STA 111: Probability & Statistical Inference

Lecture Nine – Law of Large Numbers and Central Limit Theorem

D.S. Chapter 6

Instructor: Olanrewaju Michael Akande

Department of Statistical Science, Duke University

Outline

- Questions from Last Lecture
- Linear Combinations
- Law of Large Numbers
- Central Limit Theorem
- Recap

Introduction

– We are now set to round up our discussions on probability. We will still talk about probability but in the context of statistics and how it helps with inference.

– From the next lecture, we will move to statistical inference.

– For today, we will learn about properties of linear combinations.

– Lastly, we will learn about the behavior of large random samples.

Linear Combinations

A linear combination of random variables $X_1, \ldots, X_n$ is a new random variable Y such that

$$Y = a_1 X_1 + \cdots a_n X_n = \sum_{i=1}^{n} a_i X_i$$

where the a_i's are known constants.

Some important linear combinations include:

– The sample mean, $\bar{X}$, in which each a_i equals $1/n$.

– A difference, $X_1 - X_2$, in which $a_1 = 1$ and $a_2 = -1$. This is helpful when deciding whether, say, one brand of lightbulb outlasts another brand, or whether one company outperforms another.

Distributions of Linear Combinations

Let X_i have mean μ_i and variance σ_i^2. Then

$$\mathbb{E}[Y] = \mathbb{E}[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] = \sum_{i=1}^{n} a_i \mu_i.$$

This holds even when the X_i's are dependent. It follows because integration (or sum) is a linear operator: $\int a_i x g_i(x)\, dx = a_i \int x g_i(x)\, dx = a_i \mu_i$. Also,

$$\mathbb{V}[Y] = \mathbb{V}[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}[X_i, X_j]$$

Why? $\mathbb{V}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ and $Y^2 = (a_1 X_1 + \ldots + a_n X_n)^2$ which generates the cross-product terms that define $\mathrm{Cov}[X_i, X_j]$. It takes some algebra but you should work it out to convince yourself.

Distributions of Linear Combinations Cont'd

– Looking at the definitions of variance and covariance, we can see that $\text{Cov}[X_i, X_i]$ is just the variance σ_i^2. So we can write

$$\mathbb{V}[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j \text{Cov}[X_i, X_j].$$

– In the special case when the random variables are independent, then the covariances are all zero and this simplifies to

$$\mathbb{V}[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

– Theorem: If the X_i's have (possibly different, possibly correlated) normal distributions, then Y is normally distributed.

Markov and Chebyshev Inequalities

Markov Inequality (D.S. Theorem 6.2.1): Let X be a random variable such that $\mathbb{P}(X \geq 0) = 1$. Then for every real number $t > 0$,

$$\mathbb{P}\left(X \geq t\right) \leq \frac{\mathbb{E}(X)}{t}$$

Chebyshev Inequality (D.S. Theorem 6.2.2): Let X be a random variable for which Var(X) exists. Then for every number $t > 0$,

$$\mathbb{P}\left(|X - E(X)| \geq t\right) \leq \frac{\mathbb{V}(X)}{t^2}$$

These two inequalities are extremely useful in practice since they do not require knowing the exact distribution of the random variable.

Obviously, they only give us upper bounds (and lower bounds as we will see on the next slide) for probability statements and not the exact probabilities but that is not a bad place to start.

Other forms of Chebyshev Inequality

Recall that $\mathbb{V}(X) = \sigma^2$ and $SD(X) = \sigma$. If we let $t = a\sigma$ for some $a > 0$, then Chebyshev's Inequality has a slightly different form:

$$Pr\left(|X - E(X)| \geq a\sigma\right) \leq \frac{1}{a^2}$$

This gives us an idea of the proportion of data that lie outside "a" standard deviations of the mean.

It also gives an idea of the probability that any given random variable will differ from its mean. That is, the probability that X will differ from E(X) by more than "a" standard deviations cannot exceed $1/a^2$.

Other forms of Chebyshev Inequality

If we would like to learn about the proportion of data that lie within "a" standard deviations of the mean, then we can adjust the inequality yet again to have:

$$Pr\left(|X - E(X)| < a\sigma\right) \geq 1 - \frac{1}{a^2}$$

Then the following are true:

- At least 75% of the observations must always be less than 2 standard deviations from the population mean.
- At least 89% of the observations must always be less than 3 standard deviations of the population mean.

What proportion of the observations must lie wthin 4 standard deviations of the population mean?

Law of Large Numbers

A random sample is a sample generated by making repeated draws from a box containing numbers or in the case of any random variable X, from the distribution of X. Clearly, for continuous distributions, there will be an infinite number of options/values of X.

In a random sample, draws are made with replacement, and the outcomes in a series of draws are independent. The value on the first draw does not affect the value on the second.

Recall that, by definition, the expected value is the average of the numbers in the box/distribution.

Loosely speaking, the Law of Large Numbers says that if one makes many draws from the box and averages the results, that average will converge to the expected value of the box.

Law of Large Numbers

The law is usually stated in two forms: the weak law of large numbers and the strong law of large numbers. However, because this is an introductory class, we will only focus on the general consequence of both and slightly modify the theorem from the book.

Law of Large Numbers — LLN (D.S. Theorem 6.2.4): Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the mean is μ and for which the variance is finite. Let $Y = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, denote the sample mean of the sample. Then,

$$Y = \bar{X} \to \mu$$

It turns out that this law has some interesting implications one of which is that we can think about convergence of all kinds of averages beyond the sample mean. Two examples come to mind.

Law of Large Numbers

Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the mean is μ and for which the variance is finite and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as before.

Then,

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu$$

$$\mathbb{V}[\bar{X}] = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Note that we the variance worked out because of independence!

Combining either of Markov inequality or Chebyshev inequality and the law of large numbers can be very powerful. We will revisit applications of the LLN later on but you should review the examples in section 6.2 of the textbook.

Central Limit Theorem

The Central Limit Theorem (CLT) is one of the high-water marks of mathematical thinking. It was worked upon by James Bernoulli, Abraham de Moivre, and Alan Turing. Over the centuries, the theory improved from special cases to a very general rule.

Essentially, the Central Limit Theorem allows one to describe how accurately the Law of Averages works. Most people have a good intuitive understanding of the Law of Averages, but in many cases it is important to determine whether a particular size of deviation between the sample mean and the (usually unknown) expected value is probable or improbable. That is, what is the chance that the sample average is more than some constant "d" away from the true μ?

Central Limit Theorem

Formally, the CLT says that for any "well defined function" h,

$$\sum_{i=1}^{n} h(X_i) \mathrel{\dot\sim} N\left(\mathbb{E}\left[\sum_{i=1}^{n} h(X_i)\right], \mathbb{V}\left[\sum_{i=1}^{n} h(X_i)\right]\right)$$

And in particular, for averages,

$$\bar{X} \mathrel{\dot\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \mathrel{\dot\sim} N(0,1)$$

where $\bar{X}$ is the average of n *iid* draws, μ is the expected value, and σ is the standard deviation. The approximation gets better as n gets larger.

Modifications of this formula hold for many other situations, e.g., when there is a little dependence, or when the box changes from draw to draw.

Central Limit Theorem

A version of the Central Limit holds for sums $(\sum_i^n X_i)$:

$$\sum_i^n X_i \dot{\sim} N(n\mu, n\sigma^2) \quad \text{or} \quad \frac{\sum_i^n X_i - n\mu}{\sigma\sqrt{n}} \dot{\sim} N(0,1).$$

Can you see why this is true?

This formula is useful when calculating the chance of winning a given amount of money when gambling, or getting more than a specific score on a test.

With these two central limit formulas, one can answer all sorts of practical questions.

Examples

Problem 1a: You want to estimate the average income of people in Durham. Suppose the true mean income is \$42,000 with sd of \$10,000. You draw a random sample of 100 households.

What is the probability that your sample mean is over the true value by \$500 or more?

Note that in order to solve this, we made the unreasonable assumption that we knew the true mean of the box and the sd. Later we shall relax this assumption.

Examples

Since we know μ and σ, this is relatively straightforward.

$$\begin{aligned}\mathbb{P}[\bar{X} > 42,500] &= \mathbb{P}[\bar{X} - \mu > 500] \\ &= \mathbb{P}\left[\frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} > \frac{500}{(\sigma/\sqrt{n})}\right] \\ &\doteq \mathbb{P}\left[Z > \frac{500}{(10,000/\sqrt{100})}\right] \\ &= \mathbb{P}[Z > .5].\end{aligned}$$

The CLT is used in the penultimate step.

From the standard normal table, we know this has chance 0.309. There is about a 31% chance of being too high by \$500 or more.

But this is not really the question one wants to ask in practice, nor is it the kind of information that one really has from a survey.

Examples

Problem 1b: You want to estimate the average income of people in Durham. You draw a random sample of 100 households and find the sample mean is \$42,500 and the sd of your sample is \$10,000. What is the approximate probability that you have overestimated the true average income of Durham by \$500 or more?

First, assume that the sd of your sample equals the true sd. This is an approximation, and later we shall see a way to improve this. We want to find

$$\begin{aligned}\mathbb{P}[\bar{X} - \mu > 500] &= \mathbb{P}\left[\frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} > \frac{500}{(\sigma/\sqrt{n})}\right] \\ &\doteq \mathbb{P}\left[Z > \frac{500}{(10,000/\sqrt{100})}\right] \\ &= \mathbb{P}[Z > .5].\end{aligned}$$

The probability is 0.309 that \$42,500 is \$500 too high (or more). Same as before!

Examples

Problem 2: You are playing Red and Black in roulette. (A roulette wheel has 38 pockets; 18 are red, 18 are black, and 2 are green – the house takes all the money on green).

You pick either red or black; if the ball lands in the color you pick, you win a dollar. Otherwise you lose a dollar.

Suppose you make 100 plays. What is the chance that you lose $10 or more?

Every time you make a play, there are 38 tickets, and 18 are labelled 1 and the 20 are labelled -1.

Examples

So the expected value of the box is

$$\begin{aligned}\mu &= \frac{1}{38}[1 + 1 + \cdots + 1 + (-1) + (-1) + \cdots + (-1)] \\ &= \frac{1}{38}[-2] \\ &= -1/19.\end{aligned}$$

The standard deviation of the box is

$$\begin{aligned}\sigma &= \sqrt{\left(\frac{1}{38}\sum_{i=1}^{38} X_i^2\right) - \mu^2} \\ &= \sqrt{1 - (-1/19)^2} \\ &= .998614.\end{aligned}$$

Examples

The probability of losing more than \$10 or more in 100 plays is

$$
\begin{aligned}
P[\sum_i^{100} Y_i < -10] &= P[\sum_i^{100} Y_i - n\mu < -10 - n\mu] \\
&= P[\frac{\sum_i^{100} Y_i - n\mu}{\sigma\sqrt{n}} < \frac{-10 - n\mu}{\sigma\sqrt{n}}] \\
&\doteq P[Z < \frac{-10 - n\mu}{\sigma\sqrt{n}}] \\
&= P[Z < \frac{[-10 - (100)(-1/19)]}{(10 * .998614)}] \\
&= P[Z < -0.47434].
\end{aligned}
$$

From the standard normal table, the chance of being below -0.47434 is about 0.319.

Finite Population Correction Factor

The CLT and the standard errors (deviations) of sample averages or mean are based on samples selected with replacement. However, in virtually all survey research, you sample without replacement from populations that are of a finite size, M.

If the sample size n is small compared to the population size M, one can ignore the distinction between sampling with replacement and without replacement. Then the standard deviation of an average is $\sigma/\sqrt{n}$.

If n is large relative to the population size M (say n, is more than 5% of the population size, M), then use the Finite Population Correction Factor (FPCF) to find the standard deviation of the average as

$$\frac{\sigma}{\sqrt{n}} * \sqrt{\frac{M-n}{M-1}}.$$

What will the standard deviation of sums be using the FPCF?

Recap

Today we covered:

- Linear combination of random variables
- Markov and Chebyshev Inequalities
- Law of large numbers
- The central limit theorem
- Finite population correction factor