

Week 6 Notes: Central Limit Theorem and Conditional Expectation

D. ODay

10/28/2021

Reading Notes

Law of Large Numbers

We define the sample mean as the following:

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

Strong Law of Large Numbers

The sample mean \bar{X}_n goes to the true mean μ as $n \rightarrow \infty$ with probability 1.

Weak Law of Large Numbers

Basically, the larger the sample size grows the smaller the error of the sample mean is. In other words, $\bar{X}_n - \mu$ becomes closer to 0.

Both of these laws say, in summary, the sample mean will converge on the true mean.

Central Limit Theorem

Sample mean has the following:

$$\bar{X}_n \xrightarrow{D} N\left(\mu, \frac{\sigma^2}{n}\right)$$

The Central Limit Theorem states, **for large n , the distribution of the sample mean approaches a Normal Distribution** (As shown in the formula above). \rightarrow^D means ‘converges in distribution’. The intuition is as n becomes large, the mean will be unaffected and the variance will shrink close to 0; therefore, the distribution will essentially be a constant.

Conditional Expectation

Knowing the Law of total Probability, we can get the **Law of Total Expectation** as the following:

$$E(X) = E(X|A)P(A) + E(X|A^c)P(A^c)$$

If we see $E(X|X=x)$ then we know it is asking for the expectation of the random variable X , given that we know the random variable X crystallizes to the value x . Same idea applies to $E(h(X)|X)$. Which again, this equals $h(X)$ (let $h(X)$ be a function). All we do is plug in X into $h(X)$.

Conditional Expectation for Different Random Variables

we have $E(Y|X)$ which means that we want the best prediction for Y given that we know the random variable X . In other words, does Y have a different distribution if we know the value of X ?

Example

Let's say we have $Y \sim \text{Bin}(X, 0.5)$ and $X \sim \text{DUnif}(1, 10)$. DUnif is discrete uniform. Just means we can only have integer values, not continuous. Can think of the overall structure as flipping a coin and counting heads. For $E(Y|X)$, we know that Y is Binomial; therefore, it's expectation is np . For our case, we know $p = \frac{1}{2}$. Since n is X in this distribution, we know the number of trials is random; however, since we're conditioning on X , we can say we *know* the value. Thus, since X is known, we can say $E(Y|X) = \frac{X}{2}$. **If X and Y are independent**, then then $E(X|Y) = E(Y)$. That's because knowing the random variable X does not give any information that is helpful in predicting Y .

Lecture Notes

Introduction to the Central Limit Theorem

The sample mean is also known as the *estimator* of μ . It has a distribution of its own, which is referred to as the **sampling distribution of the sample mean**. The sampling distribution depends on the following:

- The sample size n
- The population distribution of the X_i
- The method of sampling

For the Central Limit Theorem, the typical rule of thumb is $n \geq 30$. Think of using a t-test vs a z-test.

Proposition: If X_1, X_2, \dots, X_n are iid with $X_i \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$

Proposition: If X_1, X_2, \dots, X_n are independent with $X_i \sim N(\mu_i, \sigma_i^2)$ then $\sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

Example 1

Suppose you have 3 errands to do in three different stores. Let T_i be the time to make the i^{th} purchase for $i = 1, 2, 3$. Let T_4 be the total walking time between stores. Suppose $T_1 \sim N(15, 16)$, $T_2 \sim N(5, 1)$, $T_3 \sim N(8, 4)$ and $T_4 \sim N(12, 9)$. Assume T_1, T_2, T_3, T_4 are independent. If you leave at 10 a.m. and you want to tell a colleague, "I'll be back by time t ", what should t be so that you will return by that time with probability 0.99?

Let $T_0 = T_1 + T_2 + T_3 + T_4 = \text{total time}$

$E(T_0) = 15 + 5 + 8 + 12 = 40$ minutes

$V(T_0) = 16 + 1 + 4 + 9 = 30$ minutes.

Thus, $T_0 \sim N(40, 30)$. We want to find t so that $P(T_0 \leq t) = 0.99$. Now we just solve the problem using the new normal distribution we have (we know how to solve Normal Random from before.):

$$\begin{aligned} P\left(\frac{T_0 - 40}{\sqrt{30}} \leq \frac{t - 40}{\sqrt{30}}\right) &= .99 \\ \Rightarrow P\left(Z \leq \frac{t - 40}{\sqrt{30}}\right) &= \Phi\left(\frac{t - 40}{\sqrt{30}}\right) = 0.99 \\ \Rightarrow \text{Lookup : } \Phi(2.33) &= .99 \Rightarrow \frac{t - 40}{\sqrt{30}} = 2.33 \Rightarrow t = 52.76 \text{ minutes} \end{aligned}$$

Formula Summary

Sample Mean

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$
$$X_n \rightarrow^D \bar{N}(\mu, \frac{\sigma^2}{n})$$

Law of Total Expectation

$$E(X) = E(X|A)P(A) + E(X|A^c)P(A^c)$$