Solutions to Graded Homework 2

Exercise 1. a) Use $B = A \cup (B \setminus A)$, where $A$ and $B \setminus A$ are disjoint, as well as $\Omega = A \cup A^c$ and $P(\Omega) = 1$.

b) Use $A \cup B = A \cup (B \setminus (A \cap B))$, where $A$ and $B \setminus (A \cap B)$ are disjoint, as well as a).

c) Use $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$, where $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$; the $B_n$ are disjoint, so by axiom (iii') and a),

$$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

d) Applying the above formula, we obtain for $t \in D$:

$$\lim_{n \to \infty} \sum_{i=1}^{n} P(A_i \cap A_i^c) = \lim_{n \to \infty} P(\bigcup_{i=1}^{n} (A_i \cap A_i^c)) = \lim_{n \to \infty} \sum_{n=1}^{\infty} P(A_n \cap A_n^c) = \lim_{n \to \infty} P(\bigcup_{n=1}^{\infty} (A_n \cap A_n^c)).$$

e) $P(\bigcap_{n=1}^{\infty} A_n) = 1 - P(\bigcup_{n=1}^{\infty} A_n^c) = 1 - \lim_{n \to \infty} P(A_n^c) = \lim_{n \to \infty} P(A_n)$.

Exercise 2. a) As $g : \mathbb{R} \to \mathbb{R}$ is decreasing, it is invertible and its inverse $g^{-1}$ is well defined from $D = \text{range}(g)$ to $\mathbb{R}$. Moreover, we have for any $t \in D$:

$$\{x \in \mathbb{R} : g(x) \leq t\} = \{x \in \mathbb{R} : x \geq g^{-1}(t)\} = [g^{-1}(t), +\infty)$$

which is a Borel set (being the complement of an open interval).

b) For $t \in D$, we also have

$$F_Y(t) = P(\{Y \leq t\}) = P(\{g(X) \leq t\}) = P(\{X \geq g^{-1}(t)\}) = 1 - P(\{X < g^{-1}(t)\}) = 1 - \lim_{s \uparrow t} F_X(g^{-1}(s))$$

(Indeed, in general, $X$ need not be a continuous random variable, so $P(\{X < g^{-1}(t)\})$ is not necessarily equal to $F_X(g^{-1}(t))$ for all $t \in \mathbb{R}$).

c) In this case, $F_Y(t) = 1 - F_X(g^{-1}(t))$, so

$$p_Y(t) = -p_X(g^{-1}(t)) \frac{1}{g'(g^{-1}(t))}$$

Note that $p_Y$ is non-negative, as $g'(x)$ is negative for all $x \in \mathbb{R}$.

d) Applying the above formula, we obtain for $t \in D = \{t \in \mathbb{R} : t > 0\}$:

$$p_Y(t) = p_X(-\log(t)) \frac{1}{\exp(\log(t))} = \frac{1}{\sqrt{2\pi}} t^{-\mu-1} \exp\left(-\frac{(\log(t) - \mu)^2}{2}\right)$$

This pdf has a peak getting sharper and moving closer towards 0 as $\mu$ increases, or getting smoother and going towards $+\infty$ as $\mu$ decreases. It is called the log-normal distribution.

Exercise 3. a) We have, by independence of the $X_j$'s:

$$F_Y(t) = P(\{Y \leq t\}) = P(\{\max(X_1, \ldots, X_n) \leq t\}) = P(\{X_1 \leq t, \ldots, X_n \leq t\})$$

$$= \prod_{j=1}^{n} P(\{X_j \leq t\}) = (F_X(t))^n$$
b) In a similar way, we obtain:

\[ F_Z(t) = P\{Z \leq t\} = 1 - P\{Z > t\} = 1 - P(\min(X_1, \ldots, X_n) > t) = 1 - \prod_{j=1}^{n} P(\{X_j > t\}) = 1 - (1 - F_X(t))^n \]

c) In this case, we have

\[ F_Y(t) = t^n \quad \text{and} \quad F_Z(t) = 1 - (1 - t)^n \quad \text{for } t \in [0, 1] \]

from which we deduce that

\[ p_Y(t) = n t^{n-1} \quad \text{and} \quad p_Z(t) = n (1 - t)^{n-1} \quad \text{for } t \in [0, 1] \]

d) For \( t \in [0, 1] \), we have:

\[ F_{1-Y}(t) = P\{1 - Y \leq t\} = P\{Y \geq 1 - t\} = 1 - P(\{Y < 1 - t\}) = 1 - (1 - t)^n \]

and correspondingly,

\[ p_{1-Y}(t) = n (1 - t)^{n-1} \]

That is, the random variables \( 1 - Y \) and \( Z \) have the same distribution (but they are not equal).

**Exercise 4.** \( F_a, F_b \) and \( F_d \) are guaranteed to be cdfs. \( F_c \) is not, as \( \lim_{t \to -\infty} F_c(t) = F(0) \) need not be equal to 0.

**Exercise 5.**

a) The cdf is \( X \) is simply \( F_X(t) = t \) for \( t \in [0, 1] \). The computation of \( F_Y \) gives

\[ F_Y(t) = P(\{(\omega_1, \omega_2) \in [0, 1]^2 : \omega_1 + \omega_2 \leq 2t\}) = \begin{cases} 2t^2 & \text{if } 0 \leq t \leq 1/2 \\ 1 - 2(1 - t)^2 & \text{if } 1/2 \leq t \leq 1 \end{cases} \]

b) See the notebook for the code. As \( n \) gets large, we observe that the empirical cdfs \( F_X^{(n)} \) and \( F_Y^{(n)} \) approach the theoretical cdfs \( F_X \) and \( F_Y \) (no matter what sequences are drawn at random). A formal justification of this fact will be provided later in the course.

c) Given the observation made in part b), a possible test for a given data set generating an empirical cdf \( F^{(n)} \) after \( n \) draws is to look at the difference \( F^{(n)}(3/4) - F^{(n)}(1/4) \). If we believe the statement made above, when \( n \) gets large, this difference should approach 1/2 if the data points are drawn according to \( F_X \) or 3/4 if the data points are drawn according to \( F_Y \). So here is a simple way to decide between the two options:

- If \( F^{(n)}(3/4) - F^{(n)}(1/4) > 5/8 \), then conclude that the data points are drawn according to \( F_Y \);
- Else, conclude that they are drawn according to \( F_X \).

d) The dataset gives \( F^{(n)}(3/4) - F^{(n)}(1/4) = 0.68 > 0.625 = 5/8 \), from which we conclude that the data points are drawn according to \( F_Y \).