Week_6

Tuesday, February 11, 2020 3:05 PM

STA255 Week-6

Shahriar Shams

10/02/2020

Review of Week-5

- Joint distribution for two discrete variables
  - Joint probability mass function(pmf)
  - Marginal probability mass functions
- Joint distribution for two continuous variables
  - Joint probability density function (pdf)
  - Marginal probability mass functions
- Independence of two/or more variables (discrete or continuous)
  - $P[X = x, Y = y] = P[X = x] * P[Y = y]$ for every pair of (x, y)
  - $f(x, y) = f_X(x) * f_Y(y)$
- Expectation, covariance and correlation
- Conditional distributions= joint dist/marginal dist

Learning goals

- A brief introduction to Population, Sample, Parameter and Statistic
- Estimator vs Estimate
- Sampling distribution
- Mean and Variance of $\bar{X}$ under any distribution
- Sampling distribution of sample mean($\bar{X}$) under Normal distribution
- Central Limit Theorem (CLT)
- Law of large number (sample mean approaches true mean)

Population, Sample, Parameter and Statistic

- We will define all these terms using a simple example.
- Suppose we want to know the **average height** of **ALL the students of UofT** (let's call it μ).
- Due to resource constraint, a sensible approach will be to take a sample of 100 students (or some other number) and measure their average height (let's call it $\bar{X}$ which represents the sample mean).

- "Population" is defined as the collection of all subjects/individuals that we want to make a comment on.
  - in our example, **Population:** All the students of UofT (some 60,000 of us)
- "Parameter" is defined as any summary of the population. It's an unknown number believed to be a constant.
  - in our example, **Parameter:** The average height of the population, which we represent using μ.
  - Instead of population mean, we can be interested in population variance, median etc. Each of these summaries can be treated as a parameter.
- "Sample" is defined as any subset of the population that we actually get to observe.
  - in our example, **Sample:** 100 randomly selected students.
- "Statistic" is defined as any summary of the sample.
  - in our example, **Statistic:** the sample mean.
  - Instead of calculating the sample mean, we can calculate many other summaries such as sample variance, minimum value of the sample, maximum value of the sample. Each of these are called a statistic.

Estimator vs Estimate

- When a "statistic" is used to make an educated guess about the unknown "parameter", we call this "statistic" either an estimate or an estimator based on whether we have already observed the sample observations or not.
- Say sample mean is the statistic that we are using to guess the value of μ.
- If we have already observed our sample of size n, then we have $(x_1, x_2, ..., x_n)$
- We will be able to get a numeric value of the sample mean(statistic).
  - this numeric value is called an estimate. we will write it as $\bar{x}$
  - $\bar{x}$ is an **estimate** of μ
- If we haven't observed our samples yet, the observations are random variables and hence written as $(X_1, X_2, ...X_n)$.
- In theory the sample mean can vary from one set of sample of size n to another set of sample of size n.
- Hence, the sample mean is considered as a random variable and written as $\bar{X}$
  - $\bar{X}$ is an **estimator** of μ
- In summary, *Estimate* ($\bar{x}$) is a fixed numeric number, *Estimator* ($\bar{X}$) is a random variable.

Pop^n: All UofT Stu

param. ave. height

Sample: STA255 Stu.

Statistic: sample mean. = 167.5cm

$x_1, x_2 \ldots x_n$

$\boxed{\text{Estimate}}$

Estimator.

$\bar{X}$

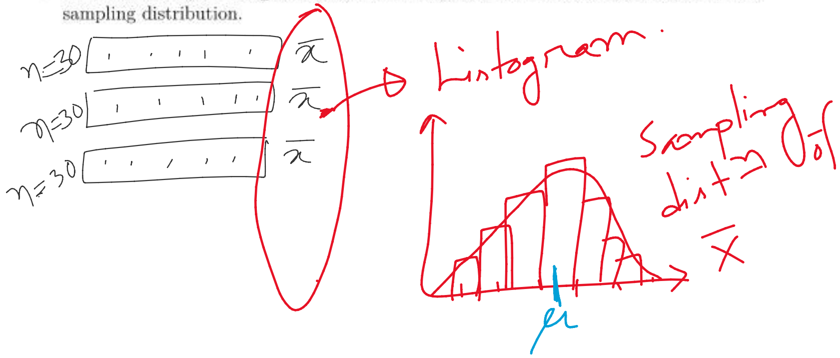
data1 $\bar{x} = 167.5$

data2 $\bar{x} = 164.2$

data3 $\bar{x} = 162.7$

Sampling distribution of an estimator $(\bar{X})$

- Sampling distribution arises from the idea of repeated sampling.
- Say we have a random sample of size n based on which we calculate an estimate.
- In theory, we can take another sample of size n and calculate another estimate.
- If we keep repeating this task over and over again we will end up with many different estimates.
- In naive words, a histogram constructed based on all these estimates is called the sampling distribution.



In this course, we will specifically learn the sampling distribution of sample mean $(\bar{X})$

Mean and Variance of $\bar{X}$

- Let $X_1, X_2, ..., X_n$ be a random sample (observations are *independent* of each other) from any distribution with mean$=\mu$ and variance $= \sigma^2$. ie. $E[X_i] = \mu$ and $V[X_i] = \sigma^2$
- Sample mean, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n$
- Recall from last week: $E[aX + bY] = aE[X] + bE[Y]$
- Mean of $\bar{X}$,

$$\begin{aligned} E[\bar{X}] &= E[\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n] \\ &= \frac{1}{n}E[X_1] + \frac{1}{n}E[X_2] + ... + \frac{1}{n}E[X_n] \\ &= \frac{1}{n}\mu + \frac{1}{n}\mu + ... + \frac{1}{n}\mu \\ &= n * \frac{1}{n}\mu = \mu \end{aligned}$$

- Recall:
  - $V[aX+b] = a^2V[X]$
  - If X and Y are independent, $V[X+Y] = V[X] + V[Y]$
- Variance of sample mean,

$$
\begin{aligned}
V[\bar{X}] &= V[\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n] \\
&= \frac{1}{n^2}V[X_1] + \frac{1}{n^2}V[X_2] + ... + \frac{1}{n^2}V[X_n] \\
&= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + ... + \frac{1}{n^2}\sigma^2 \\
&= n * \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

- In summary,
  - $\bar{X}$ is a linear combination of $X_1, X_2, ...X_n$
  - $E[\bar{X}] = \mu$ and $V[\bar{X}] = \frac{\sigma^2}{n}$ irrespective of the distribution of X_i's.

Sampling distribution of sample mean($\bar{X}$) under Normal distribution

- We will calculate the distribution of $\bar{X}$ using a general property of Normal distribution.
- Claim:
  - $X_i \sim N(\mu_i, \sigma_i^2)$ where $i = 1, 2, ...n$
  - X_i's are independent
  - For constants $a_1, a_2, ..., a_n$, let Y be a linear combination of all the X_i's with

$$Y = a_1X_1 + a_2X_2 + ...a_nX_n = \sum_{i=1}^{n} a_iX_i$$

  - Then,

$$Y \sim N\left(\sum_{i=1}^{n} a_i\mu_i \,,\, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

- In summary, any linear combination of some independent Normal random variables also follows a Normal distribution.
- In previous section, we showed that $\bar{X}$ is a linear combination of $X_1, X_2, ..., X_n$
- Combining the distribution with the mean and variance calculated in the previous section, we can write

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

$X_1 \sim N(\mu_1, \sigma_1^2)$ $\quad$ X_1 is indep of X_2

$X_2 \sim N(\mu_2, \sigma_2^2)$

$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$Y = a_1X_1 + a_2X_2 \sim N(\quad , \quad)$

$M_Y(t) = E[e^{Yt}] = E[e^{a_1X_1t + a_2X_2t}]$

$= E[e^{a_1X_1t}]\, E[e^{a_2X_2t}]$

$= M_{X_1}(ta_1) * M_{X_2}(ta_2)$

$= e^{\mu_1 a_1 t + \frac{\sigma_1^2 t^2 a_1^2}{2}} \cdot e^{\mu_2 a_2 t + \frac{\sigma_2^2 t^2 a_2^2}{2}}$

$= e^{\boxed{(a_1\mu_1 + a_2\mu_2)}t + \frac{\boxed{(a_1^2\sigma_1^2 + a_2^2\sigma_2^2)}\, t^2}{2}}$

this is a MGF of a Normal

$\therefore Y \sim N(\quad , \quad)$

Central Limit Theorem: dist of $\bar{X}$ when samples are from Non-normal distributions

- We will start with the example of heights.
- Suppose our population is all the UofT students.
- Let, X represent the height of any student.
- Let us also assume that $E[X_i] = \mu$ and $V[X_i] = \sigma^2$.

- Say we take a sample of 30 students and calculate the sample mean ($\bar{x}$). This mean ($\bar{x}$) is an estimate of μ.
- Theoretically speaking, if we continue taking 30 students randomly from our population and continue calculating sample mean we will end up with a list of means.
- It is possible to construct a density histogram using all these mean values.
- This is the density of the sample mean $\bar{X}$ for $n = 30$.

- We can repeat the task for n=60 now, we will end up with another density.
- Let's keep repeating the task by increasing n every time.

- The central limit theorem says, as $n \to \infty$, the distribution of sample mean ($\bar{X}$) will converge to a Normal distribution.

$$\bar{X} \xrightarrow{D} N(\, , \,)$$

- $\xrightarrow{D}$ stands for a distribution converging to another distribution.
- In naive words, as we increase n, the density of $\bar{X}$ starts to look like a Normal density **irrespective of what the actual distribution of X_i is.**

- Combining this with the mean and variance of $\bar{X}$, we can write

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

- Another way of writing this theorem,

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

- Exercise 19 and 21

$\boxed{19}$

$X \sim N(\mu = 2.65, \sigma = 0.85)$ $\frac{\sigma}{\sqrt{n}}$

$n = 25$ $\bar{X} \sim N\left(2.65, \frac{0.85}{\sqrt{25}}\right)$ SD

(a) $P[\bar{X} \leq 3]$

$= P\left[Z \leq \frac{3 - 2.65}{0.85/\sqrt{25}}\right]$

(b) $P\left[Z \leq \frac{3 - 2.65}{0.85/\sqrt{n}}\right] \geq 0.99$

$\boxed{21}$

$X = \#$ of tickets in one day

$X \sim Poin(\lambda = 50)$ $\mu = 50$, $\sigma^2 = 50$

$\bar{X} \longrightarrow N\left(50, \frac{50}{5}\right)$ assume 5 is big

$S = Sum$

$P[225 < S < 275]$

$P\left[\frac{225}{5} < \bar{X} < \frac{275}{5}\right]$

$= P[45 < \bar{X} < 55]$ z-transform

Law of Large Number (LLN)

$n=50$ $\bar{x}$
$n=100$ $\bar{x}$
$n=200$ $\bar{x}$

- First an intuitive idea:
- Suppose we have collected three different set of samples.
  - In one set, we have 50 randomly selected students
  - In another one, we have 100 randomly selected students
  - In the last one, we have 200 randomly selected students
- From each of these three different set of samples, we calculate average height.
- Let's call them $\bar{x}_{50}$, $\bar{x}_{100}$, $\bar{x}_{200}$
- Which one do you expect to be more closer to the value of the true mean (μ)?
- Intuitively, we will think that $\bar{x}_{200}$ will be more closer to the true mean.

- Let's think it in a different way.
- By now, we know $E[\bar{X}] = \mu$ and $V[\bar{X}] = \frac{\sigma^2}{n}$
- As n increases $E[\bar{X}]$ will still be μ but the $V[\bar{X}]$ approaches zero.
  - $V[\bar{X}] = E[(\bar{X} - \mu)^2] \to 0$
  - The mean of squared difference between $\bar{X}$ and μ goes to zero.
  - we say, $\bar{X}$ converges to μ in *mean square*.

- There are several different versions of Law of Large Numbers that are available in statistics literature. Here are two of them:
- **LLN:**
  - If $X_1, X_2, ..., X_n$ is random sample from a distribution with mean μ and variance σ^2. Then $\bar{X}$ converges to μ
  - In mean square: $E[(\bar{X} - \mu)^2] \to 0$ as $n \to \infty$
  - In probability: $P[|\bar{X} - \mu| \geq \epsilon] \to 0$ as $n \to \infty$ for any $\epsilon > 0$

- Example 6.11 (page 304)

$X_i = 1$, if the ith toss is H
0 , " " " " T

$E[X_i] = 1*0.5 + 0*0.5$
$= 0.5$

$\boxed{\frac{X_1 + X_2 + \cdots + X_{10000}}{10000}} \longrightarrow E[X_i]$
$\longrightarrow 0.5$

$\Rightarrow \bar{X} \longrightarrow 0.5$

Chapter 6.4 not needed

Homework

Chapter 6.2

14-18

Chapter 6.3

27, 28, 42, 44

Supplementary

69, 72, 74