Expectation

3.1 Expectation

Proposition 3.1. ***Expectation*** *There exists a positive linear functional* $\mathbb{E}$ *on the space of non-negative random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *such that*

(a) $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$ *for* $A \in \mathcal{F}$

(b) $\mathbb{E}(aX) = a\,\mathbb{E}(X)\ \forall\ a > 0.$

(c) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$

(d) $X \geqslant Y \Rightarrow \mathbb{E}(X) \geqslant \mathbb{E}(Y).$

Definition 3.2. *A random variable* X *is* ***integrable*** *if* $\mathbb{E}\,|X| < \infty$.

We define

$$X^+ := X \vee 0 = \max(X, 0), \quad X^- := (-X) \vee 0 = \max(-X, 0),$$

then $X = X^+ - X^-$.

Definition 3.3. *We define the expectation of an integrable random variable* X *to be*

$$\mathbb{E}(X) = \mathbb{E}(X^+) - E(X^-).$$

Four-Step Procedure to Define Expectation:

Step 1: **Indicator and Simple rvs:** Let $\mathbb{E}(\mathbb{1}_A) := \mathbb{P}(A)$ for $A \in \mathcal{F}$. We define,

Definition 3.4 (Simple random variable)**.** *A r.v.* ϕ *is a simple RV if*

$$\phi(\omega) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega),$$

where A_i*'s are disjoint measurable sets.*

We define the expectation of a simple rv, ϕ, as

$$\mathbb{E}(\phi) := \sum_{i=1}^{n} a_i\, \mathbb{P}(A_i).$$

Exercise 3.5. *Check that* $\mathbb{E}$ *is well-defined for simple rv's, i.e.,* $\mathbb{E}(\phi)$ *is the same for different representations of* ϕ*. Use the fact that*

$$\phi = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i \mathbb{1}_{A_i \cap B_j}.$$

Moreover, $\mathbb{E}$ *satisfies the conclusions in Proposition 3.1 for simple rvs.*

Step 2: **Non-negative bounded rvs.** For a non-negative bounded rv $0 \leqslant X \leqslant M$, we define

$$\mathbb{E}(X) := \sup_{\phi \leqslant X, \phi \text{ simple}} \mathbb{E}(\phi) = \inf_{\psi \geqslant X, \psi \text{ simple}} \mathbb{E}(\psi).$$

First we check that the supremum and infimum are same, so that $\mathbb{E}$ is well-defined. We divide the interval $[0, M]$ into multiple intervals, $J_0, J_1, \ldots, J_n$, where

$$J_i = \left(\frac{iM}{n}, \frac{(i+1)M}{n}\right], \quad i = 0, 1, \ldots, n.$$

We define

$$\phi_n = \sum_{i=0}^{n} \frac{iM}{n} \mathbb{1}_{X^{-1}(J_i)} \text{ and } \psi_n = \sum_{i=0}^{n} \frac{(i+1)M}{n} \mathbb{1}_{X^{-1}(J_i)}.$$

Clearly, $\phi_n \leqslant X \leqslant \psi_n \leqslant \phi_n + 1/n$. From here we have

$$\inf_{\psi \geqslant X, \psi \text{ simple}} \mathbb{E}(\psi) \leqslant \mathbb{E}(\psi_n) \leqslant \mathbb{E}(\phi_n) + 1/n \leqslant \sup_{\phi \leqslant X, \phi \text{ simple}} \mathbb{E}(\phi) + 1/n.$$

Since n is arbitrary, we have

$$\sup_{\phi \leqslant X, \phi \text{ simple}} \mathbb{E}(\phi) \geqslant \inf_{\psi \geqslant X, \psi \text{ simple}} \mathbb{E}(\psi).$$

The other direction is easy, as $\phi \leqslant X \leqslant \psi$ implies $\mathbb{E}(\phi) \leqslant \mathbb{E}(\psi)$.

Exercise 3.6. $\mathbb{E}$ *satisfies the conclusions in Proposition 3.1 for non-negative bounded rvs.*

Step 3: **Non-negative rvs.** For a non-negative rv $X \geqslant 0$, we define

$$\mathbb{E}(X) := \sup\{\mathbb{E}(h) \mid 0 \leqslant h \leqslant X, h \text{ bounded}\}.$$

Lemma 3.7. *If* $X \geqslant 0$ *then* $\mathbb{E}(X \wedge n) \uparrow \mathbb{E}(X)$ *as* $n \to \infty$.

Proof. Clearly, $\mathbb{E}(X \wedge n)$ is non-decreasing in n. Take a bounded nonnegative rv h such that $0 \leqslant h \leqslant X$. Let $h \leqslant M$ for some $M < \infty$. For $n \geqslant M$ we have $h \leqslant X \wedge n$ and thus, by definition of expectation we have

$$\mathbb{E}(h) \leqslant \mathbb{E}(X \wedge n) \text{ for all } n \geqslant M.$$

Hence, $\mathbb{E}(h) \leqslant \lim_{n\to\infty} \mathbb{E}(X \wedge n)$. Since h is arbitrary, we have

$$\mathbb{E}(X) \leqslant \lim_{n\to\infty} \mathbb{E}(X \wedge n).$$

However, $\mathbb{E}(X) \geqslant \lim_{n\to\infty} \mathbb{E}(X \wedge n)$ as $X \wedge n$ is bounded for all n. Thus equality holds. ∎

Corollary 3.8. *The lemma above implies that* $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Proof. To show that $\mathbb{E}(X)+\mathbb{E}(Y) \leqslant \mathbb{E}(X+Y)$, we note that for $0 \leqslant h \leqslant X, 0 \leqslant g \leqslant Y$, where h, g are bounded rvs, we have $0 \leqslant h + g \leqslant X + Y$ and $\mathbb{E}(h) + \mathbb{E}(g) = \mathbb{E}(h + g) \leqslant \mathbb{E}(X + Y)$. Thus,

$$\mathbb{E}(X) + \mathbb{E}(Y) \leqslant \mathbb{E}(X + Y).$$

To show the other direction, we know that $(X + Y) \wedge n \leqslant X \wedge n + Y \wedge n$ for all n. Thus $\mathbb{E}((X + Y) \wedge n) \leqslant \mathbb{E}(X \wedge n) + \mathbb{E}(Y \wedge n)$. Taking $n \to \infty$ and using Lemma 3.7 we have $\mathbb{E}(X + Y) \leqslant \mathbb{E}(X) + \mathbb{E}(Y)$ and hence the equality holds. ∎

Exercise 3.9. $\mathbb{E}$ *satisfies the conclusions in Proposition 3.1 for non-negative rvs.*

Step 4: **Integrable rvs satisfying $\mathbb{E}|X| < \infty$.** For a rv X with $\mathbb{E}(|X|) < \infty$, we define

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

where $X^+ = X \vee 0, X^- = (-X) \vee 0$.

Exercise 3.10. *For any integrable rv $\mathbb{E}$ satisfies the following:*

(i) *if $X \geqslant 0$ a.s., then $\mathbb{E}\,X \geqslant 0$,*

(ii) *for all $a \in \mathbb{R}$, $\mathbb{E}(aX) = a\,\mathbb{E}\,X$,*

(iii) $\mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y$,

(iv) *if $X \leqslant Y$ a.s., then $\mathbb{E}\,X \leqslant \mathbb{E}\,Y$,*

(v) *if $X = Y$ a.s., then $\mathbb{E}\,X = \mathbb{E}\,Y$,*

(vi) $|\mathbb{E}\,X| \leqslant \mathbb{E}\,|X|$.

Remarks:

1. $\{X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}) \mid \mathbb{E}\,|X| < \infty\} = L^1(\Omega, \mathcal{F}, \mathbb{P})$ is a Banach space.
2. $\{X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}) \mid \mathbb{E}\,|X|^p < \infty\} = L^p(\Omega, \mathcal{F}, \mathbb{P})$, where $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$, $1 \leqslant p \leqslant \infty$, is a Banach space. Here essential supremum:
$$\|X\|_\infty = \text{esssup}|X| = \inf\{k \geqslant 0 \mid \mathbb{P}(|X| > k) = 0\}.$$
3. $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with the inner product $\langle X, Y \rangle = \mathbb{E}(XY)$.

Exercise 3.11. *Show that if $\|X\|_\infty < \infty$, then $\|X\|_p \uparrow \|X\|_\infty$ as $p \uparrow \infty$.*

Definition 3.12 ($\mathbf{L^p}$ Convergence). *We say that X_n converges in L_p to X, denoted $X_n \xrightarrow{L^p} X$, if $X_n, X \in L^p$ and*

$$\|X_n - X\|_p \xrightarrow{n \to \infty} 0.$$

Lemma 3.13 (Bounded Convergence Thm (BCT)). *If $0 \leqslant X_n \leqslant M$ and $X_n \to X$ in probability, then $X_n \xrightarrow{\mathbf{L^1}} X$ and*

$$\mathbb{E}\,X_n \to \mathbb{E}\,X.$$

Proof. For $\forall \varepsilon > 0$, we have

$$\begin{aligned} |\mathbb{E}\,X_n - \mathbb{E}\,X| = |\mathbb{E}(X_n - X)| &\leqslant \mathbb{E}\,|X_n - X| \\ &= \mathbb{E}\,|X_n - X| \cdot \mathbb{1}_{|X_n - X| > \varepsilon} + \mathbb{E}\,|X_n - X| \cdot \mathbb{1}_{|X_n - X \leqslant \varepsilon|} \\ &\leqslant 2M \cdot \mathbb{P}(|X_n - X| > \varepsilon) + \varepsilon. \end{aligned}$$

Now, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$. Thus $\limsup_{n\to\infty} |\mathbb{E}\,X_n - \mathbb{E}\,X| \leqslant \varepsilon$. Since $\varepsilon > 0$ is arbitrary we have the result. ∎

Lemma 3.14 (Fatou's Lemma). *If $X_n \geqslant 0$, then*

$$\liminf_n \mathbb{E}(X_n) \geqslant \mathbb{E}(\liminf_n X_n).$$

Example 3.15. *Let $X_n = n \cdot \mathbb{1}_{(0,\frac{1}{n}]}$. Then $\mathbb{E}(X_n) = n \cdot \frac{1}{n} = 1, \liminf_n X_n \equiv 0$, so $\mathbb{E}(\liminf_n X_n) = 0$. Thus $\liminf_n \mathbb{E}(X_n) > \mathbb{E}(\liminf_n X_n)$.*

Proof of Fatou's Lemma. Let $Y_n := \inf_{m \geqslant n} X_m$. Then $X_n \geqslant Y_n$ and $\liminf_n X_n = \sup_n Y_n$. By monotonicity, we have

$$\liminf_n \mathbb{E}(X_n) \geqslant \liminf_n \mathbb{E}(Y_n).$$

Thus it suffices to show that

$$\liminf_n \mathbb{E}(Y_n) \geqslant \mathbb{E}(\sup_n Y_n).$$

For all $M > 0, Y_n \wedge M \uparrow \sup Y_n \wedge M$. By BCT,

$$\mathbb{E}(Y_n \wedge M) \to \mathbb{E}(\sup_n Y_n \wedge M).$$

So

$$\liminf_n \mathbb{E}(Y_n) \geqslant \liminf_n \mathbb{E}(Y_n \wedge M) = \mathbb{E}(\sup_n Y_n \wedge M).$$

By taking $M \to \infty$, $\liminf_n \mathbb{E}(Y_n) \geqslant \mathbb{E}(\sup_n Y_n)$. This completes the proof. ∎

3.2 Properties of Expectation

We will use Fatou's lemma to prove Monotone Convergence Theorem and Dominated Convergence Theorem.

Theorem 3.16 (Monotone Convergence Thm (MCT)). *If $X_n \geqslant 0$ for all n, and $X_n \uparrow X$, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.*

Proof. By Fatou's Lemma,

$$\liminf_n \mathbb{E}(X_n) \geqslant \mathbb{E}(\liminf X_n) = \mathbb{E}(X).$$

However, because $X_n \uparrow X$, we have $\liminf_n \mathbb{E}(X_n) \leqslant \mathbb{E}(X)$. Thus $\liminf_n \mathbb{E}(X_n) = \mathbb{E}(X)$. ∎

Theorem 3.17 (Dominated Convergence Thm (DCT)). *If $X_n \to X$ and $0 \leqslant |X_n| \leqslant Y$, with $\mathbb{E}(Y) < \infty$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

Proof. Since $X_n \to X$ and $0 \leqslant |X_n| \leqslant Y$, $X_n + Y \geqslant 0$ and converges to $X + Y$ as $n \to \infty$. By Fatou's Lemma, we have

$$\liminf_n \mathbb{E}(X_n + Y) \geqslant \mathbb{E}(X + Y)$$

which implies that

$$\liminf_n \mathbb{E}(X_n) + \mathbb{E}(Y) \geqslant \mathbb{E}(X) + \mathbb{E}(Y) \text{ and } \liminf_n \mathbb{E}(X_n) \geqslant \mathbb{E}(X).$$

Similarly, $-X_n + Y \geqslant 0$ and converges to $-X + Y$, so

$$\liminf_n -\mathbb{E}(X_n) \geqslant -\mathbb{E}(X) \text{ implies } \limsup_n \mathbb{E}(X_n) \leqslant \mathbb{E}(X).$$

Thus we have $\mathbb{E}(X_n) \to \mathbb{E}(X)$. ■

Theorem 3.18 (Change of variables formula). *Let $X : \Omega \to \mathbb{R}$ be a random variable. For any random variable $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\,|f(X)| < \infty$,*

$$\int f(X) d\,\mathbb{P} = \int f d\,\mathbb{P}_X\,.$$

Sketch of the proof: We can verify this using the 4-step procedure, checking:

1. Indicator rvs: Let $f = \mathbb{1}_A$. Then $\int f(X)d\,\mathbb{P} = \mathbb{P}(X \in A) = \mathbb{P}_X(A) = \int f d\,\mathbb{P}_X$ and simple rvs.
2. Non-negative bounded functions.
3. Non-negative functions.
4. Any measurable functions for which both sides make sense. ■

3.3 Remarks about Expectation

Exercise 3.19. *Let $\Omega = \{1, 2, ...\}$, $\mathcal{F} = 2^{\Omega}$. Let $p_i \geqslant 0$ for all $i \in \mathbb{N}$ and suppose $\sum_{i=1}^{\infty} p_i = 1$. Define $\mathbb{P}(A) = \sum_{i \in A} p_i$ for all $A \in \mathcal{F}$. Then for any random variable $X : \Omega \to \mathbb{R}$, check that*

$$\mathbb{E}\,X = \sum_{i=1}^{\infty} X(i) p_i.$$

In particular, if $\Omega = \{1, ..., N\}$ is finite, and $p_i = 1/N$ for $i = 1, 2, ..., N$, then

$$\mathbb{E}\,X = \frac{1}{N} \sum_{i=1}^{N} X(i).$$

Exercise 3.20. *Consider the probability space $((0, 1), \mathcal{B}(0, 1), \mathbb{P} =$ Lebesgue measure$)$. For any random variable $X : (0, 1) \to \mathbb{R}$, check that*

$$\mathbb{E}(X) = \int_0^1 X(x) dx.$$

Exercise 3.21. *Consider the probability space $(\Omega = \mathbb{R}, \mathcal{B}, \mathbb{P})$, where $\mathbb{P}$ has distribution function F. Check that*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} X(x) dF(x).$$

In particular, if F has density $f(x)$, i.e. $F(x) = \int_{-\infty}^{x} f(x) dx$, and F is differentiable a.e., then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} X(x) f(x) dx.$$

Not all distribution functions have densities. A measure is called **absolutely continuous** with respect to Lebesgue measure if it has a density and a measure is called **countable** if there exists a countable set with probability one.

3.4 Inequalities

Theorem 3.22 (Jensen's Inequality). *If ϕ is a convex function on the range of a random variable X, then*

$$\phi(\mathbb{E}\,X) \leqslant \mathbb{E}\,\phi(X) \tag{3.1}$$

assuming both sides exist.

Proof Sketch: A function is convex if it is the supremum of a set of linear functions (think of tangent lines), *i.e.*, there exists a collection of linear functions $A = \{a_\ell x + b_\ell \mid \ell \in \Lambda\}$ such that $\phi(x) = \sup_{\ell\in\Lambda}\{a_\ell x + b_\ell\}$. Clearly $\phi(x) \geqslant a_\ell x + b_\ell$ for all $\ell \in \Lambda$. By monotonicity and linearity of $\mathbb{E}$,

$$\mathbb{E}\,\phi(X) \geqslant a_l\,\mathbb{E}\,X + b_l \implies \mathbb{E}\,\phi(X) \geqslant \sup_{l\in\Lambda}\{a_l\,\mathbb{E}\,X + b_l\} = \phi(\mathbb{E}\,X).$$ ■

Corollary 3.23. *Let $1 \leqslant p < q < \infty$. For any random variable X, $||X||_p \leqslant ||X||_q$, i.e.,*

$$(\mathbb{E}\,|X|^p)^{\frac{1}{p}} \leqslant (\mathbb{E}\,|X|^q)^{\frac{1}{q}}. \tag{3.2}$$

Proof Sketch: Define $\phi(x) = |x|^{q/p}$. Then $\phi(x)$ is convex since $q/p > 1$. Take $Y = |X|^p$. By Jensen's inequality

$$(\mathbb{E}\,|X|^p)^{\frac{q}{p}} = \phi(\mathbb{E}(Y)) \leqslant \mathbb{E}(\phi(Y)) = \mathbb{E}\,|X|^q.$$ ■

Exercise 3.24. *Let X be a random variable and suppose $1 \leqslant p < \infty$. Prove that $||X||_p \leqslant ||X||_\infty$.*

Theorem 3.25 (Markov's Inequality). *Let X be a non-negative random variable. For $a > 0$, we have*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}\,X}{a}$$

Proof Sketch: Note that $1_{X\geqslant a} \leqslant X/a$. So

$$\mathbb{P}(X \geqslant a) = \mathbb{E}(1_{X\geqslant a}) \leqslant \mathbb{E}\left(\frac{X}{a}\right) = \frac{\mathbb{E}\,X}{a},$$

here we used monotonicity of expectation in the inequality. ■

Theorem 3.26 (Chebyshev's Inequality). *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a nondecreasing, non-negative function. Then for any $a \geqslant 0$ where $\phi(a) \neq 0$,*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}\,\phi(X)}{\phi(a)}.$$

Proof Sketch: By Markov's inequality,

$$\mathbb{P}(X \geqslant a) \leqslant \mathbb{P}\left(\phi(X) \geqslant \phi(a)\right) \leqslant \frac{\mathbb{E}\,\phi(X)}{\phi(a)}.$$ ■

Definition 3.27. *Let X be a random variable. The* ***variance*** *of X is*

$$\mathrm{Var}(X) := \mathbb{E}(X - \mathbb{E}\,X)^2$$

whenever $\mathbb{E}\,X^2 < \infty$.

Note that $\mathrm{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 \geqslant 0$. The original Chebyshev's inequality is the following:

Theorem 3.28 (Original Chebyshev Inequality). *Let X be a random variable with finite variance. For any $a > 0$,*

$$\mathbb{P}(|X - \mathbb{E}\,X| > a) \leqslant \frac{\mathrm{Var}(X)}{a^2}.$$

Below we list two other useful inequalities.

Theorem 3.29 (Holder's Inequality). *Let $1 \leqslant p, q < \infty$, where $\frac{1}{p} + \frac{1}{q} = 1$. Then for any random variables X, Y,*

$$\mathbb{E}(XY) \leqslant ||X||_p ||Y||_q = (\mathbb{E}\,|X|^p)^{\frac{1}{p}} (\mathbb{E}\,|Y|^q)^{\frac{1}{q}}.$$

whenever the RHS is finite. If $p = q = 2$, this is called the Cauchy-Schwarz inequality.

Theorem 3.30 (Minkowski's Inequality). *For any $p \geqslant 1$, and random variables X, Y,*

$$||X + Y||_p \leqslant ||X||_p + ||Y||_p.$$