STAT 7200

Introduction to Advanced Probability
Lecture 11

Taylor R. Brown

- Probability Inequalities
- Almost Sure Convergence

"A First Look at Rigorous Probability Theory" (Jeffrey Rosenthal)
Sections 5.1 and 5.2

Markov and Chebychev's Inequalities

Theorem 1 (Markov's Inequality)

For non-negative random variable X, if $\alpha > 0$, then $\mathbf{P}(X \geq \alpha) \leq \mathbf{E}(X)/\alpha$.

- **Proof** Let $A = \{\omega : X(\omega) \geq \alpha\}$, then we have $X \geq \alpha \mathbf{1}_A$. By the order-preserving property of expectation, Markov' inequality follows immediately.
- Despite its simplicity, Markov's inequality can be quite useful in practice. For example, applying it to $f(X) = (X - \mathbf{E}(X))^2$ gives us Chebychev's inequality:

Corollary 2 (Chebychev's Inequalities)

For any random variable Y with finite variance, for $\alpha \geq 0$, we have $\mathbf{P}(|Y - \mu_Y| \geq \alpha) \leq \mathbf{Var}(Y)/\alpha^2$.

The Cauchy-Schwarz Inequality

Theorem 3 (Cauchy-Schwarz Inequality)

For random variables X, Y so that $\mathbf{E}(X^2) < \infty, \mathbf{E}(Y^2) < \infty$, we have $\mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$.

- **Proof**

$$
\begin{aligned}
0 &\leq \mathbf{E}\left[\left(\frac{|X|}{\sqrt{\mathbf{E}[X^2]}} - \frac{|Y|}{\sqrt{\mathbf{E}[Y^2]}}\right)^2\right] \\
&= 1 + 1 - 2\frac{\mathbf{E}(|YX|)}{\sqrt{\mathbf{E}(X^2)}\sqrt{\mathbf{E}(Y^2)}} \\
&= 2\left[1 - \frac{\mathbf{E}(|YX|)}{\sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}}\right]
\end{aligned}
$$

The Cauchy-Schwarz Inequality

Theorem 4 (Cauchy-Schwarz Inequality)

For random variables X, Y so that $\mathbf{E}(X^2) < \infty, \mathbf{E}(Y^2) < \infty$, we have $\mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$.

- Special case 1: correlation must always be between -1 and 1. For r.v.s X and Y with finite variances, $\mathbf{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y))$. Then $|\mathbf{Cov}(X, Y)| \leq \mathbf{E}(|X - \mathbf{E}X||Y - \mathbf{E}Y|) \leq \sqrt{\mathbf{Var}(X)\mathbf{Var}(Y)}$, and $|\mathbf{Corr}(X, Y)| \leq 1$.
- Special case 2: the Cramér-Rao inequality. For esimator $\hat{\theta}$ and score function U. $\mathbf{E}[U] = 0$ and $\mathbf{Var}\,[U] = I(\theta)$.

Jensen's Inequality

Theorem 5 (Jensen's Inequality)

For random variable X with finite mean, and a convex function $\phi : \mathbf{R} \to \mathbf{R}$, we have $\mathbf{E}(\phi(X)) \geq \phi(\mathbf{E}(X))$.

- **Convex Function**: $\phi(x)$ is convex if for any $x, y \in \mathbf{R}$ and $0 < t < 1$, $t\phi(x) + (1-t)\phi(y) \geq \phi(tx + (1-t)y)$. For instance, $x^2, |x|, e^x$ are all examples of convex functions.
- One key property of convex function is that given any point $(x_0, \phi(x_0))$, you may find a straight line $g(x) = a + bx$ that is bellow $\phi(x)$ and also passes through $(x_0, \phi(x_0))$ (usually it is the tangent at x_0 if the derivative of $\phi(x)$ exists at $x = x_0$). That is, $g(x) \leq \phi(x)$ and $g(x_0) = \phi(x_0) = a + bx_0$.
- **Proof:** Apply the above property for $x_0 = \mathbf{E}(X)$, then we have $\mathbf{E}(\phi(X)) \geq \mathbf{E}(g(X)) = a + b\mathbf{E}(X) = g(\mathbf{E}(X)) = \phi(\mathbf{E}(X))$.

Almost Sure Convergence

- **Pointwise Convergence** Suppose we have random variables $Z, Z_1, Z_2, \ldots$ on probability triple $(\Omega, \mathcal{F}, \mathbf{P})$ so that for each $\omega \in \Omega$, $\lim_{n\to\infty} Z_n(\omega) = Z(\omega)$. Then we may say that $\{Z_n\}$ converges to Z pointwise,
- This type of convergence is usually unnecessarily strong in probability theory. So a slightly weaker version is more popular and useful: we would only require $Z_n(\omega)$ converges to $Z(\omega)$ with probability one. That is, $\mathbf{P}(\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\}) = 1$.
- **Almost Sure Convergence** We say that $\{Z_n\}$ converges to Z almost surely (or a.s., or with probability 1), if the above condition holds. And we usually denote it as $\mathbf{P}(Z_n \to Z) = 1$.
- **Example** Consider the uniform measure $(\Omega, \mathcal{F}, \mathbf{P})$ on $[0, 1]$. Define $Z_n(\omega) = \mathbf{1}_{[0,\frac{1}{2^n}]}(\omega)$, then for each $\omega > 0$, we have $\lim_n Z_n(\omega) = 0$ and $\lim_n Z_n(0) = 1$. So $\mathbf{P}(\lim_{n\to\infty} Z_n(\omega) = 0) = \mathbf{P}((0, 1]) = 1$. That is, $\{Z_n\}$ converges to $Z = 0$ almost surely (but not pointwise).

Almost Sure Convergence: A lemma

Lemma 6 (5.2.1)

For r.v.s $Z, Z_1, Z_2, \ldots$ so that for each $\varepsilon > 0$, $\mathbf{P}(|Z_n - Z| \geq \varepsilon \text{ } i.o.) = 0$. Then $\mathbf{P}(Z_n \to Z) = 1$. (The converse is also true).

- **Proof**: Pick $\omega \in \Omega$. $Z_n(\omega) \to Z(\omega)$ iff $\forall \epsilon > 0$, $\exists N$ such that $n \geq N$ implies $|Z_n(\omega) - Z(\omega)| < \epsilon$.
- $Z_n(\omega) \to Z(\omega)$ iff $\omega \in \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{|Z_n(\omega) - Z(\omega)| < \epsilon\}$
- $Z_n(\omega) \to Z(\omega)$ iff $\omega \in \liminf_{n \to \infty} \{|Z_n(\omega) - Z(\omega)| < \epsilon\} := \{|Z_n(\omega) - Z(\omega)| < \epsilon \text{ a.a}\}$
- $\mathbf{P}(Z_n \to Z) = \mathbf{P}(\{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| < \epsilon \text{ a.a.}\}) = 1 - \mathbf{P}(|Z_n - Z| \geq \epsilon \text{ i.o.})$

Almost Sure Convergence and Borel-Cantelli Lemma

- According to the Borel-Cantelli Lemma: for a sequence of event $A_1, A_2, \cdots$, $\sum_n \mathbf{P}(A_n) < \infty$ implies $\mathbf{P}(A_n \ i.o.) = 0$. Combining this with the lemma we obtained in previous slide:

Lemma 7

For r.v.s $Z, Z_1, Z_2, \ldots$ such that for any $\varepsilon > 0$, $\sum_n \mathbf{P}(|Z_n - Z| \geq \varepsilon) < \infty$. Then $\mathbf{P}(Z_n \to Z) = 1$.

- **Example** Let $Z_1, Z_2, \ldots$ be random variables so that $P(Z_n = 1) = \frac{1}{2^n}$ and $P(Z_n = 0) = 1 - \frac{1}{2^n}$. Then for $1 > \varepsilon > 0$, $\mathbf{P}(|Z_n| \geq \varepsilon) = \frac{1}{2^n}$ and $\sum_n \mathbf{P}(|Z_n| \geq \varepsilon) = 1 < \infty$, so we must have $Z_n \to 0$ almost surely.
- **Note** The converse of this lemma is not necessarily true. For instance, consider $Z_1, Z_2, \ldots$ defined on $([0,1], \mathcal{M}, \lambda)$, $Z_n(\omega) = \mathbf{1}_{[0, \frac{1}{n}]}(\omega)$.