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!
  - (1) If $\lambda = 3$ (say 3 buses an hour), if we use $X \sim Exp(\beta = \frac{1}{\lambda} = \frac{1}{3})$

$$\underbrace{P(0 \le X \le 1)} = \int_{0 \text{ hours}}^{1 \text{ hour}} f(x) = \int_0^1 \underbrace{\lambda e^{-\lambda x}} dx = \left[-e^{-\overset{=3}{\lambda} x}\right]_0^1 = \underbrace{1 - e^{-3}} \approx \underbrace{0.95} = \underbrace{F(1) - F(0)}$$

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

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

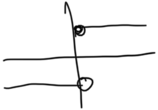
  - (3) Probability of seeing a bus in $\underbrace{\text{30 minutes}}$, $\lambda = 3$

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

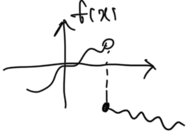
$$X \sim Exp(\beta = \frac{1}{\lambda_{60}} = \frac{1}{3}) \implies \underbrace{P(X \le \frac{1}{2})} = \int_0^{1/2} \underbrace{f(x)} dx = \underbrace{F(1/2) - \overset{0}{\overset{\|}{F(0)}}} = 1 - e^{-3 \cdot \frac{1}{2}} = \underline{0.33}$$

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

$f'(x)$

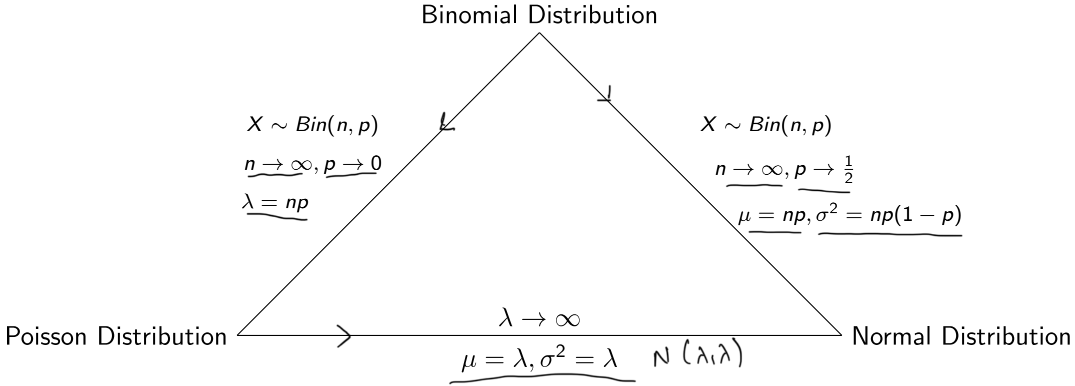


$$\frac{d}{dx}|x| = \begin{cases} -1, & x < 0 \\ +1, & x \ge 0 \end{cases}, \qquad \lim_{x \to 0^+} \frac{d}{dx}|x| = 1 \quad \text{and} \quad \lim_{x \to 0^-} \frac{d}{dx}|x| = -1$$



$|x|$

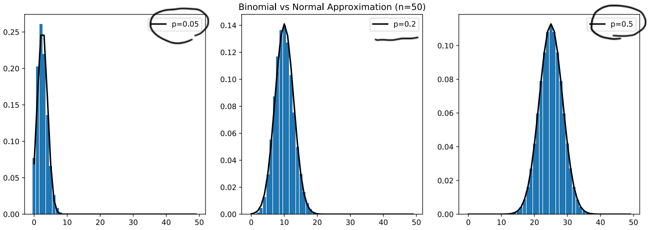
Relationship Between Distributions



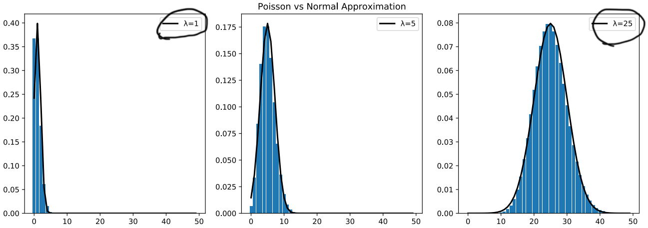
- A rough guideline to ensure the Normal approximation of the Binomial is reasonable
  - $np \geq 10$
  - $n(1-p) \geq 10$

$np \geq 10$, $p \sim 0.1$, $n \geq 100$

Relationship Between Distributions : Examples



(A)



Sample vs Population

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

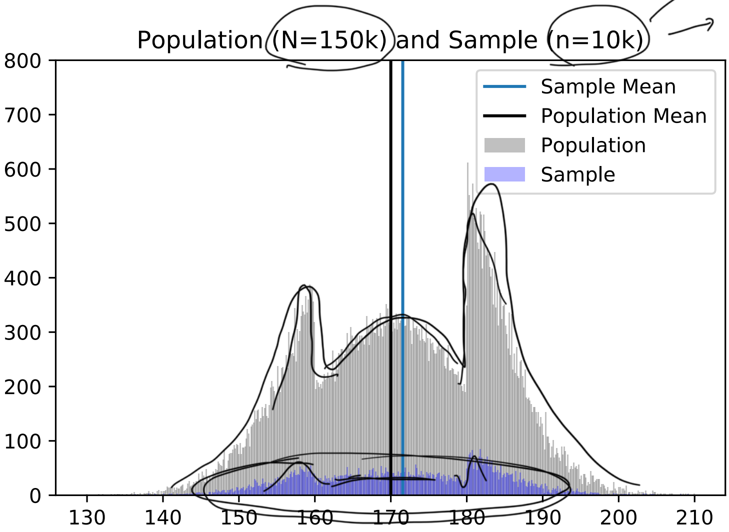
	Population	Sampling Methodology
Height of people in US	330 Mn	Selecting people from each state
Height of people in UCLA	50,000	Asking MFE/MBA/Professors
Weight of people in Japan	125 Mn	Setting up volunteer "booths" in Tokyo

- 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

(N)

$n \longrightarrow N$

Population vs Sample Data



9k : m_1

9k : m_2

⋮

m_k

Population Parameters and Sample Statistics

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

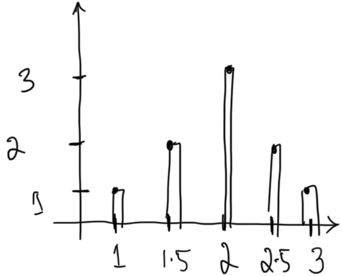
Moment	Population Parameter	Sample Statistic
Mean	$\mu = E[X] = \frac{\sum x_i}{N} = \int xf(x)dx$	$\bar{X} = m = \frac{\sum x_i}{n}$
Variance	$\sigma^2 = E[X^2] - E[X]^2 = \sum_i^N \frac{(x_i - \mu)^2}{N}$	$s^2 = \sum_i^n \frac{(x_i - m)^2}{n-1}$

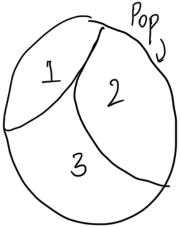
"Unbiased estimator"

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





(1,2)

(2,1)

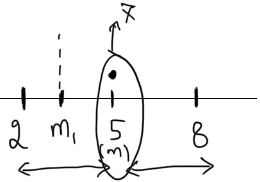
- Does the distribution look familiar? More on this next class!

Sample Variance

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

$(5-2)^2 + (5-8)^2 = 3^2 + 3^2 = 18$

$(4-2)^2 + (4-8)^2 = 4 + 16 = 20$



$s^2(\bar{X})$, $m = 5$

$m_1 = 4$

- **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 $\bar{X} =$ 20 and what is $s^2(\bar{X}) =$ NaN

$$S^2(\bar{X}) = \frac{\sum (x_i - \bar{X})^2}{n-1} \leq \sum (x_i - \mu)^2$$

$$n \quad \{1, 2, 3\} \longrightarrow \{\underbrace{1, 2}, 3 ; \underbrace{\bar{X}}\}$$

$$\bar{X} = \frac{6}{3} = 2$$

$$\begin{array}{ccc} 1, & 2, & 2 \\ x_1 & x_2 & \bar{X} \end{array}$$

$$\bar{X} = \frac{x_1 + x_2 + x_3}{3}.$$

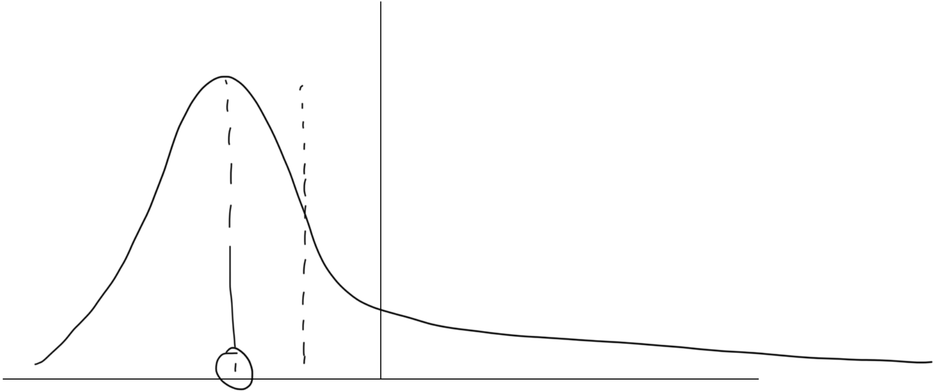
$$\Rightarrow x_3 = 3\bar{X} - x_1 - x_2 = 6 - 1 - 2 = 3$$

Skewness and Kurtosis

>0

<0

- Skewness: Which direction is the data (tail) drawn out towards? $\underbrace{\frac{1}{N}\frac{\sum(x-\mu)^3}{\sigma^3}}$
- 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 ($=$ $\underbrace{\text{kurtosis}-3}$)



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

$$E[(x-\bar{x})(y-\bar{y})] = \frac{80}{5}$$

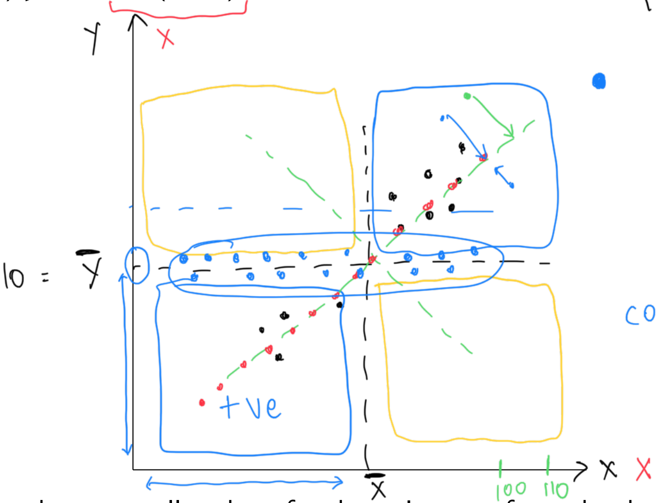
- Properties of covariance ~~terms~~
  - $cov(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] = E[(X - \bar{X})Y] - E[(X - \bar{X})\bar{Y}] = E[Y(X - \bar{X})]$
  - $cov(X, Y) = E[XY - \bar{X}Y - X\bar{Y} - \bar{X}\bar{Y}] = E[XY] - E[X]E[Y]$
  - $cov(X, c) = 0$
  - $cov(aX, X) = a \cdot cov(X, X) = a \cdot var(X)$
  - $cov(X, Y + c) = cov(X, Y)$
  - $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$

Covariance Visualization

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

$X = \{100, 102, 98, 110, 90\}$

$Y = \{10, 9, 11, 14, 6\}$



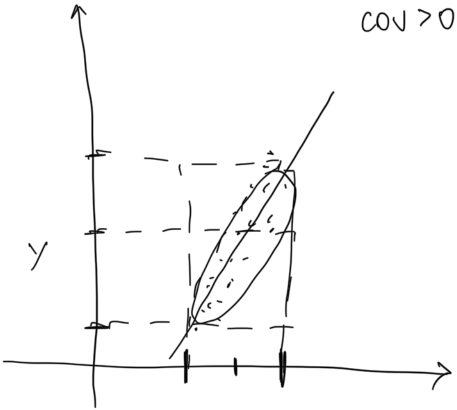
- cov > 0
- cov < 0

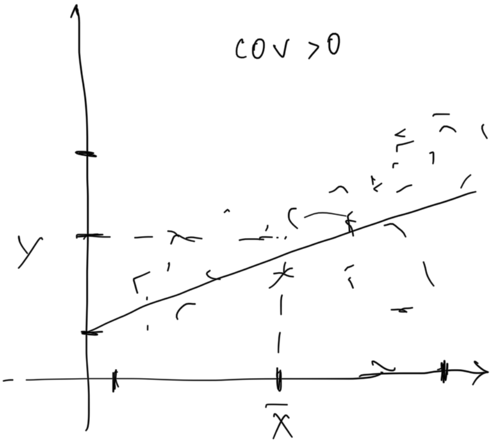
$y - \bar{y} = 0$

$cov(X, c) = 0$

$\begin{bmatrix} c \\ c \\ c \\ \vdots \\ 10000 \end{bmatrix}$

- 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


COV > 0
y


COV > 0
y
$\bar{x}$

Correlation

$Z \sim N(0,1)$: $X = \mu + \sigma Z \Rightarrow Z = \frac{X-\mu}{\sigma}$

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

$$\rho_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\overbrace{E[(X - \bar{X})(Y - \bar{Y})]}^{\text{Unit1} \times \text{Unit2}}}{\underbrace{\sigma_X}_{\cancel{\text{Unit1}}}\underbrace{\sigma_Y}_{\cancel{\text{Unit2}}}} = E\left(\overbrace{\underbrace{\frac{X - \bar{X}}{\sigma_X}}_{\tilde{X}}\underbrace{\frac{Y - \bar{Y}}{\sigma_Y}}_{\tilde{Y}}}^{H}\right)$$

- Properties of the correlation
  - Correlation is a dimensionless quantity
  - $corr(aX, Y) = corr(X, Y)$
  - $corr(X, Y + c) = corr(X, Y)$

$cov(X, Y) = E[XY] - E[X]E[Y]$

  - If X and Y are independent $\rho_{XY} = 0$, but $\rho_{XY} = 0$ does not imply independence (non-linearity)
  - If $\underbrace{X \sim N(0, 1)}_{\text{"dependence"}}$, then $cov(X, X^2) = E[(X-\bar{X})(X^2-\bar{X^2})] = E[X \cdot X^2] - E[X]\cdot E[X^2]$

$= \cancel{E[X^3]}^{0} - E[\cancel{X}^{0}]\cdot E[X^2] = 0$

$$\begin{cases} \mathrm{corr}(X, Y) = 0.75 \\ \mathrm{corr}(Y, Z) = 0.25 \end{cases}$$

$$\mathrm{corr}(X, Z) = \text{? 'x'}$$

$$\begin{array}{c|ccc} & X & Y & Z \\ \hline X & 1 & \rho_{xy} & x \\ Y & \rho_{xy} & 1 & \rho_{yz} \\ Z & x & \rho_{yz} & 1 \end{array}$$

$$\det(\mathrm{corr}) \geq 0$$

$$1(1 - \rho_{yz}^2) - \rho_{xy}(\rho_{xy} - x \cdot \rho_{yz}) + x(\underbrace{\rho_{xy}} \cdot \underbrace{\rho_{yz}} - x) \geq 0$$

$$ax^2 + bx + c \geq 0 \quad \Big| \quad x \in [\rho^L, \rho^H]$$

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?