Statistics 2 : Sampling and Covariance

Anand Systla

Masters in Financial Engineering Bootcamp
UCLA Anderson

August 16, 2022

Today's Agenda

- Discuss the previous class take home questions
- Clarifications
- Moments of Normal Distribution
- Sample vs Population
- Covariance and Correlation
- Correlation and Causation

Clarifications

- Poisson intensity λ vs Exponential intensity $\beta = \frac{1}{\lambda}$. Units become important!
  - If $\lambda = 3$ (say 3 buses an hour), if we use $X \sim Exp(\beta = \frac{1}{\lambda} = \frac{1}{3})$

$$P(0 \leq X \leq 1) = \int_0^1 f(x) = \int_0^1 \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^1 = 1 - e^{-3} \approx 0.95 = F(1) - F(0)$$

  - If $\lambda = 3/60 = 1/20$ (say 1 bus every 20 min), if we use $X \sim Exp(\beta = \frac{1}{\lambda} = 20)$

$$P(0 \leq X \leq 1) = \int_0^1 f(x) = \int_0^1 \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^1 = 1 - e^{-1/20} \approx 0.05 = F(1) - F(0)$$

  - Probability of seeing a bus in 30 minutes

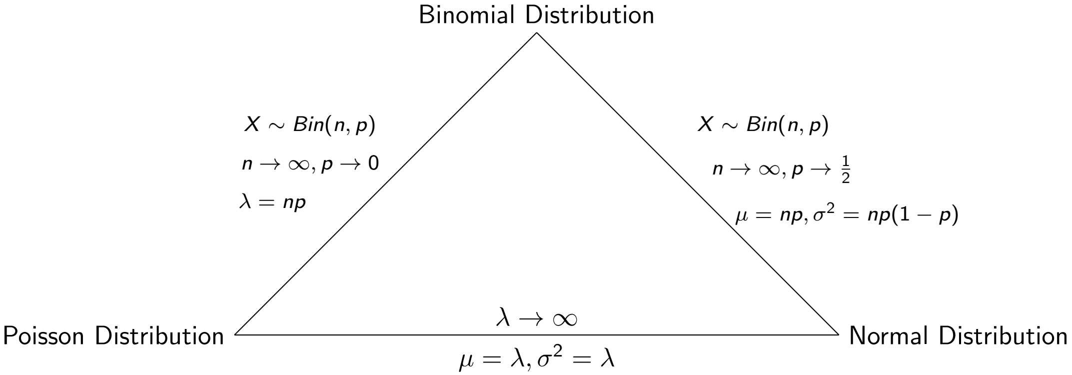
$$Y \sim Pois(\lambda_{60} = 3 \iff \lambda_{30} = 3/2) \implies P(Y = 1) = \frac{e^{-3/2}(3/2)^1}{1!} = 0.33$$

$$X \sim Exp(\beta = \frac{1}{\lambda_{60}} = \frac{1}{3}) \implies P(X \leq \frac{1}{2}) = \int_0^{1/2} f(x)dx = F(1/2) - F(0) = 1 - e^{-3\cdot\frac{1}{2}} = 0.33$$

- Modulus function $|x|$ properties
  - $|x|$ is continuous everywhere but $|x|$ is not differentiable at 0. Easy to see this for the x^2 quadratic function $\lim_{x \to 0^+} 2x = \lim_{x \to 0^-} 2x = 0$

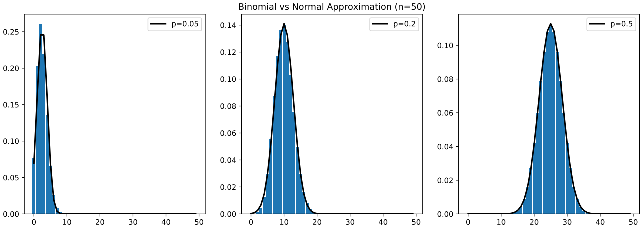
$$\frac{d}{dx}|x| = \begin{cases} -1, & x < 0 \\ +1, & x \geq 0 \end{cases}, \qquad \lim_{x \to 0^+} \frac{d}{dx}|x| = 1 \quad \text{and} \quad \lim_{x \to 0^-} \frac{d}{dx}|x| = -1$$

Relationship Between Distributions

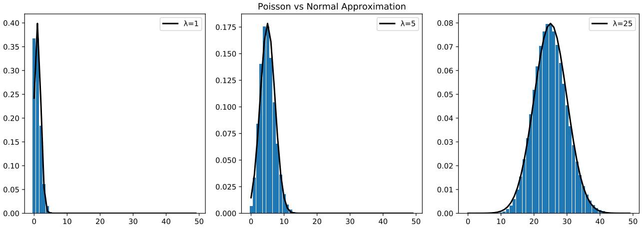


- A rough guideline to ensure the Normal approximation of the Binomial is reasonable
  - $np \geq 10$
  - $n(1-p) \geq 10$

Relationship Between Distributions : Examples



(A)



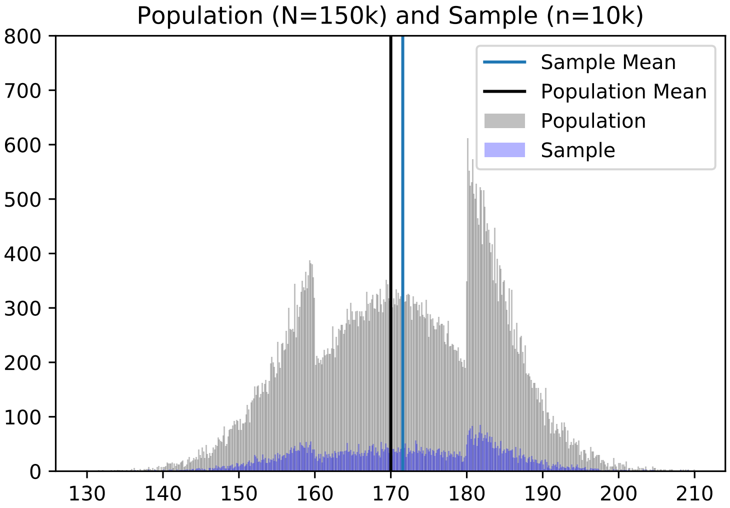
Sample vs Population

- Population includes all the data of a specified group. Sample is a subset of the population

	Population	Sampling Methodology
Height of people in US	330 Mn	Selecting people from each state
Height of people in UCLA	50,000	Asking MFE/MBA/Professors
Weight of people in Japan	125 Mn	Setting up volunteer booths in Tokyo

- Gold standard is having a random sample that is representative of the population. Generally samples suffer from sampling/selection bias. In our case - (1) non-responsiveness, (2) under-coverage, (3) location of advertising
- Population is summarized by parameters. A sample is summarized by sample statistics. As the sample size approaches the population size, the sample statistic is going to approach population parameter

Population vs Sample Data



Population Parameters and Sample Statistics

- Moments are robust ways of summarizing a RV. Common moments of interest are - mean, variance, skewness, kurtosis, quantiles, etc.

Moment	Population Parameter	Sample Statistic
Mean	$\mu = E[X] = \frac{\sum x_i}{N} = \int xf(x)dx$	$\bar{X} = m = \frac{\sum x_i}{n}$
Variance	$\sigma^2 = E[X^2] - E[X]^2 = \sum_i^N \frac{(x_i - \mu)^2}{N}$	$s^2 = \sum_i^n \frac{(x_i - m)^2}{n-1}$

Population Parameter vs Sample Statistic : Example

- Sample data is different from sample statistic. Both can have their own distributions. An example is given below where population mean $\mu = \frac{1+2+3}{3} = 2$. Generating a sample of 2 draws with replacement and ordering matters

Samples	Mean ($\bar{X}$)
(1,1)	1
(1,2)	1.5
(1,3)	2
(2,1)	1.5
(2,2)	2
(2,3)	2.5
(3,1)	2
(3,2)	2.5
(3,3)	3

- Does the distribution look familiar? More on this next class!

Sample Variance

- We have seen that the sample variance is given by $s^2 = \sum_i^n \frac{(x_i - \bar{X})^2}{n-1}$

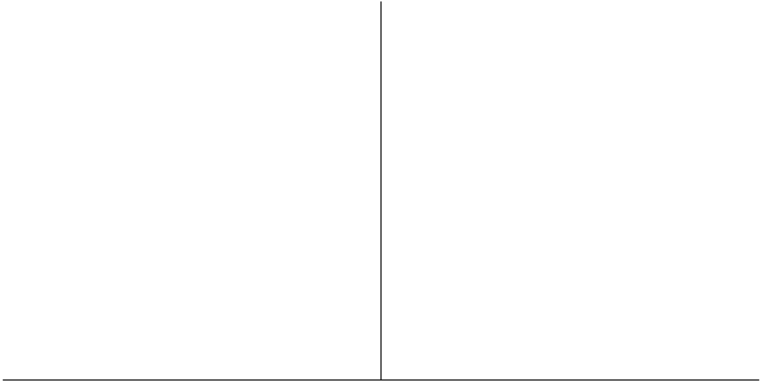
- **Intuition 1 :** We use one data point in computing the mean. If we compute the sample mean $\bar{X}$ from n-data points, we no longer have n independent data points. We can back out the nth number from $n-1$ data points and the mean.

- **Intuition 2:** The variance computes the squared deviation around the mean. s^2 computes variance centered around $\bar{X}$ and σ^2 computes the variance around μ, which is different from $\bar{X}$. Any deviation from the sample mean $\bar{X}$ will only increase the varaince. So we bump up the sample variance by dividing by a smaller number $n-1$. Dividing by $n-1$ increases $s^2(\bar{X})$, closer towards $s^2(\mu)$.

- Say you are given $X = \{20\}$, what is $\bar{X} =$ and what is $s^2(\bar{X}) =$

Skewness and Kurtosis

- Skewness: Which direction is the data (tail) drawn out towards? $\frac{1}{N}\frac{\sum(x-\mu)^3}{\sigma^3}$
- Kurtosis: How much weight do the tails have? $\frac{1}{N}\frac{\sum(x-\mu)^4}{\sigma^4}$
  - "Peakedness" has nothing to do with the kurtosis. Two distributions can have the same mean and standard deviation, but can have different weights placed on their tails
  - Kurtosis > 3 is leptokurtic (< 3 is platykurtic)
  - Kurt($N(0,1)$) $= 3$, so people are often interested in excess kurtosis ($=$ kurtosis-3)



Covariance

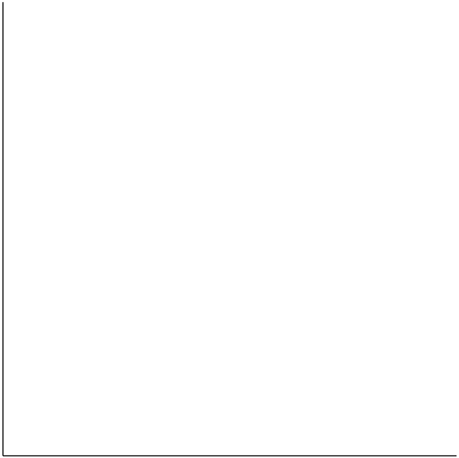
- Lets us study joint variation of two random variables and how they co-move
- In our sample below $\bar{X} = 100$ and $\bar{Y} = 10$. The covariance asks are $X - \bar{X}$ and $Y - \bar{Y}$ above and below zero together?

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
100	10	0	0	0
102	9	2	-1	-2
98	11	-2	1	-2
110	14	10	4	40
90	6	-10	-4	40

- Properties of covariance terms
  - $cov(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] = E[(X - \bar{X})Y] - E[(X - \bar{X})\bar{Y}] = E[Y(X - \bar{X})] =$
  - $cov(X, Y) = E[XY - \bar{X}Y - X\bar{Y} - \bar{X}\bar{Y}] = E[XY] - E[X]E[Y]$
  - $cov(X, c) =$
  - $cov(aX, X) =$
  - $cov(X, Y + c) =$
  - $cov(X, Y + Z) =$

Covariance Visualization

- $cov(X, Y) \lesseqgtr 0, \quad cov(X, X) > 0,$ outliers?



- Covariance does not tell us how far the points are from the dotted line. It also does not tell us how steep or flat our line is

Correlation

- Although the covariance gives us a measure of co-movement of two random variables, it scales with any constant multiplying the random variable $cov(aX, Y) = acov(X, Y)$
- Correlation does not depend on the scale of the data. It is a measure of linear dependence and $\rho_{XY} \in [-1, 1]$. Why?

$$\rho_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_X \sigma_Y} = E\left(\frac{X - \bar{X}}{\sigma_X} \frac{Y - \bar{Y}}{\sigma_Y}\right)$$

- Properties of the correlation
  - Correlation is a dimensionless quantity
  - $corr(aX, Y) = corr(X, Y)$
  - $corr(X, Y + c) = corr(X, Y)$
  - If X and Y are independent $\rho_{XY} = 0$, but $\rho_{XY} = 0$ does not imply independence (non-linearity)
  - If $X \sim N(0, 1)$, then $cov(X, X^2) =$

Correlation Visualization

- Correlation can be 1 irrespective of how spread out of narrow the data is
- How does our confidence on the correlation measure change with number of data points?