z_1 = X_1 - \pi(X_1)$ and $Z_k = X_k$ for $k \geq 2$ also has an n-dim. normal distribution, then by Claim 2, we know that Z_1 and $Y = (Z_2, \dots, Z_n) = (X_2, \dots, X_n)$ are independent. Moreover, by Claim 1, we know that $X_1 - \pi(X_1)$ is a normal random variable with mean zero. As a consequence, we have:

$$\begin{aligned} E[X_1 \mid Y] &= E[X_1 - \pi(X_1) \mid Y] + E[\pi(X_1) \mid Y] \\ &= E[X_1 - \pi(X_1)] + \pi(X_1) \\ &= \pi(X_1) = \sum_{k=2}^n \alpha_k X_k, \end{aligned}$$

which completes our proof.

Hence, we only need to determine the coefficients α_k, $k = 2, \dots, n$ for the orthogonal projection $\pi(X_1)$ and then show that $Z = (Z_1, \dots, Z_n)$ with $Z_1 = X_1 - \pi(X_1)$ and $Z_k = X_k$ for $k \geq 2$ also has an n-dim. normal distribution. The first task is easy: from basic linear algebra we see that α_k, $k = 2, \dots, n$ must satisfy the system of equations:

$$\langle X_1, X_j\rangle = \sum_{k=2}^n \alpha_k \langle X_j, X_k\rangle$$

for all $j = 2, \dots, n$. Since $\langle X_i, X_j\rangle = E[X_i X_j]$ and the latter is equal to $\sigma_{i,j}$ of the covariance matrix Σ of X, the above system of equations is equivalent to the matrix formulation:

$$\Sigma'(\alpha_k)_{k=2,\dots,n} = d,$$

where d and Σ' come from the decomposition of Σ, see the beginning of the solution. Hence, we have $(\alpha_k)_{k=2,\dots,n} = (\Sigma')^{-1}d$.

Finally, we can compute the characteristic function of Z and realize that Z has an n-dim. normal distribution with mean zero and covariance matrix S for $S = Q\Sigma Q^\top$, where

$$Q = \left(\begin{array}{c|c} 1 & -d^\top(\Sigma')^{-1} \\ \hline 0 & I_{n-1} \end{array}\right),$$

where I_{n-1} is the identity matrix of $\mathbb{R}^{(n-1)\times(n-1)}$. This finishes our first geometric approach.

Alternative 2. For notational convenience, we set

- $X = (X_1, X_2, \dots, X_n) = (X_1, Y)$,

- $A = \left(\begin{array}{c|c} a_{11} & b^T \\ \hline b & B \end{array}\right)$,
- $c := \sqrt{\det A}/(2\pi)^{n/2}$.

Now let us consider an arbitrary $D \in \mathcal{F} := \sigma(Y)$: it has to be of the form $D = Y^{-1}(G)$ for some $G \in \mathcal{B}(\mathbb{R}^{n-1})$. We can then write

$$E[X_1 1_D] = c \int_G \left(\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} x_1^2 - b^T y x_1 - \frac{1}{2} y^T B y \right\} x_1 \, \lambda(\mathrm{d}x_1) \right) \lambda^{n-1}(\mathrm{d}y).$$

By completing the square, we obtain

$$a_{11} \left(x_1 + \frac{b^T y}{a_{11}} \right)^2 = a_{11} x_1^2 + a_{11} 2 x_1 \frac{b^T y}{a_{11}} + a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 = a_{11} x_1^2 + 2 b^T y x_1 + a_{11} \left(\frac{b^T y}{a_{11}} \right)^2,$$

which implies

$$-\frac{1}{2} a_{11} x_1^2 - b^T y x_1 = -\frac{1}{2} a_{11} \left(x_1 + \frac{b^T y}{a_{11}} \right)^2 + \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2.$$

This yields in turn

$$E[X_1 1_D] = c \int_G \left(\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} \left(x_1 + \frac{b^T y}{a_{11}} \right)^2 + \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 - \frac{1}{2} y^T B y \right\} x_1 \, \lambda(\mathrm{d}x_1) \right) \lambda^{n-1}(\mathrm{d}y).$$

Using the substitution $z = x_1 + \dfrac{b^T y}{a_{11}}$, for which $\lambda(\mathrm{d}z) = \lambda(\mathrm{d}x_1)$, we get:

$$E[X_1 1_D] = c \int_G \underbrace{\left(\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} z^2 + \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 - \frac{1}{2} y^T B y \right\} \left(z - \frac{b^T y}{a_{11}} \right) \lambda(\mathrm{d}z) \right)}_{=:I} \lambda^{n-1}(\mathrm{d}y),$$

where we have

$$\begin{aligned} I = \exp\left\{ \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 - \frac{1}{2} y^T B y \right\} \int_{\mathbb{R}} z \exp\left\{ -\frac{1}{2} a_{11} z^2 \right\} \lambda(\mathrm{d}z) \\ - \frac{b^T y}{a_{11}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} z^2 + \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 - \frac{1}{2} y^T B y \right\} \lambda(\mathrm{d}z). \end{aligned}$$

On the other hand, we have

$$\int_{\mathbb{R}} z \exp\left\{ -\frac{1}{2} a_{11} z^2 \right\} \lambda(\mathrm{d}z) = 0, \text{ by symmetry.}$$

Consequently, we get, using $x_1 = z - \frac{b^T y}{a_{11}}$,

$$I = -\frac{b^T y}{a_{11}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} z^2 + \frac{1}{2} a_{11} \left(\frac{b^T y}{a_{11}} \right)^2 - \frac{1}{2} y^T B y \right\} \lambda(\mathrm{d}z)$$

$$= -\frac{b^T y}{a_{11}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} a_{11} x_1^2 - b^T y x_1 - \frac{1}{2} y^T B y \right\} \lambda(\mathrm{d}x_1)$$

$$= -\frac{b^T y}{a_{11}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} x^T A x \right\} \lambda(\mathrm{d}x_1),$$

where $x = (x_1, y) \in \mathbb{R}^n$, and thus

$$E[X_1 1_D] = c \int_G \left(-\frac{b^T y}{a_{11}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} x^T A x \right\} \lambda(\mathrm{d}x_1) \right) \lambda^{n-1}(\mathrm{d}y) = E\left[\left(-\frac{b^T Y}{a_{11}} \right) 1_D \right],$$

since we have for the density f_Y of Y:

$$f_Y(y) = c \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2} x^T A x \right\} \lambda(\mathrm{d}x_1)$$

with $x = (x_1, y) \in \mathbb{R}^n$. Finally, this implies, P-almost surely,

$$E\left[X_1 | \mathcal{F} \right] = -\frac{b^T Y}{a_{11}} = -\frac{1}{a_{11}} \sum_{j=2}^{n} a_{1j} X_j,$$

since the r.v. on the right-hand side is clearly $\sigma(Y)$-measurable and integrable.