Lecture-07: Moments

1 Properties of Expectations

Theorem 1.1 (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 $X = \sum_{x \in \mathcal{X}} x \mathbb{1}_{A_x}$ and $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{B_y}$ be simple random variables, then $(A_x \cap B_y \in \mathcal{F} : (x, y \in \mathcal{X} \times \mathcal{Y}))$ partition the sample space Ω. Hence, we can write $aX + bY = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} (ax + by) \mathbb{1}_{\{A_x \cap B_y\}}$ and from linearity of sum it follows that

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

(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.

(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, $Y = g(X)$ is a random variable. For each $y \in \mathcal{Y} = g(\mathcal{X})$, we have

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

Therefore, we can write the expectation

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

(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 $\mathbb{E}X = -\int_{x\geqslant 0} xd(1-F_X)(x) + \int_{x<0} xdF_X(x)$. Therefore, we have

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

□

2 Moments

Example 2.1 (Absolute value function). For the function $|\cdot| : \mathbb{R} \to \mathbb{R}_+$, we can compute the inverse of half open sets $(-\infty, x]$ for any $x \in \mathbb{R}$, as

$$g^{-1}(-\infty, x] = \begin{cases} \emptyset, & x < 0, \\ [-x, x], & x \geqslant 0. \end{cases}$$

Since $g^{-1}(-\infty, x] \in \mathcal{B}(R)$, it follows that $|\cdot| : \mathbb{R} \to \mathbb{R}_+$ is a Borel measurable function.

Lemma 2.2. *If $\mathbb{E}|X|$ is finite, then $\mathbb{E}X$ exists and is finite.*

Proof. The function $|\cdot| : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and hence $|X|$ is a random variable. Further $|X| \geqslant 0$, and hence the expectation $\mathbb{E}|X|$ always exists. If $\mathbb{E}|X|$ is finite, it means $\mathbb{E}X_+$ and $\mathbb{E}X_-$ are both finite, and hence $\mathbb{E}X = \mathbb{E}X - \mathbb{E}X_-$ is finite as well. □

Corollary 2.3. *Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. If $\mathbb{E}|g(X)|$ is finite, then $\mathbb{E}g(X)$ exists and is finite.*

Exercise 2.4 (Polynomial function). For any $k \in \mathbb{N}$, we define functions $g_k : \mathbb{R} \to \mathbb{R}$ such that $g_k : x \mapsto x^k$. Show that g_k is Borel measurable.

Definition 2.5 (Moments). Let $X : \Omega \to \mathbb{R}$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. We define the kth **moment** of the random variable X as $m_k \triangleq \mathbb{E}g_k(X) = \mathbb{E}X^k$. First moment $\mathbb{E}X$ is called the **mean** of the random variable.

Remark 1. If $\mathbb{E}|X|^k$ is finite, then m_k exists and is finite.

Example 2.6 (Moments). If $|X| \leqslant 1$, then $|X|^k \leqslant 1$ almost surely. Therefore, by the monotonicity of expectations $\mathbb{E}|X|^k \leqslant 1$, and the moments m_k exist and are finite for all $k \in \mathbb{N}$.

Lemma 2.7. *If m_N is finite for some $N \in \mathbb{N}$, then m_k is finite for all $k \in [N]$.*

Proof. For any random variable $X : \Omega \to \mathbb{R}$ and $k \in [N]$, we can write

$$|X|^k = |X|^k \mathbb{1}_{\left\{|X|^k \leqslant 1\right\}} + |X|^k \mathbb{1}_{\left\{|X|^k > 1\right\}} \leqslant \mathbb{1}_{\left\{|X|^k \leqslant 1\right\}} + |X|^N \mathbb{1}_{\left\{|X|^k > 1\right\}} \leqslant 1 + |X|^N.$$

The result follows from the monotonicity of expectations. □

Exercise 2.8 (Polynomials). For any $k \in \mathbb{N}$, we define functions $h_k : \mathbb{R} \to \mathbb{R}$ such that $h_k : x \mapsto (x-m)^k$. Show that h_k is Borel measurable.

Definition 2.9 (Central moments). Let $X : \Omega \to \mathbb{R}$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ with finite first moment m_1. We define the kth **central moment** of the random variable X as $\sigma_k \triangleq \mathbb{E}h_k(X) = \mathbb{E}(X - m_1)^k$. The second central moment $\sigma_2 = \mathbb{E}(X - m_1)^2$ is called the **variance** of the random variable and denoted by σ^2.

Lemma 2.10. *The first central moment $\sigma_1 = \mathbb{E}(X - m_1) = 0$ and the variance $\sigma^2 = \mathbb{E}(X - m_1)^2$ for a random variable X is always non-negative, with equality when X is a constant. That is, $m_2 \geqslant m_1^2$ with equality when X is a constant.*

Proof. Recall that h_1, h_2 are Boreal measurable functions, and hence $h_1(X) = X - m_1$ and $h_2(X) = (X - m_1)^2$ are random variables. From the linearity of expectations, it follows that $\sigma_1 = \mathbb{E}h_1(X) = \mathbb{E}X - m_1 = 0$. Since $(X - m_1)^2 \geqslant 0$ almost surely, it follows from the monotonicity of expectation that $0 \leqslant \mathbb{E}(X - m_1)^2$. From the linearity of expectation and expansion of $(X - m_1)^2$, we get $\sigma^2 = \mathbb{E}X^2 - 2m_1\mathbb{E}X + m - 1^2 = m_2 - m_1^2 \geqslant 0$. ☐

Remark 2. If second moment is finite, then the first moment is finite.

Theorem 2.11 (Markov's inequality). *Let $X : \Omega \to \mathbb{R}$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Then, for any monotonically non-decreasing function $f : \mathbb{R} \to \mathbb{R}_+$, we have*

$$P\{X \geqslant \epsilon\} \leqslant \frac{\mathbb{E}[f(X)]}{f(\epsilon)}.$$

Proof. We can verify that any monotonically increasing function $f : \mathbb{R} \to \mathbb{R}_+$ is Borel measurable. Hence, $f(X)$ is a random variable for any random variable X. Therefore,

$$f(X) = f(X)\mathbb{1}_{\{f(X) \geqslant f(\epsilon)\}} + f(X)\mathbb{1}_{\{f(X) < f(\epsilon)\}} \geqslant f(\epsilon)\mathbb{1}_{\{X \geqslant \epsilon\}}.$$

The result follows from the monotonicity of expectations. ☐

Corollary 2.12 (Markov). *Let X be a non-negative random variable, then $P\{X > x\} \leqslant \frac{\mathbb{E}X}{x}$ for all $x > 0$.*

Corollary 2.13 (Chebychev). *Let X be a random variable with finite mean μ and variance σ^2, then $P\{|X - \mu| > x\} \leqslant \frac{\operatorname{Var} X}{x^2}$ for all $x > 0$.*

Proof. Apply the Markov's inequality for random variable $Y = |X - \mu| \geqslant 0$ and increasing function $f(x) = x^2$ for $x \geqslant 0$. ☐

Corollary 2.14 (Chernoff). *Let X be a random variable with finite $\mathbb{E}[e^{\theta X}]$ for some $\theta > 0$, then $P\{X > x\} \leqslant e^{-\theta x}\mathbb{E}[e^{\theta X}]$ for all $x > 0$.*

Proof. Apply the Markov's inequality for random variable X and increasing function $f(x) = e^{\theta x} > 0$ for $\theta > 0$. ☐