Strong Law of Large Numbers & Other Inequalities 8.4-8.5

Ryan Horowitz

December 14th

1 Pre-Recorded Lecture & Readings

1.1 Strong Law of Large Numbers

The Strong Law of Large Numbers (often abbreviated as SLLN) is as follows: let $X_1, X_2, ...$ be a sequence of IID random variables with $E[X_i] = \mu < \infty$. Then, we know that $P[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu] = 1$.

We will restate the Weak Law of Large Numbers (WLLN) and then discuss key differences between the WLLN and the SLLN.

The WLLN states that for any $\epsilon > 0$, $\lim_{n\to\infty} P[|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon] = 0$.

There are some key differences between the two statements. First of all, the WLLN states that for any specified large value n^*, $(X_1 + ... + X_{n^*})/n^*$ is likely to be near μ. However, it does not say that $(X_1 + ... + X_n)/n$ is bound to stay near μ for all values of n larger than n^*. The WLLN allows for the possibility of large values of $|(X_1 + ... + X_n)/n - \mu|$ to occur infinitely often (though at infrequent intervals). The SLLN shows that this cannot occur. It implies that with probability 1 for any positive value ϵ that $|(X_1 + ... + X_n)/n - \mu|$ will be greater than ϵ only a finite number of times.

Second of all, each law requires different additional assumptions for its proof. The WLLN requires that $Var[X_i] = \sigma^2 < \infty$ while the SLLN requires that $E[X_i^4] < \infty$. Third of all, the SLLN is a "stronger" statement in that it implies the WLLN; however, the WLLN does not imply the SLLN.

The Strong Law of Large Numbers has some very important applications. One of which is as follows:

Suppose that a sequence of independent trials of some experiment is performed. Let E be a fixed event of the experiment, and denote by $P(E)$ the probability that E occurs on any particular trial. Letting $X_i = 1$ if E occurs on the ith trial and $X_i = 0$ if E does not occur on the ith trial, we have, by the SLLN, that with probability 1,

$$\frac{X_1 + X_2 + ... + X_n}{n} \to E[X] = P(E). \tag{1}$$

Since $\sum_{i=1}^{n} X_i$ is the number of times that the event E occurs in the first n trials, we may interpret equation (1) as stating that with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$. This is a key fact that we assumed earlier in the course.

1.2 Other Inequalities

We are often interested in bounding a probability of the form $P[X - \mu \geq a]$, where $a > 0$ when only the mean $\mu = E[X]$ and variance $\sigma^2 = Var(X)$ of a random variable X are known. This section provides some useful inequalities.

Trivially, since $X - \mu \geq a$, we see that $|X - \mu| \geq a$. We can now apply Chebyshev's Inequality and get that $P[X - \mu \geq a] \leq P[|X - \mu| \geq a] \leq \frac{\sigma^2}{a^2}$ for $a > 0$.

1.3 One-Sided Chebyshev's Inequality

Proposition: If $E[X] = 0, P[X \geq a] \leq \frac{\sigma^2}{\sigma^2 + a^2} \leq \frac{\sigma^2}{a^2}$.

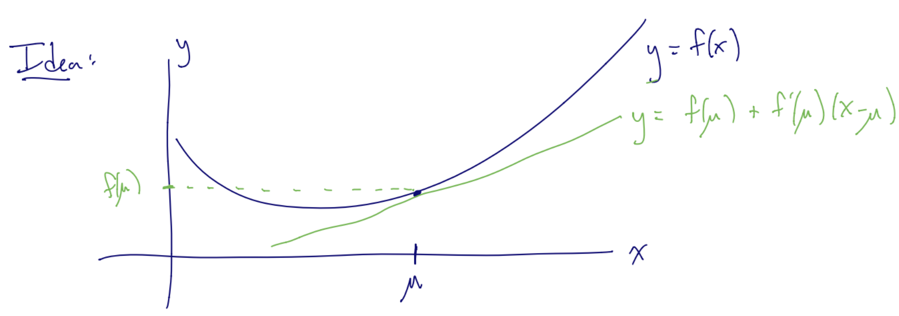
Proof. Let $b > 0$ and note that $X \geq a$ is equivalent to $X + b \geq a + b$. We see that $P[X \geq a] = P[X+b \geq a+b] \leq P[(X+b)^2 \geq (a+b)^2] \leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{\sigma^2+b^2}{(a+b)^2}$. Setting $b = \frac{\sigma^2}{a}$ (since this minimizes $\frac{\sigma^2+b^2}{(a+b)^2}$), we get that $P[X \geq a] \leq \frac{\sigma^2}{\sigma^2+a^2}$. QED

We see that this is a tighter bound than the original Chebyshev's Inequality, since $\frac{\sigma^2}{\sigma^2+a^2} \leq \frac{\sigma^2}{a^2}$. This provides us a better error bound.

1.4 Jensen's Inequality

Proposition: If f is a convex function ($\forall x, f''(x) \geq 0$), then $E[f(x)] \geq f(E[X])$, assuming $E[f(x)]$ and $E[X]$ exist and are finite.

Proof. Expanding $f(x)$ in a Taylor series expansion about $\mu = E[X]$ yields $f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi)(x-\mu)^2}{2}$, where ξ is some value between x and μ. Since $f''(\xi) \geq 0$, we obtain that $f(x) \geq f(\mu) + f'(\mu)(x - \mu)$. Thus, $f(X) \geq f(\mu) + f'(\mu)(X - \mu)$. Taking an expectation of both sides, we get that $E[f(X)] \geq E[f(\mu) + f'(\mu)(X - \mu)] = f(\mu) + f'(\mu)E[X - \mu] = f(E[X])$. A graphical interpretation can be helpful to understand Jensen's Inequality.



By graphing the curve $y = f(x)$ and the tangent line to $f(x)$ at $x = \mu$ ($y = f(\mu) + f'(\mu)(x - \mu)$), we can see that the curve $y = f(x)$ is always above or equal to the tangent line. This is due to f being a convex function. Thus, we see that $f(x) \geq f(\mu) + f'(\mu)(x - \mu)$. Replacing x with X and taking expectations on both sides yields Jensen's Inequality.

2 Synchronous Lecture

2.1 Recap

The Strong Law of Large Numbers (SLLN) states that for IID random variables $X_1, X_2, ...$ with $E[X_i] = \mu < \infty$, then $P[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu] = 1$.

The Weak Law of Large Number (WLLN) states that for any $\epsilon > 0$, $\lim_{n\to\infty} P[|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon] = 0$.

The key assumption for the proof of the WLLN is that $Var[X_i] = \sigma^2 < \infty$.

The key assumption for the proof of the SLLN is that $E[X_i^4] = K < \infty$.

2.2 Proof of the Strong Law of Large Numbers

Proof. Without loss of generality, we can assume that $E[X_i] = 0$. Let $S_n = \sum_{i=1}^{n} X_i$.

We see that $E[S_n^4] = E[(X_1 + X_2 + ... + X_n)^4] = E[\sum X_i^4 + \sum X_i^3 X_j + \sum X_i^2 X_j^2 + \sum X_i^2 X_j X_k + \sum X_i X_j X_k X_l] = E[\sum X_i^4 + \sum_{i\neq j} X_i^2 X_j^2] = nE[X_i^4] + \binom{4}{2}\binom{n}{2}E[X_1^2 X_2^2]$.

Since X_1^2 and X_2^2 are independent, then we get that $E[S_n^4] = nE[X_1^4] + \binom{4}{2}\binom{n}{2}E[X_1^2 X_2^2] = nK + \frac{6n(n-1)}{2}E[X_1^2]E[X_2^2] = nK + \frac{6n(n-1)}{2}(E[X_1^2])^2$.

Now note that $Var[X_i^2] = E[X_i^4] - (E[X_i^2])^2 \geq 0$. Thus, we know that $(E[X_1^2])^2 = E[X_1^2]E[X_2^2] \leq K = E[X_1^4]$.

Thus, we see that $E[S_n^4] \leq nK + 3n(n-1)K$.

Dividing both sides by n^4, we get that $E[\frac{S_n^4}{n^4}] \leq \frac{K}{n^3} + \frac{3(n^2-n)K}{n^4} \leq \frac{K}{n^3} + \frac{3K}{n^2} - \frac{3K}{n^3} \leq \frac{K}{n^3} + \frac{3K}{n^2}$. Since $E[\frac{S_n^4}{n^4}] \leq \frac{K}{n^3} + \frac{3K}{n^2}$, then we know that $E[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}] = \sum_{n=1}^{\infty} E[\frac{S_n^4}{n^4}] \leq \sum_{n=1}^{\infty}(\frac{K}{n^3} + \frac{3K}{n^2}) < \infty$.

Hence, we see that $\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty$ with probability 1.

Therefore, with probability 1, we know that $\lim_{n\to\infty} \frac{S_n^4}{n^4} = 0$. And therefore, with probability 1, $\lim_{n\to\infty} \frac{S_n}{n} = 0$. QED

3 Additional Examples

3.1 Example 5a

Problem: If the number of items produced in a factory during a week is a random variable with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be at least 120.

Solution. We know that $\mu = 100$ and $\sigma^2 = 400$. It follows from the one-sided Chebyshev Inequality that $P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400+20^2} = \frac{1}{2}$.

Hence, the probability that this week's production will be 120 or more is at most $\frac{1}{2}$.

If we attempted to obtain a bound by applying Markov's inequality, then we would have obtained $P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}$, which is a far weaker bound than the preceding one.

3.2 Example 5f

An investor is faced with the following choices: Either she can invest all of her money in a risky proposition that would lead to a random return X that has mean m, or she can put the money into a risk-free venture that will lead to a return of m with probability 1. What decision will she make if her decision will be made on the basis of maximizing the expected value of $u(R)$, where R is her return and u is her utility function.

By Jensen's inequality, it follows that if u is a concave function, then $E[(u(X))] \leq u(m)$, so the risk-free alternative is preferable, whereas if u is convex, then $E[u(X)] \geq u(m)$, so the risky investment alternative would be preferred.