1 Exercises

Probability, distribution theory and expectation

Exercise 1

Proof. We use the 3 probability axioms of definition 1.1.

1) Write \( A \cup B = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B) \Rightarrow P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B) \)

(from axiom 3 and since the events are mutually exclusive). Now, \( P(A) = P(A \cap B^c) + P(A \cap B) \Rightarrow P(A \cap B^c) = P(A) - P(A \cap B) \). Similarly, \( P(B) = P(B \cap A^c) + P(A \cap B) \Rightarrow P(B \cap A^c) = P(B) - P(A \cap B) \). Plugging in, you get \( P(A \cup B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B) \).

2) Consider the partition \( \Omega = A \cup A^c \), so that the 3rd axiom yields \( P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \), and from the 1rst axiom \( P(\Omega) = 1 \). Thus \( P(A) + P(A^c) = 1 \).

3) Let \( A, B \) be sets such that \( A \subseteq B \). Then either \( A = B \) or \( A \subset B \). If \( A = B \), then \( P(A) = P(B) \). If \( A \subset B \), then \( A \cap B = A \), and therefore (see the proof of part 1) \( 0 \leq P(B \cap A^c) = P(B) - P(A \cap B) \Rightarrow P(A) \leq P(B) \).

4) Let \( \emptyset \) be the empty set, with \( \Omega^c = \emptyset \). From part 2 and using axiom 1, we have \( P(\Omega) + P(\Omega^c) = 1 \Rightarrow P(\emptyset) = 0 \). ■

Exercise 2 (Total probability)

Proof. Let \( \{B_i\}_{i=1}^n \) be any partition of \( \Omega \). For any event \( A \in 2^\Omega \), we can write \( P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) = \sum_i P(A \cap B_i) \), using axiom 3. Multiplying and dividing by \( P(B_i) \), we get

\[
P(A) = \sum_i P(A \cap B_i) = \sum_i P(A \cap B_i) \frac{P(B_i)}{P(B_i)} = \sum_i \frac{P(A \cap B_i)}{P(B_i)} P(B_i),
\]

and by definition of conditional probability \( \frac{P(A \cap B_i)}{P(B_i)} = P(A|B_i) \), we have the result. ■

Exercise 3

Proof. Let \( A, B \in 2^\Omega \) be events with \( P(B) > 0 \). Then \( P(A|B) = \frac{P(A \cap B)}{P(B)} \) by definition of conditional probability. Multiplying and dividing by \( P(A) \), we obtain

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} \frac{P(A)}{P(A)} = \frac{P(B \cap A)}{P(B)} \frac{P(A)}{P(B)} = \frac{P(B|A) P(A)}{P(B)},
\]

since \( P(A \cap B) = P(B \cap A) \). ■

Exercise 4

Proof. (\( \Rightarrow \)) When \( F \) is a cdf, we have by definition that \( F(x) = P(X \leq x) \). Note that for an increasing sequence of events \( A_1 \subseteq A_2 \subseteq \ldots \), we have \( \lim_{n \to +\infty} A_n = \bigcup_n A_n \). Now letting \( B_1 = A_1, B_2 = A_2 \setminus A_1, B_2 = A_3 \setminus (A_1 \cup A_2), \) and so forth, we have that \( \{B_n\} \) is a disjoint sequence of events with \( \bigcup_n A_n = \bigcup_n B_n \), and \( A_n = \bigcup_{k=1}^n B_k \). As a result, using definition 1.1, (iii), we have

\[
P(\lim_{n \to +\infty} A_n) = P\left(\bigcup_n A_n\right) = P\left(\bigcup_{n} B_n\right) = \sum_{n=1}^{+\infty} P(B_n) = \lim_{n \to +\infty} \sum_{k=1}^{n} P(B_k) = \lim_{n \to +\infty} P\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n \to +\infty} P(A_n),
\]
so that
\[ P\left( \lim_{n \to +\infty} A_n \right) = \lim_{n \to +\infty} P(A_n). \]  
(1)

Similarly, we can show that for decreasing sequences of events, (1) still holds (i.e., consider the complements of the decreasing sequence, which yields an increasing sequence of events). This result is known as continuity of probability measure and will be rigorously established in chapter 4.

Now letting \( A_n = \{ X \leq -n \} \) and \( A = \emptyset \), we have that \( \{ A_n \} \) is a decreasing sequence with
\[ \lim_{n \to +\infty} A_n = \bigcap_{n=1}^{+\infty} A_n = A \text{ so that } \lim_{x \to -\infty} P(X \leq x) = \lim_{n \to +\infty} P(X \leq -n) = P(\emptyset) = 0. \]  
Similarly, we need to show that there is a random variable \( X \) on \( \Omega \) (a rigorous definition of the uniform distribution is given in chapters 3 and 4). We need to show that
\[ \lim_{n \to +\infty} A_n = \{ X \leq n \}, \]

\[ \text{an increasing sequence of events, with } \lim_{n \to +\infty} A_n = \bigcup_{n=1}^{+\infty} A_n = \mathcal{R}, \]  
so that the continuity of probability for \( X_n = x + \varepsilon \), yields
\[ F(x) = P(X \leq x) = \lim_{n \to +\infty} P(X \leq x + \frac{1}{n}) = \lim_{n \to +\infty} P(X \leq x + \varepsilon) = \lim_{\varepsilon \to 0} F(x + \varepsilon), \]

where \( \varepsilon = \frac{1}{n} \) for all \( x. \)

Now assume that (i)-(iii) hold. We need to show that there is a random variable \( X \), a probability (measure) \( P \) and a sample space \( \Omega \), such that \( F(x) = P(X \leq x) \) is the cdf of \( X \). Let \( \Omega = [0,1] \), define \( X(\omega) = \inf\{x : F(x) \geq \omega\} \), \( 0 < \omega < 1 \), and assume that \( P \) is the uniform probability in \( \Omega \), i.e., probability is simply the length of an interval (a rigorous definition of the uniform distribution is given in chapters 3 and 4). We need to show that \( F(x) = P(\{ \omega : X(\omega) \leq x\}) \). Since \( X \) is increasing on \( \Omega \), the event \( A = \{ \omega : X(\omega) \leq x\} \) is an interval with endpoints 0 and sup\( A \), and therefore, its probability is \( P(A) = \sup A \). Thus, we need to show that \( F(x) = \sup A \). The definition of \( X \) and the right continuity of \( F \) imply \( F(X(\omega)) \geq \omega \), so that if \( \omega \in A \), then \( F(x) \geq F(X(\omega)) \geq \omega \). Therefore, \( F(x) \) is an upper bound of \( A \), with \( F(x) \in A \), since \( X(F(x)) \leq x \). Thus, we must have \( F(x) = \sup A \).

Exercise 5

Proof. Let \( Z = \frac{X_1}{Y} = \frac{X_1}{X_1 + X_2} \). Solving for \( X_1, X_2 \), we get \( X_1 = ZY \) and \( X_2 = Y - X_1 \Rightarrow X_2 = Y - ZY = (1 - Z)Y \). The Jacobian matrix is given by \( J = \begin{bmatrix} 1 & -1 \\ z & y \end{bmatrix} \), so that the absolute value of the determinant of the Jacobian is \( |J| = y \). Now the joint of \( X_1 \) and \( X_2 \) is \( f_{X_1, X_2}(x, y) = \theta e^{-x\theta} e^{-y\theta} = \theta^2 e^{-\theta(x + zy)} \) and so that the pdf of \( Y \) and \( Z \) is of the form \( f_{Y,Z}(y, z) = f_{X_1, X_2}(zy, y - zy) |J| = y \theta^2 e^{-\theta y} \). Note that the original space \( S = \{(x_1, x_2) : x_1, x_2 > 0\} \) is now transformed into
\[ T = \{(y, z) : zy > 0, (1 - z)y > 0\} = \{(y, z) : zy > 0, 0 < y < 1\}. \]

As a result we can write
\[ f_{Y,Z}(y, z) = f_Y(y) f_Z(z), \]

where \( Y \sim Gamma(2, \theta) \) and \( Z \sim Uniform(0, 1) \), and they are independent.

Exercise 6
First note that for any positive numbers \( a \) and \( b \), the discrete uniform distribution.

\[
m(y) = \frac{1}{n} f_{Y,U}(y,u)du = \frac{1}{n} P(Y = y|U = u)f_U(u)du
\]

\[
= \frac{1}{n} C^n_y u^n(1-u)^{n-y}du = C^n_y \frac{1}{n} w^y+1(1-u)^{n-y+1-1}du
\]

\[
= C^n_y Bc(y+1,n-y+1) = C^n_y \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(y+1+n-y+1)}
\]

\[
= C^n_y y!(n-y)! = \frac{n!}{y!(n-y)!} \frac{1}{(n+1)!}.
\]


Proof. We have

\[
m(y) = \frac{1}{0} f_{Y,U}(y,u)du = \frac{1}{0} P(Y = y|U = u)f_U(u)du
\]

\[
= \frac{1}{0} C^n_y u^n(1-u)^{n-y}du = C^n_y \frac{1}{0} w^y+1(1-u)^{n-y+1-1}du
\]

\[
= C^n_y Bc(y+1,n-y+1) = C^n_y \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(y+1+n-y+1)}
\]

\[
= C^n_y y!(n-y)! = \frac{n!}{y!(n-y)!} \frac{1}{(n+1)!}.
\]

the discrete uniform distribution. ■

Exercise 7

Proof. Starting with the left inequality, let \( A \) be the event that \( X \leq x \) and \( B \) be the event that \( Y \leq y \). We know that

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B) \Rightarrow
\]

\[
P(A \cap B) = P(A) + P(B) - P(A \cup B),
\]

where \( 0 \leq P(A \cup B) \leq 1 \). Then

\[
P(X \leq x, Y \leq y) = P(X \leq x) + P(Y \leq y) - P(A \cup B) \Rightarrow
\]

\[
P(X \leq x, Y \leq y) \geq P(X \leq x) + P(Y \leq y) - 1,
\]

since we are subtracting the maximum value of \( P(A \cup B) \). Thus, \( F_{X,Y}(x,y) \geq F_X(x) + F_Y(y) - 1 \).

For the inequality on the right, first note that

\[
F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = E[I(X \leq x)I(Y \leq y)],
\]

where \( I(.) \) represents the indicator function. An easy way to show the upper bound is to use the Cauchy-Schwarz Inequality, i.e., \( [E(XY)]^2 \leq E(X^2)E(Y^2) \). Then we have

\[
[E(I(X \leq x)I(Y \leq y))]^2 \leq E[I(X \leq x)]E[I(Y \leq y)]^2 \Rightarrow
\]

\[
E[I(X \leq x)I(Y \leq y)]^2 \leq E[I(X \leq x)]E[I(Y \leq y)]
\]

since the square of an indicator function is the indicator function. Thus, \( E[I(X \leq x)I(Y \leq y)] \leq \sqrt{E[I(X \leq x)]E[I(Y \leq y)]} \) which reduces to \( F_{X,Y}(x,y) \leq \sqrt{F_X(x)F_Y(y)} \).

Proof of Hölder’s and Cauchy-Schwarz Inequalities:

First note that for any positive numbers \( p \) and \( q \) with \( 1/p + 1/q = 1 \), then for any positive real numbers \( a \) and \( b \) we have

\[
\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab,
\]

(2)

(take the function \( g(a) = \frac{1}{p}a^p + \frac{1}{q}b^q - ab \), with \( \frac{d}{da}g(a) = a^{p-1} - b \), so that it’s minimum is attained at \( a^* = b^{1/p-1} \), with \( g(a^*) = 0 \). Now letting \( a = \frac{|X|}{|E[X^p]|^{1/p}} \) and \( b = \frac{|Y|}{|E[Y^p]|^{1/p}} \) for any random variables \( X \) and \( Y \), applying the inequality (2) and integrating on both sides leads to

\[
E|XY| \leq \frac{|X|}{|E[X^p]|^{1/p}} |E[Y^q]|^{1/q}.
\]

(3)

Taking \( p = q = 2 \) above leads to the Cauchy-Schwarz inequality

\[
[E(XY)]^2 \leq E(X^2)E(Y^2).
\]

(4)
Exercise 8

Proof.
(i) We have
\[ F_{X_1}(x) = P(X_1 \leq x) = 1 - P(\min_i X_i > x) = 1 - P(X_i > x, \forall i) \]
\[ = 1 - P(X_i > x, \forall i) = 1 - \prod_{i=1}^n P(X_i > x) = 1 - \prod_{i=1}^n [1 - P(X_i \leq x)] \]
\[ = 1 - [1 - F(x)]^n, \]
so that taking derivative with respect to \( x \) yields
\[ f_{X_1}(x) = n[1 - F(x)]^{n-1} f(x). \]

(ii) Similarly, we have
\[ F_{X_n}(x) = P(X_n \leq x) = P(\max_i X_i \leq x) = P(X_i \leq x, \forall i) = \prod_{i=1}^n P(X_i \leq x) \]
\[ = [F(x)]^n, \]
so that taking derivative with respect to \( x \) leads to
\[ f_{X_n}(x) = n[F(x)]^{n-1} f(x). \]

Exercise 9

Proof. Let \( f(x, y) = f(x)f(y) \) be the joint distribution of the random variables \( X, Y \). We have
\[ P(X < Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f(x, y)dx\,dy = \int_{-\infty}^{y} \left[ \int_{-\infty}^{+\infty} f(x)dx \right] f(y)dy \]
\[ = \int_{-\infty}^{+\infty} F(y)dF(y) = \left[ \frac{F(y)^2}{2} \right]_{-\infty}^{+\infty} = \frac{1}{2}. \]

Exercise 10

Proof. First note that if \( W \sim f(w) \), then for \( |t| < 1 \), we have
\[ m_W(t) = E(e^{tW}) = \frac{1}{2} \int_\mathbb{R} e^{t|w|}\,dw = \frac{1}{2} \int_{-\infty}^{0} e^{tw}\,dw + \frac{1}{2} \int_{0}^{+\infty} e^{tw}\,dw \]
\[ = \frac{1}{2} \int_{-\infty}^{0} e^{(t+1)w}\,dw + \frac{1}{2} \int_{0}^{+\infty} e^{(t-1)w}\,dw = \frac{1}{2} \left[ \frac{1}{t+1} e^{(t+1)w} \right]_{-\infty}^{0} + \frac{1}{2} \left[ \frac{1}{t-1} e^{(t-1)w} \right]_{0}^{+\infty} \]
\[ = 1 + \frac{1}{2} + \frac{1}{2(t+1)} + \frac{1}{2(t-1)} = \frac{1}{1-t^2}. \]

Now consider the identity
\[ (X_1 + X_2)^2 - (X_1 - X_2)^2 = 4X_1X_2 \Rightarrow \]
\[ \left( \frac{X_1 + X_2}{\sqrt{2}} \right)^2 - \left( \frac{X_1 - X_2}{\sqrt{2}} \right)^2 = 2X_1X_2, \]
with
\[ X_1 \pm X_2 \sim N(0, 2) \Rightarrow \frac{X_1 \pm X_2}{\sqrt{2}} \sim N(0, 1) \Rightarrow \left( \frac{X_1 \pm X_2}{\sqrt{2}} \right)^2 \sim \chi^2_1, \]
where
\[
\text{Cov}(X_1 - X_2, X_1 + X_2) = \text{Cov}(X_1, X_1) + \text{Cov}(X_1, X_2) - \text{Cov}(X_2, X_1) - \text{Cov}(X_2, X_2) = 1 + 0 - 0 - 1 = 0,
\]
and therefore \( X_1 - X_2 \) independent of \( X_1 + X_2 \), since they are both normally distributed. Letting \( Z_1 = \left( \frac{X_1 - X_2}{\sqrt{2}} \right)^2, Z_2 = \left( \frac{X_1 + X_2}{\sqrt{2}} \right)^2 \), we have that \( Z_1 \) and \( Z_2 \) are independent \( \chi^2_1 \) random variables, so that
\[
m_{X_1X_2}(t) = m_{\frac{1}{2}(Z_1-Z_2)}(t) = E(\exp(\frac{1}{\theta}Z_1)\exp(-\frac{1}{\theta}Z_2)) = \left( \frac{1}{1-\frac{1}{\theta}} \right) \left( \frac{1}{1+\frac{1}{\theta}} \right) \left( \frac{1}{1-t^2} \right),
\]
and since \( X_1X_2 \) independent of \( X_3X_4 \), we can write
\[
m_Y(t) = E(e^{tY}) = E(e^{t(X_1X_2-\chi^2)}) = E(e^{tX_1X_2})E(e^{(-t)\chi^2}) = \left( \frac{1}{1-t^2} \right) \left( \frac{1}{1-(-t)^2} \right) = \frac{1}{1-t^2},
\]
which is the mgf of a Laplace distribution.

**Exercise 11**

**Proof.** We have
\[
m_W(t) = E(e^{tW}) = E(e^{t(X_1X_2)}) = E(e^{tX_1})E(e^{(-t)X_2}) = \frac{1}{1-t/\theta} \cdot \frac{1}{1+t/\theta} = \frac{1}{1-(\frac{t}{\theta})^2},
\]
which is the mgf of a Laplace distribution with scale parameter \( \theta \).

**Exercise 12**

**Proof.** First note that in order for \( f(x|\theta_1, \theta_2) \) to be a proper cdf we must have
\[
\int_{\theta_1}^{\theta_2} \frac{1}{h(\theta_1, \theta_2)} dx = 1 \Rightarrow h(\theta_1, \theta_2) = \theta_2 - \theta_1.
\]
As a result, the cdf of \( X \) is \( F(x|\theta_1, \theta_2) = \frac{x-\theta_1}{\theta_2-\theta_1}, \) and since \( Q = \frac{X_{(n)} - X_{(1)}}{h(\theta_1, \theta_2)} = \frac{X_{(n)} - X_{(1)}}{\theta_2 - \theta_1} \), we can write
\[
Q = F(X_{(n)}|\theta_1, \theta_2) - F(X_{(1)}|\theta_1, \theta_2) = Y_{(n)} - Y_{(1)},
\]
so that \( Q = Y_{(n)} - Y_{(1)} \) is the range of a sample from a \( \text{Unif}(0, 1) \) rv, since \( Y_i = F(X_i|\theta_1, \theta_2) \sim \text{Unif}(0, 1) \), and \( F \) is an increasing function. Now the joint distribution of \( Y_{(1)} \) and \( Y_{(n)} \) is given by
\[
f_{Y_{(1)}, Y_{(n)}}(y_1, y_n) = n(n-1) (y_n - y_1)^{n-2}, \quad 0 < y_1 < y_n < 1.
\]
Let \( Y = Y_{(1)} \), so that \( Y_{(1)} = Y \), and \( Y_{(n)} = Q + Y \), and the Jacobian of the transformation \((Y_{(1)}, Y_{(n)}) \rightarrow (Y, Q)\) becomes \( J = \begin{vmatrix} \frac{\partial(Y_{(1)}, Y_{(n)})}{\partial(Y, Q)} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \), so that
\[
(Y, Q) \in \{ (y, q) : 0 \leq y \leq 1, 0 \leq q + y \leq 1 \}
= \{ (y, q) : 0 \leq q \leq 1, 0 \leq y \leq 1 - q \}.
\]
and the distribution of \((Y, Q)\) becomes
\[
f(y, q) = n(n - 1)q^{n-2}, \quad 0 \leq q \leq 1, 0 \leq y \leq 1 - q.
\]
Now the distribution of the range \(Q\) is given by
\[
f_Q(q) = \int_0^{1-q} n(n - 1)q^{n-2} dy = n(n - 1)q^{n-2}(1 - q),
\]
and therefore, \(Q \sim \text{Beta}(n - 1, 2)\) since
\[
\frac{1}{\beta(n - 1, 2)} = \frac{\Gamma(n+1)}{\Gamma(n-1)\Gamma(2)} = \frac{n!}{(n-2)!} = n(n - 1).
\]

Exercise 13

**Proof.** In order for the event \(\{u < X_{(r)} < u + du, v < X_{(s)} < v + dv\}\) to be realized, it is required that \(r\) observations are less than \(u\), one is in \((u, u + du)\), \(s - r - 1\) are in \((u + du, v)\), one is in \(v < X_{(s)} < v + dv\), and the remaining \(n - s\) are above \(v + dv\). Since the data are iid, the corresponding probabilities of these events are \(F(u)^{r-1}, f(y), [F(v) - F(u)]^{s-r-1}, f(v),\) and \([1 - F(v)]^{n-s}\). Therefore, we have a multinomial probability describing the distribution of \(X_{(r)}\) and \(X_{(s)}\), which leads to equation
\[
f_{X_{(r)},X_{(s)}}(u, v) = C^n_{r-1,s-r-1,n-s}f(u)f(v)[F(u)]^{r-1}[F(v) - F(u)]^{s-r-1}[1 - F(v)]^{n-s}.
\]

Exercise 14

**Proof.** We must have
\[
\int_{\theta_1}^{\theta_2} \frac{g(x)}{h(\theta_1, \theta_2)} dx = 1 \Rightarrow \int_{\theta_1}^{\theta_2} g(x) dx = h(\theta_1, \theta_2),
\]
and as a result, the cdf of \(X\) is
\[
F(x|\theta_1, \theta_2) = \int_{\theta_1}^{x} \frac{g(x)}{h(\theta_1, \theta_2)} dx = \frac{h(\theta_1, x)}{h(\theta_1, \theta_2)},
\]
and since
\[
Q = \frac{h(X_{(r)}, X_{(s)})}{h(\theta_1, \theta_2)} = -\frac{h(\theta_1, X_{(r)}) + h(\theta_1, X_{(s)})}{h(\theta_1, \theta_2)},
\]
we can write
\[
Q_{rs} = F(X_{(s)}|\theta_1, \theta_2) - F(X_{(r)}|\theta_1, \theta_2) = Y_{(s)} - Y_{(r)},
\]
so that \(Q_{rs} = Y_{(s)} - Y_{(s)}\) is based on order statistics from a sample from a \(\text{Unif}(0, 1)\) rv, since \(Y_i = F(X_i|\theta_1, \theta_2) \sim \text{Unif}(0, 1)\), and \(F\) is an increasing function. Now the joint distribution of \(Y_{(r)}\) and \(Y_{(s)}\) is given by
\[
f_{Y_{(r)},Y_{(s)}}(u, v) = C^n_{r-1,s-r-1,n-s}u^{r-1}(v - u)^{s-r-1}(1 - v)^{n-s},
\]
\(u < v, 1 \leq r < s \leq n\).

Now we perform a transformation \((U = Y_{(r)}, V = Y_{(s)})) \rightarrow \((Y, Q_{rs})\), where \(Q_{rs} = V - U\) and \(Y = U\). Then \(V = Q_{rs} + Y\) and \(U = Y\), so that the Jacobian is \(|J| = 1\), and so that
\[
(Y, Q_{rs}) \in \{(y, q) : 0 \leq q \leq 1, 0 \leq q + y \leq 1\} = \{(y, q) : 0 \leq q \leq 1, 0 \leq y \leq 1 - q\}.
\]
Then, the distribution of \((Y, Q_{rs})\) becomes
\[
f(y, q) = C^n_{r-1,s-r-1,n-s}y^{r-1}q^{s-r-1}(1 - q - y)^{n-s}, \quad 0 \leq q \leq 1, 0 \leq y \leq 1 - q,
\]
and we need to integrate over $Y$ to get the marginal distribution of $Q$. We have

$$f_Q(q) = C^n_{r-1,s-r-1,n-s} q^{s-r-1} (1-q)^{n-s-1} \int_0^y y^{r-1} \left(1 - \frac{y}{1-q}\right)^{n-s} dy,$$

so that letting $w = \frac{y}{1-q} \Rightarrow y = (1-q)w$, we can write

$$f_Q(q) = C^n_{r-1,s-r-1,n-s} q^{s-r-1} (1-q)^{n-s} (1-q)^{r-1}(1-q) \int_0 w^{r-1} (1-w)^{n-s} dw.$$

Since

$$\int_0^y w^{r-1} (1-w)^{n-s} dw = \int_0^w (1-w)^{n-s-1} dw = Be(r, n-s+1) = \frac{(r-1)!(n-s)!}{(n-s-r)!},$$

we have

$$f_Q(q) = \frac{1}{Be(s-r, n-s+r+1)} q^{s-r-1} (1-q)^{n-s+r+1-1},$$

$$0 \leq q \leq 1,$$

where

$$C^n_{r-1,s-r-1,n-s} Be(r, n-s+1) = \frac{n!}{(r-1)!(s-r-1)!(n-s)! (n-s+r+1-1)!}$$

$$= \frac{n!}{(s-r-1)!(n-s+r)!} = Be(s-r, n-s+r+1).$$

Thus, $Q_{rs} \sim Beta(s-r, n-s+r+1).$ ■

Exercise 15

**Proof.** We have

$$E \left[ \frac{X_1 + \cdots + X_n}{X_1 + \cdots + X_n} \right] = 1 \Rightarrow E \left[ \frac{X_1}{X_1 + \cdots + X_n} + \frac{X_2}{X_1 + \cdots + X_n} + \cdots + \frac{X_n}{X_1 + \cdots + X_n} \right] = 1,$$

and since the $X_i's$ are iid

$$E \left[ \frac{X_1}{X_1 + \cdots + X_n} \right] + E \left[ \frac{X_2}{X_1 + \cdots + X_n} \right] + \cdots + E \left[ \frac{X_n}{X_1 + \cdots + X_n} \right] = nE \left[ \frac{X_1}{X_1 + \cdots + X_n} \right],$$

which gives the result. ■

Exercise 16

**Proof.** First note that $P(X > a) = P(e^{tX} > e^{ta})$. By Markov’s inequality, $P(e^{tX} > e^{ta}) \leq \frac{e^{tx}}{e^{ta}} = m_{X-a}(t)$, for any $t > 0$, so that

$$P(e^{tX} > e^{ta}) \leq \inf_{t > 0} \{m_{X-a}(t)\}.$$ ■

Exercise 17

**Proof.** For $X \sim \mathcal{N}(0,1)$, the mgf is $m_X(t) = e^{\frac{t^2}{2}}$, so that

$$m_{X-a}(t) = e^{-at} e^{\frac{t^2}{2}} = e^{\frac{t^2-2at}{2}} = e^{\frac{(t-a)^2}{2}} e^{-\frac{t^2}{2}}.$$

which is minimized at $t = a$. Using the previous exercise, we have

$$P(X > a) \leq \inf \{m_{X-a}(t), t > 0\} = \inf \{e^{\frac{(t-a)^2}{2}} e^{-\frac{t^2}{2}}, t > 0\} = e^{-\frac{a^2}{2}}.$$ ■
Exercise 18

Proof.  
(i) Since $|\Sigma| = 4 - 4 = 0$, the rank is 1.
(ii) We have $X \sim N(0, \Sigma)$. Assume that $Y \sim N(\mu, \sigma^2)$, so that $X = aY = \begin{bmatrix} a_1 Y \\ a_2 Y \end{bmatrix}$, with mean $\mu_X = E(X) = \begin{bmatrix} a_1 \mu \\ a_2 \mu \end{bmatrix} = 0$, so that we must have $\mu = 0$. Moreover,

$$
\Sigma = \text{Cov}(X) = E((aY)^T aY) = Var(Y^2)a^T = \sigma^2 a a^T
$$

so that $\sigma^2 a_1^2 = 4$, $\sigma^2 a_1 a_2 = 2$, and $\sigma^2 a_2^2 = 1$. Thus, $a_1 = \frac{2}{\sigma}$, $a_2 = \frac{1}{\sigma}$ and so that $a = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and $Y \sim N(0, \sigma^2)$, for any $\sigma > 0$. ■

Exercise 19

Proof. Let $c = \sum_{j=1}^{p} z_j^2$, and note that $X_i \sim N(\beta + \gamma z_i, \sigma^2)$, with

$$
\bar{X} = \frac{1}{p} \sum_{i=1}^{p} X_i \sim N(\beta + \gamma \sum_{i=1}^{p} z_i, \frac{1}{p} \sigma^2) = N(\beta, \frac{1}{p} \sigma^2),
$$

since $\sum_{i=1}^{p} z_i = 0$, and

$$
U = \sum_{i=1}^{p} \frac{z_i}{c} X_i \sim N(\frac{1}{c} \sum_{i=1}^{p} z_i (\beta + \gamma z_i), \frac{1}{c^2} \sum_{i=1}^{p} z_i^2 \sigma^2) = N(\gamma, \frac{1}{c} \sigma^2).
$$

Since $[\bar{X}, U]^T = AX$, a linear transformation of $X$, with

$$
A = \begin{bmatrix}
\frac{1}{p} & \cdots & \frac{1}{p}
\end{bmatrix},
$$

we only to find the covariance between $\bar{X}$ and $U$, and $[\bar{X}, U]^T \sim N_2([\beta, \gamma]^T, \Sigma)$, with $\Sigma = [(\sigma_{ij})]$, $\sigma_{11} = \frac{1}{p} \sigma^2$, $\sigma_{22} = \frac{1}{c} \sigma^2$, and $\sigma_{12} = \text{Cov}(\bar{X}, U)$. We have

$$
\text{Cov}(\bar{X}, U) = \text{Cov} \left( \sum_{i=1}^{p} \frac{z_i}{c} X_i, \sum_{j=1}^{p} \frac{z_j}{c} X_j \right) = \sum_{i=1}^{p} \frac{z_i}{pc} \text{Cov}(X_i, X_j)
$$

$$
= \frac{1}{pc} \sum_{i=1}^{p} z_i \text{Cov}(X_i, X_i) = \frac{\sigma^2}{pc} \sum_{i=1}^{p} z_i = 0,
$$

and therefore the joint distribution is one of independence, with $\bar{X} \sim N(\beta, \frac{1}{p} \sigma^2)$ and $U \sim N(\gamma, \frac{1}{c} \sigma^2)$.

■

Exercise 20 (Chebyshev inequality)

Proof. Let $A = \{ \omega : |X(\omega) - \mu| \geq \varepsilon \}$. Then

$$
\text{Var}(X) = E[(X - \mu)^2] \geq E[(X - \mu)^2 I_A(X)] \geq \varepsilon^2 E(I_A(X)) = \varepsilon^2 P(A),
$$

and rearranging gives the result. ■

Exercise 21 (Markov inequality)
Proof. Assume that \( r < +\infty \), otherwise the result holds trivially. Define the rv \( Y_r = X - r I(X \geq r) \), and note that by definition it is non-negative. Then, taking expectation yields

\[
E(Y_r) \geq 0 \iff E(X) \geq r E[I(X \geq r)] = r P(X \geq r),
\]

and the inequality is established. Now equality is attained if and only if \( E(Y_r) = 0 \), and since \( Y_r \geq 0 \), this is equivalent to \( P(Y_r = 0) = 1 \). But \( Y_r = 0 \) if and only if \( X = 0 \) or \( X = r \), and the proof is complete.

Exercise 22 (Jensen inequality)

Proof. Let \( l(x) = a + bx \), be the tangent line to \( g(x) \) at \( x = E[X] \), (for some \( a \) and \( b \)) so that \( g(E[X]) = l(E[X]) \). By definition of convexity we have \( g(x) \geq l(x) \), and taking the expected value of both sides we have

\[
E[g(X)] \geq E[l(X)] = E[a + bX] = a + bE(X) = l[E(X)] \Rightarrow E[g(X)] \geq g(E(X)),
\]

and the inequality is established.

Now if \( g(x) \) is linear, equality follows from linearity of expectation. Now let \( \mu = E(X) \) and assume that \( E(g(X)) = g(EX) = g(\mu) \), where \( g \) is convex but not linear. But \( g(\mu) = l(\mu) \), at \( x = \mu \), so that \( E(g(X)) = l(\mu) \). Define \( Y = g(X) - l(X) \), and note that \( Y \geq 0 \) (by convexity), and at the same time

\[
E(Y) = E(g(X) - l(X)) = E(g(X)) - E(l(X))
\]

so that we have a nonnegative rv with mean zero. As a result, we must have \( Y = 0 \) wp 1 (all the mass of \( Y > 0 \) should be concentrated at 0 otherwise \( E(Y) \) would be positive not 0), which concludes the proof.

Simulation and computation: use your favorite language to code the functions

Exercise 23

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 24

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 25

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 26

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 27

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 28

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.

Exercise 29

Proof. See the file Chapter1Vol1RCodeSolutions.R for the code.
Exercise 30

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■

Exercise 31

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■

Exercise 32

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■

Exercise 33

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■

Exercise 34

Proof. Consider the discrete case for simplicity, that is, we want to simulate from a discrete rv X with \( P(X = x) = p_x, \) \( x = 0, 1, \ldots, \) and let \( Y \sim \{q_x\} \) denote the proposal distribution. Let \( c = \max \{p_x / q_x\} \). Recall that iterations are independent and the probability of acceptance does not depend on the iteration number.

(i) The probability that a single iteration of the rejection algorithm results in a realization that is accepted is given by

\[
P(Y = x \text{ and it is accepted}) = P(Y = x)P(\text{accept} | Y = x)
\]

\[
= \frac{p_x}{cq_x} = \frac{p_x}{c},
\]

so that the probability that we accept a realization is

\[
P(\text{accept}) = \sum_x \frac{p_x}{c} = \frac{1}{c}.
\]

(ii) Let \( Z \) denote the first iteration that we accept the generated value. Then, from part (i), \( Z \) follows a geometric distribution with probability of success (acceptance) \( p = \frac{1}{c} \). Therefore, the expected number of iterations until the first accepted proposed value is \( \frac{1}{p} = c. \) ■

Exercise 35

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■

Exercise 36

Proof. See the file Chapter1Vol1RCODESolutions.R for the code. ■