Lecture-08: Moments

1 Moments

Example 1.1 (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 1.2. *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 1.3. *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 1.4 (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 1.5 (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 1.6 (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 1.7. *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. □

2 L^p spaces

Remark 2. The set of random variables is a vector space.

Definition 2.1. For a probability space $(\Omega, \mathcal{F}, P)$, and $p \geqslant 1$, we define the set of random variables with finite absolute pth moment as the vector space

$$L^p \triangleq \left\{ X : (\mathbb{E}|X|^p)^{1/p} < \infty \right\}.$$

Example 2.2. The L^1 space consists of random variables with bounded absolute mean. The L^2 space consists of random variables with bounded second moment.

Remark 3. For any real numbers $1 \leqslant p \leqslant q$, we have $L^q \subseteq L^p$.

3 Moment generating functions

Suppose that $X : \Omega \to \mathbb{R}$ is a continuous random variable on the probability space $(\Omega, \mathcal{F}, P)$ with distribution function $F_X : \mathbb{R} \to [0,1]$.

Example 3.1. A function $g_\theta : \mathbb{R} \to \mathbb{R}$ defined by $g_\theta(x) \triangleq e^{\theta x}$ is Borel measurable for all $\theta \in \mathbb{R}$. Therefore, $g_\theta(X)$ is a positive random variable on this probability space. We can show that $h_\theta : \mathbb{R} \to \mathbb{C}$ defined by $h_\theta(x) \triangleq e^{j\theta x} = \cos(\theta x) + j\sin(\theta x)$ is also Borel measurable for all $\theta \in \mathbb{R}$, where $j = \sqrt{-1}$. Thus, $h_\theta(X)$ is a complex valued random variable on this probability space.

Definition 3.2 (Moment generating function). For a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the moment generating function $M_X : \mathbb{R} \to \mathbb{R}$ is defined by $M_X(\theta) = \mathbb{E}e^{\theta X}$ for all $\theta \in \mathbb{R}$ where $M_X(\theta)$ is finite.

Definition 3.3 (Characteristic function). For a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the **characteristic function** $\Phi_X : \mathbb{R} \to \mathbb{C}$ is defined by $\Phi_X(\theta) = \mathbb{E}e^{j\theta X}$ for all $\theta \in \mathbb{R}$.

Theorem 3.4. *Two random variables have the same probability distribution iff they have the same characteristic function.*

Proof. It is easy to see the necessity and the sufficiency is difficult. □

Lemma 3.5. *If $\mathbb{E}[X^k]$ exists and is finite for an integer $k \in \mathbb{N}$, then the derivatives of Φ_X up to order k exist and are continuous, and $\Phi_X^{(k)}(0) = j^k \mathbb{E}[X^k]$.*

Definition 3.6. For a non-negative integer-valued random variable X it is often more convenient to work with the z-transform of the PMF, defined by $\Psi_X(z) = \mathbb{E}z^X = \sum_{k \geqslant 0} z^k p_X(k)$, for real or complex z with $|z| \leqslant 1$.

Theorem 3.7. *Two non-negative integer-valued random variables have the same probability distribution iff their z-transforms are equal. If $\mathbb{E}[X^k]$ is finite it can be found from the derivatives of Ψ_X up to the kth order at $z = 1$, $\Psi_X^{(k)}(1) = \mathbb{E}[X(X-1)\dots(X-k+1)]$.*

Proof. The necessity is clear. For sufficiency, we see that $\Psi_X^{(k)}(0) = k!p_X(k)$. Further, interchanging the derivative and the summation (by dominated convergence theorem), we get the second result. □

4 Central Moments

Exercise 4.1 (Polynomials). For any $k \in \mathbb{N}$, we define functions $h_k : \mathbb{R} \to \mathbb{R}$ such that $h_k : x \mapsto (x - m_1)^k$. Show that h_k is Borel measurable.

Definition 4.2 (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 4.3. *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 4. If second moment is finite, then the first moment is finite. That is, $L^2 \subseteq L^1$.

5 Inequalities

Theorem 5.1 (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,

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

The result follows from the monotonicity of expectations. ☐

Corollary 5.2 (Markov). *Let X be a non-negative random variable, then*

$$P\{X > \epsilon\} \leqslant \frac{\mathbb{E}X}{\epsilon}, \text{ for all } \epsilon > 0.$$

Corollary 5.3 (Chebychev). *Let X be a random variable with finite mean μ and variance σ^2, then*

$$P\{|X - \mu| > \epsilon\} \leqslant \frac{\operatorname{Var} X}{\epsilon^2}, \text{ for all } \epsilon > 0.$$

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 5.4 (Chernoff). *Let X be a random variable with finite $\mathbb{E}[e^{\theta X}]$ for some $\theta > 0$, then*

$$P\{X > \epsilon\} \leqslant e^{-\theta\epsilon}\mathbb{E}[e^{\theta X}], \text{ for all } \epsilon > 0.$$

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