Limit theorems

Sayan Mukherjee

Limit theorems

We will learn various law of large number results. These results will also be used to motivate issues such as integration and switching limits. These are usually taught before limit theorems but we have reversed the order. We will also look at examples of how to use law of large numbers. We will discuss the strong law and weak law of large numbers in the notes on convergence.

Law of Large Numbers

In this lecture, we will look at concentration inequalities or law of large numbers for a fixed function. Let $(\Omega, \mathcal{L}, \mu)$ be a probability space. Let $X_1, ..., X_n$ be real random variables on Ω. A sequence of random variables Y_n converges almost surely to a random variable Y iff $\mathbf{P}(Y_n \to Y) = 1$. A sequence of random variables Y_n converges in probability to a random variable Y iff for every $\epsilon > 0$, $\lim_{n\to\infty} \mathbf{P}(|Y_n - Y| > \epsilon) = 0$. Let $\hat{\mu}_n := n^{-1}\sum_{i=1}^n X_n$. The sequence $X_1, ..., X_n$ satisfies the strong law of large numbers if for some constant c, $\hat{\mu}_n$ converges to c almost surely. The sequence $X_1, ..., X_n$ satisfies the weak law of large numbers iff for some constant c, $\hat{\mu}_n$ converges to c in probability. In general the constant c will be the expectation of the random variable $\mathbf{E}X$.

A given function $f(X)$ of random variables X concentrates if the deviation between its empirical average, $n^{-1}\sum_{i=1}^n f(X_i)$ and expectation, $\mathbf{E}f(X)$, goes to zero as n goes to infinity. That is $f(X)$ satisfies the law of large numbers.

Polynomial inequalities

Theorem 0.0.1 (Jensen) *If ϕ is a convex function then $\phi(\mathbf{E}x) \leq \mathbf{E}\phi(x)$.*

Theorem 0.0.2 (Bienaymé-Chebyshev) *For any random variable X, $\epsilon > 0$*

$$\mathbf{P}(|X| \geq \epsilon) \leq \frac{\mathbf{E}X^2}{\epsilon^2}.$$

Proof.

$$\mathbf{E}X^2 \geq E(X^2 I_{\{|x|\geq\epsilon\}}) \geq \epsilon^2\mathbf{P}(|X| > \epsilon). \ \square$$

Theorem 0.0.3 (Markov) *For any random variable X, $\epsilon > 0$*

$$\mathbf{P}(|X| \geq \epsilon) \leq \frac{\mathbf{E}e^{\lambda X}}{e^{\lambda\epsilon}}$$

and

$$\mathbf{P}(|X| \geq \epsilon) \leq \inf_{\lambda<0} e^{-\lambda\epsilon}\mathbf{E}e^{\lambda X}.$$

Proof.

$$\mathbf{P}(X > \epsilon) = \mathbf{P}(e^{\lambda X} > e^{\lambda\epsilon}) \leq \frac{\mathbf{E}e^{\lambda X}}{e^{\lambda\epsilon}}. \ \square$$

Exponential inequalities

For the sums or averages of independent random variables the above bounds can be improved from polynomial in $1/\epsilon$ to exponential in ϵ.

Theorem 0.0.4 (Bennet) *Let $X_1, ..., X_n$ be independent random variables with $\mathbf{E}X = 0$, $\mathbf{E}X^2 = \sigma^2$, and $|X_i| \leq M$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^n X_i| > \epsilon\right) \leq 2e^{\frac{-n\sigma^2}{M^2}\phi\left(\frac{\epsilon M}{n\sigma^2}\right)},$$

where

$$\phi(z) = (1+z)\log(1+z) - z.$$

Proof. We will prove a bound on one-side of the above theorem

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i > \epsilon\right).$$

$$\begin{aligned}\mathbf{P}\left(\sum_{i=1}^{n} X_i > \epsilon\right) \le e^{-\lambda\epsilon}\mathbf{E}e^{\lambda\sum X_i} &= e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbf{E}e^{\lambda X_i} \\ &= e^{-\lambda\epsilon}(\mathbf{E}e^{\lambda X})^n.\end{aligned}$$

$$\begin{aligned}\mathbf{E}e^{\lambda X} = \mathbf{E}\sum_{k=0}^{\infty}\frac{(\lambda X)^k}{k!} &= \sum_{k=0}^{\infty}\lambda^k\frac{\mathbf{E}X^k}{k!} \\ = 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}\mathbf{E}X^2x^{k-2} &\le 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}M^{k-2}\sigma^2 \\ = 1 + \frac{\sigma^2}{M^2}\sum_{k=2}^{\infty}\frac{\lambda^k M^k}{k!} &= 1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M) \\ &\le e^{\frac{\sigma^2}{M^2}(e^{\lambda M} - \lambda M - 1)}.\end{aligned}$$

The last line holds since $1 + x \le e^x$.

Therefore,

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i > \epsilon\right) \le e^{-\lambda\epsilon}e^{\frac{\sigma^2}{M^2}(e^{\lambda M} - \lambda M - 1)}. \tag{1}$$

We now optimize with respect to λ by taking the derivative with respect to λ

$$\begin{aligned}0 &= -\epsilon + \frac{n\sigma^2}{M^2}(Me^{\lambda M} - M), \\ e^{\lambda M} &= \frac{\epsilon M}{n\sigma^2} + 1, \\ \lambda &= \frac{1}{M}\log\left(1 + \frac{\epsilon M}{n\sigma^2}\right).\end{aligned}$$

The theorem is proven by substituting λ into equation (1). $\square$

The problem with Bennet's inequality is that it is hard to get a simple expression for ϵ as a function of the probability of the sum exceeding ϵ.

Theorem 0.0.5 (Bernstein) *Let $X_1, ..., X_n$ be independent random variables with $\mathbf{E}X = 0$, $\mathbf{E}X^2 = \sigma^2$, and $|X_i| \le M$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} X_i| > \epsilon\right) \le 2e^{-\frac{\epsilon^2}{2n\sigma^2 + \frac{2}{3}\epsilon M}}.$$

Proof.

Take the proof of Bennet's inequality and notice

$$\phi(z) \ge \frac{z^2}{2 + \frac{2}{3}z}. \ \square$$

Remark 0.0.1 *With Bernstein's inequality a simple expression for ϵ as a function of the probability of the sum exceeding ϵ can be computed*

$$\sum_{i=1}^{n} x_i \le \frac{2}{3}uM + \sqrt{2n\sigma^2 u}.$$

Outline.

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i > \epsilon\right) \le 2e^{-\frac{\epsilon^2}{2n\sigma^2+\frac{2}{3}\epsilon M}} = e^{-u},$$

where

$$u = \frac{\epsilon^2}{2n\sigma^2 + \frac{2}{3}\epsilon M}.$$

we now solve for ϵ

$$\epsilon^2 - \frac{2}{3}\epsilon M - 2n\sigma^2\epsilon = 0$$

and

$$\epsilon = \frac{1}{3}uM + \sqrt{\frac{u^2M^2}{9} + 2n\sigma^2 u}.$$

Since $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$

$$\epsilon = \frac{2}{3}uM + \sqrt{2n\sigma^2 u}.$$

So with large probability

$$\sum_{i=1}^{n} X_i \le \frac{2}{3}uM + \sqrt{2n\sigma^2 u}. \ \triangle$$

If we want to bound

$$|n^{-1}\sum_{i=1}^{n} f(X_i) - \mathbf{E}f(X)|$$

we consider

$$|f(X_i) - \mathbf{E}f(X)| \le 2M.$$

Therefore

$$\sum_{i=1}^{n}(f(X_i) - \mathbf{E}f(X)) \le \frac{4}{3}uM + \sqrt{2n\sigma^2 u}$$

and

$$n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbf{E}f(x) \le \frac{4}{3}\frac{uM}{n} + \sqrt{\frac{2\sigma^2 u}{n}}.$$

Similarly,

$$\mathbf{E}f(x) - n^{-1}\sum_{i=1}^{n} f(x_i) \ge \frac{4}{3}\frac{uM}{n} + \sqrt{\frac{2\sigma^2 u}{n}}.$$

In the above bound

$$\sqrt{\frac{2\sigma^2 u}{n}} \ge \frac{4uM}{n}$$

which implies $u \le \frac{n\sigma^2}{8M^2}$ and therefore

$$|n^{-1}\sum_{i=1}^{n} f(X_i) - \mathbf{E}f(X)| \lesssim \sqrt{\frac{2\sigma^2 u}{n}} \text{ for } u \lesssim n\sigma^2,$$

which corresponds to the tail probability for a Gaussian random variable and is predicted by the Central Limit Theorem (CLT) Condition that $\lim_{n\to\infty} n\sigma^2 \to \infty$. If $\lim_{n\to\infty} n\sigma^2 = C$, where C is a fixed constant, then

$$|n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbf{E}f(x)| \lesssim \frac{C}{n}$$

which corresponds to the tail probability for a Poisson random variable.

We now look at an even simpler exponential inequality where we do not need information on the variance.

Theorem 0.0.6 (Hoeffding) *Let $X_1, ..., X_n$ be independent random variables with $\mathbf{E}X = 0$ and $|X_i| \leq M_i$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} X_i| > \epsilon\right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} M_i^2}}.$$

Proof.

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i > \epsilon\right) \leq e^{-\lambda\epsilon}\mathbf{E}e^{\lambda \sum_{i=1}^{n} X_i} = e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbf{E}e^{\lambda X_i}.$$

It can be shown (Homework problem)

$$\mathbf{E}(e^{\lambda X_i}) \leq e^{\frac{\lambda^2 M_i^2}{8}}.$$

The bound is proven by optimizing the following with respect to λ

$$e^{-\lambda\epsilon}\Pi_{i=1}^{n}e^{\frac{\lambda^2 M_i^2}{8}}. \quad \square$$

Applying Hoeffding's inequality to

$$n^{-1}\sum_{i=1}^{n} f(X_i) - \mathbf{E}f(X)$$

we can state that with probability $1 - e^{-u}$

$$n^{-1}\sum_{i=1}^{n} f(X_i) - \mathbf{E}f(X) \leq \sqrt{\frac{2Mu}{n}},$$

which is a sub-Gaussian as in the CLT but without the variance information we can never achieve the $\frac{1}{n}$ rate we achieved when the random variable has a Poisson tail distribution.

Theorem 0.0.7 (Hoeffding) *Let $X_1, ..., X_n$ be independent random variables with $\mathbf{P}(X_i = M_i) = 1/2$ and $\mathbf{P}(X_i = -M_i) = 1/2$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} X_i| > \epsilon\right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} M_i^2}}.$$

Proof.

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \leq e^{-\lambda\epsilon}\mathbf{E}e^{\lambda \sum_{i=1}^{n} x_i} = e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbf{E}e^{\lambda x_i}.$$

$$\begin{aligned}
\mathbf{E}(e^{\lambda X_i}) &= \frac{1}{2}e^{\lambda M_i} + \frac{1}{2}e^{-\lambda M_i}, \\
\frac{1}{2}e^{\lambda M_i} + \frac{1}{2}e^{-\lambda M_i} &= \sum_{k=0}^{\infty} \frac{(M_i\lambda)^{2k}}{(2k)!} \leq e^{\frac{\lambda^2 M_i^2}{2}}.
\end{aligned}$$

Optimize the following with respect to λ

$$e^{-\lambda\epsilon}\Pi_{i=1}^{n}e^{\frac{\lambda^2 M_i^2}{2}}. \quad \square$$

Martingale inequalities

In the previous section we stated some concentration inequalities for sums of independent random variables. We now look at more complicated functions of independent random variables and introduce a particular Martingale inequality to prove concentration.

Let $(\Omega, \mathcal{L}, \mu)$ be a probability space. Let $X_1, ..., X_n$ be real random variables on Ω. Let the function $Z(X_1, ..., X_n) : \Omega^n \to \mathbb{R}$ be a map from the random variables to a real number.

The function Z concentrates if the deviation between $Z(X_1, ..., X_n)$ and $\mathbf{E}_{X_1,..,X_n} Z(X_1, .., X_n)$ goes to zero as n goes to infinity.

Theorem 0.0.8 (McDiarmid) *Let $X_1, ..., X_n$ be independent random variables let $Z(X_1, ..., X_n) : \Omega^n \to \mathbb{R}$ such that*

$$\forall X_1, ..., X_n, X_1', ..., X_n' \quad |Z(X_1, .., X_n) - Z(X_1, ..., X_{i-1}, X_i', X_{i+1}, X_n)| \leq c_i,$$

then

$$\mathbf{P}\left(Z - \mathbf{E}Z > \epsilon\right) \leq e^{-\frac{\epsilon^2}{2\sum_{i=1}^n c_i^2}}.$$

Proof.

$$\mathbf{P}(Z - \mathbf{E}Z > \epsilon) = \mathbf{P}(e^{\lambda(Z-\mathbf{E}Z)} > e^{\lambda\epsilon}) \leq e^{-\lambda\epsilon}\mathbf{E}e^{\lambda(Z-\mathbf{E}Z)}.$$

We will use the following very useful decomposition

$$\begin{aligned} Z(X_1, ..., X_n) - \mathbf{E}_{X_1',..,X_n'} Z(X_1', .., X_n') &= [Z(X_1, ..., X_n) - \mathbf{E}_{X_1'} Z(X_1', X_2, ..., X_n)] \\ &+ [E_{X_1'} Z(X_1', X_2, ..., X_n) - E_{X_1', X_2'} Z(X_1', X_2', X_3, ..., X_n)] \\ &+ ... \\ &+ [E_{X_1', ..., X_{n-1}'} Z(X_1', X_2', ...X_{n-1}', x_n) - E_{X_1', ..., X_n'} Z(X_1', ..., X_n')]. \end{aligned}$$

We denote the random variable

$$Z_i(X_i, ..., X_n) := \mathbf{E}_{X_1', ..., X_{i-1}'} Z(X_1', ..., X_{i-1}', X_i, ..., X_n) - \mathbf{E}_{X_1', ..., X_i'} Z(X_1', ..., X_i', X_{i+1}, ..., X_n),$$

and

$$Z(X_1, ..., X_n) - \mathbf{E}_{X_1',..,X_n'} Z(X_1', .., X_n') = Z_1 + ... + Z_n.$$

The following inequality is true (see the following Lemma for a proof)

$$\mathbf{E}_{X_i} e^{\lambda Z_i} \leq e^{\lambda^2 c_i^2/2} \ \ \forall \lambda \in \mathbb{R}.$$

$$\begin{aligned} \mathbf{E}e^{\lambda(Z-\mathbf{E}Z)} &= \mathbf{E}e^{\lambda(Z_1+...,+Z_n)} \\ \mathbf{E}\mathbf{E}_{X_1} e^{\lambda(Z_1+...,+Z_n)} &= \mathbf{E}e^{\lambda(Z_2+...,+Z_n)}\mathbf{E}_{X_1} e^{\lambda Z_1} \\ &\leq \mathbf{E}e^{\lambda(Z_2+...,+Z_n)} e^{\lambda c_1^2/2}, \end{aligned}$$

by induction

$$\mathbf{E}e^{\lambda(Z-\mathbf{E}Z)} \leq e^{\lambda^2 \sum_{i=1}^n c_i^2/2}.$$

To derive the bound we optimize with respect to λ

$$e^{-\lambda\epsilon + \lambda^2 \sum_{i=1}^n c_i^2/2}. \ \square$$

Lemma 0.0.1 *For all $\lambda \in \mathbb{R}$*

$$\mathbf{E}_{X_i} e^{\lambda Z_i} \leq e^{\lambda^2 c_i^2/2}.$$

Proof.

For any $t \in [-1, 1]$ the function $e^{\lambda t}$ is convex with respect to λ.

$$\begin{aligned} e^{\lambda t} &= e^{\lambda(\frac{1+t}{2}) - \lambda(\frac{1-t}{2})} \\ &\leq \frac{1+t}{2} e^{\lambda} + \frac{1-t}{2} e^{-\lambda} \\ &= \frac{e^{\lambda} + e^{-\lambda}}{2} + t\frac{e^{\lambda} - e^{-\lambda}}{2} \\ &\leq e^{\lambda^2/2} + t\,\mathrm{sh}(\lambda). \end{aligned}$$

Set $t = \frac{Z_i}{c_i}$ and notice that $\frac{Z_i}{c_i} \in [-1, 1]$ so,

$$e^{\lambda z_i} = e^{\lambda c_i \frac{z_i}{c_i}} \leq e^{\lambda^2 c_i^2/2} + \frac{z_i}{c_i} \mathrm{sh}(\lambda c_i),$$

and

$$\mathbf{E}_{X_i} e^{\lambda Z_i} \leq e^{\lambda^2 c_i^2/2}. \ \square$$