Learning Objectives

By the end of this lecture, you should be able to:

- Be able to understand some important inequalities.
- Be able to understand the Law of Large Number

Introduction

The most important theoretical results in probability are limit theorems, especially the law of large numbers and central limit theorem. Both tell us what happens to the sample mean as we obtain more and more data. Limit theorems let us make approximations which are likely to work well when we have a large number of data points. Before we introduce the law of large numbers and central limit theorem, we will start with some important inequalities in statistics.

Inequalities

Jensen's Inequalities. Let X be a random variable. If g is a convex function, then $E[g(X)] \geq g[E(X)]$. If g is a concave function, then $E[g(X)] \leq g[E(X)]$.

Remarks: If we forget the direction of Jensen's inequality, these simple cases make it easy to recover the correct direction.

- $E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}$ for positive random variables X.
- $E[\log(X)] \leq \log[E(X)]$ for positive random variables X.

Markov's Inequality. For any random variable X, $X > 0$, and a constant $a > 0$ we have

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof:

Chebyshev Inequality. Let X have mean μ and variance σ^2. Then for any $a > 0$ we have

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Proof:

Chernoff Inequality. For any random variable X and constants $a > 0$ and $t > 0$,

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$$

Proof:

Example 1.1. Let X be a random variable representing a test score, following normal distribution with mean of 70 and standard deviation of 5. We know that $P(X > 80) = 0.00135$ by converting to Z-value and find out the exact probability. Now compute the upper bounds obtained from Markov's and Chebyshev's $P(|X - 70| \geq 15)$.

Remark of Three Inequalities

In real life, we almost never know the true distribution (Normal, Gamma, etc.). We propose these inequalities because they provide universal safety guarantees that work for any distribution. They allow us to prove that our sample estimator, most of time would be sample mean, will eventually converge to the truth.

Law of Large Numbers

We now turn to the law of large numbers (LLN). The LLN states that as the sample size n grows, the sample mean $\bar{X}_n$

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots X_n}{n}$$

converges to the true mean μ. There are two versions of LLN, Strong LLN (SLLN) and Weak LLN (WLLN), which use slightly different definitions of what it means for a sequence of random variables to converge to a number.

Strong Law of Large Number. The sample mean $\bar{X}_n$ converges to the true mean μ pointwise, with probability 1 as $n \to \infty$. In other words, $P(\bar{X}_n \to \mu) = 1$ as $n \to \infty$, the event $\bar{X}_n \to \mu$ has probability 1. The event of the convergence has probability 1

Weak Law of Large Number. The sample mean $\bar{X}_n$ is "far" (by distance ϵ) from true mean μ with probability approaches 0 as n grows. For all $\epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$. (This form of convergence is called *Convergence in probability.*). The probability of the error being large approaches 0.

Learning Objectives

By the end of this lecture, you should be able to:

- Understand the difference between SLLN and WLLN.
- Understand the Central Limit Theorems.

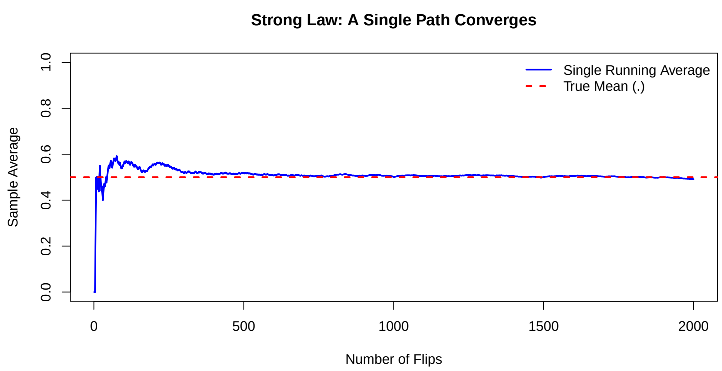
SLLN vs WLLN

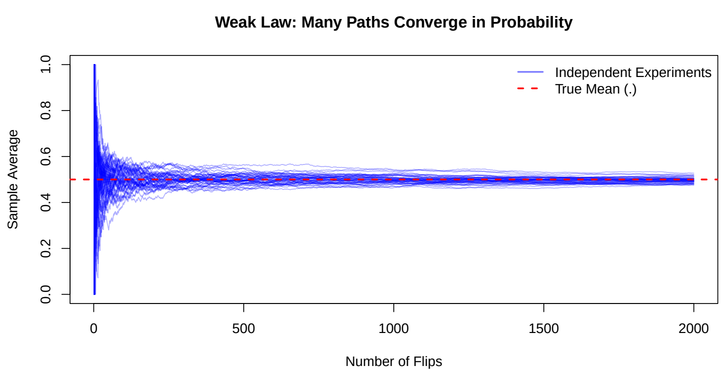
Strong Law of Large Number. The sample mean $\bar{X}_n$ converges to the true mean μ pointwise, with probability 1 as $n \to \infty$. In other words, $P(\bar{X}_n \to \mu) = 1$ as $n \to \infty$.

Weak Law of Large Number. For all $\epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$. (This form of convergence is called *Convergence in probability.*)

Remarks:

- WLLN says that the sample mean is likely close to μ for large n.
- SLLN states that with probability 1 the sample mean will eventually stay close to μ.





Central Limit Theorem (CLT)

Suppose $X_1, X_2, \ldots, X_n$ represents a random sample of Normal(μ, σ^2) random variables. Find the distribution of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

What if X_i's were non-Normal random variables?

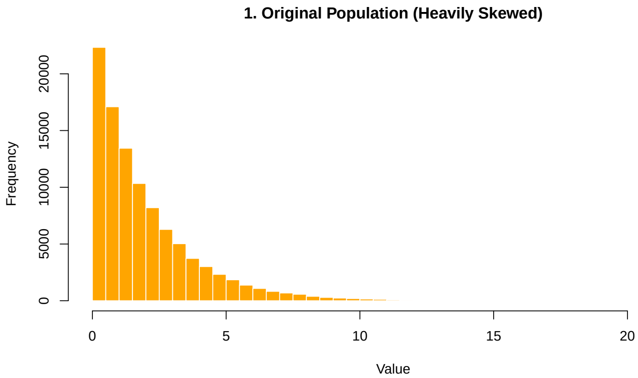
Central Limit Theorem. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Define

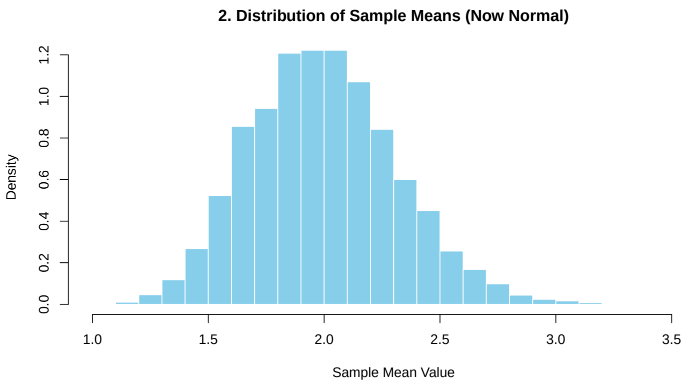
$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then the distribution of Z_n converges to the standard Normal distribution as $n \to \infty$.

Proof

Example 2.1. Suppose we have a heavily skewed exponential distribution. Let's take 5,000 samples of 40 values each and plotted the average of every sample.





Remarks: Note that the CLT holds not only for $\bar{X}_n$ but also for a sum of iid random variables with finite variance.

Example 2.2. Let $Y \sim \text{Bin}(n, p)$, we can consider Y to be a sum of n iid $\text{Bern}(p)$ random variables. Therefore for large n,

$$Y \dot{\sim} N(np, np(1-p))$$

Example 2.3. Suppose that X_i are iid Bern(p) random variables with $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{n} X_i$. Determine the expected value and variance of $\bar{Y} = \frac{Y}{n}$ and find the approximate distributions of Y.

Learning Objectives

By the end of this lecture, you should be able to:

- Understand the Central Limit Theorems.
- Understand the Normal Approximation.

Central Limit Theorem

Example 3.1. Candidate A believes that she can win a city election if she can earn at least 55% of the votes in precinct 1. She also believes that about 50% of the city's voters favor her. If $n = 100$ voters show up to vote at precinct 1, what is the probability that candidate A will receive at least 55% of their votes?

Example 3.2. Suppose that X has a binomial distribution with $n = 25$ and $p = 0.4$. Find the exact probabilities that $X \leq 8$ and $X = 8$ and compare these to the corresponding values found by using the normal approximation.

Remarks: The CLT requires that the mean and variance of the X_i to be finite, and our proof of the WLLN relied on the same conditions. The Cauchy distribution has no mean or variance so the Cauchy distribution obeys either the LLN or the CLT not matter how large n gets. So the sample mean never approaches a Normal distribution, contrary to the behavior seen in the CLT. There is also no true mean for $\bar{X}_n$ to converge to, so the LLN does not apply either.

Example 3.3. For iid $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, the sample variance is the random variable

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

Show that

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Student t distribution Let

$$T = \frac{Z}{\sqrt{V/n}},$$

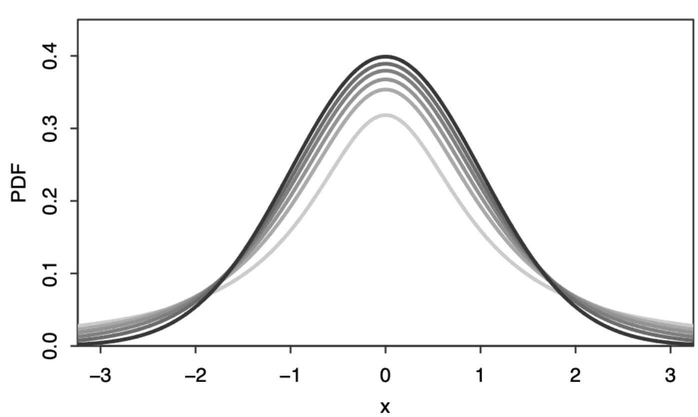
where $Z \sim N(0,1)$, $V \sim \chi^2_n$, and Z is independent of V. Then T is said to have the Student t distribution with n degrees of freedom. We write as $T \sim t_n$. In short, we call it t distribution. The pdf is

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

This is a very complicated form.

Remarks:

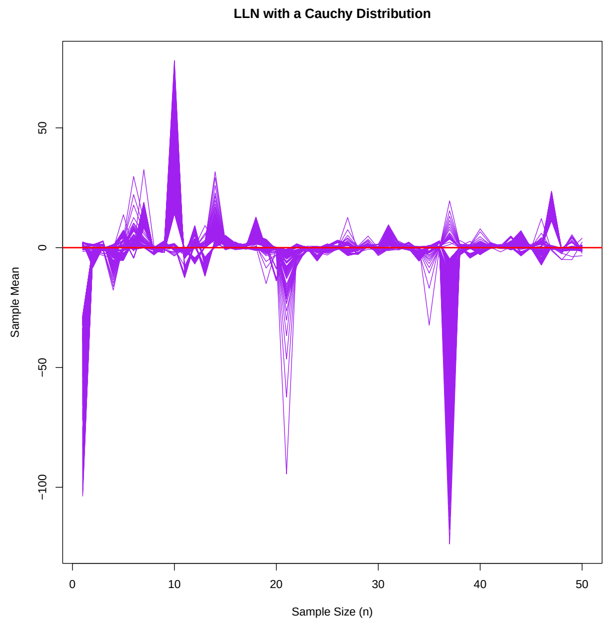
- The t distribution is used in hypothesis testing procedures, known as t-test.
- The pdf of t distribution is heavy tails, and much heavier if n is small.
- Cauchy distribution is a special case of t distribution with one degree of freedom.
- As $n \to \infty$, the t_n distribution approaches to the standard Normal distribution.



Instructions

This Homework contains *4 questions* and *20 points* in total. Show all work for full credit. Partial credit will be given if your process is clearly shown and mathematically reasonable. If the question is theoretical, please provide a written explanation.
Due date: Friday, December 12th by 11:59pm.

1. A machine produces light bulbs whose lifetimes (in hours) are modeled by an exponential random variable $X \sim \text{Expo}\left(\lambda = \frac{1}{500}\right)$.
   (a) (2 points) Compute $P(X \geq 1000)$ using Markov's inequality.
   (b) (2 points) Compute $P(X \geq 1000)$ using Chebyshev's inequality.
   (c) (2 points) Compute $P(X \geq 1000)$ using the exact exponential cdf and compare the results with the previous two.
2. (4 points) Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables defined as $Y_i = e^{X_i}$ for each i, where $X_i \overset{iid}{\sim} \text{Expo}(3)$. Find the approximated distribution of the sample mean of Y_i, that is, $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
3. From the past experience, we know that the test score of a student taking the final examination is a random variable with mean 70 (out of 100) and standard deviation 10.
   (a) (2 points) Find an upper bound for the probability that a student's test score will exceed 85. *Hint: Use one of the well-known inequality.*
   (b) (2 points) Find the lower bound for the probability that a student will score between 55 and 85. *Hint: The two event $\{55 < X < 85\}$ and $\{|X - 70|| < 15\}$ are equivalent. Use one of a well-known inequality to find the probability.*
   (c) (3 points) How many students would have to take the examination to ensure that the class average would be within 65 of 75 with probability at least 0.9? That is, find n satisfying that $P(|\bar{X}_n - 70| < 5) \geq 0.9$. *Hint: Use the inequality in (b). Keep in mind, the random variable we are interested here is $\bar{X}_n$ not X. You need to compute $Var(\bar{X}_n)$.*
4. (3 points) In the Lecture we introduced that mean of Cauchy distribution is not finite. Now let's see the following plot. The data is generated from Cauchy distribution.



Based on your observation, explain why having a finite mean is a crucial condition for the Law of Large Numbers.