2 Random Variables

What We Need for Random Variables

- In many problems, we are not just interested in the experiment itself, but rather the consequence of the random outcome. Therefore, it makes sense to assign each outcome a real value via a *random variable* $X : \Omega \to \mathbb{R}$.
- For example for the experiment where a fair coin is tossed twice, we may have $X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$ or $W(HH) = 4, W(HT) = W(TH) = W(TT) = 0$.
- What if we want to describe the distribution of possible values of X? Two approaches
  (1) $f(x)$ =probability that X is equal to x; this is OK for the above example, e.g., $f(2) = \frac{1}{4}$, but in general not appropriate;
  (2) a more appropriate way is to use the **distribution function** $F : \mathbb{R} \to [0, 1]$ that describes "probability that X does not exceed x". Formally, for any $x \in \mathbb{R}$, we have

$$F(x) = P(A(x)), \text{ where } A(x) = \{\omega \in \Omega : X(w) \leq x\}$$

- However, P is a function defined on $\mathcal{F}$; this requires that $A(x)$ has to belong to $\mathcal{F}$. Otherwise, this does not make any sense! Here we provide a general definition requiring that Borel set of $\mathbb{R}$ is measurable under X; we will see later that it is equivalent to $A(x)$.

Definition: $\mathcal{F}$-Measurable & Random Variables

A function $X : \Omega \to \mathbb{R}$ is said to be **$\mathcal{F}$-measurable** if for any Borel set $B \in \mathcal{B}(\mathbb{R})$, we have $\{X \in B\} \in \mathcal{F}$, where $\{X \in B\}$ is the abbreviation of event $\{\omega \in \Omega : X(w) \in B\}$ or $X^{-1}(B)$, which is the pre-image of Borel set B. If $(\Omega, \mathcal{F}, P)$ is a probability space, then such a $\mathcal{F}$-measurable function X is called a **random variable**.

What is a Borel Set

Definition: Borel σ-field & Borel Set

The Borel σ-field $\mathcal{B}(\mathbb{R})$ is the unique smallest σ-field containing all subsets of the form $(a, b] \subseteq \mathbb{R} = (-\infty, \infty)$. An element in $\mathcal{B}(\mathbb{R})$ is a **Borel set**.

- The reason why we consider $\mathcal{B}(\mathbb{R})$ is that there does not exist a probability measure for $2^{\mathbb{R}}$. Therefore, we compromise and consider a smaller σ-field that contains certain "nice" subsets of the sample space $\mathbb{R}$. These "nice" subsets are the intervals.
- If $\mathcal{F}$ and $\mathcal{G}$ are two σ-fields, then $\mathcal{F} \cap \mathcal{G}$ is also a σ-field. Then for any collections of events $\mathcal{F} \subseteq 2^{\Omega}$, there exists a unique **smallest σ-field containing $\mathcal{F}$** denoted by $\sigma(\mathcal{F})$.
- One of the most important properties of $\mathcal{B}(\mathbb{R})$ is that $(a, b), [a, b], \{a\}, (-\infty, b], (-\infty, b), (a, \infty), [a, \infty)$ and their unions are all Borel sets. For example,

$$(a, b) = \bigcup_{n\geq 1}(a, b - \frac{1}{n}], \quad [a, b] = \bigcap_{n\geq 1}(a - \frac{1}{n}, b] \quad \text{and} \quad \{a\} = \bigcap_{n\geq 1}(a - \frac{1}{n}, a].$$

Examples of Random Variable

- **Example 1: A Valid Random Variable**
  Consider the following probability space $(\Omega, \mathcal{F}, P)$, where
  - $\Omega = \{\omega : |\omega^2| \leq 100\} \subseteq \mathbb{C}$
  - $\mathcal{F} = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, C \cup A\}$
  - $A = \{\omega : |\omega^2| \leq 1\}$, $B = \{\omega : 1 < |\omega^2| \leq 4\}$ and $C = \{\omega : 4 < |\omega^2| \leq 100\}$
  - $P(A) = \frac{1}{9}, P(B) = \frac{3}{9}$ and $P(C) = \frac{5}{9}$

  Now, consider a function $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = \begin{cases} 10 & \text{if } \omega \in A \\ 5 & \text{if } \omega \in B \\ 1 & \text{if } \omega \in C \end{cases}$.

  This is a random variable. For example, if we consider $\{5, 10\} \in \mathcal{B}(\mathbb{R})$, then we have $X^{-1}(\{5, 10\}) = A \cup B \in \mathcal{F}$. If we consider $\{1\} \in \mathcal{B}(\mathbb{R})$, then we have $X^{-1}(\{1\}) = C \in \mathcal{F}$.

- **Example 2: Not A Valid Random Variable**
  Consider the following probability space $(\Omega, \mathcal{F}, P)$, where
  - $\mathcal{F} = \{\emptyset, \Omega, A, B, C' \cup D, A \cup B, A \cup C' \cup D, B \cup C' \cup D\}$
  - $D = \{\omega : |(\omega - 5i)^2| \leq 1\}, C' = C \setminus D$ and $P(D)$ is arbitrary

  Now, consider a function $X : \Omega \to \mathbb{R}$ same as the above and $X(\omega) = 2$ if $\omega \in D$.

  This is a NOT random variable since for $\{2\} \in \mathcal{B}(\mathbb{R})$, we have $X^{-1}(\{2\}) = D \notin \mathcal{F}$.

More About Random Variables

- For any random variable $X : \Omega \to \mathbb{R}$, it provides a probability measure (also called the **probability law**) $P_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ defined by: for any Borel set $B \in \mathcal{B}(\mathbb{R})$, we have

$$P_X(B) := P(X \in B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

  Essentially, the probability law can be seen as the composition of $P^{-1}(\cdot)$ with the inverse image $X^{-1}(\cdot)$, i.e., $P_X(\cdot) = P \circ X^{-1}(\cdot)$.

- Moreover, X may not take value in the entire $\mathbb{R}$ but just a subset $Y = \{X(\omega) \in \mathbb{R} : \omega \in \Omega\} \subseteq \mathbb{R}$. Therefore, we also do not need to measure the entire Borel sets $\mathcal{B}(\mathbb{R})$. In fact, we just need to consider a smaller σ-field $\mathcal{Y} = \{X(A) : A \in \mathcal{F}\} \subseteq \mathcal{B}(\mathbb{R})$.

- Therefore, in some textbooks, you will see a random variable on probability space $(\Omega, \mathcal{F}, P)$ is defined as a function

$$X : (\Omega, \mathcal{F}, P) \to (Y, \mathcal{Y}, P_X)$$

- The above notation is particular useful when we handle discrete random variable with countable values Y. In this case, it suffices to consider $\mathcal{Y} = 2^Y$. In the above Example 1, we have $Y = \{1, 5, 10\}$ and $\mathcal{Y} = 2^Y = \{\emptyset, \{1\}, \{5\}, \{10\}, \{1, 5\}, \{5, 10\}, \{1, 10\}, \{1, 5, 10\}\}$.

- If we look from the other direction, we define $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ as the **σ-field generated by random variables**, which is the smallest σ-field X needed to be measurable.

Cumulative Distribution Function (CDF)

- Now, let us summarize what we have so far:
  - First, we need to have a probability space $(\Omega, \mathcal{F}, P)$.
  - Then we have a random variable $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable, which induces a new measure $P_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$.
  - Therefore, it makes sense to discuss $P_X(B) = P(X \in B)$ for any Borel set $B \in \mathcal{B}(\mathbb{R})$.
- In fact, $\forall B \in \mathcal{B}(\mathbb{R}) : X^{-1}(B) \in \mathcal{F}$ **if and only if** $\forall a \in \mathbb{R} : X^{-1}((-\infty, a]) \in \mathcal{F}$.
- To describe how a random variable looks like, it it useful to consider a particular Borel set $B = (-\infty, a]$, which leads to the definition of the *cumulative distribution function*.

Definition: Cumulative Distribution Function (CDF)

The cumulative distribution function of a random variable $X : \Omega \to \mathbb{R}$ is a function $F_X : \mathbb{R} \to [0, 1]$ defined by:

$$\forall x \in \mathbb{R} : F_X(x) = P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\}) = P_X((-\infty, x])$$

- **Example 1: Constant Variables**

  Let $c \in \mathbb{R}$ and define $X : \Omega \to \mathbb{R}$ by $\forall \omega \in \Omega : X(\omega) = c,$. So $F_X(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \ge c \end{cases}$.

- **Example 2: Indicator Functions**

  Let $A \in \mathcal{F}$ be an event and define $\mathbf{1}_A : \Omega \to \mathbb{R}$ by $\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$. What is $F_{\mathbf{1}_A}$?

Properties of Cumulative Distribution Function

(P1) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
Proof: Let $B_n = \{\omega \in \Omega : X(\omega) \le -n\} = \{X \le -n\}$. The sequence $B_1, B_2, \ldots$ is decreasing as $\emptyset$ as the limit. Thus, we have $P(B_n) \to P(\emptyset) = 0$.

(P2) F_X is non-decreasing, i.e., $\forall x < y : F_X(x) \le F_X(y)$.
Proof: Let $A(x) = \{X \le x\}$ and $A(x, y) = \{x < X \le y\}$. Then we can write $A(y) = A(x) \dot{\cup} A(x, y)$. Therefore, $P(A(y)) = P(A(x)) + P(A(x, y))$, which gives $F_X(y) = F_X(x) + P(A(x, y)) \ge F_X(x)$.

(P3) F_X is right-continuous, i.e., $\lim_{x \downarrow x_0} F(x) = F(x_0)$.
Proof: leave as a homework.

- In many problems, we can just work on the CDF of a random variable because the probability law P_X is **uniquely specified** by its CDF $F_X(\cdot)$
- Also, $F_X(\cdot)$ is a CDF for some random variable *if and only if* it satisfies the above three conditions.

Discrete Random Variable

- Depending on whether X takes value in countable set or uncountable set, random variables can be classified as discrete random variables and continuous random variables.

Definition: Discrete Random Variable

A random variable X is said to be **discrete** if there exists a (finite or infinite) sequence of distinct real numbers $x_1, x_2, \ldots$ such that for any Borel set $B \in \mathcal{B}(\mathbb{R})$ we have

$$P(X \in B) = \sum_{x_i \in B} P(X = x_i)$$

- Let us discuss the implications of the above definition:
  1. If X is discrete, then set $D = \{x \in \mathbb{R} : P(X = x) \neq 0\}$ must be countable.
  2. In fact, we further have $\sum_{x \in D} P(X = x) = 1$.
  3. It is useful to define $p_X(x) = P(X = x)$ as the **probability mass function** (PMF).
  4. Therefore, a discrete random variable can be written as $X : (\Omega, \mathcal{F}, P) \to (D, 2^D, P_X)$.
- The above implications actually comes from the Borel–Cantelli lemma saying that

$$\sum_{n=1}^{\infty} P(E_n) < \infty \quad \Rightarrow \quad P\left(\limsup_{n\to\infty} E_n\right) = 0, \text{ where } \limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n}^{\infty} E_k$$

Examples of Discrete Random Variables

- **The Uniform Random Variable:**
  Let $X : (\Omega, \mathcal{F}, P) \to (D, 2^D, P_X)$, where $D = \{x_1, x_2, \ldots, x_n\}$, such that

$$\boxed{\forall i = 1, \ldots, n : p_X(x_i) = P(X = x_i) = \frac{1}{n}}$$

- **The Bernoulli Random Variable:**
  Let $X : (\Omega, \mathcal{F}, P) \to (\{0, 1\}, 2^{\{0,1\}}, P_X)$ such that

$$\boxed{\forall i = 0, 1 : p_X(i) = P(X = i) = (1 - p)^{1-i} p^i},$$

  i.e., $p_X(0) = 1 - p$ and $p_X(1) = p$ for some $0 < p < 1$.

- **The Geometric Random Variable with Parameter p:**
  Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{N}^+, \mathbb{N}^+, P_X)$ such that

$$\boxed{\forall n = 1, 2, \cdots : p_X(n) = P(X = n) = (1 - p)^{n-1} p}$$

- **The Possion Random Variable with Parameter λ:**
  Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{N}_0, 2^{\mathbb{N}_0}, P_X)$ such that

$$\boxed{\forall n = 0, 1, 2, \cdots : p_X(n) = P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}}$$

  Note that $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}$, so $\sum_{n=0}^{\infty} P(X = n) = 1$.

Continuous Random Variable

- Depending on whether X takes value in countable set or uncountable set, random variables can be classified as discrete random variables and continuous random variables.

Definition: Continuous Random Variable

A random variable X is said to be **(absolutely) continuous** if there exists an integrable function $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$\forall B \in \mathcal{B}(\mathbb{R}) : P(X \in B) = P_X(B) = \int_B f_X(x)dx = \int_{\mathbb{R}} \mathbf{1}_B(x) f_X(x)dx$$

- The above function $f_X : \mathbb{R} \to \mathbb{R}$ is called the **probability density function** (PDF) and it has the following properties

  1. $\int_{\mathbb{R}} f_X(x)dx = P(X \in \mathbb{R}) = P_X(\mathbb{R}) = 1$, so f_X is non-negative almost everywhere.
  2. For any $\{a\} \in \mathcal{B}(\mathbb{R})$, we have $P(X = a) = P_X(\{a\}) = \int_{\{a\}} f(x)dx = 0$.
  3. A random variable is continuous iff every countable set has probability zero.
  4. If $\forall x \in \mathbb{R} : P(X = x) = 0$, then this does not mean that it is absolutely continuous. Counter-example: the Cantor random variable.
  5. PDF and CDF can be related by: $F_X(a) = P_X((-\infty, a]) = \int_{-\infty}^{a} f_X(x)dx$. Then by the Lebniz's Rule, we also have $\frac{d}{da}F_X(a) = \frac{d}{da}\int_{-\infty}^{a} f_X(x)dx = f_X(a)$.
  6. For continuous random variable, we have $\lim_{x\downarrow x_0} F(x) = F(x_0) = \lim_{x\uparrow x_0} F(x)$. (proof this as a homework)

Examples of Continuous Random Variables

- **The Uniform Random Variable with Interval $[\mathbf{a}, \mathbf{b}]$:**

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad \Rightarrow \quad F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x \end{cases}$$

- **The Exponential Random Variable with Parameter λ:**

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \Rightarrow \quad F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

- **The Gaussian Random Variable $N(\mu, \sigma^2)$:**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **The Cauchy Random Variable with Parameter λ:**

$$f_X(x) = \frac{1}{\pi}\frac{\lambda}{\lambda^2 + x^2} \quad \Rightarrow \quad F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x}{\lambda}\right) + \frac{1}{2}$$

Binomial Random Variable/Poisson Random Variable/Exponential Random Variable

- Suppose that we take a sequence of n independent experiments asking a yes–no question: success (with probability p) or failure (with probability $q = 1 - p$). Then the number of successes satisfies the **binomial distribution**. Note that, when $n = 1$, the binomial distribution is actually a Bernoulli distribution.

- The PMF of a binomial random variable $X : (\Omega, \mathcal{F}, P) \to (\{0, 1, \ldots, n\}, 2^{\{0,1,\ldots,n\}}, P_X)$ is

$$\boxed{\forall k = 0, 1, \ldots, n : p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

- **Poisson distribution** describes random variable k based on the following assumptions:
  - $k = 0, 1, 2 \ldots$ is the number of times an event occurs in an interval
  - The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
  - The average rate at which events occur is independent of any occurrences.
  - Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

  In fact, the Poisson distribution can be seen as the limit of the binomial distribution. Suppose that you divide the interval into n same sub-intervals and the probability that something happens in each interval is p. Then the probability that something happens in k intervals is actually

$$P_X(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  If we want to count the totally number of occurrences of something in the entire interval, we can compute

$$P_X(X = k) = \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k}$$

  However, when we increase n to get more smaller intervals, the probability p should go to zero. So what is p? Note that, for any $X \sim B(n, p)$, we have $E(X) = np$, which should be a fixed number not matter how many sub-intervals we divide. Therefore, we choose this expectation as the "rate" λ so that $p = \lambda/n$. This gives

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} &= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-k+1}{n} \left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n}}_{\to e^{-\lambda}} = \frac{\lambda^k}{k!} e^{-\lambda}\end{aligned}$$

- If one is interested in random variable X="the time between two successive events", then this is the **exponential distribution**. This can be derived by

$$F(t) = 1 - P(X \geq t) = 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t} \quad \Rightarrow \quad f(x) = F'(t) = -\frac{d}{dt} P(X \geq t) = \lambda e^{-\lambda t}$$