# Lecture-06: Expectation

## 1 Expectation

Consider $N$ trials of a random experiment, and a random variable $X$ defined on this probability space. Corresponding to the $i$-th outcome $\omega_i \in \Omega$, there exists a number $X(\omega_i)$. The empirical mean of random variable $X$ can be written as

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{N} X(\omega_i).$$

If $X \in \mathcal{X}$ was a discrete random variable with probability mass function $P$, then the PMF can be estimated for each $x \in \mathcal{X}$ as empirical PMF

$$\hat{P}_X(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{X(\omega_i)=x\}}.$$

That is, we can write the empirical mean in terms of the empirical PMF as

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{N}\sum_{x\in\mathcal{X}} x\mathbb{1}_{\{X(\omega_i)=x\}} = \sum_{x\in\mathcal{X}} x\hat{P}_X(x).$$

This example motivates the following definition of mean for simple random variables.

**Definition 1.1 (Expectation of simple random variable).** Consider a discrete random variable $X : \Omega \to \mathcal{X} \subseteq \mathbb{R}$ taking finitely many values $\mathcal{X}$ and having PMF $P_X : \mathcal{X} \to [0,1]$ is called a **simple random variable**. The **mean** or **expectation** of a simple random variable $X$ is denoted by $\mathbb{E}[X]$ and defined as

$$\mathbb{E}[X] \triangleq \sum_{x\in\mathcal{X}} xP_X(x).$$

*Remark* 1. Recall that a simple random variable can be written as $X = \sum_{x\in\mathcal{X}} x\mathbb{1}_{\{X=x\}}$, where $(A_x(\omega) = X^{-1}\{x\} : x \in \mathcal{X})$ is a partition of the sample space $\Omega$ and $P_X(x) = P(A_x)$. Hence, the expectation can be written as an integral

$$\mathbb{E}[X] = \int_\Omega X(\omega)P(d\omega) = \int_\Omega \sum_{x\in\mathcal{X}} x\mathbb{1}_{\{A_x\}}(\omega)P(d\omega) = \sum_{x\in\mathcal{X}} x\int_\Omega \mathbb{1}_{\{A_x\}}(\omega)P(d\omega) = \sum_{x\in\mathcal{X}} x\mathbb{E}[\mathbb{1}_{\{A_x\}}] = \sum_{x\in\mathcal{X}} xP_X(x).$$

That is, the expectation of an indicator function is the probability of the indicated set.

**Definition 1.2 (Expectation of a non-negative random variable).** For a non-negative random variable $X$ defined on a probability space $(\Omega, \mathcal{F}, P)$, there exists a sequence of non-decreasing simple random variables $(X_n : n \in \mathbb{N})$ such that for all $\omega \in \Omega$

$$X_n(\omega) \leqslant X_{n+1}(\omega), \text{ for all } n \in \mathbb{N}, \text{ and } \lim_n X_n(\omega) = X(\omega).$$

Then $\mathbb{E}[X_n]$ is defined for each $n \in \mathbb{N}$ and $(\mathbb{E}[X_n] : n \in \mathbb{N})$ is a non-decreasing sequence, so the limit $\lim_n \mathbb{E}[X_n] \in \mathbb{R} \cup \{\infty\}$ exists, and is independent of the choice of the sequence and depends only on the probability space $(\Omega, \mathcal{F}, P)$. The **expectation** of the non-negative random variable $X$ is defined as

$$\mathbb{E}[X] \triangleq \lim_n \mathbb{E}[X_n].$$

**Definition 1.3 (Expectation of a real random variable).** For a real-valued random variable $X$ defined on a probability space $(\Omega, \mathcal{F}, P)$, we can define the following functions

$$X_+ \triangleq \max\{X, 0\}, \qquad X_- \triangleq \max\{0, -X\}.$$

We can verify that $X_+, X_-$ are non-negative random variables and hence their expectations are well defined. We observe that $X(\omega) = X_+(\omega) - X_-(\omega)$ for each $\omega \in \Omega$. If at least one of the $\mathbb{E}[X_+]$ and $\mathbb{E}[X_-]$ is finite, then the **expectation** of the random variable $X$ is defined as

$$\mathbb{E}[X] \triangleq \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

**Theorem 1.4 (Expectation as an integral with respect to the distribution function).** *For a random variable $X$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the expectation is given by*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dF_X(x).$$

*Proof.* It suffices to show this for a non-negative random variable $X$. We define simple random variables $(X_n : n \in \mathbb{N})$ in the following fashion

$$X_n(\omega) \triangleq \sum_{k=0}^{2^{2n}-1} k2^{-n} \mathbb{1}_{\{[k2^{-n},(k+1)2^{-n})\}}(\omega) = \begin{cases} k2^{-n}, & k2^{-n} \leqslant X(\omega) < (k+1)2^{-n}, k \in \{0, \dots, 2^{2n}-1\}, \\ 0, & X(\omega) \geqslant 2^n. \end{cases}$$

We observer that $X_n$ is a quantized version of $X$, and its value is the left end-point $k2^{-n}$ when $X \in [k2^{-n}, (k+1)2^{-n})$ for each $k \in \{0, \dots, 2^{2n}-1\}$. As $n$ grows larger, we cover the non-negative real line and the step size grows smaller. Thus, the limiting random variable $X$ can take all possible non-negative real values. We see that $X_n(\omega) \leqslant X_{n+1}(\omega)$ and $\lim_n X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. Let $F_X(x^-) = P\{X < x\}$, then we can write the expectation

$$\mathbb{E}[X_n] = \sum_{k=0}^{2^{2n}-1} k2^{-n}[F_X(((k+1)2^{-n})^-) - F_X((k2^{-n})^-)],$$

is completely specified by the distribution function $F_X$. It follows that the expectation of the random variable $X$ is given by

$$\mathbb{E}[X] = \lim_n \mathbb{E}[X_n] = \lim_n \frac{1}{2^n} \sum_{k=0}^{2^{2n}-1} k2^{-n} \frac{[F_X((k2^{-n} + 2^{-n})^-) - F_X((k2^{-n}))^-)]}{2^{-n}} = \int_{\mathbb{R}} x dF_X(x).$$

□

## 1.1 Linearity of expectations

**Theorem 1.5 (Linearity of expectations).** *Suppose $X : \Omega \to \mathbb{R}^n$ is a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$. Then, for constants $\alpha_i \in \mathbb{R}$ for all $i \in [n]$, we have*

$$\mathbb{E}[\sum_{i=1}^n \alpha_i X_i] = \sum_{i=1}^n \alpha_i \mathbb{E}[X_i].$$

*Proof.* It suffices to show that $\int_{x \in \mathbb{R}^n} x_1 dF_X(x) = \int_{x_1 \in \mathbb{R}} x_1 dF_{X_1}(x_1)$. The result follows from the observation that

$$\int_{x \in \mathbb{R}^n} x_1 dF_X(x) = \int_{x_1 \in \mathbb{R}} x_1 \int_{x_2, \dots, x_n} dF_X(x) = \int_{x_1 \in \mathbb{R}} x_1 dF_{X_1}(x_1).$$

□

## 1.2 Functions of random variables

Consider a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Suppose $g : \mathbb{R} \to \mathbb{R}$ is function such that $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$, then $g(X)$ is a random variable.

**Example 1.6 (Monotone function of random variables).** Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing function such that $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$. Consider a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, then $Y \triangleq g(X)$ is a random variable with distribution function

$$F_Y(y) = P\{g(X) \leqslant y\} = P\left\{X \leqslant g^{-1}(y)\right\} = F_X(g^{-1}(y)).$$

**Example 1.7 (Independence of function of random variables).** Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be functions such that $g^{-1}(-\infty, x]$ and $h^{-1}(-\infty, x]$ are Borel sets for all $x \in \mathbb{R}$. Consider independent random variables $X$ and $Y$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Can you show that $g(X)$ and $h(Y)$ are independent random variables?

**Theorem 1.8 (Fundamental theorem of expectations).** *Suppose $X$ is a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$, and $g : \mathbb{R} \to \mathbb{R}$ is a function such that $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$. Then $Y = g(X)$ is also a random variable, and*

$$\mathbb{E}[Y] = \mathbb{E}[g(X)].$$

*Proof.* It suffices to show this is true for simple random variables $X : \Omega \to \mathcal{X} \subseteq \mathbb{R}$. Let $A_x \triangleq \{\omega \in \Omega : X(\omega) = x\}$ for each $x \in \mathcal{X}$. Then $(A_x = X^{-1}\{x\} \in \mathcal{F} : x \in \mathcal{X})$ partitions the sample space $\Omega$, and we can write its expectation as

$$\mathbb{E}[X] = \mathbb{E}[\sum_{x \in \mathcal{X}} x \mathbb{1}_{\{A_x\}}] = \sum_{x \in \mathcal{X}} x P_X(x).$$

It follows that $Y : \Omega \to \mathcal{Y} = g(\mathcal{X})$ is also a discrete random variable, and we can write

$$B_y \triangleq \{\omega \in \Omega : (g \circ X)(\omega) = y\} = \cup_{x \in \mathcal{X}} \{X(\omega) = x, g(x) = y\} = \cup_{x \in g^{-1}\{y\}} A_x.$$

Therefore, we can write the expectation

$$\mathbb{E}[Y] = \mathbb{E}[\sum_{y \in \mathcal{Y}} y \mathbb{1}_{B_y}] = \mathbb{E}[\sum_{y \in \mathcal{Y}} y \sum_{x \in g^{-1}\{y\}} \mathbb{1}_{A_x}] = \mathbb{E}[\sum_{x \in \mathcal{X}} g(x) \mathbb{1}_{A_x}] = \sum_{x \in \mathcal{X}} g(x) P_X(x).$$

□

# A Indicator functions

For a sequence of disjoint events $(A_n \in \mathcal{F} : n \in \mathbb{N})$, we have

$$\mathbb{1}_{\{\cup_{n \in \mathbb{N}} A_n\}} = \sum_{n \in \mathbb{N}} \mathbb{1}_{A_n}.$$

For any sequence of events $(A_n \in \mathcal{F} : n \in \mathbb{N})$, we have

$$\mathbb{1}_{\{\cap_{n \in \mathbb{N}} A_n\}} = \prod_{n \in \mathbb{N}} \mathbb{1}_{A_n}.$$

## A.1 Law of total probability

Let $(A_n \in \mathcal{F} : n \in \mathbb{N})$ be a partition of the sample space, i.e. $A_n \cap A_m = \emptyset$ for $n \neq m$ and $\cup_{n \in \mathbb{N}} A_n = \Omega$. Then, any event $B = B \cap \Omega = B \cap (\cup_{n \in \mathbb{N}} A_n) = \cup_{n \in \mathbb{N}} (B \cap A_n)$. Therefore, we can write

$$\mathbb{1}_B = \sum_{n \in \mathbb{N}} \mathbb{1}_{B \cap A_n}.$$