Statistics 3 : Central Limit Theorem and Law of Large Numbers

Anand Systla

Masters in Financial Engineering Bootcamp
UCLA Anderson

August 18, 2022

Today's Agenda

- Correlation and Causation
- Moment Generating Function
- Central Limit Theorem
- Law of Large Numbers
- t-Test
- ANOVA

Correlation and Causation

- Studying causal inference is the fundamental problem of econometric inference. Causation goes a step further than correlation and asks how does change in X **cause** a change in Y?

$$Y \quad \sim \quad X$$

- Establishing causality
  - Experiments
  - Cause occurs before effect
  - If X occurs, then Y also occurs

Moment Generating Function (MGF)

- MGF lets us compute the moments quickly. Moments allow us to summarize the distribution of the RV. We are usually interested in the first few moments of any RV. Moments can either be central of non-central
- Non-central moments are $E[X], E[X^2], E[X^3]$ and so on
- The k-th central moments is $E[(X - E[X])^k]$. How is a RV distributed around a particular value (like the mean)

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tX}$$

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + ..$$

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + ..$$

$$\frac{d}{dt}E[e^{tX}]|_{t=0} = 0 + E[X] + tE[X^2] + .. = E[X] \implies E[X^n] = \frac{d^n}{dt^n}M_X(t)|_{t=0}$$

MGF Example

- Say $X \sim Bin(n, p)$, the MGF is given by

$$M_X(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx} \ {}^nC_x p^x (1-p)^{n-x} = \sum_{x=0}^{n} {}^nC_x (e^t \cdot p)^x (1-p)^{n-x} = (e^t \cdot p + (1-p))^n$$

$$E[X] = \frac{dM_t(X)}{dt} = \frac{d}{dt}(e^t \cdot p + (1-p))^n = n \cdot (e^t \cdot p + (1-p))^{n-1} \cdot p \cdot e^t|_{t=0} = np$$

$$E[X^2] = \frac{d^2}{dt^2} M_X(t)|_{t=0} =$$

- Properties of the MGF
  - $M_X(t) = M_Y(t)$ then $F_X(x) = F_Y(x)$
  - If $Y = X_1 + X_2 + .. + X_n$, sum of independent RVs, then

$$M_Y(t) = E[e^{tY}] = E[e^{t(X_1+X_2+..+X_n)}] = M_{X_1}(t) M_{X_2}(t) .. M_{X_n}(t)$$

LOLN and CLT

- If $X_1, X_2, ...$ are i.i.d and drawn from some population P with a defined mean (μ) and variance (σ^2), then

$$\text{LOLN: } \bar{X}_n \to^p \mu \quad \text{as } n \to \infty$$

$$\text{CLT: } \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to^d \mathcal{N}(0,1)$$

$$\text{CLT: } \bar{X}_n \to^d \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

- Why is this important?
- How can we visualize CLT and LOLN in Python?

LOLN and CLT Visualization

Why Does CLT Work?

- To see the power of the CLT, defining a RV X constructed from various distributions

$$X = f_{\text{Exponential}} + f_{\text{Gamma}} + f_{\text{Normal}} + f_{\text{Chi-squared}} + \dots$$

$$E[X] = \mu, \qquad \text{and} \qquad V[X] = \sigma^2$$

$$\bar{X}_1, \bar{X}_2, \bar{X}_3, .., \bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$

- Why does it work?
  - $\bar{X}_i = \mu + \epsilon_i$
  - Low probability values are drawn less frequently
- Applications of CLT
  - Hypothesis testing
  - Bootstrapping
  - Monte-carlo simulations
  - Motivation to use portfolios (more stable than individual stocks!)

Hypothesis Test

- There are usually two hypotheses. The **null** hypothesis is a boring/uninteresting statement about a parameter. The **alternate** hypothesis is a more interesting that we would like to conclude from the data. If we were flipping a coin 20 times and we observe 14 heads

$$H_0(\text{Null Hypothesis}) : \text{Coin is fair}$$

$$H_1(\text{Alternate Hypothesis}) : \text{Coin is unfair}$$

- Why do we need tests?
- A common measure of probabilistic significance is the p-value. A low p-value (say 5%) tells us that there is a 5% chance of observing a test statistic result as extreme as the one we got

T-Test

- t-distribution was developed to solve the problem of "small samples" in statistics. When the underlying population distribution is normal, but we only have a small sample, CLT might fail to kick in
- In the examples below, we are using the sample mean to test the population mean. The population variance (stddev) is unknown.

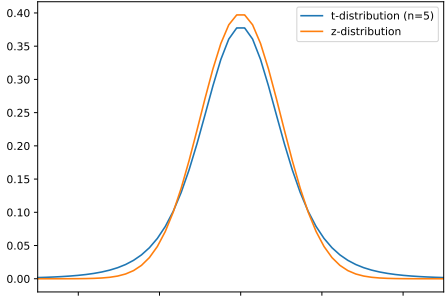
$$t \sim \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \text{or} \quad \frac{b - \beta}{s/\sqrt{n}} \quad \text{or} \quad \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}}}} \quad \sim \quad \frac{\text{Signal}}{\text{Noise}}$$

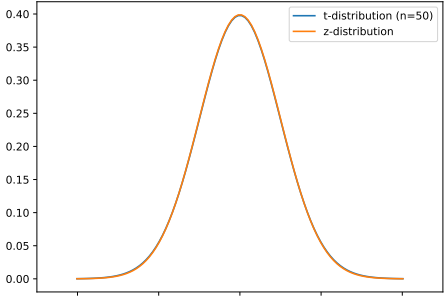
- Example 1: Say you are interning in a hedge fund. Your role is to come up with a trading strategy that beats the market. Historically, the market has given a mean return of $\mu \approx 5\%$ YoY. Your 5Y back-tested returns are

$$X = \{9.5\%, 6\%, 3\%, 8\%, 9\%\}, \quad \bar{X} = 7.1\%, \quad \mu = 5\%, \quad s^2 = \frac{\sum(x_i - \bar{X})}{n-1} = 2.655\%$$

$$t_4 = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{7.1 - 5}{2.655} \approx 1.58$$

PDF of T-distribution vs Normal Distribution





Analysis of Variance (ANOVA)

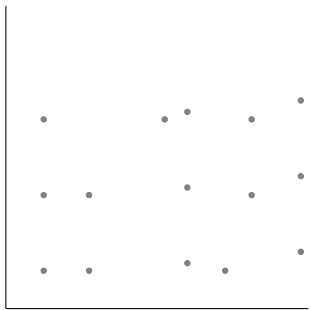
- ANOVA tests if there are statistically significant differences between 3 or more groups. It is an extension of the $t-$test between 2 or more groups. Example - 3 fund managers who generate daily returns. Is there a difference between them? Another example would be age of people using Python, R, and Fortran. Difference in income of 3 groups?
- Question asked by ANOVA: Is there a difference between individual groups (independent variable) and their characteristics (dependent variable)?
- It does **not** give us a causal estimate of how the dependent variable changes the independent variable. For example, it does not tell us if older people use Fortran. It tells us that there is a significant difference in the age of people who use Python/R/Fortran.
- Null hypothesis is that there is no difference in the mean of the individual groups. Alternate hypothesis is that mean of atleas 2 groups differ
- ANOVA decomposes the total variation into variation within groups and variation across groups to develop a test statistic and draw a conclusion

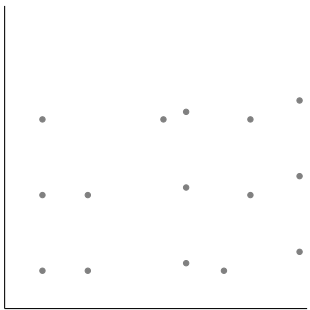
$$\text{SST} = \sum_i (x_i - \bar{X})^2$$

$$\text{SS Total} = \text{SS Within} + \text{SS Across}$$

ANOVA Visualization

- Say we have 15 data points, say the salary of 15 individuals. ANOVA lets us differentiate between the means of different groups of data. For example -





ANOVA Example 1

- Say we have a set whose mean $\mu = \frac{45}{9} = 5$ and $SST = \sum_i (x_i - \mu)^2 = 28$

$$X = \{4, 5, 6, 3, 5, 7, 2, 5, 8\}$$

$$X_1 = \{4, 5, 6\}, \qquad X_2 = \{3, 5, 7\}, \qquad X_3 = \{2, 5, 8\}$$

$$\bar{X}_1 = \qquad \bar{X}_2 = \qquad \bar{X}_3 =$$

$$\text{SSW}_1 = \qquad \text{SSW}_2 = \qquad \text{SSW}_3 =$$

$$\text{SSA} =$$

$$F = \frac{\text{Mean Sq. Across}}{\text{Mean Sq. Within}} = \frac{SSA/k - 1}{SSW/n - k}$$

ANOVA Example 2

- Say we have a set whose mean $\mu = \frac{45}{9} = 5$ and $SST = \sum_i (x_i - \mu)^2 = 28$

$$X = \{4, 5, 6, 3, 5, 7, 2, 5, 8\}$$

$$X_1 = \{5, 6, 7\}, \qquad X_2 = \{3, 4, 5\}, \qquad X_3 = \{2, 5, 8\}$$

$$\bar{X}_1 = \qquad \bar{X}_2 = \qquad \bar{X}_3 =$$

$$\text{SSW}_1 = \qquad \text{SSW}_2 = \qquad \text{SSW}_3 =$$

$$\text{SSA} =$$

$$F = \frac{\text{Mean Sq. Across}}{\text{Mean Sq. Within}} = \frac{SSA/k-1}{SSW/n-k}$$