Probability Lecture Notes

December 10

1 Large Deviation Estimates (LDE) for Martingales

1.1 Recall: Hoeffding's Inequality

Let $(X_n)_n$ be a sequence of i.i.d. random variables with finite exponential moments. Then $\forall \epsilon > 0$, there exists $\tilde{c}(\epsilon)$ such that:

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

where $\tilde{c}(\epsilon) = \sup_{t>0}(t\epsilon - c(t))$ is the Lagendre transform of $c(t)$, and $c(t) = \log M(t)$ is the cumulant generating function, with $M(t) = \mathbb{E}[e^{tX}]$.

Theorem 1 (Hoeffding's Inequality). *Assume $X_1, \ldots, X_n$ are independent with $X_i \in [a_i, b_i]$. Let $S_n = \sum X_i$ and $\mu = \mathbb{E}[S_n/n]$. Then:*

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

where $k = \sum_{i=1}^{n}(b_i - a_i)^2$.

Lemma 1 (Hoeffding's Lemma). *If X is a random variable such that $X \in [a, b]$ a.s. and $\mathbb{E}[X] = 0$, then:*

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

Moreover if $\mathcal{F}_0$ is a sub σ-algebra then

$$\mathbb{E}[e^{tX}|\mathcal{F}_0] \leq e^{\frac{t^2(b-a)^2}{8}}$$

2 Azuma-Hoeffding Inequality

Theorem 2 (Azuma-Hoeffding). *Let $M = (M_n)_{n\geq 0}$ be a martingale (with respect to a filtration $\mathcal{F}_n$) such that $\mathbb{E}[M_n] < \infty$. Assume the increments are bounded, i.e., there exist constants c_n such that:*

$$|M_n - M_{n-1}| \leq c_n \quad a.s.\ \forall n$$

Then for any $\epsilon > 0$:

$$\mathbb{P}(|M_n - M_0| > \epsilon) \leq 2e^{\left(\frac{-\epsilon^2}{2\sum_{i=1}^{n} c_i^2}\right)}$$

2.1 Martingale Differences

A martingale difference sequence (X_n) satisfies $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = 0$.

(a) If (M_n) is a martingale, define $X_n = M_n - M_{n-1}$ (with $X_0 = 0$). Then (X_n) is a martingale difference.

(b) Conversely, if (X_n) is a martingale difference, null at 0, then $M_n = \sum_{i=1}^n X_i$ is a martingale.

Theorem 3 (Azuma-Hoeffding Difference Version). *If (X_n) is a martingale difference sequence with $|X_n| \le c_n$ a.s., then:*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \le 2e^{\left(\frac{-t^2}{2\sum_{i=1}^n c_i^2}\right)}$$

2.2 Proof of Azuma-Hoeffding

Let $X_n = M_n - M_{n-1}$. We have $M_n - M_0 = \sum_{i=1}^n X_i$. We want to bound $\mathbb{P}(M_n - M_0 > \lambda)$. Using the Chernoff bound technique:

$$\mathbb{P}(M_n - M_0 > \lambda) \le e^{-t\lambda}\mathbb{E}[e^{t(M_n - M_0)}]$$

Decompose the expectation:

$$\mathbb{E}[e^{t(M_n - M_0)}] = \mathbb{E}\left[e^{t(M_{n-1} - M_0)} \cdot e^{tX_n}\right]$$

Use the tower property of conditional expectation (conditioning on $\mathcal{F}_{n-1}$):

$$\mathbb{E}\left[e^{t(M_{n-1} - M_0)}\mathbb{E}[e^{tX_n}|\mathcal{F}_{n-1}]\right]$$

Since $|X_n| \le c_n$ and $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = 0$, by Hoeffding's Lemma (applied conditionally):

$$\mathbb{E}[e^{tX_n}|\mathcal{F}_{n-1}] \le e^{\frac{t^2(2c_n)^2}{8}} = e^{\frac{t^2c_n^2}{2}}$$

Iterating this process backwards from n to 1:

$$\mathbb{E}[e^{t(M_n - M_0)}] \le \prod_{i=1}^n e^{\frac{t^2c_i^2}{2}} = e^{\frac{t^2}{2}\sum c_i^2}$$

Optimizing over t yields the result.

3 McDiarmid's Inequality

Let $X_1, \dots, X_n$ be independent random variables taking values in a set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ be a function satisfying the **Bounded Difference Property**:

$$\sup_{x_1,\dots,x_n,x_i'} |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n)| \le c_i$$

for all $1 \le i \le n$. Then:

$$\mathbb{P}(|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| > \epsilon) \le 2e^{\left(\frac{-2\epsilon^2}{\sum_{i=1}^n c_i^2}\right)}$$

Remark 1. *Hoeffding inequality is deducable from McDiarmid by taking*

$$f(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i$$

3.1 Proof of McDiarmid

Let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$ for $1 \leq i \leq n$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Define the Doob Martingale:

$$M_i = \mathbb{E}[f(X_1, \ldots, X_n)|\mathcal{F}_i]$$

- $M_0 = \mathbb{E}[f(X)]$
- $M_n = f(X)$ (since $f(X)$ is $\mathcal{F}_n$-measurable)

We must show the increments are bounded.

$$M_i - M_{i-1} = \mathbb{E}[f(X)|\mathcal{F}_i] - \mathbb{E}[f(X)|\mathcal{F}_{i-1}]$$

Let $X^{(i)} = (X_1, \ldots, X_i', \ldots, X_n)$ where X_i' is an independent copy of X_i.

$$|M_i - M_{i-1}| \leq \sup |f(x) - f(x')| \leq c_i$$

Applying Azuma-Hoeffding to M_n yields the result.

4 Application: Balls into Bins

Consider throwing n balls into m bins uniformly at random. Let $Z_{n,m}$ be the number of empty bins. We want a good estimate for $\mathbb{E}[Z_{n,m}]$ and concentration around the mean.

Let $\mathbf{1}_{B_i}$ be the indicator that bin i is empty.

$$Z_{n,m} = \sum_{i=1}^{m} \mathbf{1}_{B_i}$$

Expectation:

$$\mathbb{E}[Z_{n,m}] = \sum_{i=1}^{m} \mathbb{P}(\text{Bin } i \text{ is empty}) = m\left(1 - \frac{1}{m}\right)^n \approx me^{-n/m}$$

To use McDiarmid's inequality, let X_k be the index of the bin where the k-th ball falls ($1 \leq k \leq n$). $X_1, \ldots, X_n$ are i.i.d. variables. We can write $Z_{n,m} = f(X_1, \ldots, X_n)$. Changing one ball's position (one X_k) can change the number of empty bins by at most 1. Thus, the bounded difference condition holds with $c_k = 1$.

$$\sum_{k=1}^{n} c_k^2 = n$$

By McDiarmid:

$$\mathbb{P}(|Z_{n,m} - \mathbb{E}[Z_{n,m}]| > \epsilon) \leq 2e^{-\frac{2\epsilon^2}{n}}$$

5 Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Assume $\mu = 0, \sigma^2 = 1$ for simplicity. Let $S_n = \sum_{i=1}^{n} X_i$.

5.1 Comparison with WLLN

By the Weak Law of Large Numbers (WLLN):

$$\frac{S_n}{n} \xrightarrow{p} 0$$

This implies $S_n = o(n)$ with high probability. Question: What is the typical size of S_n? Answer: $S_n = O(\sqrt{n})$.

By Chebyshev's Inequality:

$$\mathbb{P}\left(\left|\frac{S_n}{\sqrt{n}}\right| > \lambda\right) \leq \frac{\mathrm{Var}(S_n/\sqrt{n})}{\lambda^2} = \frac{1}{\lambda^2}$$

Since $\mathrm{Var}(S_n/\sqrt{n}) = \frac{1}{n}\sum \mathrm{Var}(X_i) = 1$.

5.2 Convergence Questions

Does $\frac{S_n}{\sqrt{n}}$ converge almost surely or in probability to a random variable? **Answer: No.**

Proposition 1. *Let X_i be i.i.d. with mean 0 and variance 1. Then $\frac{S_n}{\sqrt{n}}$ does not converge almost surely (nor in probability) to any random variable.*

Proof Sketch. Assume by contradiction that $\frac{S_n}{\sqrt{n}} \to Y$ almost surely. Then $(\frac{S_n}{\sqrt{n}})^2 \to Y^2$ a.s. Since the variances are uniformly bounded, by the Bounded Moment Convergence Theorem (BMCT), $\mathbb{E}[(\frac{S_n}{\sqrt{n}})^2] \to \mathbb{E}[Y^2]$. We know $\mathbb{E}[(\frac{S_n}{\sqrt{n}})^2] = 1$, so $\mathbb{E}[Y^2] = 1$.

However, Y must be measurable with respect to the tail σ-algebra

$$\mathcal{T} = \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \dots)$$

By Kolmogorov's 0-1 Law, Y must be constant almost surely. Since $\mathbb{E}[Y] = \lim \mathbb{E}[S_n/\sqrt{n}] = 0$, we must have $Y = 0$ a.s. This implies $\mathbb{E}[Y^2] = 0$, which contradicts $\mathbb{E}[Y^2] = 1$. □

5.3 Statement of CLT

Although it does not converge in probability, $\frac{S_n}{\sqrt{n}}$ converges in **distribution** to the standard normal distribution $\mathcal{N}(0, 1)$.

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \in [a, b]\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$