Lecture 10: Sampling Distributions and Limits

MATH 697

Omidali Jazi

October 4, 2018

McGill University

Goals for this Chapter

- Preview
- Convergence in probability
  - Weak law of large numbers
- Convergence with probability 1
  - Strong Law of Large Numbers
  - Monte Carlo Integration
- convergence in distribution
  - The Central Limit Theorem

Preview

- This lecture makes the **transition between probability and inferential statistics**.
- Given a sample of n observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters).
- Prior to obtaining data, there is uncertainty as to which of all possible samples will occur.
- Because of this, estimates such as $\bar{x}$ (the sample mean) will vary from one sample to another

Preview

- The behavior of such estimates in repeated sampling is described by what are called **sampling distributions**.
- Any particular sampling distribution will give an indication of how close the estimate is likely to be to the value of the parameter being estimated.

Preview

- We will use probability results to study sampling distributions.
- A particularly important result is the **Central Limit Theorem**, which shows how the behavior of the sample mean can be described by a particular normal distribution when the sample size is large.

Statistics and Their Distributions

Two random samples will be different

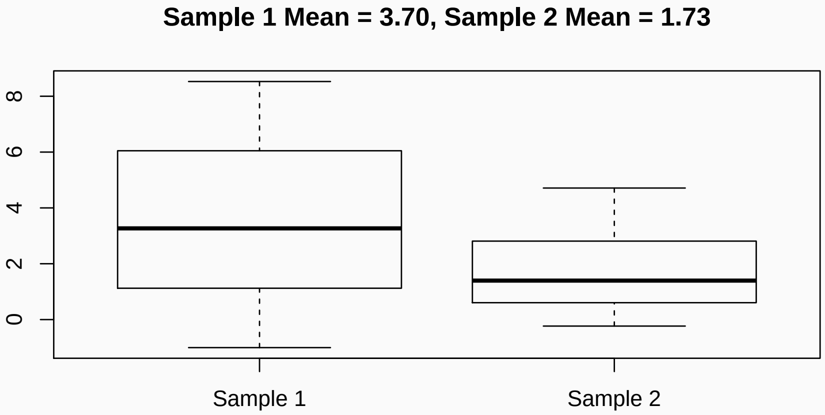
- The observations in a single sample are denoted by $x_1, x_2, \dots, x_n$
- Consider selecting two different samples of size n from the same population distribution.
- The x_i's in the second sample will virtually always differ at least a bit from those in the first sample.

Uncertainty in Summary Measures of the Random Samples

- This variation in observed values in turn implies that the value of any function of the sample observations - such as the sample mean or sample standard deviation also varies from sample to sample.
- That is, prior to obtaining $x_1, \dots, x_n$, there uncertainty as to the value of $\bar{x}$ and s (the sample standard deviation)

Two Random Samples from a N(2,4) Distribution

```
x1 <- rnorm(10, 2, 2) ; x2 <- rnorm(10, 2, 2)
boxplot(x1,x2, main = sprintf("Sample 1 Mean = %0.2f, Sample 2 Mean = %0.2f",
        mean(x1), mean(x2)), names = c("Sample 1","Sample 2"))
```



A Statistic

Definition 1 (Statistic)

- A **statistic** is any quantity whose value can be calculated from sample data
- Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result.
- A **statistic is a random variable** and will be denoted by an uppercase letter (e.g. $\bar{X}$)
- A lowercase letter is used to represent the calculated or observed value of the statistic (e.g. $\bar{x}$)

Sample Mean is a Statistic

- Suppose a drug is given to a sample of patients, another drug is given to a second sample, and the cholesterol levels are denoted by $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively.
- The statistic $\bar{X} - \bar{Y}$, i.e., the difference between the two sample mean cholesterol levels, may be important.

Any Statistic has a Probability Distribution

- Suppose, for example, that $n = 2$ components are randomly selected and the number of breakdowns while under warranty is determined for each one.
- Possible values for the sample mean number of breakdowns $\bar{X}$ are

X_1	X_2	$\bar{X}$
0	0	0
0	1	0.5
1	0	0.5
0	2	1
2	0	1
$\vdots$	$\vdots$	$\vdots$

Probability Distribution of Statistic is its Sampling Distribution

- The probability distribution of $\bar{X}$ specifies $P(\bar{X} = 0)$, $P(\bar{X} = 0.5)$, $P(\bar{X} = 1)$ and so on
- From these, other probabilities such as $P(1 \leq \bar{X} \leq 3)$ and $P(\bar{X} \geq 2.5)$ can be calculated
- The probability distribution of a statistic is referred to as its **sampling distribution** to emphasize that it describes how the statistic varies in value across all samples that might be selected.

Random Samples

Definition 2 (Random Sample)

The random variables $X_1, X_2, \ldots, X_n$ are said to form a **random sample** of size n is

- The X_i's are independent random variables
- Every X_i has the same probability distribution

These two conditions can be paraphrased by saying that the X_i's are *independent and identically distributed* (iid).

Deriving the Sampling Distribution of a Statistic

- Probability rules can be used to obtain the distribution of a statistic provided that
  - it is a fairly simple function of the X_i's and
  - either there are relatively few different X values in the population or the population distribution has a nice form
- The next examples illustrate such a situation and provides a motivation for finding an approximation of the sampling distribution

Example 3 (MP3 Players)

A certain brand of MP3 player comes in three configurations:

memory	2 GB	4 GB	8 GB
x (cost)	80	100	120
$p(x)$	0.20	0.30	0.50

With $\mu = 106, \sigma^2 = 244$. Suppose only two MP3 players are sold today: X_1 and X_2 representing the cost of the 1st and 2nd player, respectively. When $n = 2$, $s^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	(.2)(.2) = .04	80	0
80	100	(.2)(.3) = .06	90	200
80	120	(.2)(.5) = .10	100	800
100	80	(.3)(.2) = .06	90	200
100	100	(.3)(.3) = .09	100	0
100	120	(.3)(.5) = .15	110	200
120	80	(.5)(.2) = .10	100	800
120	100	(.5)(.3) = .15	110	200
120	120	(.5)(.5) = .25	120	0

Example (MP3 Players) cont 1

Example 4 (MP3 Players)

To obtain the probability distribution of $\bar{X}$, the sample average cost per MP3 player, we must consider each possible value $\bar{x}$ and compute its probability, e.g., $P(\bar{x} = 100) = 0.10 + 0.09 + 0.10 = 0.29$, $P(S^2 = 800) = 0.10 + 0.10 = 0.20$. The complete sampling distributions of $\bar{X}$ and S^2 are given below:

$\bar{x}$	80	90	100	110	120
$p(\bar{x})$	0.04	0.12	0.29	0.30	0.25

s^2	0	200	800
$p(s^2)$	0.38	0.42	0.20

- $E(\bar{X}) = \sum \bar{x} p_{\bar{X}}(\bar{x}) = 106 = \mu$
- $V(\bar{X}) = \sum (\bar{x} - \mu)^2 = \sum (\bar{x} - 106)^2 p_{\bar{X}}(\bar{x}) = 122 = 244/2 = \sigma^2/2$ (half the population variance: why?)
- $E(S^2) = \sum s^2 p_{S^2}(s^2) = 0(0.38) + 200(0.42) + 800(0.20) = 244 = \sigma^2$

Example (MP3 Players) cont 2

Example 5 (MP3 Players)

The probability histogram for both the original distribution X (a) and the $\bar{X}$ (b) distribution. We see that the mean of $\bar{X}$ (denoted by $E(\bar{X})$) is equal to the mean of the original distribution. We also see that the $\bar{X}$ distribution has **smaller spread** than the original distribution, since the values of $\bar{x}$ are **more concentrated toward the mean**. The $\bar{X}$ sampling distribution is centered at the population mean μ.

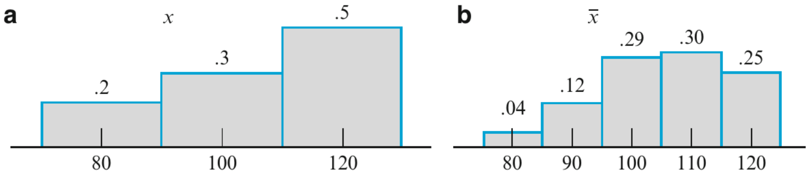


Figure 6.2 Probability histograms for (a) the underlying population distribution and (b) the sampling distribution of $\overline{X}$ in Example 6.2

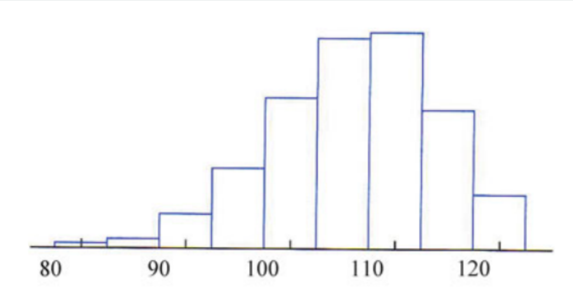
Example (MP3 Players) cont 3

Example 6 (MP3 Players)

If four MP3 players had been purchased on the day of interest, the sample average cost $\bar{X}$ would be based on a random sample of four X_i's. More calculation eventually yields the distribution of $\bar{X}$ for $n = 4$ as

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\bar{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625

From this, $E(\bar{X}) = 106 = \mu$ and $V(\bar{X}) = 61 = \sigma^2/4$



Some Remarks

- The previous example showed us that the computation of $p_{\bar{X}}(\bar{x})$ and $p_{S^2}(s^2)$ can be tedious
- This example should also suggest that there are some general relationships between $E(\bar{X})$, $V(\bar{X})$, $E(S^2)$ and the population mean μ and variance σ^2.
- Sampling distributions can sometimes be computed by direct computation or by approximations such as the **central limit theorem (CLT)**
- Techniques for deriving such approximations will be discussed next

Convergence in Probability

Convergence in Probability

Definition 7 (Convergence in Probabilty)

Let $X_1, X_2, \ldots$ be an infinite sequence of random variables, and let Y be another random variable. Then the sequence $\{X_n\}$ converges in probability to Y, if for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - Y| \geq \epsilon) = 0 \tag{1}$$

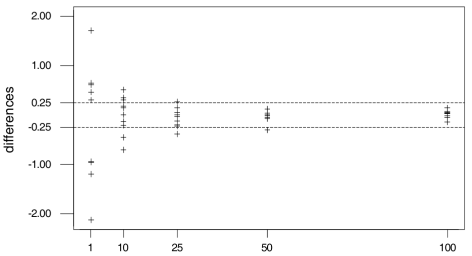
or equivalently

$$\lim_{n\to\infty} P(|X_n - Y| < \epsilon) = 1 \tag{2}$$

In short notation we write $X_n \xrightarrow{p} Y$

Convergence in Probability

We plot the differences $X_n - Y$ for selected values of n, for 10 generated sequences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y in probability. We have also plotted the horizontal lines at $\pm\epsilon$ for $\epsilon = 0.25$. From this we can see the increasing concentration of the distribution of $X_n - Y$ about 0, as n increases, as required by Definition (7). In fact, the 10 observed values of $X_{100} - Y$ all satisfy the inequality $|X_{100} - Y| < 0.25$.



Convergence in Probability Example

Example 8 (Identical Random Variables)

Let X be any random variable, and let $X_1 = X_2 = X_3 = \cdots = X$, i.e., the random variables are all identical to each other.

Example: Functions of Uniforms

- Let $U \sim Uniform(0, 1)$. Define X_n by

$$X_n = \begin{cases} 3 & U \leq 2/3 - 1/n \\ 8 & otherwise \end{cases}$$

  and define X by

$$X = \begin{cases} 3 & U \leq 2/3 \\ 8 & otherwise \end{cases}$$

- Then $X_n \xrightarrow{p} X$

Example: Functions of Uniforms

- Solution: It follows from the definition that for every $\epsilon > 0$, we can take a sufficiently large n such that

$$P(|X_n - X| > \epsilon) = P(|-1/n| \geq \epsilon) \to 0$$

Example : Exponential and a Constant

Let $Y_n \sim Exp(1/n)$ and let $Y = 0$.

- Show that $Y_n \xrightarrow{p} 0$
- Solution:

$$
\begin{aligned}
P(|Y_n - 0| \leq \epsilon) &= P(-\epsilon \leq Y_n \leq \epsilon) \\
&= P(Y_n \leq \epsilon) - P(Y_n \leq -\epsilon) \\
&= F_{Y_n}(\epsilon) - 0 = 1 - e^{-n\epsilon} \to 1
\end{aligned}
$$

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**
- Suppose $X_1, X_2, \cdots$ is a sequence of independent random variables that each have the same mean μ and variance σ^2
- For all positive integers n, let

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

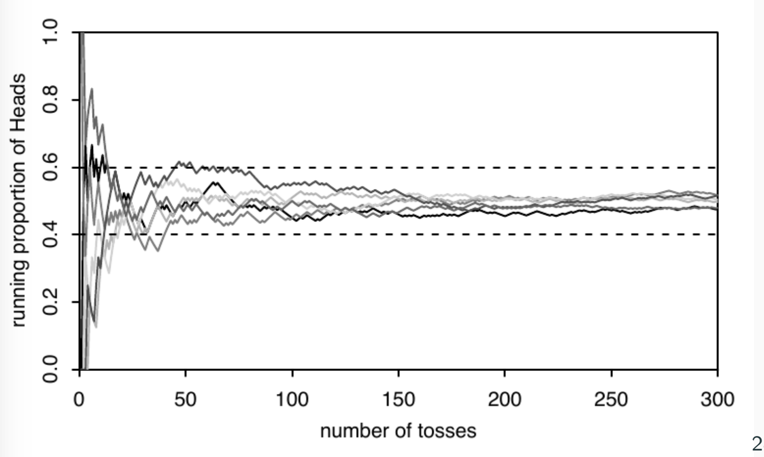
  be the *sample average*, or *sample mean*, for $X_1, \cdots, X_n$
- When the sample size n is fixed, we will often use $\bar{X}$ as a notation for sample mean instead of $\bar{X}_n$.
- The sample mean is itself a random variable with mean μ and variance σ^2/n (why?)

Coin Flips

- If we flip a sequence of fair coins, and if $X_i = 1$ or $X_i = 0$ as the ith coin comes up heads or tails, then $\bar{X}_n$ represents the fraction of the first n coins that came up heads
- We might expect that for large n, this fraction will be close to $1/2$, i.e., to the expected value of the X_i
- The **weak law of large numbers** provides a precise sense in which average values $\bar{X}_n$ tend to get close to $E(X_i)$, for large n

Running proportion of Heads in 6 sequences of fair coin tosses

- Coin Flips: Dashed lines at 0.6 and 0.4 are plotted for reference. As the number of tosses increases, the proportion of Heads approaches 1/2.



Weak Law of Large Numbers

Theorem 9 (Weak Law of Large Numbers (WLLN))

Let $X_1, X_2, \cdots,$ be a sequence of independent random variables, each having the same mean μ and each having finite variance $\sigma^2 < \infty$. Then for all $\epsilon > 0$,

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0 \tag{3}$$

That is, the averages converge in probability to the common mean μ or $\bar{X}_n \xrightarrow{p} \mu$

Consistency of Sample Variance

Example 10 (Sample Variance)

Let $X_1, X_2, \cdots,$ be a sequence of iid random variables, each having the same mean $E(X_i) = \mu$ and each having variance $V(X_i) = \sigma^2 < \infty$. If we define

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \tag{4}$$

can we prove a WLLN for S_n^2?

Consistency of Sample Variance

Example 10 (Sample Variance)

Let $X_1, X_2, \cdots,$ be a sequence of iid random variables, each having the same mean $E(X_i) = \mu$ and each having variance $V(X_i) = \sigma^2 < \infty$. If we define

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \tag{4}$$

can we prove a WLLN for S_n^2?

Using Chebyshev's Inequality we have,

$$P(|S_n^2 - \sigma^2| \geq \epsilon) \leq \frac{E(S_n^2 - \sigma^2)^2}{\epsilon^2} = \frac{Var(S_n^2)}{\epsilon^2}$$

and thus, a sufficient condition that S_n^2 converges in probability to σ^2 is that $Var(S_n^2) \to 0$ as $n \to \infty$

WLLN Applications

Example 11 (Fair coins)

Consider flipping a sequence of identical fair coins. Let $\bar{X}_n$ be the fraction of the first n coins that are heads. Then $\bar{X}_n = (X_1 + \cdots + X_n)/n$, where $X_i = 1$ if the ith coin is heads, otherwise $X_i = 0$.

Visualization of the Law of Large Numbers

Exercise in `R`: To plot the running proportion of Heads in a sequence of independent fair coin tosses, we first generate the coin tosses themselves:

```
nsim <- 300
p <- 1/2
x <- rbinom(nsim, 1, p)
```

Then we compute $\bar{X}_n$ for each value of n and store the results in `xbar`:

```
# what is this code doing?
xbar <- cumsum(x) / (1:nsim)
```

Finally, plot $xbar$ against the number of coin tosses. What do you notice?

WLLN Importance

- The law of large numbers (LLN) is essential for **simulations, statistics, and science!!**.
- Consider generating data from a large number of independent replications of an experiment, performed either by computer simulation or in the real world
- Every time we use the proportion of times that something happened as an approximation to its probability, we are implicitly appealing to LLN.
- Every time we use the average value in the replications of some quantity to approximate its theoretical average, we are implicitly appealing to LLN.

Summary

- A sequence $\{X_n\}$ of random variables converges in probability to Y if

$$\lim_{n\to\infty} P(|X_n - Y| \geq \epsilon) = 0$$

- The Weak Law of Large Numbers (WLLN) says that if $\{X_n\}$ is iid, then

$$\bar{X}_n = (X_1 + \cdots + X_n)/n \xrightarrow{p} E(X_i)$$

Convergence with Probability 1

Convergence with Probability 1

Definition 12 (Convergence with Probability 1)

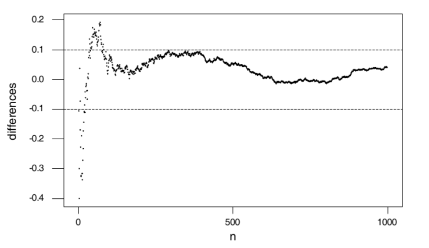
Let $X_1, X_2, \ldots,$ be an infinite sequence of random variables. We shall say that the sequence $\{X_i\}$ **converges with probability 1** (or **converges almost surely** (a.s)) to a random variable Y if

$$P(\lim_{n \to \infty} X_n = Y) = 1 \tag{5}$$

we write this as $X_n \overset{a.s.}{\to} Y$

Convergence with Probability 1

- Graph of the sequence of differences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y with probability 1.



- Definition (12) indicates that for any given $\varepsilon > 0$, there will exist a value N_ε such that $|X_n - Y| < \varepsilon$ for every $n \geq N_\varepsilon$.
- Contrast this with the situation depicted for convergence in probability, which only says that the probability distribution $X_n - Y$ concentrates about 0 as n grows and not that the individual values of $X_n - Y$ will necessarily all be near 0

Strong Law of Large Numbers

The following is a strengthening of the weak law of large numbers because it concludes convergence with probability 1 instead of just convergence in probability.

Theorem 13 (Strong Law of Large Numbers (SLLN))

Let $X_1, X_2, \cdots,$ be a sequence of independent random variables, each having finite mean μ. Then

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1 \tag{6}$$

That is, the averages converges with probability 1 to the common mean μ or $\bar{X}_n \overset{a.s}{\to} \mu$

SLLN Applications: Monte Carlo Integration

- Suppose we want to evaluate the integral $\mathcal{J} = \int_a^b h(x)dx$ for some function h. If h is complicated there may be no known closed form expression for $\mathcal{J}$.
- **Monte Carlo integration** is an approach for approximating $\mathcal{J}$ which is notable for its simplicity, generality and scalability.
- Let us begin by writing

$$\mathcal{J} = \int_a^b h(x)dx = \int_a^b w(x)f(x)dx$$

where $w(x) = h(x)(b-a)$ and $f(x) = 1/(b-a)$.

SLLN Applications: Monte Carlo Integration

- Notice that f is the PDF for a $Uniform(a, b)$. Hence

$$\mathcal{J} = E_f[w(X)]$$

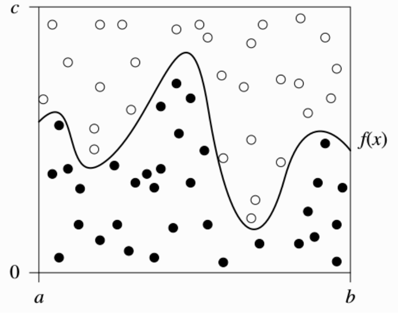
where $X \sim Uniform(a, b)$.

- If we generate $X_1, \ldots, X_N \sim Uniform(a, b)$, then by the **Strong Law of Large Numbers**

$$\hat{\mathcal{J}} \equiv \frac{1}{N} \sum_{i=1}^{N} w(X_i) \rightarrow E(w(X)) = \mathcal{J}$$

Monte Carlo Integration Visual

- We thus transform the function integration problem into an area estimation problem.
- This Monte Carlo integration method as known as the "hit or miss" approach, because the approximation is based on the hit-or-miss estimate of the area.



Monte Carlo Integration Exercise

Use Monte Carlo integration to solve for these integrals and see that its close to the actual value.

Exercise 14 (Monte Carlo Integration)

1. *Let* $h(x) = x^3$. *Then* $\mathcal{I} = \int_0^1 x^3 dx = 1/4$.
2. $\mathcal{I} = \Phi(1.25) = \int_{-\infty}^{1.25} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$. *Verify your answer with* `pnorm`
3. $\mathcal{I} = \int_{0.25}^{0.75} \frac{4}{1+x^2} dx$. *Verify your answer with* `integrate`

MGFs to Determine Distribution of the Sample Mean

Theorem 15 (MGF of Sample Mean)

Let $X_1, X_2, \cdots,$ be a sequence of independent random variables with MGF $M_X(t)$. Then the MGF of the sample mean is

$$M_{\bar{X}}(t) = [M_X(t/n)]^n \tag{7}$$

Proof: on board

This theorem is only useful if the expression for $M_{\bar{X}}(t)$ is a familiar MGF.

Example: Distribution of the Mean of Normal RVs

- Let $X_1, X_2, \ldots$ be iid with distribution $N(\mu, \sigma^2)$. Using MGFs, find the distribution of the sample mean $\bar{X}_n$
- Solution: By Theorem 15, we have

$$\begin{aligned} M_{\bar{X}_n}(t) &= [M_X(t/n)]^n = [e^{\mu(t/n)+\sigma^2(t/n)^2}]^n \\ &= e^{\mu t+\sigma^2 t^2/n} \end{aligned}$$

- That is, $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

Convergence in Distribution

Definition 16 (Convergence in Distribution)

Let $X_1, X_2, ...$ be a sequence of random variables. Then we say that the **sequence converges in distribution** to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous and we write $X_n \xrightarrow{D} X$

Convergence in Distribution

- Intuitively, $X_1, X_2, \ldots$ converges in distribution to X if for large n, the distribution of X_n is close to that of X
- The importance of this, is that often the distribution of X_n is difficult to work with, while that of X is much simpler
- With X_n converging in distribution to X, however, we can approximate the distribution of X_n by that of X.

Convergence in Probability Implies in Distribution

Theorem 17 (Convergence in Probability Implies in Convergence in Distribution)

If the sequence of random variables $X_1, X_2, ...$ converges to a random variable X, the sequence also converges in distribution to X. That is if $X_n \overset{P}{\to} X$ then $X_n \overset{D}{\to} X$

Poisson Approximation to the Binomial

Example 18 (Poisson Approximation to the Binomial)

Suppose that $X_n \sim Binomial(n, \lambda/n)$ and $X \sim Poisson(\lambda)$. We have previously seen that as $n \to \infty$

$$P(X_n = j) = \binom{n}{j} \left(\frac{\lambda}{n}\right)^j \left(1 - \frac{\lambda}{n}\right)^{n-j} \to e^{-\lambda} \frac{\lambda^j}{j!}$$

This implies that $F_{X_n}(x) \to F_X(x)$ and thus $X_n \xrightarrow{D} X$

Central Limit Theorem

Introduction

- Let $X_1, X_2, \dots,$ be iid with mean μ and finite variance σ^2. The law of large numbers says that as $n \to \infty$, $\bar{X}_n$ converges to the constant μ (with probability 1).
- But what is its distribution along the way to becoming a constant?
- This is addressed by the central limit theorem (CLT)

Central Limit Theorem (CLT)

- The CLT states that for large n, the distribution of $\bar{X}_n$ (after standardization) approaches a standard Normal distribution.
- By standardization, we mean that we subtract μ, the expected value of $\bar{X}_n$, and divide by $\sigma/\sqrt{n}$, the standard deviation of $\bar{X}_n$.

Theorem 19 (CLT)

As $n \to \infty$

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{D} \mathcal{N}(0,1)$$

From this we can also say that for large n

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

- In words, the CDF of the left-hand side approaches Φ, the CDF of a standard Normal distribution.
- Starting from virtually no assumptions (other than independence and finite variances), **we end up with normality**

Running Proportion of Heads Revisited

Before, we used the law of large numbers to conclude that $\bar{X}_n \to 1/2$ as $n \to \infty$. Now, using the central limit theorem, we can say more: $E(\bar{X}_n) = 1/2$ and $Var(\bar{X}_n) = 1/(4n)$, so for large n,

$$\bar{X}_n \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right)$$

Central Limit Theorem (CLT) for the Sum

- The CLT says that the sample mean $\bar{X}_n$ is approximately Normal
- But since the sum $S_n = X_1 + \cdots + X_n = n\bar{X}_n$ is just a scaled version of $\bar{X}_n$, the CLT also implies that S_n is approximately Normal
- If X_j have mean μ and variance σ^2, S_n has mean $n\mu$ and variance $n\sigma^2$.

Theorem 20 (CLT for the Sum)

for large n

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2)$$

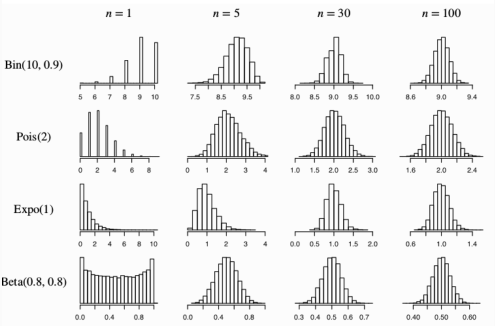
Binomial Convergence to Normal

Let $Y \sim Binomial(n, p)$. Using MGFs it can be shown that Y can be expressed as a sum of n iid Bernoulli(p) random variables. Therefore by CLT, for large n,

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

CLT Visual

Histograms of the distribution of $\bar{X}_n$ for different starting distributions of the X_j (indicated by the rows) and increasing values of n (indicated by the columns). Each histogram is based on 10,000 simulated values of $\bar{X}_n$. Regardless of the starting distribution of the X_j, the distribution of $\bar{X}_n$ approaches a Normal distribution as n grows.

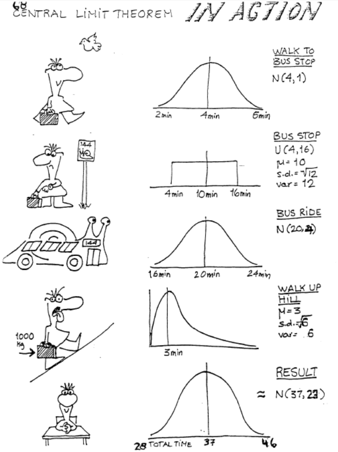


CLT

One way to visualize the CLT for a distribution of interest is to plot the distribution of $\bar{X}_n$ for various values of n. To do this, we first have to generate iid $X_1, \ldots, X_n$ a bunch of times from our distribution of interest. For example, suppose that our distribution of interest is $Unif(0, 1)$, and we are interested in the distribution of $\bar{X}_{12}$, i.e., we set $n = 12$. In the following code, we create a matrix of iid standard Uniforms. The matrix has 12 columns, corresponding to X_1 through X_{12}. Each row of the matrix is a different realization of X_1 through X_{12}.

```
nsim <- 10^4
n <- 12
x <- matrix(runif(n*nsim), nrow=nsim, ncol=n)
xbar <- rowMeans(x)
hist(xbar)
# change runif for rexp. what do you notice?
```

CLT In Action



1 http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-607/notes.pdf

CLT In Action

1. Set $n = 10$
2. Simulate data from various distributions, each representing a different part of the journey to McGill. Create a `data.frame` of the simulated times and add a column for the total transit time

```
walk <- rnorm(n, 4, 1) ; bus <- runif(n, 4, 16)
ride <- rpois(n, 8); climb <- rgamma(n, shape = 6, scale = 0.5)
fall <- rexp(n, rate = 4)
DT <- data.frame(walk, bus, ride, climb, fall)
DT$transit_time <- apply(DT,1,sum)
```

3. Calculate the theoretical means and variances for each of the distributions. Use the CLT to determine the mean and variance of the total transit time.

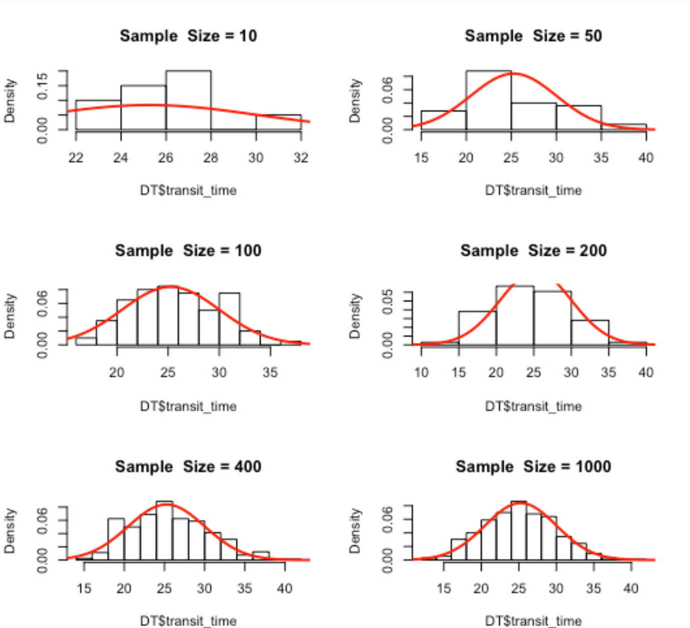
```
mean.walk=4; mean.bus=(4+16)/2; mean.ride=8; mean.climb=6*0.5; mean.fall=1/4
var.walk=1; var.bus=(16-4)^2/12; var.ride=8; var.climb=6*0.5^2; var.fall=1/4^2
mean.transit <- mean.walk+mean.bus+mean.ride+mean.climb+mean.fall
var.transit <- var.walk+var.bus+var.ride+var.climb+var.fall
```

CLT In Action

4. Plot a histogram of the total transit times and superimpose the density of the theoretical distribution of the sum. Repeat the whole exercise for $n = 10, 50, 100, 200, 400, 1000$.

```
hist(DT$transit_time, freq=FALSE)
curve(dnorm(x,mean=mean.transit,sd=sqrt(var.transit)),0,50,
      add=TRUE,lwd=2,col="red")
```

Simulation Results



What We Have Just Learned

- Preview
- Convergence in probability
  - Weak law of large numbers
- Convergence with probability 1
  - Strong Law of Large Numbers
  - Monte Carlo Integration
- convergence in distribution
  - The Central Limit Theorem