Probability Lecture Notes

December 08

1 The Moment Method and the LLN

1.1 Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random variables. Assume $\sigma^2 = \mathbb{E}[X_1^2] < \infty$ and $\mathbb{E}[X_1] = 0$.

Then, for any $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} \to 0$$

Proof (p = 2). Using Chebyshev's inequality:

$$\mathbb{P}(|S_n| > n\varepsilon) \leq \frac{\mathbb{E}[S_n^2]}{n^2\varepsilon^2} = \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Recall that for $S_n = \sum_{i=1}^n X_i$:

$$S_n^2 = \sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j$$

Taking expectations (using independence and $\mathbb{E}[X_i] = 0$):

$$\mathbb{E}[S_n^2] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i]\mathbb{E}[X_j] = n\mathbb{E}[X_1^2] = n\sigma^2$$

□

1.2 Connection between Convergence in Probability and A.S.

Lemma 1. If $X_n \to X$ in probability at a rate $\sum r_n < \infty$, i.e.,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq r_n \quad \text{where } \sum r_n < \infty,$$

then $X_n \to X$ almost surely (a.s.).

Exercise 1. Proof. Hint: Use Borel-Cantelli lemma. If $\sum \mathbb{P}(E_n) < \infty$, then $\mathbb{P}(\limsup E_n) = 0$ (meaning E_n happens infinitely often with probability 0).

1.3 Strong Law of Large Numbers (SLLN) with 4th Moment

Theorem 1 (SLLN)**.** If $\mathbb{E}[X^4] < \infty$ (and $\mathbb{E}[X_i] = 0$), then $\dfrac{S_n}{n} \to 0$ a.s.

Proof. We use the 4th moment method and Markov's inequality:

$$\mathbb{P}(|S_n| \geq n\varepsilon) \leq \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4}$$

Expanding $S_n^4 = (\sum_{i=1}^n X_i)^4$:

$$S_n^4 = \sum_{i=1}^{n} X_i^4 + \sum_{\substack{i,j,k,l \\ \text{indices distinct}}} X_i X_j X_k X_l + \sum_{i \neq j} X_i^2 X_j^2 + \dots$$

When taking the expectation $\mathbb{E}[S_n^4]$, any term with a singleton index (like X_i, $X_i^3 X_j$, etc.) vanishes because $\mathbb{E}[X_i] = 0$. The only surviving terms are of the form X_i^4 and $X_i^2 X_j^2$.

$$\mathbb{E}[S_n^4] = n\mathbb{E}[X_1^4] + 3n(n-1)(\mathbb{E}[X_1^2])^2$$

Thus, $\mathbb{E}[S_n^4] \leq Cn^2$. Substituting this back into the probability bound:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{Cn^2}{n^4\varepsilon^4} = O\left(\frac{1}{n^2}\right)$$

Since $\sum \dfrac{1}{n^2} < \infty$, by the Borel-Cantelli lemma (or the previous Lemma), $\dfrac{S_n}{n} \to 0$ a.s. □

1.4 Remark on Truncation

The WLLN and SLLN actually hold with just $\mathbb{E}|X| < \infty$. To derive this stronger version, we use **truncation**. Let $M > 0$. Define:

$$X = X\mathbb{I}_{\{|X| \leq M\}} + X\mathbb{I}_{\{|X| > M\}} = X_{\leq M} + X_{>M}$$

We use the fact that:

$$\mathbb{P}(|X| > M) \leq \frac{\mathbb{E}|X|}{M}$$

and $\mathbb{E}[X_{>M}]$ relates to the tail probabilities.

2 Martingales and Large Deviations

2.1 SLLN for Martingales in L^2

Recall: Let $M = (M_n)_n$ be a martingale. Let $W = (W_n)_n$ be a martingale, null at 0 ($W_0 = 0$). Under an appropriate scaling, we analyze the structure. We have the decomposition $W^2 = P + C$, where:

- P is a martingale.
- $C = \langle W \rangle$ is the **Predictable Increasing Process** (Quadratic Variation).

Example 1. Let (X_n) be independent RVs in L^2 with $\mathbb{E}[X_i] = 0$. Let $S_n = X_1 + \cdots + X_n$. Then S_n is a martingale in L^2. The quadratic variation is $\langle S \rangle_n = \sum_{i=1}^n \sigma_i^2$.

Theorem 2 (SLLN for Martingales)**.** Let W be a martingale in L^2, null at 0. On the set $\{\langle W \rangle_\infty = \infty\}$:

$$\frac{W_n}{\langle W \rangle_n} \to 0 \quad \text{a.s. as } n \to \infty$$

2.1.1 Proof

Recall $\langle W \rangle_n = C_n = \sum_{k=1}^n \mathbb{E}(W_k^2 - W_{k-1}^2 \mid \mathcal{F}_{k-1}) = \sum_{k=1}^n \mathbb{E}((W_k - W_{k-1})^2 \mid \mathcal{F}_{k-1})$.

The process $(1+C)^{-1} = (\dfrac{1}{1+C_n})_n$ is bounded between 0 and 1 and is predictable. Consider the martingale transform $M = (1+C)^{-1} \bullet W$:

$$M_n = \sum_{k=1}^n \frac{1}{1+C_k}(W_k - W_{k-1})$$

M is a martingale in L^2. By the previous theorem (Doob's Convergence), $\lim M_n$ exists if the quadratic variation is bounded. Let $A = \langle M \rangle$.

$$A_n = \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1})$$

Claim 1. $A_n \leq \dfrac{1}{1+C_0} - \dfrac{1}{1+C_n} \leq 1.$

Proof of claim:

$$(M_n - M_{n-1})^2 = \frac{1}{(1+C_n)^2}(W_n - W_{n-1})^2$$

Taking conditional expectation:

$$\begin{aligned}
\mathbb{E}((M_n - M_{n-1})^2 \mid \mathcal{F}_{n-1}) &= (1+C_n)^{-2}\mathbb{E}((W_n - W_{n-1})^2 \mid \mathcal{F}_{n-1}) \\
&= (1+C_n)^{-2}(C_n - C_{n-1}) \\
&\leq \Big(\frac{1}{1+C_n}\Big)\Big(\frac{1}{1+C_{n-1}}\Big)\big[(C_n + 1) - (C_{n-1} + 1)\big] \\
&= \frac{1}{1+C_{n-1}} - \frac{1}{1+C_n}
\end{aligned}$$

Since C_n is non-deacreasing, we sum this up to get the result. Therefore, $\lim M_n$ exists a.s. This implies $\sum_n \frac{1}{1+C_n}(W_n - W_{n-1})$ converges a.s.

Kronecker's Lemma: If $\sum \frac{x_n}{b_n}$ converges (where $b_n \uparrow \infty$), then $\frac{1}{b_n} \sum_{i=1}^n x_i \to 0$.

Applying this with $x_n = W_n - W_{n-1}$ (so $\sum x_i = W_n$) and $b_n = 1 + C_n$: If $C_n(\omega) \to \infty$, then $\frac{W_n}{1+C_n} \to 0$, which implies $\frac{W_n}{C_n} \to 0$.

2.2 Large Deviation Estimates (LDE)

Let $X_1, X_2, \ldots$ be i.i.d. RVs with mean $\mathbb{E}[X_1] = \mu$. We want to show that for $\varepsilon > 0$, $\mathbb{P}(|\frac{S_n}{n} - \mu| > \varepsilon)$ decays exponentially fast to 0 as $n \to \infty$.

2.2.1 Bernstein's Trick / Chernoff Bounding Technique

For $t > 0$:

$$X \geq \lambda \iff tX \geq t\lambda \iff e^{tX} \geq e^{t\lambda}$$

By Markov's inequality:

$$\mathbb{P}(X \geq \lambda) = \mathbb{P}(e^{tX} \geq e^{t\lambda}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\lambda}}$$

Applying to S_n:

$$\mathbb{P}(S_n \geq n\varepsilon) \leq e^{-tn\varepsilon}\mathbb{E}[e^{tS_n}]$$

Using independence: $\mathbb{E}[e^{tS_n}] = (\mathbb{E}[e^{tX_1}])^n = (M(t))^n = e^{nc(t)}$, where $M(t)$ is the moment generating function and $c(t) = \log M(t)$ is the cumulative generating function.

2.2.2 Maclaurin Series Expansion of The Genrating Function

$$c(0) = \log M(0) = \log 1 = 0$$

$$c'(0) = \frac{M'(0)}{M(0)} = \mathbb{E}[X] = 0 \quad \text{(assuming centered)}$$

$$c''(0) = \frac{M''(0)M(0) - (M'(0))^2}{M(0)^2} = \mathbb{E}[X^2] = \underbrace{\sigma^2 > 0}_{\text{Verify this}}$$

We get:

$$c(t) = \frac{\sigma^2}{2}t^2 + O(t^3)$$

And so:

$$\mathbb{P}(S_n \geq n\varepsilon) \leq e^{-n(t\varepsilon - c(t))} \leq e^{-n\tilde{c}(t)}$$

We optimize over $t > 0$. Let $\tilde{c}(\varepsilon) = \sup_{t>0}(t\varepsilon - c(t))$. This is the **Legendre Transform** of $c(t)$. Since $c(t) \sim \frac{\sigma^2}{2}t^2$, we find the optimal rate by finding a local maxima and by the definition of the function it will be a global maxima.

Theorem 3 (Cramer's Inequality)**.** Assume X_n has exponential moments ($\mathbb{E}[e^{tX}] < \infty$ for some t). Then:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq 2e^{-n\tilde{c}(\varepsilon)}$$

Remark: $\tilde{c}(\varepsilon) \approx c_0\varepsilon^2$. Asymptotically, this is the correct deviation rate (Large Deviation Principle - LDP).

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = -\tilde{c}(\varepsilon)$$

2.3 Hoeffding Inequalities

Let $X_1, \dots, X_n$ be independent (not necessarily identically distributed) RVs such that $X_i \in [a_i, b_i]$ a.s. Then:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| > \varepsilon\right) \leq 2e^{-\frac{2n^2\varepsilon^2}{K}}$$

where $K = \sum_{i=1}^{n}(b_i - a_i)^2$. In particular, if $|X_i| \leq L$ a.s., then $K = n(2L)^2 = 4L^2n$, and:

$$\mathbb{P}(\dots) \leq 2e^{-\frac{n\varepsilon^2}{2L^2}}$$

Lemma 2 (Hoeffding)**.** If X is a centered RV in $[a, b]$, then:

$$\mathbb{E}[e^{tX}] \leq e^{t^2(b-a)^2/8}$$

Proof Sketch of Lemma. Since $x \mapsto e^{tx}$ is convex:

$$e^{tX} \leq \frac{b - X}{b - a} e^{ta} + \frac{X - a}{b - a} e^{tb}$$

Taking expectations (with $\mathbb{E}[X] = 0$) yields the result. □