Business Modeling and Simulation

Lecture 2
Refresh on Probability, Statisitcs and Linear Algebra

Yi Ding

Faculty of Business Administration, University of Macau

Spring 2024

Outline

Outline

Random variable

- random outcome: e.g., head/tail in coin tossing
- random variable: numerical function on random space: e.g., head $\to X = 1$; tail $\to X = 0$
- discrete random variable: takes on discrete number of values, e.g., $0, 1, ...,$
- continuous random variable: takes on continuum of possible values

"**The Merchant of Venice**" by William Shakespeare:
This contract doesn't give you any blood at all. The words expressly specify "a pound of flesh." So take your penalty of a pound of flesh, but if you shed one drop of Christian blood when you cut it, the state of Venice will confiscate your land and property under Venetian law.–Portia

Probability distributions of random variables

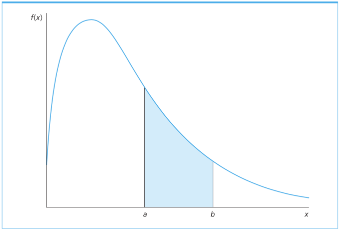
- Probability distribution of discrete variable: e.g., coin tossing
  $P(X = 1) = 1/2$, $P(X = 0) = 1/2$

Probability distributions of random variables

- Probability distribution of discrete variable: e.g., coin tossing $P(X = 1) = 1/2$, $P(X = 0) = 1/2$
- Cumulative probability distribution (cdf): $F(x) = P(X \leq x)$

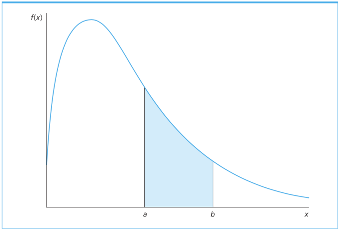
Probability distributions of random variables

- Probability distribution of discrete variable: e.g., coin tossing $P(X = 1) = 1/2$, $P(X = 0) = 1/2$
- Cumulative probability distribution (cdf): $F(x) = P(X \leq x)$
- Probability density function (pdf) of continuous variable: $P(X \leq x) = \int^{x} f(x)dx$



Probability distributions of random variables

- Probability distribution of discrete variable: e.g., coin tossing $P(X = 1) = 1/2$, $P(X = 0) = 1/2$
- Cumulative probability distribution (cdf): $F(x) = P(X \leq x)$
- Probability density function (pdf) of continuous variable: $P(X \leq x) = \int^{x} f(x)dx$



The probability of $P(X = c)$ for continuous variable is ?

Moments of random variables

- Expected value (or mean/expectation) of a random variable X is weighted average of all possible values of X
  - Discrete variable:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

  - Continuous variable: $E(X) = \int xf(x)dx$

Moments of random variables

- Expected value (or mean/expectation) of a random variable X is weighted average of all possible values of X
  - Discrete variable:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

  - Continuous variable: $E(X) = \int x f(x) dx$
- Properties of expected value:

$$E(aY + bX) = aE(Y) + bE(X)$$

where a, b are constants; X, Y are two random variables.

Moments of random variables

- Expected value (or mean/expectation) of a random variable X is weighted average of all possible values of X
  - Discrete variable:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

  - Continuous variable: $E(X) = \int xf(x)dx$
- Properties of expected value:

$$E(aY + bX) = aE(Y) + bE(X)$$

where a, b are constants; X, Y are two random variables.

- Variance: Measures of Variability

$$\mathrm{Var}(X) = E((X - E(X))^2)$$

Moments of random variables

- Expected value (or mean/expectation) of a random variable X is weighted average of all possible values of X
  - Discrete variable:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

  - Continuous variable: $E(X) = \int x f(x) dx$
- Properties of expected value:

$$E(aY + bX) = aE(Y) + bE(X)$$

  where a, b are constants; X, Y are two random variables.
- Variance: Measures of Variability

$$\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

Moments of random variables

- Expected value (or mean/expectation) of a random variable X is weighted average of all possible values of X
  - Discrete variable:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

  - Continuous variable: $E(X) = \int xf(x)dx$
- Properties of expected value:

$$E(aY + bX) = aE(Y) + bE(X)$$

where a, b are constants; X, Y are two random variables.
- Variance: Measures of Variability

$$\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

- Standard deviation: $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$.

Example: normal distribution

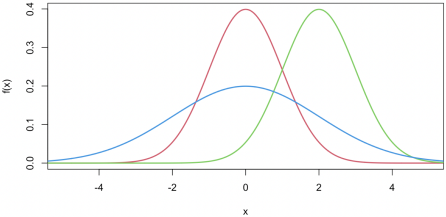
normal distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

- bell shaped
- $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$

Example: normal distribution

normal distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

- bell shaped
- $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$

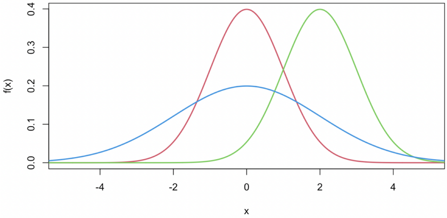


- Guess which is which? $(\mu, \sigma^2) \in \{(0,1), (0,2^2), (2,1)\}$

Example: normal distribution

normal distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

- bell shaped
- $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$



- Guess which is which? $(\mu, \sigma^2) \in \{(0,1), (0, 2^2), (2,1)\}$
  - BLUE: ____________
  - RED: ____________
  - GREEN: ____________

Joint distribution

For two variables X and Y;

Joint distribution

For two variables X and Y; Joint cdf: $F(x, y) = P(X < x, Y < y)$

Joint distribution

For two variables X and Y; Joint cdf: $F(x, y) = P(X < x, Y < y)$

- Covariance: measure of association

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Joint distribution

For two variables X and Y; Joint cdf: $F(x, y) = P(X < x, Y < y)$

- Covariance: measure of association

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Property 1	$\lvert\mathrm{Cov}(X, Y)\rvert \leq \mathrm{sd}(X) \cdot \mathrm{sd}(Y)$
Property 2	if X and Y are independent, i.e., $F(x, y) = F(x) \times F(y)$, then $\mathrm{Cov}(X, Y) = 0$
Property 3	$\mathrm{Cov}(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 \mathrm{Cov}(X, Y)$

Joint distribution

For two variables X and Y; Joint cdf: $F(x, y) = P(X < x, Y < y)$

- Covariance: measure of association

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Property 1	$\lvert\mathrm{Cov}(X, Y)\rvert \leq \mathrm{sd}(X) \cdot \mathrm{sd}(Y)$
Property 2	if X and Y are independent, i.e., $F(x, y) = F(x) \times F(y)$, then $\mathrm{Cov}(X, Y) = 0$
Property 3	$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X, Y)$

- $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$
- Correlation

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{sd}X \cdot \mathrm{sd}Y}$$

Joint distribution

For two variables X and Y; Joint cdf: $F(x,y) = P(X < x, Y < y)$

- Covariance: measure of association

$$\mathrm{Cov}(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Property 1	$\lvert\mathrm{Cov}(X,Y)\rvert \leq \mathrm{sd}(X)\cdot\mathrm{sd}(Y)$
Property 2	if X and Y are independent, i.e., $F(x,y) = F(x)\times F(y)$, then $\mathrm{Cov}(X,Y) = 0$
Property 3	$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X,Y)$

- $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$
- Correlation

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{sd}X\cdot\mathrm{sd}Y}$$

Properties:

$\lvert\mathrm{Corr}(X,Y)\rvert \leq 1$;

Joint distribution

For two variables X and Y; Joint cdf: $F(x, y) = P(X < x, Y < y)$

- Covariance: measure of association

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Property 1	$\lvert\mathrm{Cov}(X, Y)\rvert \leq \mathrm{sd}(X) \cdot \mathrm{sd}(Y)$
Property 2	if X and Y are independent, i.e., $F(x, y) = F(x) \times F(y)$, then $\mathrm{Cov}(X, Y) = 0$
Property 3	$\mathrm{Cov}(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 \mathrm{Cov}(X, Y)$

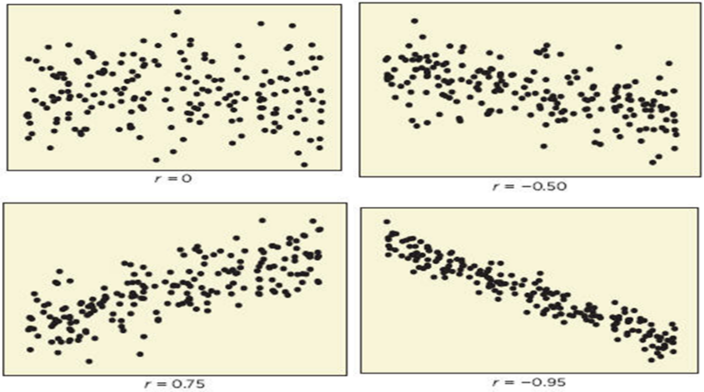
- $\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$
- Correlation

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{sd}X \cdot \mathrm{sd}Y}$$

Properties:

$\lvert\mathrm{Corr}(X, Y)\rvert \leq 1$; $\mathrm{Corr}(a_1 X + b_1, a_2 Y + b_2) = \mathrm{Sign(a_1 a_2) Corr(X, Y)}$

Illustration of different correlations



Conditional expectation

- Conditional expectation is the distribution of a random variable conditional on another random variable taking on a specific value; e.g.,
  - $P(\text{Grade in Statistics} = A | \text{Grade in Calculas} = A)$
  - $P(\text{child's height} \leq x | \text{mother's height} \leq y)$
- Bayesian formula

$$P(X \leq x | Y \leq y) = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)}$$

- Conditional Expectation $E(Y|X = x) = \sum_i y_i P(Y = y_i | X = x)$
- Tower property (iterated law) of conditonal expectation:

$$E(Y) = E(E(Y|X))$$

Independence and Uncorrelatedness

- Two variables X and Y are independent when $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all x, y.
- If X and Y are independent, $P(X \leq x | Y \leq y) = P(X \leq x)$ and hence $E(X|Y) = E(X)$.
- Uncorrelatedness: $\mathrm{Corr}(X, Y) = 0$.
- Independence implies corelation zero.
- If X and Y are uncorrelated this does not necessarily imply independence!

Normal and related distributions

- standard normal Z, pdf $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$
- Chi-Square distribution(χ^2_m): $X = Z_1^2 + Z_2^2 + Z_m^2$, $(Z_i)_{i=1}^m$ are independent standard normal distributed, m is the degree of freedom.
- T distribution (t_m): $T = \frac{Z}{\sqrt{X/m}}$, where Z and X are independent.
- F distribution (F_{m_1,m_2}): $F = \frac{X_1/m_1}{X_2/m_2}$, where X_1 and X_2 are independent Chi-square with df m_1 and m_2 resp.

Simple random sampling

- If $Y_1, ... Y_n$ are independent random variables with the common probability density $f(y; \theta)$, then $\{Y_1, ..., Y_n\}$ is said to ba a random sample from $f(y; \theta)$
- When $\{Y_1, ..., Y_n\}$ is a random sample from $f(y; \theta)$, we also say they are independent, identically distributed (i.i.d.) random variables. e.g.,
    - repeated experiment of coin tossing.

Descriptive statistics

Given a set of data $(x_i)_{i=1}^n$

- **sample mean** (sample average): $\overline{x} = \sum_{i=1}^n x_i/n$

Descriptive statistics

Given a set of data $(x_i)_{i=1}^n$

- **sample mean** (sample average): $\overline{x} = \sum_{i=1}^n x_i/n$
- Let $d_i = x_i - \overline{x}$; d_i is deviation from average.

Descriptive statistics

Given a set of data $(x_i)_{i=1}^n$

- **sample mean** (sample average): $\overline{x} = \sum_{i=1}^n x_i/n$
- Let $d_i = x_i - \overline{x}$; d_i is deviation from average. Important properties:
  - $$\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \overline{x}) = 0$$

Descriptive statistics

Given a set of data $(x_i)_{i=1}^n$

- **sample mean** (sample average): $\overline{x} = \sum_{i=1}^n x_i/n$
- Let $d_i = x_i - \overline{x}$; d_i is deviation from average. Important properties:
  - $$\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \overline{x}) = 0$$
  - $$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n(\overline{x})^2$$

Descriptive statistics

Given a set of data $(x_i)_{i=1}^n$

- **sample mean** (sample average): $\overline{x} = \sum_{i=1}^n x_i / n$
- Let $d_i = x_i - \overline{x}$; d_i is deviation from average. Important properties:
  - $$\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \overline{x}) = 0$$
  - $$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n(\overline{x})^2$$
  - Given two data sets $(x_i)_{i=1}^n$ and $(y_i)_{i=1}^n$

$$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{x}) = \sum_{i=1}^n x_i y_i - n\overline{x}\,\overline{y}$$

Estimator and statistical properties

- an estimator W of parameter θ can be written as $W = h(Y_1, ..., Y_n)$.
- W is also a random variable (sample dependent).

Estimator and statistical properties

- an estimator W of parameter θ can be written as $W = h(Y_1, ..., Y_n)$.
- W is also a random variable (sample dependent).
- Unbiasness:

$$E(W) = \theta$$

e.g.,

Estimator and statistical properties

- an estimator W of parameter θ can be written as $W = h(Y_1, ..., Y_n)$.
- W is also a random variable (sample dependent).
- Unbiasness:

$$E(W) = \theta$$

e.g.,

  - sample mean $\overline{Y}$ is an unbiased estimator of $E(Y)$.
  - For i.i.d. variables, an unbiased estimator of $\mathrm{Var}(Y)$ is

$$S^2 = \frac{1}{n-1} \sum_{i}^{n} (Y_i - \overline{Y})^2.$$

Estimator and statistical properties

- an estimator W of parameter θ can be written as $W = h(Y_1, ..., Y_n)$.
- W is also a random variable (sample dependent).
- Unbiasness:

$$E(W) = \theta$$

e.g.,

  - sample mean $\overline{Y}$ is an unbiased estimator of $E(Y)$.
  - For i.i.d. variables, an unbiased estimator of $\mathrm{Var}(Y)$ is

$$S^2 = \frac{1}{n-1}\sum_i^n (Y_i - \overline{Y})^2.$$

- Bias of an Estimator: $\text{Bias}(W) = E(W) - \theta$

Efficiency of unbiased estimator and MSE

- an unbiased estimator W_1 is efficient relative to an unbiased estimator W_2 if $\mathrm{Var}(W_1) \leq \mathrm{Var}(W_2)$ for all θ with strict inequality for at least one θ.
- e.g., for i.i.d. variables with $\mathrm{Var}(Y) = \sigma^2$:

$$\mathrm{Var}(\overline{Y}) = \frac{\sigma^2}{n}$$

  $\overline{Y}$ is efficient relative to Y_1, because $\mathrm{Var}(\overline{Y}) < \mathrm{Var}(Y_1) = \sigma^2$.
- Mean squared error (MSE): If we do not restrict our attention to unbiased estimators, then comparing variances is meaningless. Instead, we can look at mean squared error (MSE)

$$E((W - \theta)^2) = \mathrm{Var}(W) + (\mathrm{bias}(W))^2$$

Consistency

- W_n depends on a sample $Y_1, ..., Y_n$. W_n is consistent estimator of θ if

$$P(|W_n - \theta| \geq \varepsilon) \to 0, \text{ as } n \to \infty,$$

write $\mathrm{plim(W_n)} = \theta$ (convergence in probability).

Consistency

- W_n depends on a sample $Y_1, ..., Y_n$. W_n is consistent estimator of θ if
$$P(|W_n - \theta| \geq \varepsilon) \to 0, \text{ as } n \to \infty,$$
write $\mathrm{plim(W_n)} = \theta$ (convergence in probability).
- Law of large numbers (LNN): for i.i.d. sample $Y_1, ..., Y_n$ with expectation μ
$$\mathrm{plim}(\overline{\mathrm{Y}}) = \mu$$
- Slustky' Theorem

If $\mathrm{plim(T_n) = T}$*, and* $\mathrm{plim}(\overline{\mathrm{U}}_\mathrm{n}) = \beta$*, where T is a random variable and* β *is a constant, then* $\mathrm{plim(T_n + U_n) = T} + \beta$*;*
$\mathrm{plim(T_n \cdot U_n) = T}\beta$*;*
$\mathrm{plim(T_n/U_n) = T}/\beta$ *provided that* $\beta \neq 0$

Confidence interval

- What direct statement can we make about the true value θ?
- Confidence interval: e.g.
- $(Y_i)_{i=1}^n$ are i.i.d. normal $\mathcal{N}(\mu, 1)$.

Confidence interval

- What direct statement can we make about the true value θ?
- Confidence interval: e.g.
- $(Y_i)_{i=1}^n$ are i.i.d. normal $\mathcal{N}(\mu, 1)$. Then $\overline{Y} \sim \mathcal{N}(\mu, 1/n)$,

Confidence interval

- What direct statement can we make about the true value θ?
- Confidence interval: e.g.
- $(Y_i)_{i=1}^n$ are i.i.d. normal $\mathcal{N}(\mu, 1)$. Then $\overline{Y} \sim \mathcal{N}(\mu, 1/n)$,

$$P(-1.96/\sqrt{n} < (\overline{Y} - \mu) < 1.96/\sqrt{n}) = 0.95$$

Confidence interval

- What direct statement can we make about the true value θ?
- Confidence interval: e.g.
- $(Y_i)_{i=1}^n$ are i.i.d. normal $\mathcal{N}(\mu, 1)$. Then $\overline{Y} \sim \mathcal{N}(\mu, 1/n)$,

$$P(-1.96/\sqrt{n} < (\overline{Y} - \mu) < 1.96/\sqrt{n}) = 0.95$$

The 95% confidence interval of μ is $[\overline{Y} - 1.96/\sqrt{n}, \overline{Y} + 1.96/\sqrt{n}]$.

Confidence interval

- What direct statement can we make about the true value θ?
- Confidence interval: e.g.
- $(Y_i)_{i=1}^n$ are i.i.d. normal $\mathcal{N}(\mu, 1)$. Then $\overline{Y} \sim \mathcal{N}(\mu, 1/n)$,

$$P(-1.96/\sqrt{n} < (\overline{Y} - \mu) < 1.96/\sqrt{n}) = 0.95$$

The 95% confidence interval of μ is $[\overline{Y} - 1.96/\sqrt{n}, \overline{Y} + 1.96/\sqrt{n}]$. It means the **random interval** contains μ with probability 0.95. In other words, **before the random sample** is drawn, there is a 95% chance that the CI contains μ.

Hypothesis test

- Election example:
  - Two candidates A and B got election votes 42% v.s. 58%, resp.
  - Candidate A, however, think that he/she is more popular.

Hypothesis test

- Election example:
  - Two candidates A and B got election votes 42% v.s. 58%, resp.
  - Candidate A, however, think that he/she is more popular.
  - **Question**: Is the election rigged?

Hypothesis test

- Election example:
  - Two candidates A and B got election votes 42% v.s. 58%, resp.
  - Candidate A, however, think that he/she is more popular.
  - **Question**: Is the election rigged?
  - Candidate A hires a consulting agency to randomly sample 100 voters to record whether or not each person voted for him. For the sample collected, 53 people voted for Candidate A.
  - Should Candidate A conclude that the election was indeed a fraud?
- Let θ denote the true proportion of population voting.

Hypothesis test

- Election example:

  - Two candidates *A* and *B* got election votes 42% v.s. 58%, resp.
  - Candidate *A*, however, think that he/she is more popular.
  - **Question**: Is the election rigged?
  - Candidate *A* hires a consulting agency to randomly sample 100 voters to record whether or not each person voted for him. For the sample collected, 53 people voted for Candidate A.
  - Should Candidate A conclude that the election was indeed a fraud?

- Let θ denote the true proportion of population voting. Hypothesis test: $H_0 : \theta \leq 0.42$ v.s. $H_1 : \theta > 0.42$
- In order to conclude that H_0 is false and that H_1 is true, we must have strong evidence.

Hypothesis test

- Election example:
  - Two candidates A and B got election votes 42% v.s. 58%, resp.
  - Candidate A, however, think that he/she is more popular.
  - **Question**: Is the election rigged?
  - Candidate A hires a consulting agency to randomly sample 100 voters to record whether or not each person voted for him. For the sample collected, 53 people voted for Candidate A.
  - Should Candidate A conclude that the election was indeed a fraud?
- Let θ denote the true proportion of population voting. Hypothesis test: $H_0 : \theta \leq 0.42$ v.s. $H_1 : \theta > 0.42$
- In order to conclude that H_0 is false and that H_1 is true, we must have strong evidence.
- How "strong" is strong? Is 100 sample sufficient?

Hypothesis test, cont'd

- Type I error: Reject H_0 when it is in fact true
- Type II error: failing to reject H_0 when it is actually false

Hypothesis test, cont'd

- Type I error: Reject H_0 when it is in fact true
- Type II error: failing to reject H_0 when it is actually false
- Want to get small Type I error
- Significance level: α is probability of a Type I error.

$$\alpha = P(\text{Reject } H_0 | H_0)$$

- Commonly used levels $\alpha = 0.01, 0.05, 0.1$

Test statistic and p-values

- Construct test statistic:

$$T = \frac{\overline{Y} - \theta_0}{s/\sqrt{n}}$$

For the **election example**:

- $\theta_0 = 42\%$, $n = 100$, $\overline{Y} = 53\%$,
- s is estimated standard deviation of Y_i where Y_i is binary variable (indicator of voting for candidate A), $s = \sqrt{\overline{Y}(1 - \overline{Y})}$.
- observed value of T is $t = \frac{0.53 - 0.42}{\sqrt{100 \times 0.42 \times 0.58}} \approx 2.229$

Test statistic and p-values

- Central Limit Theorem (CLT): Under H_0,

$$T \to \mathcal{N}(0, 1) \text{ in distribution.}$$

Test statistic and p-values

- Central Limit Theorem (CLT): Under H_0,

$$T \to \mathcal{N}(0,1) \text{ in distribution.}$$

For the **election example**:

- Will reject H_0 for large T value

Test statistic and p-values

- Central Limit Theorem (CLT): Under H_0,

$$T \to \mathcal{N}(0, 1) \text{ in distribution.}$$

For the **election example**:

- Will reject H_0 for large T value
- p-value$= P(T > 2.229|H_0) = 1 - \Phi(2.229) = 0.013$, where $\Phi(\cdot)$ is cdf of standard normal.
- Given significance level $\alpha = 0.05$,
- p-value=0.013<α;

Test statistic and p-values

- Central Limit Theorem (CLT): Under H_0,

$$T \to \mathcal{N}(0, 1) \text{ in distribution.}$$

For the **election example**:

- Will reject H_0 for large T value
- p-value$= P(T > 2.229|H_0) = 1 - \Phi(2.229) = 0.013$, where $\Phi(\cdot)$ is cdf of standard normal.
- Given significance level $\alpha = 0.05$,
- p-value=0.013<α; reject H_0.

Outline

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

- Multiplication: $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have dimension $m \times n$ and $n \times p$, resp.

$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik} b_{kj}\right] = \left[a_{i1} b_{1j} + a_{21} b_{2j} + ..., + a_{in} b_{nj}\right]$$

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

- Multiplication: $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have dimension $m \times n$ and $n \times p$, resp.

$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik} b_{kj}\right] = \left[a_{i1} b_{1j} + a_{21} b_{2j} + ..., + a_{in} b_{nj}\right]$$

e.g.

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}$$

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,
$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

- Multiplication: $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have dimension $m \times n$ and $n \times p$, resp.
$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik} b_{kj}\right] = \left[a_{i1} b_{1j} + a_{21} b_{2j} + \ldots, + a_{in} b_{nj}\right]$$

  e.g.

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}$$

- Transpose: for matrix $\mathbf{A} = [a_{ij}]$, $\mathbf{A}^t = [a_{ji}]$

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

- Multiplication: $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have dimension $m \times n$ and $n \times p$, resp.

$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik} b_{kj}\right] = [a_{i1} b_{1j} + a_{21} b_{2j} + \ldots, + a_{in} b_{nj}]$$

  e.g.

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}$$

- Transpose: for matrix $\mathbf{A} = [a_{ij}]$, $\mathbf{A}^t = [a_{ji}]$
- Trace: for square matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A}) = \mathrm{a}_{11} + \mathrm{a}_{22} +, \ldots, + \mathrm{a}_{\mathrm{nn}}$.

Matrix Operations

- Addition: two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ that have same dimension $m \times n$,
$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$
- Multiplication: $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have dimension $m \times n$ and $n \times p$, resp.
$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik} b_{kj}\right] = \left[a_{i1} b_{1j} + a_{21} b_{2j} + ..., + a_{in} b_{nj}\right]$$
e.g.
$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}$$
- Transpose: for matrix $\mathbf{A} = [a_{ij}]$, $\mathbf{A}^t = [a_{ji}]$
- Trace: for square matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A}) = \mathrm{a}_{11} + \mathrm{a}_{22} +, ..., + \mathrm{a}_{\mathrm{nn}}$.
- Inverse: $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where $\mathbf{I} = \mathrm{diag}(1, ..., 1)$.

Quadratic forms and Positive definite matrices

- quadratic form

$$\mathbf{x}^t\mathbf{A}\mathbf{x} = \sum_{ij} x_i x_j a_{ij} = \sum_i x_i^2 a_{ii}^2 + 2\sum_{i<j} a_{ij} x_i x_j,$$

  where $\mathbf{x}$ is $n \times 1$ vector, and $\mathbf{A}$ is $n \times n$ symetric matrix, $a_{ij} = a_{ji}$.

- Positive Definite and Positive Semi-Definite:
  symetric matrix $\mathbf{A}$ is called positive definite (or semi-positive definite) when

$$\mathbf{x}^t\mathbf{A}\mathbf{x} > 0 (\text{or} \geq 0)$$

Quadratic forms and Positive definite matrices

- quadratic form

$$\mathbf{x}^t\mathbf{A}\mathbf{x} = \sum_{ij} x_i x_j a_{ij} = \sum_i x_i^2 a_{ii}^2 + 2\sum_{i<j} a_{ij}x_i x_j,$$

where $\mathbf{x}$ is $n \times 1$ vector, and $\mathbf{A}$ is $n \times n$ symetric matrix, $a_{ij} = a_{ji}$.

- Positive Definite and Positive Semi-Definite:
symetric matrix $\mathbf{A}$ is called positive definite (or semi-positive definite) when

$$\mathbf{x}^t\mathbf{A}\mathbf{x} > 0(\text{or} \geq 0)$$

e.g. $\mathbf{X}^t\mathbf{X}$ is (semi-) positive definite for any $n \times k$ matrix $\mathbf{X}$

- Idempotent: If $\mathbf{A}\mathbf{A} = \mathbf{A}$, $\mathbf{A}$ is called idempotent.

Quadratic forms and Positive definite matrices

- quadratic form
$$\mathbf{x}^t\mathbf{A}\mathbf{x} = \sum_{ij} x_i x_j a_{ij} = \sum_i x_i^2 a_{ii}^2 + 2\sum_{i<j} a_{ij} x_i x_j,$$
where $\mathbf{x}$ is $n \times 1$ vector, and $\mathbf{A}$ is $n \times n$ symetric matrix, $a_{ij} = a_{ji}$.
- Positive Definite and Positive Semi-Definite:
symetric matrix $\mathbf{A}$ is called positive definite (or semi-positive definite) when
$$\mathbf{x}^t\mathbf{A}\mathbf{x} > 0(\text{or} \geq 0)$$
e.g. $\mathbf{X}^t\mathbf{X}$ is (semi-) positive definite for any $n \times k$ matrix $\mathbf{X}$
- Idempotent: If $\mathbf{A}\mathbf{A} = \mathbf{A}$, $\mathbf{A}$ is called idempotent. e.g. if $\mathbf{X}^t\mathbf{X}$ is positive definite, then
$$\mathbf{P} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t, \quad \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$$
are both idempotent.

Quadratic forms and Positive definite matrices

- quadratic form

$$\mathbf{x}^t\mathbf{A}\mathbf{x} = \sum_{ij} x_i x_j a_{ij} = \sum_i x_i^2 a_{ii}^2 + 2\sum_{i<j} a_{ij} x_i x_j,$$

  where $\mathbf{x}$ is $n \times 1$ vector, and $\mathbf{A}$ is $n \times n$ symetric matrix, $a_{ij} = a_{ji}$.

- Positive Definite and Positive Semi-Definite:
  symetric matrix $\mathbf{A}$ is called positive definite (or semi-positive definite) when

$$\mathbf{x}^t\mathbf{A}\mathbf{x} > 0(\text{or} \geq 0)$$

  e.g. $\mathbf{X}^t\mathbf{X}$ is (semi-) positive definite for any $n \times k$ matrix $\mathbf{X}$

- Idempotent: If $\mathbf{AA} = \mathbf{A}$, $\mathbf{A}$ is called idempotent. e.g. if $\mathbf{X}^t\mathbf{X}$ is positive definite, then

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t, \quad \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$$

  are both idempotent.
  property of idempotent matrix: $\mathrm{tr}(\mathbf{A}) = \mathrm{rank}(\mathbf{A})$.

Outline

Summary

Reference: 

- GC Chapter 3; LP Chapter 4–7

Summary

Reference: 

- GC Chapter 3; LP Chapter 4–7

Come next: 

- Refresh on Python
- Random number generator

Survey and Q&A time