Lecture-08: Expectation

1 Expectation

Example 1.1. Consider a probability space $(\Omega,\mathcal{F},P)$. We consider N trials of a random experiment, and define a random vector $X:\Omega\to\mathcal{X}^N$ such that $X_i\triangleq\pi_i\circ X:\Omega\to\mathcal{X}$ is a discrete random variable associated with the trial $i\in[N]$. If the marginal distributions of random variables $(X_1,\dots,X_N)$ are identical with the common probability mass function $P_{X_1}:\mathcal{X}\to[0,1]$, then the empirical mean of random variable X_1 can be written as

$$\hat{m}\triangleq\frac{1}{N}\sum_{i=1}^{N}X_i.$$

For a random variable $X_1:\Omega\to\mathcal{X}$, we can define events $E_{X_1}(x)\triangleq X_1^{-1}\{x\}$ for each value $x\in\mathcal{X}$. The probability mass function $P_{X_1}:\mathcal{X}\to[0,1]$ for the discrete random variable X_1 can be estimated for each $x\in\mathcal{X}$ as the empirical probability mass function

$$\hat{P}_{X_1}(x)\triangleq\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{E_{X_i}(x)}.$$

Recall that a simple random variable X_1 can be written as $X_1=\sum_{x\in\mathcal{X}}x\mathbb{1}_{E_{X_1}(x)}$, where $E_{X_1}\triangleq(E_{X_1}(x)\in\mathcal{F}:x\in\mathcal{X})$ is a finite partition of the sample space Ω and $P_{X_1}(x)=P(E_{X_1}(x))$. That is, we can write the empirical mean in terms of the empirical PMF as

$$\hat{m}=\frac{1}{N}\sum_{i=1}^{N}\sum_{x\in\mathcal{X}}x\mathbb{1}_{E_{X_i}(x)}(\omega)=\sum_{x\in\mathcal{X}}x\hat{P}_{X_1}(x).$$

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

Definition 1.2 (Expectation of simple random variable). The **mean** or **expectation** of a simple random variable $X:\Omega\to\mathcal{X}\subseteq\mathbb{R}$ defined on a probability space $(\Omega,\mathcal{F},P)$, is denoted by $\mathbb{E}[X]$ and defined as

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

Remark 1. For an indicator random variable $\mathbb{1}_A$, we have $\mathbb{E}\mathbb{1}_A=P(A)$. That is, the expectation of an indicator function is the probability of the indicated set.

Remark 2. Since a simple random variable can be written as $X=\sum_{x\in\mathcal{X}}x\mathbb{1}_{E_X(x)}$ where $E_X(x)\triangleq X^{-1}\{x\}$ for all $x\in\mathcal{X}$, we can write the expectation of a simple random variable as an integral

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

Theorem 1.3. *Consider a non-negative random variable $X:\Omega\to\mathbb{R}_+$ defined on a probability space $(\Omega,\mathcal{F},P)$. There exists a sequence of non-decreasing non-negative simple random variables $Y:\Omega\to\mathbb{R}_+^{\mathbb{N}}$ such that for all $\omega\in\Omega$*

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

Then $\mathbb{E}[Y_n]$ is defined for each $n \in \mathbb{N}$, the sequence $(\mathbb{E}[Y_n] \in \mathbb{R}_+ : n \in \mathbb{N})$ is non-decreasing, and the limit $\lim_n \mathbb{E}[Y_n] \in \mathbb{R}_+ \cup \{\infty\}$ exists. This limit is independent of the choice of the sequence and depends only on the probability space.

Proof. For each $n \in \mathbb{N}$ and $k \in \{0, \dots, 2^{2n} - 1\}$, we define half-open sets $B_{n,k} \triangleq (k2^{-n}, (k+1)2^{-n}]$. Then, the collection of sets $B_n \triangleq (B_{n,k} : k \in \{0, \dots, 2^{2n} - 1\})$ partitions the set $(0, 2^n]$ for each $n \in \mathbb{N}$. Further, we observe that $\cup_{n \in \mathbb{N}} (0, 2^n] = \mathbb{R}^+$ and that $B_{n+1,2k} \cup B_{n+1,2k+1} = B_{n,k}$ for all $n \in \mathbb{N}$ and k.

For a non-negative random variable $X : \Omega \to \mathbb{R}_+$, we define events $A^X_{n,k} = X^{-1}(B_{n,k}) \in \mathcal{F}$, and a sequence of simple non-negative random variables $Y : \Omega \to \mathbb{R}_+^{\mathbb{N}}$ in the following fashion

$$Y_n(\omega) \triangleq \sum_{k=0}^{2^{2n}-1} \mathbb{1}_{A^X_{n,k}}(\omega) \left(\inf_{\omega \in A^X_{n,k}} X(\omega) \right) = \sum_{k=0}^{2^{2n}-1} \mathbb{1}_{A^X_{n,k}}(\omega) \left(\inf_{X(\omega) \in B_{n,k}} X(\omega) \right) = \sum_{k=0}^{2^{2n}-1} k2^{-n} \mathbb{1}_{A^X_{n,k}}(\omega).$$

We observe that Y_n is a quantized version of X, and its value is the left end-point $k2^{-n}$ when $X \in B_{n,k}$ for each $k \in \{0, \dots, 2^{2n} - 1\}$. Since $\cup_{k=0}^{2^{2n}-1} A^X_{n,k} = X^{-1}(0, 2^n]$, it follows that we cover the positive real line as n grows larger and the step size grows smaller. Thus, the limiting random variable can take all possible non-negative real values. We observe that

$$\begin{aligned}
Y_{n+1}(\omega) &= \sum_{k=0}^{2^{2(n+1)}-1} \mathbb{1}_{A^X_{n+1,k}}(\omega) \left(\inf_{X(\omega) \in B_{n+1,k}} X(\omega) \right) \\
&\geqslant \sum_{k=0}^{2^{2n}-1} \left(\mathbb{1}_{A^X_{n+1,2k}}(\omega) \left(\inf_{X(\omega) \in B_{n+1,2k}} X(\omega) \right) + \mathbb{1}_{A^X_{n+1,2k+1}}(\omega) \left(\inf_{X(\omega) \in B_{n+1,2k+1}} X(\omega) \right) \right) \\
&\geqslant \sum_{k=0}^{2^{2n}-1} \mathbb{1}_{A^X_{n,2k}}(\omega) \left(\inf_{X(\omega) \in B_{n,k}} X(\omega) \right) = Y_n(\omega).
\end{aligned}$$

We see that $Y_n(\omega) \leqslant Y_{n+1}(\omega) \leqslant X(\omega)$ and $\lim_n Y_n(\omega) = X(\omega)$ for all $\omega \in \Omega$.

Since $Y_n : \Omega \to \mathbb{R}$ is a simple random variable for all $n \in \mathbb{N}$, the expectation $\mathbb{E}[Y_n]$ is defined for all n, and can be written as

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

We observe that this expectation is completely specified by the distribution function F_X, and we can write the limit

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

□

Definition 1.4 (Expectation of a non-negative random variable). For a non-negative random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, consider the sequence of non-decreasing simple random variables $Y : \Omega \to \mathbb{R}_+^{\mathbb{N}}$ such that $\lim_n Y_n = X$. The **expectation** of the non-negative random variable X is defined as

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

Remark 3. From the definition, it follows that $\mathbb{E}[X] = \int_{\mathbb{R}_+} x dF_X(x)$.

Definition 1.5 (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 1.6 (Expectation as an integral with respect to the distribution function). *For a random variable $X : \Omega \to \mathbb{R}$ 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, and the result follows from the definition of expectation of a non-negative random variable as the limit of expectation of approximating simple functions. □

2 Properties of Expectations

Theorem 2.1 (Properties). *Let $X : \Omega \to \mathbb{R}$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$.*

(i) ***Linearity:*** *Let $a, b \in \mathbb{R}$ and X, Y be random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. If $\mathbb{E}X, \mathbb{E}Y$, and $a\mathbb{E}X + b\mathbb{E}Y$ are well defined, then $\mathbb{E}(aX + bY)$ is well defined and*

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

(ii) ***Monotonicity:*** *If $P\{X \geqslant Y\} = 1$ and $\mathbb{E}[Y]$ is well defined with $\mathbb{E}[Y] > -\infty$, then $\mathbb{E}[X]$ is well defined and $\mathbb{E}[X] \geqslant \mathbb{E}[Y]$.*

(iii) ***Functions of random variables:*** *Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function, then $g(X)$ is a random variable with $\mathbb{E}[g(X)] = \int_{x \in \mathbb{R}} g(x) dF(x)$.*

(iv) ***Continuous random variables:*** *Let $f_X : \mathbb{R} \to [0, \infty)$ be the density function, then $\mathbb{E}X = \int_{x \in \mathbb{R}} x f_X(x) dx$.*

(v) ***Discrete random variables:*** *Let $P_X : \mathcal{X} \to [0, 1]$ be the probability mass function, then $\mathbb{E}X = \sum_{x \in \mathcal{X}} x P_X(x)$.*

(vi) ***Integration by parts:*** *The expectation $\mathbb{E}X = \int_{x \geqslant 0} (1 - F_X(x)) dx - \int_{x < 0} F_X(x) dx$ is well defined when at least one of the two parts is finite on the right hand side.*

Proof. It suffices to show properties $(i) - (iii)$ for simple random variables.

(i) Let $X = \sum_{x \in \mathcal{X}} x \mathbb{1}_{E_X(x)}$ and $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{E_Y(y)}$ be simple random variables, then $(E_X(x) \cap E_Y(y) \in \mathcal{F} : (x, y) \in \mathcal{X} \times \mathcal{Y})$ partition the sample space Ω. Hence, we can write $aX + bY = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} (ax + by) \mathbb{1}_{\{E_X(x) \cap E_Y(y)\}}$ and from linearity of sum it follows that

$$\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} (ax + by) P\{E_X(x) \cap E_Y(y)\} \\
&= a \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} P\{E_X(x) \cap E_Y(y)\} + b \sum_{y \in \mathcal{Y}} y \sum_{x \in \mathcal{X}} P\{E_X(x) \cap E_Y(y)\} \\
&= a \sum_{x \in \mathcal{X}} x P(E_X(x)) + b \sum_{y \in \mathcal{Y}} y P(E_Y(y)) = a\mathbb{E}X + b\mathbb{E}Y.
\end{aligned}$$

(ii) From the fact that $X - Y \geqslant 0$ almost surely and linearity of expectation, it suffices to show that $\mathbb{E}X \geqslant 0$ for non-negative random variable X. It can easily be shown for simple non-negative random variables, and follows for general non-negative random variables by taking limits.

(iii) It suffices to show this holds true for simple random variables $X : \Omega \to \mathcal{X} \subset \mathbb{R}$. Since $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, $Y \triangleq g(X) : \Omega \to \mathcal{Y} \triangleq g(\mathcal{X})$ is a random variable. It follows that, we can write the following disjoint union $\mathcal{X} = \cup_{y \in \mathcal{Y}} \cup_{x \in g^{-1}\{y\}} \{x\}$. Further, for each $y \in \mathcal{Y}$, we have

$$E_Y(y) = \{\omega \in \Omega : (g \circ X)(\omega) = y\} = X^{-1} \circ g^{-1}\{y\} = \cup_{x \in g^{-1}\{y\}} E_X(x).$$

Since $E_X(x)$ are disjoint for all $x \in \mathcal{X}$, we get $P(E_Y(y)) = \sum_{x \in g^{-1}\{y\}} P(E_X(x))$. Using the above two facts, we can write the expectation

$$\mathbb{E}[Y] = \sum_{y \in \mathcal{Y}} y P(E_Y(y)) = \sum_{y \in g(\mathcal{X})} \sum_{x \in g^{-1}(y)} g(x) P(E_X(x)) = \sum_{x \in \mathcal{X}} g(x) P(E_X(x)).$$

(iv) For continuous random variables, we have $dF_X(x) = f_X(x)dx$ for all $x \in \mathbb{R}$.

(v) For discrete random variables $X : \Omega \to \mathcal{X}$, we have $dF_X(x) = P_X(x)$ for all $x \in \mathcal{X}$ and zero otherwise.

(vi) We can write $\mathbb{E}X = \int_{x<0} x dF_X(x) - \int_{x\geqslant 0} x d(1 - F_X)(x)$. We apply integration by parts to the first term on the right, to get

$$\int_{x<0} x dF_X(x) = xF_X(x)|_{-\infty}^{0} - \int_{x<0} F_X(x)dx.$$

Similarly, we apply integration by parts to the second term on the right, to get

$$-\int_{x\geqslant 0} x d(1 - F_X)(x) = -x(1 - F_X(x))|_0^{\infty} + \int_{x\geqslant 0} (1 - F_X(x))dx.$$

□