STAT 7200

Introduction to Advanced Probability
Lecture 9

Taylor R. Brown

1. Foundations of Probability
   - Expectations of Simple Random Variable
   - Expectations of Non-Negative Random Variables

"A First Look at Rigorous Probability Theory" (Jeffrey Rosenthal)
Sections 4.1 and 4.2

Expectations of Simple Random Variable

- Over a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, if we have an indicator random variable $\mathbf{1}_A$ on $A \in \mathcal{F}$, so that $\mathbf{1}_A = 1$ when $\omega \in A$, and $\mathbf{1}_A = 0$ when $\omega \notin A$. Then we can define expectation of this indicator random variable as: $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$.
- Similarly, we can extend this definition to *simple random variables*. A random variable X is simple if it only takes a finite number of values. If we list the possible values that X may take as $x_1, x_2, \cdots, x_n$, we should be able to represent X as: $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$, where $A_1, A_2, \cdots, A_n$ forms a partition of Ω.
- Then we define the expectation of simple random variable as: $\mathbf{E}(X) = \sum_{i=1}^{n} x_i \mathbf{P}(A_i)$.

Property of Expectations: Linearity

- **Linearity** The expectation of a simple random variable is linear. That is, for two simple random variables X, Y and $a, b \in \mathbf{R}$, we have $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$.
- **Proof**: Let us denote $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$ and $Y = \sum_{j=1}^{m} y_j \mathbf{1}_{B_j}$. Since $\{A_i\}$ forms a partition of Ω, $\{B_i\}$ forms a partition of Ω, $\{A_i \cap B_j\}$ also forms a partition of Ω. Then $aX+bY = \sum_{i=1}^{n} ax_i \mathbf{1}_{A_i} + \sum_{j=1}^{m} by_j \mathbf{1}_{B_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_j) \mathbf{1}_{A_i \cap B_j}$.

- So

$$
\begin{aligned}
\mathbf{E}(aX + bY) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_j) \mathbf{P}(A_i \cap B_j) \\
&= \sum_{i=1}^{n} ax_i \left[\sum_{j=1}^{m} \mathbf{P}(A_i \cap B_j) \right] + \sum_{j=1}^{m} by_j \left[\sum_{i=1}^{n} \mathbf{P}(A_i \cap B_j) \right] \\
&= \sum_{i=1}^{n} ax_i \mathbf{P}(A_i) + \sum_{j=1}^{m} by_j \mathbf{P}(B_j) = a\mathbf{E}(X) + b\mathbf{E}(Y)
\end{aligned}
$$

Property of Expectations: Others

- **Consequence of Linearity** By linearity of expectation, for $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$ where $A_1, \cdots A_n$ may not form a partition of Ω, we still have $\mathbf{E}(X) = \sum_{i=1}^{n} x_i \mathbf{P}(A_i)$.
- **Order Preserving** The expectation of simple random variable preserves the order, that is, for simple random variables X, Y, if $X \leq Y$ for every ω, then we have $\mathbf{E}(X) \leq \mathbf{E}(Y)$.
- **Proof**: This property is quite obvious since $X \leq Y$ implies $Y - X \geq 0$, then $\mathbf{E}(Y - X) \geq 0$ and we have $\mathbf{E}(X) \leq \mathbf{E}(Y)$.
- A direct consequence of order preservation is the **triangle inequality**: since $-|X| \leq X \leq |X|$, we have $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$.
- **Functions of Simple Random Variables** Suppose X is simple random variable $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$. Given any function $f : \mathbf{R} \to \mathbf{R}$, $f(X) = \sum_{i=1}^{n} f(x_i) \mathbf{1}_{A_i}$ is also a simple random variable and $\mathbf{E}(f(X)) = \sum_{i=1}^{n} f(x_i) \mathbf{P}(A_i)$.

Expectation and Independence

- **Expectation and Independence** If X, Y are simple random variables and $X \perp Y$, then $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.
- **Proof**: Denote $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$ and $Y = \sum_{j=1}^{m} y_j \mathbf{1}_{B_j}$, and without loss of generality, suppose $\{x_i\}$ are distinct and $\{y_j\}$ are distinct.

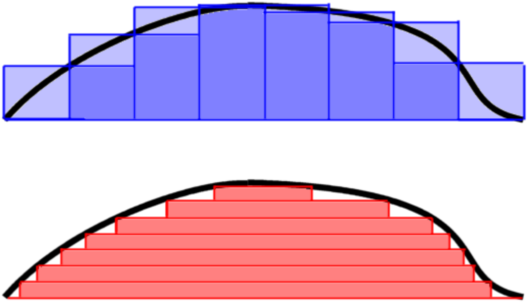
- Since $X \perp Y$, $\mathbf{P}(X = x_i, Y = y_j) = \mathbf{P}(X = x_i)\mathbf{P}(Y = y_j)$, then we have $\mathbf{P}(A_i \cap B_j) = \mathbf{P}(A_i)\mathbf{P}(B_j)$
- $XY = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \mathbf{1}_{A_i \cap B_j}$ is a simple random variable and

$$\begin{aligned}
\mathbf{E}(XY) &= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \mathbf{P}(A_i \cap B_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \mathbf{P}(A_i)\mathbf{P}(B_j) \\
&= \left[\sum_{i=1}^{n} x_i \mathbf{P}(A_i) \right] \left[\sum_{j=1}^{m} y_j \mathbf{P}(B_j) \right] = \mathbf{E}(X)\mathbf{E}(Y)
\end{aligned}$$

Expectations of Non-Negative Random Variables

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

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



Expectations of Non-Negative Random Variables

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

- First, this definition does not contradict the definition for simple random variables since $\mathbf{E}(X) = \sup\{\mathbf{E}(Y) : \text{Y } \textit{simple}, Y \leq X\}$ if X is simple.
- Second, this definition still preserves orderings: If X_1 and X_2 are two non-negative random variables so that $X_1 \leq X_2$, then $\mathbf{E}(X_1) \leq \mathbf{E}(X_2)$.
- Example: $E[X^k] < \infty \implies E[X^{k-1}] < \infty$ because $x^{k-1} \leq \max(x^k, 1) \leq 1 + x^k$
- Third, the expectation might be infinite. Example: $X(\omega) = \sum_{n=1}^{\infty} 2^n 1(2^{-n} \leq \omega < 2^{-(n-1)})$ on $([0,1], \mathcal{F}, \mathbf{P})$
- Proving linearity requires another result...

The Monotone Convergence Theorem

Theorem 1 (The Monotone Convergence Theorem)

If $X_1, X_2, \dots$ are non-negative random variables such that $\{X_n\} \nearrow X$. Then X is a random variable and $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

$$\{X_n\} \nearrow X \text{ means } X_1 \leq X_2 \leq \dots \text{ and } \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Proof on next slide...

The Monotone Convergence Theorem

Proof.

For any real x, $\{X \leq x\} = \cap_n \{X_n \leq x\}$, so the limit of random variables is still a rv.

By monotonicity, $\mathbf{E}[X_n] \leq \mathbf{E}[X]$ for all. Taking the limit yields $\lim_{n\to\infty} \mathbf{E}[X_n] \leq \mathbf{E}[X]$. The limit exists because it is a monotonic sequence, and it may be infinite.

Last, pick $Y = \sum_{i=1}^m y_i 1_{A_i}$ be a simple rv such that $Y \leq X$ and such that $\{A_i\}$ partitions Ω. Pick an $0 < \epsilon$, and for each i, define $A_{in} = \{\omega \in A_i : X_n(\omega) \geq y_i - \epsilon\}$ (not a partition). Clearly $\{A_{in}\} \nearrow A_i$ for any i. For a fixed n,
$\mathbf{E}[X_n] \geq \sum_{i=1}^m (y_i - \epsilon)\mathbf{P}(A_{in}) = \sum_{i=1}^m y_i \mathbf{P}(A_{in}) - \epsilon \mathbf{P}(\cup_{i=1}^m A_{in})$. Taking the limit: $\lim_{n\to\infty} \mathbf{E}[X_n] \geq \sum_{i=1}^m y_i \mathbf{P}(A_i) - \epsilon$. This is true for any epsilon, and any simple random variable $Y \leq X$, and so the result holds.

$\square$

The Monotone Convergence Theorem

You can't always move the limit inside and outside of the expectation operator.

Example, on $([0,1], \mathcal{F}, \mathbf{P})$ consider $X_n(\omega) = n1_{(0,n^{-1})}$.

Non-Negative Random Variables as a Limit of Simple Random Variables

- Given any non-negative random variable X, we will construct a sequence of simple random variable $\Psi_n(X)$, such that the expectation of $\Psi_n(X)$ would approach the expectation of X.
- To construct $\Psi_n(X)$, for each n:
- If $X \geq n$, $\Psi_n(X) = n$.
- When $X < n$, we divide the region $[0, n)$ evenly into $n2^n$ intervals.
  - For instance, if $n = 1$, we will divide $[0, 1)$ into $[0, 1/2), [1/2, 1)$;
  - If $n = 2$, we divide $[0, 2)$ into $[0, 1/4), [1/4, 1/2), \cdots, [7/4, 2))$.
- If $k/2^n \leq X < (k+1)/2^n$ $(0 \leq k \leq n2^n - 1)$, $\Psi_n(X) = k/2^n$.

- This definition ensures that 1) $\Psi_n(X)$ is simple, as it only takes at most $n2^n + 1$ different values; 2) $\Psi_n(X) \leq X$; 3) $\Psi_n(X)$ forms a sequence of increasing random variables, and 4.) $\Psi_n(x) \to x$ as $n \to \infty$.

Property of Expectations of Non-Negative Random Variables

- As $\Psi_n(X) \to X$ as $n \to \infty$, by the monotone convergence theorem, we have $\lim_{n\to\infty} \mathbf{E}(\Psi_n(X)) = \mathbf{E}(X)$. Then we may prove the following properties for non-negative random variables based on the similar properties for simple random variables.
- **Linearity** For non-negative random variables X, Y, and $a, b > 0$, we have $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$.

  Proof We may construct $\Psi_n(X) \to X$ and $\Psi_n(Y) \to Y$, then $a\Psi_n(X) + b\Psi_n(Y)$ are an increasing sequence of non-negative random variables that converge to $aX + bY$. By the monotone convergence theorem:

$$\begin{aligned}\mathbf{E}(aX + bY) &= \lim_{n\to\infty} \mathbf{E}(a\Psi_n(X) + b\Psi_n(Y)) \\ &= \lim_{n\to\infty} [a\mathbf{E}(\Psi_n(X)) + b\mathbf{E}(\Psi_n(Y))] = a\mathbf{E}(X) + b\mathbf{E}(Y)\end{aligned}$$

Property of Expectations of Non-Negative Random Variables

- **Expectation and Independence** For non-negative random variables $X \perp Y$, we have $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. ($\Psi_n(X)$ is a function of X, $\Psi_n(Y)$ is a function of Y, then if $X \perp Y$, $\Psi_n(X) \perp \Psi_n(Y)$)
- **Proof** We may construct $\Psi_n(X) \to X$ and $\Psi_n(Y) \to Y$, then $\Psi_n(X)\Psi_n(Y)$ are an increasing sequence of non-negative random variables that converge to $aX + bY$. Furthermore, as $\Psi_n(X)$ is a function of X, $\Psi_n(Y)$ is a function of Y, then if $X \perp Y$, $\Psi_n(X) \perp \Psi_n(Y)$.

\- By the monotone convergence theorem:

$$\begin{aligned}
\mathbf{E}(XY) &= \lim_{n\to\infty} \mathbf{E}(\Psi_n(X)\Psi_n(Y)) \\
&= \lim_{n\to\infty} [\mathbf{E}(\Psi_n(X))\mathbf{E}(\Psi_n(Y))] = \mathbf{E}(X)\mathbf{E}(Y)
\end{aligned}$$