Lecture-09: Moments and L^p spaces

1 Moments

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

Example 1.1 (Absolute value function). For the function $|\cdot| : \mathbb{R} \to \mathbb{R}_+$, we can write the inverse image of half open sets $(-\infty, x]$ for any $x \in \mathbb{R}$ as $A_g(x) = g^{-1}(-\infty, x]$. It follows that $A_g(x) = \emptyset \in \mathcal{B}(\mathbb{R})$ for $x < 0$ and $A_g(x) = [-x, x] \in \mathcal{B}(\mathbb{R})$ for $x \in \mathbb{R}_+$. 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 for all $k \in \mathbb{N}$.

Definition 1.5 (Moments). 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.

Remark 2. If $P\{|X| \leqslant 1\} = 1$, then $P\left\{|X|^k \leqslant 1\right\} = 1$. Therefore, by the monotonicity of expectations $\mathbb{E}|X|^k \leqslant 1$, and the moments m_k exist and are finite for all $k \in \mathbb{N}$.

2 L^p spaces

Definition 2.1. A vector space over field $\mathbb{F}$ is denoted by a set V has (a) vector addition $+ : V \times V \to V$ that satisfies associativity and commutativity, and has an identity and inverse element for each vector $v \in V$, and (b) scalar multiplication $\cdot : \mathbb{F} \times V \to V$ that satisfies compatibility of field multiplication, distributivity with vector and field addition, and has an identity element.

Remark 3. We can verify that the set of random variables is a vector space over reals, and we denote it by V. To this end, we can verify the associativity and commutativity of addition of random variables. For any random variable X, the constant function 0 is the identity random variable, and $-X$ is the additive inverse for vector additions. We can verify the compatibility of scalar multiplication with random variables, existence of unit scalar 1, and distributivity of scalar multiplications with respect to field addition and vector addition.

Definition 2.2. For a vector space V over field $\mathbb{F}$, a **norm** on the vector space is a map $f : V \to \mathbb{R}_+$ that satisfies

homogeneity: $f(av) = |a| f(v)$ for all $a \in \mathbb{F}$ and $v \in V$,

sub-additivity: $f(v + w) \leqslant f(v) + f(w)$ for all $v, w \in V$, and

point-separating: $f(v) \geqslant 0$ for all $v \in V$.

Definition 2.3. 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 $L^p \triangleq \left\{ X \in V : (\mathbb{E}|X|^p)^{\frac{1}{p}} < \infty \right\}$.

Definition 2.4. We define a function $\|\|_p : L^p \to \mathbb{R}_+$ defined by $\|\|_p (X) = \|X\|_p \triangleq (\mathbb{E}|X|^p)^{\frac{1}{p}}$ for any $X \in L^p$ and real $p \geqslant 1$.

Remark 4. For $p = 1$, the map $\|\|_p$ is norm. Therefore, L^1 is a normed vector space consisting of random variables with bounded absolute mean.

Remark 5. We will show that $\|X\|_\infty = \sup\{|X(\omega)| : \omega \in \Omega\}$, and hence L^∞ is a normed vector space of bounded random variables.

Remark 6. We will also show that $\|\|_p$ is a norm for all $p \in (1, \infty)$, and hence L^p is a normed vector space of random variables with bounded $\|\|_p$ norm. In particular, the L^2 space consists of random variables with bounded second moment.

Remark 7. If $\mathbb{E}|X|^N$ is finite for some $N \in \mathbb{N}$, then $\mathbb{E}|X|^k$ is finite for all $k \in [N]$. This follows from the linearity and monotonicity of expectations, and the fact that

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

This implies that $L^N \subseteq L^k$ for all $k \in [N]$. We will show that $L^q \subseteq L^p$ for any real numbers $1 \leqslant p \leqslant q$.

3 Central Moments

Example 3.1 (Shifted polynomial functions). For any $k \in \mathbb{N}$, we define functions $h_k : \mathbb{R} \to \mathbb{R}$ such that $h_k : x \mapsto (x - m_1)^k$. Then, $h_k = g_k(x - m_1) = g_k \circ f$ where $f : \mathbb{R} \to \mathbb{R}$ is defined as $f(x) = x - m_1$ for all $x \in \mathbb{R}$. Since g_k and f are measurable, so is h_k.

Definition 3.2 (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 3.3. *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 8. If second moment is finite, then the first moment is finite. That is, $L^2 \subseteq L^1$.

4 Inequalities

Theorem 4.1 (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 non-decreasing 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 4.2 (Markov). *Let X be a non-negative random variable, then $P\{X \geqslant \epsilon\} \leqslant \frac{\mathbb{E}X}{\epsilon}$ for all $\epsilon > 0$.*

Corollary 4.3 (Chebychev). *Let X be a random variable with finite mean m_1 and variance σ^2, then*

$$P\{|X - m_1| \geqslant \epsilon\} \leqslant \frac{\sigma_2}{\epsilon^2}, \text{ for all } \epsilon > 0.$$

Proof. Apply the Markov's inequality for random variable $Y = |X - m_1| \geqslant 0$ and increasing function $f(x) = x^2$ for $x \geqslant 0$. □

Corollary 4.4 (Chernoff). *Let X be a random variable with finite $\mathbb{E}[e^{\theta X}]$ for some $\theta > 0$, then*

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

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

Definition 4.5 (Convex function). A real-valued function $f : \mathbb{R} \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}$ and $\theta \in [0,1]$, we have

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y).$$

Theorem 4.6 (Jensen's inequality). *For any convex function $f : \mathbb{R} \to \mathbb{R}$ and random variable X, we have*

$$f(\mathbb{E}X) \leqslant \mathbb{E}f(X).$$

Proof. It suffices to show this for simple random variables $X : \Omega \to \mathcal{X}$. We show this by induction on cardinality of alphabet $\mathcal{X}$. The inequality is trivially true for $|\mathcal{X}| = 1$. We assume that the inductive hypothesis is true for $|\mathcal{X}| = n$.

Let $X \in \mathcal{X}$, where $|\mathcal{X}| = n+1$. We can denote $\mathcal{X} = \{x_1, \dots, x_{n+1}\}$ with $p_i \triangleq P\{X = x_i\}$ for all $i \in [n+1]$. We observe that $(\frac{p_j}{1-p_1} : j \geqslant 2)$ is a probability mass function for some random variable $Y \in \mathcal{Y} = \{x_2, \dots, x_{n+1}\}$ with cardinality n. Hence, by inductive hypothesis, we have

$$f\left(\sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\right) = f(\mathbb{E}Y) \leqslant \mathbb{E}f(Y) = \sum_{i=2}^{n+1} \frac{p_i}{1-p_1} f(x_i).$$

Applying the convexity of f to $\theta = p_1, x = x_1, y = \sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i$, we get

$$f(\mathbb{E}X) = f(\sum_{i=1}^{n+1} p_i x_i) = f\Big(p_1 x_1 + (1-p_1)\sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\Big) \leqslant p_1 f(x_1) + (1-p_1) f\Big(\sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\Big).$$

From the inductive step, it follows that the RHS is upper bounded by $\mathbb{E}f(X)$, and the result follows. □

Theorem 4.7. *For any real numbers $1 \leqslant p \leqslant q < \infty$, we have $L^q \subseteq L^p$.*

Proof. Let $1 \leqslant p \leqslant q < \infty$ and consider a convex function $g : \mathbb{R}_+ \to \mathbb{R}_+$ defined by $g(x) \triangleq x^{q/p}$ for all $x \in \mathbb{R}_+$. It follows that $g(|X|^p) = |X|^q$ and hence from the Jensen's inequality, we get

$$\|X\|_p^q = g(\mathbb{E}|X|^p) \leqslant \mathbb{E}g(|X|^p) = \|X\|_q^q.$$

□