

STAT 7200

Introduction to Advanced Probability

Lecture 10

Taylor R. Brown

- Expectations
 - Expectations of Simple Random Variables
 - Expectation of Non-Negative Random Variable
 - Expectations of Arbitrary Random Variables
 - Expectations and Integrals
 - Expectations and Integrals

“A First Look at Rigorous Probability Theory” (Jeffrey Rosenthal)
Sections 4.3 and 4.4

Expectations of Simple Random Variables

- A random variable X is simple if it only takes finite number of values: $X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$, where A_1, A_2, \dots, A_n forms a partition of Ω .
- Then we can define the expectation of simple random variable as:
$$\mathbf{E}(X) = \sum_{i=1}^n x_i \mathbf{P}(A_i).$$
- **Linearity** For simple random variables X, Y , we have
$$\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y) \quad (a, b \in \mathbf{R}).$$
- **Order Preserving** For simple random variables X, Y , if $X \leq Y$ for every ω , then we have $\mathbf{E}(X) \leq \mathbf{E}(Y)$.
- **Expectation and Independence** If X, Y are simple random variables and $X \perp Y$, then $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.

Expectation of Non-Negative Random Variable

- For a non-negative random variable X , we define its expectation as the supremum of the expectations of all the simple random variables Y not greater than X . In other words:

$$\mathbf{E}(X) = \sup\{\mathbf{E}(Y) : Y \text{ simple}, Y \leq X\}$$

- Given any non-negative random variable X , we can find a sequence of increasing simple random variables $\Phi_n(X)$, so that $\Phi_n(X) \rightarrow X$ as $n \rightarrow \infty$.
- We used this to show that expectations are still order-preserving.
- The monotone convergence theorem gives us a sufficient condition for $\lim_n \mathbf{E}(\Phi_n(X)) = \mathbf{E}(X)$. We can use that to derive linearity and factoring properties for independent random variables.

The Monotone Convergence Theorem

Recall the MCT:

Theorem 1 (The Monotone Convergence Theorem)

If X_1, X_2, \dots are non-negative random variables so that $X_1 \leq X_2 \leq \dots$. Suppose that $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ for each $\omega \in \Omega$, then X is a random variable and $\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

Property of Expectations of Non-Negative Random Variables

- Last class we showed that expectations are linear for *finite* linear combinations. We can extend this to countable linear combinations:
- **Countable Linearity** If $X_1, X_2, \dots \geq 0$, then define $M_n = \sum_{i=1}^n X_i$. We have:

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n\right) = \mathbf{E}\left(\lim_{n \rightarrow \infty} M_n\right) = \lim_{n \rightarrow \infty} \mathbf{E}(M_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{E}(X_i) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$$

Expectations of Arbitrary Random Variables

- For any random variable X , we may split it into the positive part and the negative part: $X^+(\omega) = \max(X(\omega), 0)$, $X^-(\omega) = \max(-X(\omega), 0)$, then $X = X^+ - X^-$, and $X^+, X^- \geq 0$.
- We can then define the expectation of arbitrary random variables as

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

- Note that the above expectation is only defined when at least one of $\mathbf{E}(X^+)$, $\mathbf{E}(X^-)$ is finite, if not, the expectation is undefined ($\infty - \infty$ is undefined).
- **Order Preserving** For arbitrary random variables X, Y with well defined expectation, if $X \leq Y$ then $\mathbf{E}(X) \leq \mathbf{E}(Y)$. (See Exercise 4.3.2)
- **Independence and Expectation** For arbitrary random variables X, Y with finite expectation, if $X \perp Y$ then $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. (See Exercise 4.3.4)

Properties of Expectation: Finite Linearity

- **Linearity** $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ for any random variables X, Y with finite expectation.
- **Proof:**
 - First, if $X, Y \geq 0$, $X - Y \geq 0$, then $\mathbf{E}(X - Y) = \mathbf{E}(X) - \mathbf{E}(Y)$. This is true since $\mathbf{E}(X) = \mathbf{E}(X - Y + Y) = \mathbf{E}(X - Y) + \mathbf{E}(Y)$ (Linearity of non-negative random variables).
 - Second, if $X, Y \geq 0$, then $\mathbf{E}(X - Y) = \mathbf{E}(X) - \mathbf{E}(Y)$.
To prove this, let us denote $A = \{\omega : X(\omega) \geq Y(\omega)\}$. Then $(X - Y)^+ = (X - Y)\mathbf{1}_A \geq 0$ and $(X - Y)^- = (Y - X)\mathbf{1}_{A^c} \geq 0$:
$$\begin{aligned}\mathbf{E}(X - Y) &= \mathbf{E}[(X - Y)\mathbf{1}_A] - \mathbf{E}[(Y - X)\mathbf{1}_{A^c}] \\ &= \mathbf{E}(X\mathbf{1}_A) - \mathbf{E}(Y\mathbf{1}_A) - \mathbf{E}(Y\mathbf{1}_{A^c}) + \mathbf{E}(X\mathbf{1}_{A^c}) \\ &= \mathbf{E}[X(\mathbf{1}_A + \mathbf{1}_{A^c})] - \mathbf{E}[Y(\mathbf{1}_A + \mathbf{1}_{A^c})] = \mathbf{E}(X) - \mathbf{E}(Y)\end{aligned}$$

Properties of Expectation: Linearity

- **Proof (continued):** Now let us return to general X, Y .

Denote $A = \{\omega : X(\omega) + Y(\omega) \geq 0\}$. Then

$$(X + Y)^+ = (X + Y)\mathbf{1}_A = (X^+ + Y^+ - X^- - Y^-)\mathbf{1}_A \text{ and}$$

$$(X + Y)^- = (X^- + Y^- - X^+ - Y^+)\mathbf{1}_{A^c}.$$

$$\text{ - } \mathbf{E}(X + Y) = \mathbf{E}[(X^+ + Y^+ - X^- - Y^-)\mathbf{1}_A] - \mathbf{E}[(X^- + Y^- - X^+ - Y^+)\mathbf{1}_{A^c}]$$

(If $X, Y \geq 0$, then $\mathbf{E}(X - Y) = \mathbf{E}(X) - \mathbf{E}(Y)$)

$$\begin{aligned} &= \mathbf{E}[(X^+ + Y^+)\mathbf{1}_A] - \mathbf{E}[(X^- + Y^-)\mathbf{1}_A] - \mathbf{E}[(X^- + Y^-)\mathbf{1}_{A^c}] + \\ &\mathbf{E}[(X^+ + Y^+)\mathbf{1}_{A^c}] \end{aligned}$$

(By the linearity of non-negative random variables)

$$\begin{aligned} &= \mathbf{E}[(X^+ + Y^+)(\mathbf{1}_A + \mathbf{1}_{A^c})] - \mathbf{E}[(X^- + Y^-)(\mathbf{1}_A + \mathbf{1}_{A^c})] \\ &= \mathbf{E}[X^+ + Y^+] - \mathbf{E}[X^- + Y^-] = \mathbf{E}(X^+) + \mathbf{E}(Y^+) - \mathbf{E}(X^-) - \mathbf{E}(Y^-) \\ &= \mathbf{E}(X) + \mathbf{E}(Y) \end{aligned}$$

Properties of Expectation: Linearity

For arbitrary random variables with finite expectations X , Y , and for any two real numbers a, b , we have $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$.

From the previous result, it's clear that

$\mathbf{E}[aX + bY] = \mathbf{E}[aX] + \mathbf{E}[bY]$. All we need to show is that the nonrandom real numbers can be pulled out of the expectation.

WLOG assume $a < 0$: $\mathbf{E}[aX] = \mathbf{E}[(aX)^+] - \mathbf{E}[(aX)^-] = -\mathbf{E}[a(X)^-] + \mathbf{E}[a(X)^+] = a\mathbf{E}[(X)^+] - a\mathbf{E}[(X)^-] = a\mathbf{E}[X]$.

Expectations and Integrals

- Our previous discussion allow us to define the expectation $\mathbf{E}(X)$ of any random variable X on probability triple $(\Omega, \mathcal{F}, \mathbf{P})$. The expectation we defined is also called as *Lebesgue integral*, and can be denoted as:

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) \mathbf{P}(d\omega)$$

- On the real space \mathbf{R} the expectation or Lebesgue integral can be viewed as a generalization of the usual Riemann integral.
- This is the reason why we can calculate the expectations of discrete random variables as the summation with respect to the probability mass function, and the expectations of continuous random variables as Riemann integrals with respect to the probability density functions.
- Also, there are examples of random variables that are not Riemann integrable, but are Lebesgue integrable.

Expectations and Integrals

Before we get into this, we need to review two properties from the previous chapters, that you might not have shown:

- Remark 4.2.3: Since expected values are unchanged if we modify the random variable values on sets of probability 0, we still have $\lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \mathbf{E}[X]$ if $\{X_n\} \nearrow X$ almost surely.
- Proof: let $A = \{\omega : \{X_n\} \nearrow X\}$. By the standard MCT: $\lim_{n \rightarrow \infty} \mathbf{E}[X_n 1_A] = \mathbf{E}[X 1_A]$. Then $\lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n 1_A] + \lim_{n \rightarrow \infty} \mathbf{E}[X_n 1_{A^c}] = \mathbf{E}[X 1_A] + 0 = \mathbf{E}[X]$
- Remark 3.1.9: "If $(\Omega, \mathcal{F}, \mathbf{P})$ is complete, and if X is a random variable, and $Y : \Omega \rightarrow \mathbf{R}$ such that $P(X = Y) = 1$, then Y must also be a random variable.
- Proof: The only set we are guaranteed is measurable is $\{Y = X\}$. Let $A \in \mathcal{F}$ be arbitrary. Then $\{Y \in A\} = \{Y = X, X \in A\} \cup \{Y \in A, Y \neq X\}$. $\{Y \in A, Y \neq X\}$ is measurable because it is a subset of $\{X \neq Y\}$. $\{Y = X, X \in A\}$ is the intersection of two measurable sets. Therefore $\{Y \in A\}$ is measurable.

Expectations and Riemann Integrals

- Without loss of generality, let us consider any Riemann integrable function $X(t)$ over $[0, 1]$. We will show that X is a random variable over the uniform measure (Lebesgue measure) $(\Omega, \mathcal{F}, \mathbf{P})$ on $\Omega = [0, 1]$, and $\mathbf{E}(X) = \int_0^1 X(t)dt$, the Riemann integral.
- First, recall the definition of Riemann integral $\int_0^1 X(t)dt$. For any $\mathbf{s} = (s_0, s_1, \dots, s_n)$ such that $0 = s_0 < s_1 < \dots < s_n = 1$, let step functions $L_{\mathbf{s}}(t) = \inf_{s_{i-1} \leq t \leq s_i} X(t)$ and $U_{\mathbf{s}}(t) = \sup_{s_{i-1} \leq t \leq s_i} X(t)$ for $s_{i-1} \leq t < s_i$. Then $L_{\mathbf{s}}(t) \leq X(t) \leq U_{\mathbf{s}}(t)$.
- Then we may define the lower integral L and upper integral U as:
$$L = \sup\{\sum_i (s_i - s_{i-1}) \inf_{s_{i-1} \leq t \leq s_i} X(t) \text{ for all } \mathbf{s}\},$$
$$U = \inf\{\sum_i (s_i - s_{i-1}) \sup_{s_{i-1} \leq t \leq s_i} X(t) \text{ for all } \mathbf{s}\}$$
- Since we assume that Riemann integral $\int_0^1 X(t)dt$ exists so we have $L = U = \int_0^1 X(t)dt$.

Expectations and Riemann Integrals

- We may also view L_s, U_s as simple random variables on $(\Omega, \mathcal{F}, \mathbf{P})$:
$$\mathbf{E}(L_s) = \sum_i (s_i - s_{i-1}) \inf_{s_{i-1} \leq t \leq s_i} X(t),$$
$$\mathbf{E}(U_s) = \sum_i (s_i - s_{i-1}) \sup_{s_{i-1} \leq t \leq s_i} X(t).$$
Lower integral $L = \sup_s \mathbf{E}(L_s)$, and upper integral $U = \inf_s \mathbf{E}(U_s)$
- As $L = \sup_s \mathbf{E}(L_s)$, for any n , we should be able to find a simple random variable $L_n \leq X$, so that $\mathbf{E}(L_n) \geq L - \frac{1}{n}$. Similarly, we can find simple random variable $U_n \geq X$ so that $\mathbf{E}(U_n) \leq U + \frac{1}{n}$.
- We then let $A_n = \max(L_1, \dots, L_n)$, $B_n = \min(U_1, \dots, U_n)$. It is easy to see that A_n is an increasing sequence of random variables and B_n is a decreasing sequence of random variables.
- Since $L_n \leq A_n \leq X \leq B_n \leq U_n$, we have
$$L \leq \lim_{n \rightarrow \infty} \mathbf{E}(A_n) \leq \lim_{n \rightarrow \infty} \mathbf{E}(B_n) \leq U.$$
Thus $\lim_{n \rightarrow \infty} \mathbf{E}(A_n) = \lim_{n \rightarrow \infty} \mathbf{E}(B_n) = L = U = \int_0^1 X(t) dt$.
- If we can show that A_n converges to X , then X must be a random variable. Furthermore, by the monotone convergence theorem,
$$\lim_{n \rightarrow \infty} \mathbf{E}(A_n) = \mathbf{E}(X) = \int_0^1 X(t) dt.$$

Expectations and Riemann Integrals

- To see why A_n converges to X . Let us denote $S_k = \{\omega : \lim_{n \rightarrow \infty} (B_n - A_n) \geq 1/k\}$. That is, S_k represents the set on which the limit of B_n and A_n differs at least by $1/k$. If we could prove that the probability of S_k equals 0 for all k , we can then conclude that the set on which the limit of B_n and A_n differs has probability 0. But since $A_n \leq X \leq B_n$, we can conclude that both A_n and B_n should converge to X with probability 1.
- As $0 = \lim_{n \rightarrow \infty} \mathbf{E}(B_n - A_n) \geq \frac{1}{k} \mathbf{P}(S_k)$. So $\mathbf{P}(S_k) = 0$.
- Then $\mathbf{P}(\lim_{n \rightarrow \infty} A_n < X) \leq \mathbf{P}(\lim_{n \rightarrow \infty} A_n < \lim_{n \rightarrow \infty} B_n) \leq \mathbf{P}(\bigcup_k S_k) \leq \sum_k \mathbf{P}(S_k) = 0$. So $\{A_n\} \nearrow X$ almost surely.
- **Conclusion:** Any Riemann integrable function $X(t)$ over $[0, 1]$ is also a random variable over the uniform measure (Lebesgue measure) $(\Omega, \mathcal{F}, \mathbf{P})$ on $\Omega = [0, 1]$, and $\mathbf{E}(X) = \int_0^1 X(t) dt$.