STAT 7200

Introduction to Advanced Probability
Lecture 8

Taylor R. Brown

"A First Look at Rigorous Probability Theory" (Jeffrey Rosenthal)
Sections 3.3, 3.4, and 4.1

Uniqueness of Extension Theorem: continued

Corollary 1

Let $\mathcal{J}$ be a semi-algebra and $\mathcal{F}$ be the $\sigma -$ algebra generated by $\mathcal{J}$. Let $\mathbf{P}$ *and* $\mathbf{Q}$ *be two probability measures over $\mathcal{F}$, so that* $\mathbf{P}(A) = \mathbf{Q}(A)$ *for any* $A \in \mathcal{J}$. *Then* $\mathbf{P}(A) = \mathbf{Q}(A)$ *for any* $A \in \mathcal{F}$.

Corollary 2

Let $\mathbf{P}$ *and* $\mathbf{Q}$ *be two probability measures over $\mathcal{B}$, the collection of Borel sets, so that* $\mathbf{P}((-\infty, x]) = \mathbf{Q}((\infty, x])$ *for any* $x \in \mathbf{R}$. *Then* $\mathbf{P}(A) = \mathbf{Q}(A)$ *for any* $A \in \mathcal{B}$.

Independence and Uniqueness of Measure

Proposition 3

Let X and Y be random variables jointly defined on the probability triple $(\Omega, \mathcal{F}, \mathbf{P})$. Then X and Y are independent if and only if $\mathbf{P}(X \leq x, Y \leq y) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y)$ for all the real numbers x and y.

- **proof** The "only if" part is trivial. To prove the "if" part, we need to show $\mathbf{P}(X \leq x, Y \leq y) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y)$ implies $\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B)$ for all Borel sets A and B, which requires the uniqueness of extension theorem.

- Define $\mathbf{Q}_x(S) = \mathbf{P}(X \leq x, Y \in S)/\mathbf{P}(X \leq x)$ for fixed x with $\mathbf{P}(X \leq x) > 0$. We can verify $\mathbf{Q}_x$ is a proper probability measure over $\mathcal{B}$. Furthermore, since $\mathbf{P}(X \leq x, Y \leq y) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y)$, we have $\mathbf{Q}_x((-\infty, y]) = \mathbf{P}(Y \leq y)$. Then the uniqueness of the extension theorem ensures that $\mathbf{Q}_x(B) = \mathbf{P}(Y \in B)$ for all Borel set B, which further implies $\mathbf{P}(X \leq x, Y \in B) = \mathbf{P}(X \leq x)\mathbf{P}(Y \in B)$.

Independence and Uniqueness of Measure: continued

- **Proof: continued**: As $\mathbf{P}(X \leq x, Y \in B) = \mathbf{P}(X \leq x)\mathbf{P}(Y \in B)$ for any x and Borel set B. Let us define another function $\mathbf{R}_B(X \in S) = \mathbf{P}(X \in S, Y \in B)/\mathbf{P}(Y \in B)$ for fixed B. Again, $\mathbf{R}_B$ is a proper probability measure.
- We can also verify that $\mathbf{R}_B(X \leq x) = \mathbf{P}(X \leq x)$. Then by the uniqueness of extension theorem again, $\mathbf{R}_B(X \in (-\infty, x]) = P(X \leq x)$ guarantees $\mathbf{P}(X \in S, Y \in B) = \mathbf{R}_B(X \in S)\mathbf{P}(Y \in B) = \mathbf{P}(X \in S)\mathbf{P}(Y \in B)$.
- If either A or B is empty, then the equation would've held trivially.

Continuity of Probability (Proposition 3.3.1)

- **Limit of Increasing and Decreasing Events**: The limit of an increasing sequence of events $A_1, A_2, \cdots$ where $A_1 \subseteq A_2 \subseteq \cdots$ is defined as the union: $\lim_{n\to\infty} A_n = \bigcup_n A_n$. For a decreasing sequence $A_1, A_2, \ldots$ where $A_1 \supseteq A_2 \supseteq \cdots$, the limit is defined as the intersection: $\lim_{n\to\infty} A_n = \bigcap_n A_n$.
- **Continuity of Probabilities** *In the above scenarios*, by countable additivity, we have $\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A)$. (See of the textbook).
- If $\{A_n\} \nearrow A$, define $B_k = A_k \cap A_{k-1}^c$:

$$\mathbf{P}(\lim_n A_n) = \mathbf{P}(\cup_{k\geq 1} A_k) = \mathbf{P}(\cup_{k\geq 1} B_k) = \sum_{k=1}^{\infty} \mathbf{P}(B_k)$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{P}(B_k) = \lim_{n\to\infty} \mathbf{P}(\cup_{k=1}^{n} B_k) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

Continuity of Probability

- If $\{A_n\} \searrow A := \lim_n A_n$, then $\{A_n^c\} \nearrow A^c := \lim_{n\to\infty} A_n^c$ and $\lim_{n\to\infty} \mathbf{P}(A_n^c) = \mathbf{P}(\lim_{n\to\infty} A_n^c)$ and

$$\mathbf{P}(\lim_n A_n) = \mathbf{P}(\cap_n A_n) = 1 - \mathbf{P}(\cup_n A_n^c) = 1 - \lim_{n\to\infty} \mathbf{P}(A_n^c)$$
$$= \lim_{n\to\infty} [1 - \mathbf{P}(A_n^c)] = \lim_n \mathbf{P}(A_n)$$

- Just as is true for sequences of real numbers, not all sequences of events have a limit. E.g. on $([0,1], \mathcal{M}, \mathbf{P})$ consider $A_n = [0, \frac{3}{4})$ if n is even, and $[\frac{1}{3}, 1]$ if n is odd.
- Recall, for real-valued sequences:
$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}(\inf_{m\geq n} x_m)$
$\limsup_{n\to\infty} x_n = \lim_{n\to\infty}(\sup_{m\geq n} x_m)$.
- We'll apply the same idea to events.

Limits of Events

- **Limit Infimum and Limit Supremum of Events**: For *any* sequence of events $A_1, A_2, \ldots$ define:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k;$$

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

- $C_n := \bigcup_{k=n}^{\infty} A_k$ is decreasing, so its limit is an intersection.
- $D_n := \bigcap_{k=n}^{\infty} A_k$ is increasing, so its limit is a union.

Limits of Events

- $$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k;$$
  $$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$
- $[\limsup_{n\to\infty} A_n]^c = \liminf_{n\to\infty} A_n^c$
- $\liminf_n A_n \subseteq \limsup_n A_n$.

Limits of Events

- **Limit Infimum and Limit Supremum of Events**: For *any* sequence of events $A_1, A_2, \ldots$ define:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k;$$

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

- The event $\limsup_{n\to\infty} A_n$ occurs if for any $n \in \mathbf{N}$, there is at least one $k \geq n$ such that A_k occurs. Thus, this event is also referred as "A_n infinitely often", or "A_n i.o.".
- The event $\liminf_{n\to\infty} A_n$ occurs if there is a $n \in \mathbf{N}$, and all the event A_k with $k \geq n$ occur. Thus, this event is also referred as "A_n almost always", or "A_n a.a.".

Limits of Events

:

- Consider $([0,1], \mathcal{B}, \mathbf{P})$ and define $A_n = [0, \frac{3}{4})$ if n is even, and $[\frac{1}{3}, 1]$ if n is odd. What is $\{A_n \text{ i.o.}\}$? What is $\{A_n \text{ a.a}\}$?
- Consider $([0,1], \mathcal{B}, \mathbf{P})$ and define $A_n = [0, 1/2)$ if n is even, and $[1/2, 1]$ if n is odd. What is $\{A_n \text{ i.o.}\}$? What is $\{A_n \text{ a.a}\}$?
- Consider $(\Omega, \mathcal{F}, \mathbf{P})$ for the infinite fair coin tossing experiment and define H_n is the event that the nth coin-toss is a heads. What is $\{A_n \text{ i.o.}\}$? What is $\{A_n \text{ a.a.}\}$?

Limits of Events

Proposition 4 (3.4.1)

$$\mathbf{P}(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} \mathbf{P}(A_n) \le \limsup_{n\to\infty} \mathbf{P}(A_n) \le \mathbf{P}(\limsup_{n\to\infty} A_n)$$

- **Proof:** By the continuity of probability, $\mathbf{P}(\limsup_{n\to\infty} A_n) = \mathbf{P}(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) = \lim_{n\to\infty} \mathbf{P}(\bigcup_{k=n}^{\infty} A_k) = \limsup_{n\to\infty} \mathbf{P}(\bigcup_{k=n}^{\infty} A_k) \ge \limsup_{n\to\infty} \mathbf{P}(A_n)$
- Similarly, we may prove $\mathbf{P}(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} \mathbf{P}(A_n)$. And $\liminf_{n\to\infty} \mathbf{P}(A_n) \le \limsup_{n\to\infty} \mathbf{P}(A_n)$ by the definition of the limit infimum and limit supremum of real numbers.

Borel-Cantelli Lemma

Theorem 5

For $A_1, A_2, \ldots \in \mathcal{F}$,
(i) If $\sum_n \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(\limsup_n A_n) = 0$.
(ii) If $\sum_n \mathbf{P}(A_n) = \infty$, $\{A_n\}$ are independent, then $\mathbf{P}(\limsup_n A_n) = 1$.

- **Proof:** For (i), $\mathbf{P}(\limsup_n A_n) = \mathbf{P}(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k) \leq \mathbf{P}(\bigcup_{k=m}^{\infty} A_k) \leq \sum_{k=m}^{\infty} \mathbf{P}(A_k)$,
  which converges to 0 as $m \to \infty$ if $\sum_n \mathbf{P}(A_n) < \infty$.
  For (ii), we need to show $\mathbf{P}((\limsup_n A_n)^c) = \mathbf{P}(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^C) = 0$.
  It would be sufficient if we can show $\mathbf{P}(\bigcap_{k=n}^{\infty} A_k^C) = 0$ for any $n \in \mathbf{N}$:
  By independence, $\mathbf{P}(\bigcap_{k=n}^{\infty} A_k^C) \leq \mathbf{P}(\bigcap_{k=n}^{m} A_k^C) = \prod_{k=n}^{m}(1 - \mathbf{P}(A_k))$
  Since for any real number, $1 - x \leq e^{-x}$, and $\sum_{k=n}^{\infty} \mathbf{P}(A_k) = \infty$:
  $\prod_{k=n}^{m}(1 - \mathbf{P}(A_k)) \leq \prod_{k=n}^{m} e^{-\mathbf{P}(A_k)} = e^{-\sum_{k=n}^{m} \mathbf{P}(A_k)} \to 0$ as $m \to \infty$.
  Thus $\mathbf{P}(\bigcap_{k=n}^{\infty} A_k^C) = 0$ for any $n \in \mathbf{N}$, and $\mathbf{P}(\limsup_n A_n) = 1$.

Borel-Cantelli Lemma: Example

- Going back to the infinite coin tossing example, let H_n be the event that the nth toss comes up "Heads." Assume that $\mathbf{P}(H_n) = p > 0$. Then, as these events are independent, and as $\sum_n \mathbf{P}(H_n) = \infty$, $\mathbf{P}(\limsup_n H_n) = 1$. That is, if you toss a coin infinite times, if the chance of obtaining heads is not zero, then with probability 1, an infinite sequence of tosses would contain infinite number of heads.
- Another example: define $B_n : H_n \cap \cdots \cap H_{n+999}$. Although $\sum_n \mathbf{P}(B_n) = \infty$, the Borel-Cantelli lemma is not *directly* applicable as $\{B_n\}$ are not independent.
- However, we can focus on a subsequence: $\{B_{k_n}\}$, where $k_n = 1 + (n-1) * 1000$. $\{B_{k_n}\}$ would be a sequence of independent events and we can use Borel-Cantelli lemma to show that B_{k_n} occurs infinitely often, so B_n must also occur infinitely often as well.
- This lemma will be very useful when we discuss the relationship between convergence almost surely and convergence in probability.

Expectations

- You are already familiar with the concept of expectation from introductory classes. In the next lecture, we will develop the rigorous definition of expected value!
- Next lecture's roadmap:
  - define it for *simple* random variables;
  - define it for nonnegative random variables;
  - define it for general random variables.

  In this process, we will show that the expectation (still) has all the properties that you know and love, such as linearity and the fact that it is order preserving.

Expectations and Integrals

- One key idea behind the rigorous definition of an expectation is: when we calculate the average of a random variable X, rather than summing up (or integrating) the $X(\omega)$ for all $\omega \in \Omega$, we partition the sample space based on the values of $X(\omega)$. Outcomes with similar values of $X(\omega)$ will be grouped together, and we calculate the average of $X(\omega)$ based on such partitions.
- This idea is fundamentally different from standard Riemann integration described in a calculus class.
- Here is an illustration: suppose I have a pile of bills with values: \$1, \$5, \$1, \$5, \$100. To calculate the total amount of money in this pile, I can either
  1. proceed in whatever order it's in: calculate \$1+ \$ 5 + \$1 + \$5 + \$100 (i.e. Riemann)
  2. organize the bills by value: 2 \$1 bills, 2 \$ 5 bills, 1 \$ 100 bills (i.e. Lebesgue)