3. Week 3

Remark 3.1 (Shorthand notation for Pre-images)*.* In the setting of Notation 2.39, we shall suppress the symbols ω and use the following notation for convenience, viz.

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = (X \in A).$$

For specific sets A, other notations, again for convenience, may be used. For example for

(a) If $A = (-\infty, x]$, then we write

$$X^{-1}(A) = (X \in A) = \{\omega \in \Omega : X(\omega) \in (-\infty, x]\} = \{\omega \in \Omega : X(\omega) \leq x\} = (X \leq x).$$

For $A = (-\infty, x), (x, \infty), [x, \infty)$, we shall write $X^{-1}(A)$ to be equal to $(X < x), (X > x), (X \geq x)$ respectively.

(b) If $A = \{x\}$, then we write

$$X^{-1}(A) = (X \in A) = \{\omega \in \Omega : X(\omega) \in \{x\}\} = \{\omega \in \Omega : X(\omega) = x\} = (X = x).$$

Remark 3.2 (Properties of pre-images)*.* Let $X : \Omega \to \mathbb{R}$ be a function. The following are some properties of the pre-images under X, which follow from the fact that X is a function.

(a) $X^{-1}(\mathbb{R}) = \Omega$.
(b) $X^{-1}(\emptyset_{\mathbb{R}}) = \emptyset_{\Omega}$, where $\emptyset_{\mathbb{R}}$ and $\emptyset_{\Omega}$ denote the empty sets under $\mathbb{R}$ and Ω, respectively. When there is no chance of confusion, we simply write $X^{-1}(\emptyset) = \emptyset$.
(c) For any two subsets A, B of $\mathbb{R}$ with $A \cap B = \emptyset$, we have $X^{-1}(A) \cap X^{-1}(B) = \emptyset$.
(d) For any subset A of $\mathbb{R}$, we have $X^{-1}(A^c) = \left(X^{-1}(A)\right)^c$.
(e) Let $\mathcal{I}$ be an indexing set. For any collection $\{A_i : i \in \mathcal{I}\}$ of subsets of $\mathbb{R}$, we have

$$X^{-1}\left(\bigcup_{i \in \mathcal{I}} A_i\right) = \bigcup_{i \in \mathcal{I}} X^{-1}(A_i), \quad X^{-1}\left(\bigcap_{i \in \mathcal{I}} A_i\right) = \bigcap_{i \in \mathcal{I}} X^{-1}(A_i).$$

The above properties shall be used frequently throughout the course.

Note 3.3. As discussed in Note 2.38, we now look at real valued functions defined on Ω, where Ω is the sample space of a random experiment $\mathcal{E}$. We shall also assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 3.4 (Random variable or RV)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Any real valued function $X : \Omega \to \mathbb{R}$ shall be referred to as a random variable or simply, an RV. In this case, we shall say that X is an RV defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Note 3.5. If $\mathcal{F}$ is taken to be 2^Ω, we immediately have

$$X^{-1}(A) = (X \in A) \in \mathcal{F}$$

for any subset A of $\mathbb{R}$. If $\mathcal{F}$ is taken to be a smaller collection of subsets of Ω, then the above observation may not hold for any arbitrary function X. Given such $\mathcal{F}$, we then restrict our attention to the class of functions X satisfying the above property and refer to them as RVs. It is therefore important to specify $\mathcal{F}$ before we discuss RVs X.

Note 3.6. The probability function/measure $\mathbb{P}$ has not been used in the definition of an RV X. We now discuss the role of $\mathbb{P}$ in analysis of RVs X.

Notation 3.7. We write $\mathbb{B}$ to denote the power set of $\mathbb{R}$.

Notation 3.8. Let X be an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then for all $A \in \mathbb{B}$, we have $X^{-1}(A) \in \mathcal{F}$ and hence $\mathbb{P}(X^{-1}(A))$ is well defined. We denote this in terms of a set function $\mathbb{P} \circ X^{-1} : \mathbb{B} \to [0, 1]$ given by $\mathbb{P} \circ X^{-1}(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A), \forall A \in \mathbb{B}$. A shorthand notation $\mathbb{P}_X$ shall also be used to refer to $\mathbb{P} \circ X^{-1}$.

Notation 3.9. Similar to the discussion in Remark 3.1, we shall write $\mathbb{P}(X \leq x), \mathbb{P}(X = x)$ etc. for $\mathbb{P} \circ X^{-1}(A)$ where $A = (-\infty, x], \{x\}$ etc. respectively.

Proposition 3.10. *Let X be an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the set function $\mathbb{P} \circ X^{-1}$ is a probability function/measure defined on the collection $\mathbb{B}$.*

Proof. We verify the axioms/properties of a probability function/measure as mentioned in Definition 1.33.

We have $\mathbb{P} \circ X^{-1}(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1$. Since $\mathbb{P}$ is a probability measure on $\mathcal{F}$, we also have $\mathbb{P} \circ X^{-1}(A) = \mathbb{P}(X^{-1}(A)) \geq 0, \forall A \in \mathbb{B}$.

If $\{A_n\}_n$ is a sequence of pairwise disjoint sets in $\mathbb{B}$, then $\{X^{-1}(A_n)\}_n$ is a sequence of pairwise disjoint events in $\mathcal{F}$. Hence,

$$\mathbb{P} \circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(A_n)\right) = \sum_{n=1}^{\infty} \mathbb{P}(X^{-1}(A_n)) = \sum_{n=1}^{\infty} \mathbb{P} \circ X^{-1}(A_n).$$

This proves countable additivity property for $\mathbb{P} \circ X^{-1}$ and the proof is complete. $\square$

Definition 3.11 (Induced Probability Space and Induced Probability Measure)**.** If X is an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the probability function/measure $\mathbb{P} \circ X^{-1}$ on $\mathbb{B}$ is referred to as the induced probability function/measure induced by X. In this case, $(\mathbb{R}, \mathbb{B}, \mathbb{P} \circ X^{-1})$ is referred to as the induced probability space induced by X.

Example 3.12. Recall from Remark 2.37, that if we toss a fair coin twice independently, then the sample space is $\Omega = \{HH, HT, TH, TT\}$ with $\mathbb{P}(\{HH\}) = \mathbb{P}(\{HT\}) = \mathbb{P}(\{TH\}) = \mathbb{P}(\{TT\}) = \frac{1}{4}$. Consider the RV $X : \Omega \to \mathbb{R}$ which denotes the number of heads. Here,

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

Consider the induced probability measure $\mathbb{P} \circ X^{-1}$ on $\mathbb{B}$. We have

$$\mathbb{P} \circ X^{-1}(\{0\}) = \mathbb{P}(X^{-1}(\{0\})) = \mathbb{P}(\{TT\}) = \frac{1}{4},$$

$$\mathbb{P} \circ X^{-1}(\{1\}) = \mathbb{P}(X^{-1}(\{1\})) = \mathbb{P}(\{HT, TH\}) = \frac{1}{2},$$

$$\mathbb{P} \circ X^{-1}(\{2\}) = \mathbb{P}(X^{-1}(\{2\})) = \mathbb{P}(\{HH\}) = \frac{1}{4}.$$

More generally, for any $A \in \mathbb{B}$, we have

$$\mathbb{P} \circ X^{-1}(A) = \mathbb{P}(\{\omega : X(\omega) \in A\}) = \sum_{i \in \{0,1,2\} \cap A} \mathbb{P} \circ X^{-1}(\{i\}).$$

Remark 3.13. If we know the probability function/measure $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ for any RV X, then we get the information about all the probabilities $\mathbb{P}(X \in A), A \in \mathbb{B}$ for events $X^{-1}(A) = (X \in A), A \in \mathbb{B}$

involving the RV X. In what follows, our analysis of RV X shall be through the understanding of probability function/measure $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ on $\mathbb{B}$.

Definition 3.14 (Law/Distribution of an RV)**.** If X is an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the probability function/measure $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ on $\mathbb{B}$ is referred to as the law or distribution of the RV X.

We now discuss some properties of a probability function/measure. To do this, we first introduce a concept involving sequence of sets.

Definition 3.15 (Increasing and decreasing sequence of sets)**.** Let $\{A_n\}_n$ be a sequence of subsets of a non-empty set Ω.

(a) If $A_n \subseteq A_{n+1}, \forall n = 1, 2, \cdots$, we say that the sequence $\{A_n\}_n$ is increasing. In this case, we say A_n increases to A, denoted by $A_n \uparrow A$, where $A = \bigcup_{n=1}^{\infty} A_n$.

(b) If $A_n \supseteq A_{n+1}, \forall n = 1, 2, \cdots$, we say that the sequence $\{A_n\}_n$ is decreasing. In this case, we say A_n decreases to A, denoted by $A_n \downarrow A$, where $A = \bigcap_{n=1}^{\infty} A_n$.

Remark 3.16. $A_n \uparrow A$ if and only if $A_n^c \downarrow A^c$.

Proposition 3.17 (Continuity of a probability measure)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space.*

(a) (Continuity from below) Let $\{A_n\}_n$ *be sequence in* $\mathcal{F}$*, such that* $A_n \uparrow A$*. Then*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

(b) (Continuity from above) Let $\{A_n\}_n$ *be sequence in* $\mathcal{F}$*, such that* $A_n \downarrow A$*. Then*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

Proof. To prove the first statement. Since $\{A_n\}_n$ is an increasing sequence of sets, we have

$$A_n \cap (A_1 \cup A_2 \cup \cdots \cup A_{n-1})^c = A_n \cap A_{n-1}^c, \forall n \geq 2.$$

Then using a hint from practice problem set 1, we have

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \left(\bigcup_{n=2}^{\infty} (A_n \cap A_{n-1}^c) \right).$$

Since the sets $A_1, A_2 \cap A_1^c, A_3 \cap A_2^c, \cdots$ are pairwise disjoint, using the countable additivity of $\mathbb{P}$, we have

$$\begin{aligned}\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}(A_1) + \sum_{n=2}^{\infty} \mathbb{P}(A_n \cap A_{n-1}^c) = \mathbb{P}(A_1) + \lim_{k\to\infty} \sum_{n=2}^{k} \mathbb{P}(A_n \cap A_{n-1}^c) \\ &= \mathbb{P}(A_1) + \lim_{k\to\infty} \sum_{n=2}^{k} \left[\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\right] \\ &= \mathbb{P}(A_1) + \lim_{k\to\infty} \left[\mathbb{P}(A_k) - \mathbb{P}(A_1)\right] = \lim_{k\to\infty} \mathbb{P}(A_k).\end{aligned}$$

This completes the proof of the first statement.

To prove the second statement. First observe that $A_n^c \uparrow A^c$ with $A = \bigcap_{n=1}^{\infty} A_n$. Using the first statement, we have

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c) = \lim_{n\to\infty} [1 - \mathbb{P}(A_n^c)] = \lim_{n\to\infty} \mathbb{P}(A_n).$$

The proof is complete. $\square$

Definition 3.18 (Distribution function of an RV)**.** Let X be an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with law/distribution $\mathbb{P}_X$. Consider the function $F_X : \mathbb{R} \to \mathbb{R}$ defined by $F_X(x) := \mathbb{P} \circ X^{-1}((-\infty, x]) = \mathbb{P}(X \leq x), \forall x \in \mathbb{R}$. The function F_X is called the cumulative distribution function (CDF) or simply, the distribution function (DF) of the RV X.

Remark 3.19 (RVs equal in law/distribution). Let X be an RV defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let Y be an RV defined on a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$. If $\mathbb{P} \circ X^{-1} = \mathbb{P}' \circ Y^{-1}$, i.e. $\mathbb{P} \circ X^{-1}(A) = \mathbb{P}' \circ Y^{-1}(A), \forall A \in \mathbb{B}$, then we say that X and Y are equal in law/distribution. In this case, $F_X = F_Y$, i.e. $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$.

Remark 3.20. Let X and Y be two RVs, possibly defined on different probability spaces. If $F_X = F_Y$, then it can be shown that X and Y are equal in law/distribution. The proof of this statement is beyond the scope of this course. This statement is often restated as 'the DF of an RV uniquely determines the law/distribution of the RV'.

Example 3.21 (The DF of a constant RV)**.** Let $c \in \mathbb{R}$ and let $X : \Omega \to \mathbb{R}$ be given by $X(\omega) := c, \forall \omega \in \Omega$. Then, for all $x \in \mathbb{R}$, we have

$$(X \le x) = \{\omega \in \Omega : X(\omega) \le x\} = \begin{cases} \emptyset, \text{ if } x < c, \\ \Omega, \text{ if } x \ge c. \end{cases}$$

Therefore, the DF of the RV X is given by

$$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} \mathbb{P}(\emptyset), \text{ if } x < c, \\ \mathbb{P}(\Omega), \text{ if } x \ge c \end{cases} = \begin{cases} 0, \text{ if } x < c, \\ 1, \text{ if } x \ge c. \end{cases}.$$

Example 3.22 (The DF of a two-valued RV)**.** Let $A \subset \Omega$ and let $X : \Omega \to \mathbb{R}$ be given by $X(\omega) := 1_A(\omega) = \begin{cases} 1, \text{ if } \omega \in A, \\ 0, \text{ if } \omega \in A^c \end{cases}$. Then, for all $x \in \mathbb{R}$, we have

$$(X \le x) = \{\omega \in \Omega : X(\omega) \le x\} = \begin{cases} \emptyset, \text{ if } x < 0, \\ A^c, \text{ if } 0 \le x < 1, \\ \Omega, \text{ if } x \ge 1. \end{cases}$$

Therefore, the DF of the RV X is given by

$$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} \mathbb{P}(\emptyset), \text{ if } x < 0, \\ \mathbb{P}(A^c), \text{ if } 0 \le x < 1, \\ \mathbb{P}(\Omega), \text{ if } x \ge 1 \end{cases} = \begin{cases} 0, \text{ if } x < 0, \\ \mathbb{P}(A^c), \text{ if } 0 \le x < 1, \\ 1, \text{ if } x \ge 1 \end{cases}.$$

Example 3.23. Consider X as in Example 3.12. Then for all $x \in \mathbb{R}$, we have

$$F_X(x) = \mathbb{P}_X((-\infty, x])) = \sum_{i \in \{0,1,2\} \cap (-\infty, x]} \mathbb{P}_X(\{i\}) = \begin{cases} 0, \text{ if } x < 0, \\ \mathbb{P}_X(\{0\}), \text{ if } 0 \le x < 1, \\ \mathbb{P}_X(\{0\}) + \mathbb{P}_X(\{1\}), \text{ if } 1 \le x < 2, \\ \mathbb{P}_X(\{0\}) + \mathbb{P}_X(\{1\}) + \mathbb{P}_X(\{2\}), \text{ if } x \ge 2. \end{cases}$$

Therefore,

$$F_X(x) = \begin{cases} 0, \text{ if } x < 0, \\ \frac{1}{4}, \text{ if } 0 \leq x < 1, \\ \frac{3}{4}, \text{ if } 1 \leq x < 2, \\ 1, \text{ if } x \geq 2. \end{cases}$$