Hypothesis testing

Andrew Thangaraj

IIT Madras

Section 1

Introduction

What is hypothesis testing?

Motivating example: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

What is hypothesis testing?

Motivating example: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Hypothesis testing

- Using samples, decide between a *null hypothesis* denoted H_0 and an *alternative hypothesis* denoted H_A

What is hypothesis testing?

Motivating example: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Hypothesis testing

- Using samples, decide between a *null hypothesis* denoted H_0 and an *alternative hypothesis* denoted H_A
  - Counterfeit coin example: H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

What is hypothesis testing?

Motivating example: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Hypothesis testing

- Using samples, decide between a *null hypothesis* denoted H_0 and an *alternative hypothesis* denoted H_A
  - Counterfeit coin example: H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$
- One of the most important statistical analysis methods with a wide range of applications

Accepting or Rejecting the Null Hypothesis

Example: Is a coin authentic or counterfeit?

- Suppose we toss the coin 3 times
  - Possible outcomes are $HHH, HHT, \ldots, TTT$
  - For some outcomes, we will accept H_0 and the others, we will reject H_0
  - Let A be the set of all outcomes for which we accept H_0
- Every *acceptance* subset A corresponds to a *test*

Accepting or Rejecting the Null Hypothesis

Example: Is a coin authentic or counterfeit?

- Suppose we toss the coin 3 times
  - Possible outcomes are $HHH, HHT, \ldots, TTT$
  - For some outcomes, we will accept H_0 and the others, we will reject H_0
  - Let A be the set of all outcomes for which we accept H_0
- Every *acceptance* subset A corresponds to a *test*

Acceptance set and test

$X_1, \ldots, X_n \sim$ iid X, H_0: null hypothesis, H_A: alternative

- Suppose $X \in \mathcal{X}$. Then, the samples $X_1, \ldots, X_n \in \mathcal{X}^n$
- Subset $A \subseteq \mathcal{X}^n \leftrightarrow$ a hypothesis test

If $X_1, \ldots, X_n \in A$, we accept H_0; otherwise, we reject H_0

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true
  - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{is true})$

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true
  - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{is true})$
- Metric 2: *Power* of a test, $1 - \beta$

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true
  - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{is true})$
- Metric 2: *Power* of a test, $1 - \beta$
  - Type II error: Accept H_0 when H_A is true

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true
  - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{is true})$
- Metric 2: *Power* of a test, $1 - \beta$
  - Type II error: Accept H_0 when H_A is true
  - $\beta = P(\text{Type II error}) = P(\text{Accept } H_0 | H_A \text{is true})$

Metrics for hypothesis testing

Example: Is a coin authentic or counterfeit?

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

- Suppose we toss the coin 3 times: 8 outcomes
  - $2^8 = 256$ subsets $\leftrightarrow$ 256 tests
- How to define a *good* acceptance set or a test?

Size and power of a test

- Metric 1: *Significance level* (also called *size*) of a test, denoted α
  - Type I error: Reject H_0 when H_0 is true
  - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{is true})$
- Metric 2: *Power* of a test, $1 - \beta$
  - Type II error: Accept H_0 when H_A is true
  - $\beta = P(\text{Type II error}) = P(\text{Accept } H_0 | H_A \text{is true})$
  - Power $= 1 - \beta = P(\text{Reject } H_0 | H_A \text{is true})$

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \dots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
  - $\alpha = P(A^c | P(H) = 0.5) = 2/8 = 0.25$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \dots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
  - $\alpha = P(A^c | P(H) = 0.5) = 2/8 = 0.25$
  - $\beta = P(A | P(H) = 0.6) = 3\underbrace{(0.4)^2(0.6)}_{2T,\,1H} + \underbrace{3(0.4)(0.6)^2}_{2H,\,1T} = 0.72$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
  - $\alpha = P(A^c | P(H) = 0.5) = 2/8 = 0.25$
  - $\beta = P(A | P(H) = 0.6) = 3(0.4)^2(0.6) + 3(0.4)(0.6)^2 = 0.72$
- $A = \{TTT, TTH, THT, HTT\}$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
  - $\alpha = P(A^c | P(H) = 0.5) = 2/8 = 0.25$
  - $\beta = P(A | P(H) = 0.6) = 3(0.4)^2(0.6) + 3(0.4)(0.6)^2 = 0.72$
- $A = \{TTT, TTH, THT, HTT\}$
  - $\alpha = 4/8 = 0.5$

There is a tradeoff between α and β

Counterfeit coin: Computing α, β

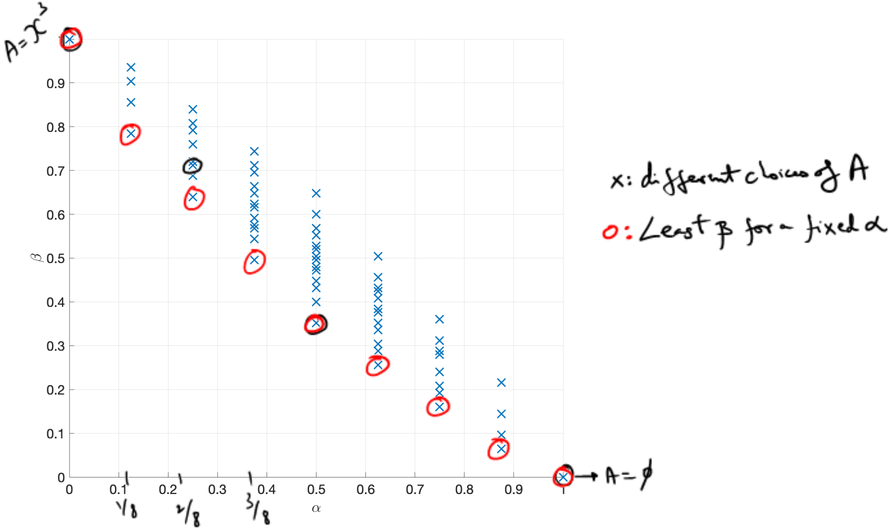
H_0: $P(H) = 0.5$ and H_A: $P(H) = 0.6$

Toss 3 times. $\mathcal{X}^3 = \{HHH, HHT, \ldots, TTT\}$

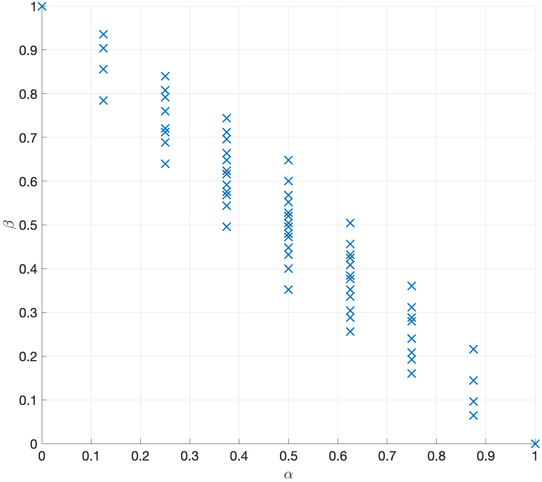
- $A = \phi$
  - Always reject H_0
  - $\alpha = 1$, $\beta = 0$
- $A = \mathcal{X}^3$
  - Always accept H_0
  - $\alpha = 0$, $\beta = 1$
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
  - $\alpha = P(A^c | P(H) = 0.5) = 2/8 = 0.25$
  - $\beta = P(A | P(H) = 0.6) = 3(0.4)^2(0.6) + 3(0.4)(0.6)^2 = 0.72$
- $A = \{TTT, TTH, THT, HTT\}$
  - $\alpha = 4/8 = 0.5$
  - $\beta = 0.4^3 + 3(0.4)^2(0.6) = 0.352$

There is a tradeoff between α and β

Counterfeit coin: α, β for all 256 tests



Counterfeit coin: α, β for all 256 tests



What if we toss 100 times? What about other distributions?

Neyman-Pearson paradigm of hypothesis testing

$$X_1, \dots, X_n \sim \text{iid } X$$

- H_0: null hypothesis on distribution of X, H_A: alternative hypothesis
- Test: defined by an acceptance set A
  - If samples fall in A, accept H_0; otherwise, reject H_0
- Two errors
  - Type I error: Reject H_0 when H_0 is true
  - Type II error: Accept H_0 when H_A is true
- Two metrics
  - *Significance level*, α
    - $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0)$
  - *Power* of a test, $1 - \beta$
    - $\beta = P(\text{Type II error}) = P(\text{Accept } H_0 | H_A)$

Section 2

Problems

Problem 1

Consider 100 tosses of a coin, which could be either authentic with probabiltiy of heads equal to 0.5, or counterfeit with probability of heads 0.6. Suppose T is the number of heads seen. Consider a test that rejects H_0 if $T > c$ for some constant c. What is the significance level of the test? What is the power of the test?

$A = \{\text{outcomes}: T \leq c\}$ $\quad T \sim \text{Binomial}(100, P(H))$

$$\alpha = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_0}_{P(H)=0.5}) = P(A^c \mid P(H)=1/2) = \sum_{k=c+1}^{100} \binom{100}{k} \left(\tfrac{1}{2}\right)^k \left(\tfrac{1}{2}\right)^{100-k}$$

→ Fix an α and find c

$$1-\beta = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_A}_{P(H)=0.6}) = P(A^c \mid P(H)=0.6) = \sum_{k=c+1}^{100} \binom{100}{k} 0.6^k \, 0.4^{100-k}$$

→ lower $c \Rightarrow$ higher power.

→ Fix α. Find lowest possible c.

Problem 2

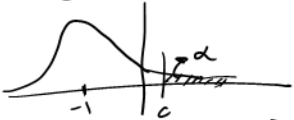
Consider one sample $X \sim \text{Normal}(\mu, 1)$. Let the null and alternative hypothesis be $H_0 : \mu = -1$ and $H_A : \mu = 1$. Consider a test that rejects H_0 if $X > c$ for some constant c. What is the significance level of the test? What is the power of the test?

$$A = \{X \le c\}$$

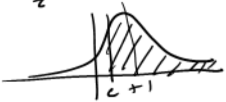
$$\alpha = P(A^c \mid \mu = -1) = P(N(-1,1) > c)$$

$$= P\Big(\underbrace{\frac{N(-1,1)+1}{1}}_{Z \sim N(0,1)} > \frac{c+1}{1}\Big) = P(Z > c+1)$$

$$= 1 - F_Z(c+1)$$



F_Z : CDF of $N(0,1)$



$$1-\beta = P(A^c \mid \mu = 1) = P(N(1,1) > c)$$

$$= P\Big(\frac{N(1,1)-1}{1} > \frac{c-1}{1}\Big) = P(Z > c-1)$$

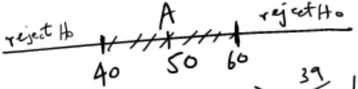
$$= 1 - F_Z(c-1)$$

$\underline{\underline{c = -1}}$ $\alpha = 1/2$, $\beta = 1 - F_Z(-2)$

$\underline{\underline{c = 0}}$ $\alpha = 1 - F_Z(1)$, $\beta = 1 - F_Z(-1)$

Problem 3

Consider one sample $X \sim$ Binomial$(100, p)$. Let the null and alternative hypothesis be $H_0 : p = 0.5$ and $H_A : p \neq 0.5$. Consider a test that rejects H_0 if $|X - 50| > 10$. What is the significance level of the test? What is the power of the test as a function of p? Use the normal approximation.



$A = \{40, 41, \ldots, 60\}$

$$\alpha = P(A^c \mid p = 0.5) = \sum_{k=0}^{39} \binom{100}{k} \left(\frac{1}{2}\right)^{100} + \sum_{k=61}^{100} \binom{100}{k} \left(\frac{1}{2}\right)^{100}$$

$$1 - \beta = P(A^c \mid p) = P(|X-50| > 10 \mid p) = P(X > 60 \mid p) + P(X < 40 \mid p)$$

$$= P\left(\frac{X - 100p}{\sqrt{100p(1-p)}} > \frac{60 - 100p}{\sqrt{100p(1-p)}}\right) + P\left(\frac{X - 100p}{\sqrt{100p(1-p)}} < \frac{40 - 100p}{\sqrt{100p(1-p)}}\right)$$

$$\simeq P\left(N(0,1) > \frac{6 - 10p}{\sqrt{p(1-p)}}\right) + P\left(N(0,1) < \frac{4 - 10p}{\sqrt{p(1-p)}}\right)$$

$$= 1 - F_Z\left(\frac{6 - 10p}{\sqrt{p(1-p)}}\right) + F_Z\left(\frac{4 - 10p}{\sqrt{p(1-p)}}\right)$$

mean $= 100p$, Var $= 100p(1-p)$

$X \approx$ Normal$(100p, 100p(1-p))$

$Z \sim N(0,1)$

Problem 4

Consider 100 samples $X_1, \ldots, X_{100} \sim$ iid Normal$(\mu, 1)$. Let the null and alternative hypothesis be $H_0 : \mu = -1$ and $H_A : \mu = 1$. Suppose $T = (X_1 + \cdots + X_{100})/100$. Consider a test that rejects H_0 if $T > c$ for some constant c. What is the significance level of the test? What is the power of the test?

$T \sim N\left(\mu, \frac{1}{100}\right) \qquad A = \{T \le c\}$

$$\alpha = P(T > c \mid \mu = -1) = P\left(\underbrace{\frac{N(-1, \frac{1}{100}) + 1}{1/10}}_{Z \sim N(0,1)} > \frac{c+1}{1/10}\right) = P(Z > 10(c+1))$$

$$= 1 - F_Z(10(c+1))$$

$$\overset{c=0}{=} \alpha = 1 - \underbrace{F_Z(10)}_{\approx 1}$$

$$1 - \beta = P(T > c \mid \mu = 1) = P(Z > 10(c-1))$$

$$= 1 - F_Z(10(c-1))$$

$$\overset{c=0}{=} 1 - \beta = 1 - \underbrace{F_Z(-10)}_{\approx 0}$$

Section 3

Types of hypothesis testing

Simple hypothesis

Definition (Simple hypothesis)

A hypothesis that completely specifies the distribution of the samples is called a simple hypothesis.

Simple hypothesis

Definition (Simple hypothesis)

A hypothesis that completely specifies the distribution of the samples is called a simple hypothesis.

Examples of simple hypothesis

- Coin toss
  - P(Heads) $= 0.5$, P(Heads) $= 0.9$ etc.
- Normal$(\mu, 3)$ samples
  - $\mu = 1$, $\mu = -1$ etc.

Simple hypothesis

Definition (Simple hypothesis)

A hypothesis that completely specifies the distribution of the samples is called a simple hypothesis.

Examples of simple hypothesis

- Coin toss
  - P(Heads) $=$ 0.5, P(Heads) $=$ 0.9 etc.
- Normal$(\mu, 3)$ samples
  - $\mu = 1$, $\mu = -1$ etc.

Simple null vs simple alternative

- Very well understood, best approach is known
- Rarely occurs

Composite hypothesis

Definition (Composite hypothesis)

A hypothesis that does not completely specify the distribution of the samples is called a composite hypothesis.

Composite hypothesis

Definition (Composite hypothesis)

A hypothesis that does not completely specify the distribution of the samples is called a composite hypothesis.

Examples

- Coin toss
  - Null: P(Heads) $= 0.5$ (coin is fair), simple
  - Alternative: P(Heads) $\neq 0.5$ (coin is unfair), composite
- Normal$(\mu, 3)$ samples
  - Null: $\mu = 0$ (some effect is not present), simple
  - Alternative: $\mu > 1$ (effect is present), composite

Composite hypothesis

Definition (Composite hypothesis)

A hypothesis that does not completely specify the distribution of the samples is called a composite hypothesis.

Examples

- Coin toss
  - Null: P(Heads) $= 0.5$ (coin is fair), simple
  - Alternative: P(Heads) $\neq 0.5$ (coin is unfair), composite
- Normal$(\mu, 3)$ samples
  - Null: $\mu = 0$ (some effect is not present), simple
  - Alternative: $\mu > 1$ (effect is present), composite

Simple/composite null vs composite alternative

- Well studied, but multiple approaches are possible
- Most common

Standard tests: One sample

$$X_1, \ldots, X_n \sim iid X, E[X] = \mu, \mathrm{Var}(X) = \sigma^2$$

Standard tests: One sample

$$X_1, \ldots, X_n \sim iid X, E[X] = \mu, \mathrm{Var}(X) = \sigma^2$$

- Testing for mean, null $H_0 : \mu = c$
  - Alternative
    - Right tail test, $H_A : \mu > c$
    - Left tail test, $H_A : \mu < c$
    - Two tail test, $H_A : \mu \neq c$
  - Two cases: known or unknown variance

Standard tests: One sample

$$X_1, \ldots, X_n \sim iid X, E[X] = \mu, \text{Var}(X) = \sigma^2$$

- Testing for mean, null $H_0 : \mu = c$
  - Alternative
    - Right tail test, $H_A : \mu > c$
    - Left tail test, $H_A : \mu < c$
    - Two tail test, $H_A : \mu \neq c$
  - Two cases: known or unknown variance
- Testing for variance
  - Null $H_0 : \sigma = c$
  - Alternative $H_A : \sigma > c$

Standard tests: Two samples

$$X_1, \ldots, X_{n_1} \sim iid X, E[X] = \mu_1, \mathrm{Var}(X) = \sigma_1^2$$

$$Y_1, \ldots, Y_{n_2} \sim iid Y, E[X] = \mu_2, \mathrm{Var}(X) = \sigma_2^2$$

Standard tests: Two samples

$$X_1, \ldots, X_{n_1} \sim iid X, E[X] = \mu_1, \text{Var}(X) = \sigma_1^2$$

$$Y_1, \ldots, Y_{n_2} \sim iid Y, E[X] = \mu_2, \text{Var}(X) = \sigma_2^2$$

- Testing to compare means
  - Null $H_0 : \mu_1 = \mu_2$
  - Alternative $H_A : \mu_1 \neq \mu_2$

Standard tests: Two samples

$$X_1, \ldots, X_{n_1} \sim iid X, E[X] = \mu_1, \mathrm{Var}(X) = \sigma_1^2$$

$$Y_1, \ldots, Y_{n_2} \sim iid Y, E[X] = \mu_2, \mathrm{Var}(X) = \sigma_2^2$$

- Testing to compare means
  - Null $H_0 : \mu_1 = \mu_2$
  - Alternative $H_A : \mu_1 \neq \mu_2$
- Testing to compare variances
  - Null $H_0 : \sigma_1 = \sigma_2$
  - Alternative $H_A : \sigma_1 \neq \sigma_2$

Goodness of fit testing

Samples: $X_1, \ldots, X_n$

Problem: Do the samples follow a certain distribution?

Goodness of fit testing

Samples: $X_1, \ldots, X_n \sim \text{iid } X$

Problem: Do the samples follow a certain distribution?

Examples

- Integer samples $X_i \in \{0, 1, 2, \ldots\}$. Is the distribution Poisson?

  H_0: $X \sim \text{Poisson}(\lambda)$

  H_A: X : not Poisson

Goodness of fit testing

Samples: $X_1, \ldots, X_n$

Problem: Do the samples follow a certain distribution?

Examples

- Integer samples $X_i \in \{0, 1, 2, \ldots\}$. Is the distribution Poisson?
- Continuous positive samples $X_i \in [0, \infty)$. Is the distribution Gamma?

Goodness of fit testing

Samples: $X_1, \ldots, X_n$

Problem: Do the samples follow a certain distribution?

Examples

- Integer samples $X_i \in \{0, 1, 2, \ldots\}$. Is the distribution Poisson?
- Continuous positive samples $X_i \in [0, \infty)$. Is the distribution Gamma?
- Continuous samples $X_i \in (-\infty, \infty)$. Is the distribution normal?

Goodness of fit testing

Samples: $X_1, \ldots, X_n$

Problem: Do the samples follow a certain distribution?

Examples

- Integer samples $X_i \in \{0, 1, 2, \ldots\}$. Is the distribution Poisson?
- Continuous positive samples $X_i \in [0, \infty)$. Is the distribution Gamma?
- Continuous samples $X_i \in (-\infty, \infty)$. Is the distribution normal?
- Multinomial $X_i \in \{1, 2, \ldots, M\}$. Is the distribution $\{f_1(\theta), \ldots, f_M(\theta)\}$?

$P(X_i = 1) \quad P(X_i = M)$

Ex: $\frac{2}{3}(1-\theta) \quad \ldots \quad \frac{1}{3}\theta$

Section 4

Answering questions using data

Questions

- In most cases, useful questions can be posed as the testing of a hypothesis. Several classes of testing problems arise.
- Breaking down the question so that it becomes a hypothesis test is an important *design* step

Questions

- In most cases, useful questions can be posed as the testing of a hypothesis. Several classes of testing problems arise.
- Breaking down the question so that it becomes a hypothesis test is an important *design* step

Examples

- A person claims magical powers in being able to predict something. How will you design a statistical hypothesis test?

Questions

- In most cases, useful questions can be posed as the testing of a hypothesis. Several classes of testing problems arise.
- Breaking down the question so that it becomes a hypothesis test is an important *design* step

Examples

- A person claims magical powers in being able to predict something. How will you design a statistical hypothesis test?
- A company claims a new treatment method for a disease. How will you test for the effectiveness of the treatment?

Questions

- In most cases, useful questions can be posed as the testing of a hypothesis. Several classes of testing problems arise.
- Breaking down the question so that it becomes a hypothesis test is an important *design* step

Examples

- A person claims magical powers in being able to predict something. How will you design a statistical hypothesis test?
- A company claims a new treatment method for a disease. How will you test for the effectiveness of the treatment?
- Data of accidental deaths in a country: Is there a seasonal or monthly pattern in this data?

Questions

- In most cases, useful questions can be posed as the testing of a hypothesis. Several classes of testing problems arise.
- Breaking down the question so that it becomes a hypothesis test is an important *design* step

Examples

- A person claims magical powers in being able to predict something. How will you design a statistical hypothesis test?
- A company claims a new treatment method for a disease. How will you test for the effectiveness of the treatment?
- Data of accidental deaths in a country: Is there a seasonal or monthly pattern in this data?
- Data on hiring by an organization: Is there any gender or geographical bias in the hiring?

Example 1: Magical powers

Suppose a person claims magical powers to predict the throw of a die. Here is one possible way to post it as a hypothesis testing problem.

Example 1: Magical powers

Suppose a person claims magical powers to predict the throw of a die. Here is one possible way to post it as a hypothesis testing problem.

- Throw the die n times and record the predictions. Let T be the number of correct predictions.
- Null $H_0 : T \sim \text{Binomial}(n, 1/6)$ No magical power; random guess.
- Alternative $H_A : T \sim$ any other distribution

Example 1: Magical powers

Suppose a person claims magical powers to predict the throw of a die. Here is one possible way to post it as a hypothesis testing problem.

- Throw the die n times and record the predictions. Let T be the number of correct predictions.
- Null $H_0 : T \sim \text{Binomial}(n, 1/6)$
- Alternative $H_A : T \sim$ any other distribution

In the above test, we need to measure or quantify the *confidence* of our conclusion and justify its *statistical significance*. The number of trials n will be an important factor to decide.

Example 2: New medical treatment

A company claims a new drug is effective in reducing heart attacks in a certain segment of the population. Here is a common way in which drugs are tested.

Example 2: New medical treatment

A company claims a new drug is effective in reducing heart attacks in a certain segment of the population. Here is a common way in which drugs are tested.

- n volunteers are chosen randomly for the trial from the population segment. About $n/2$ of them (Group I) are chosen randomly and given the drug, and the remaining (Group II) are given a placebo. Volunteers are not told what they got.
- Over a time period, the volunteers are observed for heart attacks. Suppose the fraction of volunteers who got a heart attack in Group I is f_1, and the same fraction in Group II is f_2
- Null $H_0 : f_1 \approx f_2$
- Alternative $H_A : \dfrac{f_2 - f_1}{f_2} = c$

Example 2: New medical treatment

A company claims a new drug is effective in reducing heart attacks in a certain segment of the population. Here is a common way in which drugs are tested.

- n volunteers are chosen randomly for the trial from the population segment. About $n/2$ of them (Group I) are chosen randomly and given the drug, and the remaining (Group II) are given a placebo. Volunteers are not told what they got.
- Over a time period, the volunteers are observed for heart attacks. Suppose the fraction of volunteers who got a heart attack in Group I is f_1, and the same fraction in Group II is f_2
- Null $H_0 : f_1 \approx f_2$
- Alternative $H_A : \dfrac{f_2 - f_1}{f_2} = c$

It is important to find c, and to quantify *confidence* and *statistical significance*. Once again, n will be an important factor.

Example 3: Pattern in accidental deaths

The number of accidental deaths in a country are tabulated every month over a year. Here is one way to test if there is a constant number of deaths per day, i.e. a constant rate.

Example 3: Pattern in accidental deaths

The number of accidental deaths in a country are tabulated every month over a year. Here is one way to test if there is a constant number of deaths per day, i.e. a constant rate.

- Estimate the overall rate, $\theta =$ Total deaths / Total number of days
- Estimated monthly deaths: $\{31\theta, 28\theta, 31\theta, \ldots, 31\theta\}$
- Null H_0 : Estimated deaths fits the observed deaths
- Alternative H_A: Estimated and observed deaths do not fit for some months

Example 3: Pattern in accidental deaths

The number of accidental deaths in a country are tabulated every month over a year. Here is one way to test if there is a constant number of deaths per day, i.e. a constant rate.

- Estimate the overall rate, $\theta =$ Total deaths / Total number of days
- Estimated monthly deaths: $\{31\theta, 28\theta, 31\theta, \ldots, 31\theta\}$
- Null H_0 : Estimated deaths fits the observed deaths
- Alternative H_A: Estimated and observed deaths do not fit for some months

Assessing goodness of fit is an important ingredient here. We need to quantify the *confidence* in the fit.

Example 4: Gender bias in hiring

Consider the following cross-tabulation of hires made by a company.

	Female	Male	Total
Hired	6	12	18
Not hired	9	25	34
Total	15	37	52

Is there a gender bias in the hiring? Here is an approach.

Example 4: Gender bias in hiring

Consider the following cross-tabulation of hires made by a company.

	Female	Male	Total
Hired	6	12	18
Not hired	9	25	34
Total	15	37	52

$T = 12 - 6 = 6$

Is there a gender bias in the hiring? Here is an approach.

- Pick 18 out of 52 uniformly at random, $T = M - F \rightarrow$ Distribution
- Null H_0 : Distribution of T is as given above
- Alternative H_A : Any other distribution

Example 4: Gender bias in hiring

Consider the following cross-tabulation of hires made by a company.

	Female	Male	Total
Hired	6	12	18
Not hired	9	25	34
Total	15	37	52

Is there a gender bias in the hiring? Here is an approach.

- Pick 18 out of 52 uniformly at random, $T = M - F$
- Null H_0 : Distribution of T is as given above
- Alternative H_A : Any other distribution

Is 6 a reasonable value for T? How to quantify this?

Observations

- In all examples, the question seems to be reasonably posed in a statistical hypothesis testing framework
- In most cases, the null and/or alternative are composite
- In all cases, the *confidence* of the testing is very important
- How do you quantify *confidence*?
  - We use ideas from confidence interval of estimation
  - A notion called P-value is used to quantify confidence

Section 5

Standard testing methods: z-test

General methodology of testing

$$X_1, \ldots, X_n \sim \text{iid } X$$

- *Test statistic*, denoted T
  - Some function of the samples
  - Example: Sample mean $\overline{X}$, Sample variance S^2 etc
- Acceptance and rejection regions are specified through T
  - Examples
    - Reject H_0 if $T > c$ (right)
    - Reject H_0 if $T < -c$ (left)
    - Reject H_0 if $|T| > c$ (two-sided)
- Significance level α depends on c and the distribution of $T|H_0$
  - Right-sided: $\alpha = P(T > c|H_0)$ (similar for others)
  - Fix α and find c

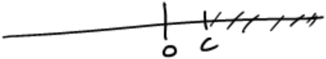
Testing for mean (normal samples, known variance)

$$X_1, \ldots, X_n \sim \text{iid } N(\mu, 4^2)$$

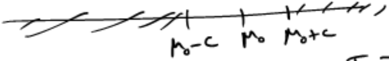
- Test statistic $T = \overline{X} \triangleq \frac{1}{n}(X_1 + \cdots + X_n)$
- Null $H_0 : \mu = 0$, Alternative $H_A : \mu > 0$
- Test: Reject H_0 if $T > c$

(Could be $\overset{H_0}{\mu=\mu_0}$ & $\overset{H_A}{\mu>\mu_0}$ (or) $\mu<\mu_0$ (or) $\mu\neq\mu_0$)

$H_0: \mu=0 \quad H_A: \mu>0$



$H_0: \mu=\mu_0$, $H_A: \mu\neq\mu_0$



Reject H_0: $T>\mu_0+c$ (or) $T<\mu_0-c$

Testing for mean (normal samples, known variance)

$$X_1, \ldots, X_n \sim \text{iid } N(\mu, 4^2)$$

- Test statistic $T = \overline{X} \triangleq \frac{1}{n}(X_1 + \cdots + X_n)$
- Null $H_0 : \mu = 0$, Alternative $H_A : \mu > 0$
- Test: Reject H_0 if $T > c$
- Different samplings, $n = 10$
  - [-6.9, 0.6, -0.6, -4.8, -1.9, -5.1, 7.5, 6.1, 0.5, 3.3], $T = -0.14$ Acc
  - [-1.8, -1.8, 4.1, 3.4, 1.9, 0.6, 1.7, -6.9, 0.3, -4.0], $T = -0.25$ Acc
  - [-5.8, 2.0, 2.5, 1.7, -2.8, 0.9, -0.4, 0.6, -8.5, -2.9], $T = -1.25$ Acc
  - [4.2, 14.2, 7.1, -5.1, -2.3, -3.9, -3.2, -0.9, -1.4, -6.4], $T = 0.23$ Acc
  - [1.0, 3.6, 5.9, -2.2, 2.3, 6.9, 1.7, 0.1, 6.3, 4.0], $T = 2.96$ Rej ?
  - [1.7, 3.9, -1.6, 3.8, 4.0, 1.9, -1.8, 10.3, 4.2, 4.6], $T = 3.10$ Rej
  - [9.4, 2.2, 13.8, 3.1, 6.3, 7.0, 5.8, 1.0, 7.6, 5.7], $T = 6.20$ Rej

1.9...?

Stronger

Testing for mean (normal samples, known variance)

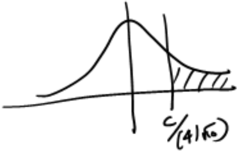
$$X_1, \ldots, X_n \sim \text{iid } N(\mu, 4^2)$$

- Test statistic $T = \overline{X} \triangleq \frac{1}{n}(X_1 + \cdots + X_n)$
- Null $H_0 : \mu = 0$, Alternative $H_A : \mu > 0$
- Test: Reject H_0 if $T > c$
- Different samplings, $n = 10$
  - [-6.9, 0.6, -0.6, -4.8, -1.9, -5.1, 7.5, 6.1, 0.5, 3.3], $T = -0.14$
  - [-1.8, -1.8, 4.1, 3.4, 1.9, 0.6, 1.7, -6.9, 0.3, -4.0], $T = -0.25$
  - [-5.8, 2.0, 2.5, 1.7, -2.8, 0.9, -0.4, 0.6, -8.5, -2.9], $T = -1.25$
  - [4.2, 14.2, 7.1, -5.1, -2.3, -3.9, -3.2, -0.9, -1.4, -6.4], $T = 0.23$
  - [1.0, 3.6, 5.9, -2.2, 2.3, 6.9, 1.7, 0.1, 6.3, 4.0], $T = 2.96$
  - [1.7, 3.9, -1.6, 3.8, 4.0, 1.9, -1.8, 10.3, 4.2, 4.6], $T = 3.10$
  - [9.4, 2.2, 13.8, 3.1, 6.3, 7.0, 5.8, 1.0, 7.6, 5.7], $T = 6.20$

Higher values of T give us more confidence in rejecting null

Testing for mean: Significance level

$$X_1, \ldots, X_{10} \sim \text{iid } N(\mu, 4^2)$$



- Significance level $\alpha = P(\overline{X} > c \mid \mu = 0)$
- Since $(\overline{X} \mid \mu = 0) \sim \text{Normal}(0, 4^2/10)$, we have

$Z \sim$ Standard normal

$\sim$ Normal$(0,1)$

$$\alpha = P\left(\frac{\overline{X}}{4/\sqrt{10}} > \frac{c}{4/\sqrt{10}}\right) = 1 - F_Z(\sqrt{10}c/4)$$

$\to \sqrt{n}$

CDF of Z

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001

depends on $n=10$

very common

Testing for mean: Significance level

$$X_1, \ldots, X_{10} \sim \text{iid } N(\mu, 4^2)$$

- Significance level $\alpha = P(\overline{X} > c | \mu = 0)$
- Since $(\overline{X} | \mu = 0) \sim \text{Normal}(0, 4^2/10)$, we have

critical value

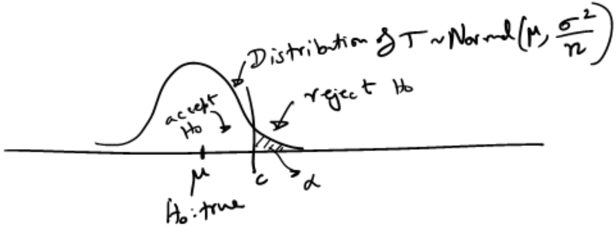
$$\alpha = P\left(\frac{\overline{X}}{4/\sqrt{10}} > \frac{c}{4/\sqrt{10}}\right) = 1 - F_Z(\sqrt{10}c/4)$$

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001

z-**test at significance level** α: Reject H_0 if $T > c$, where c is as above.

Testing for mean: Results and P-value

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001
$T = 0.23$	Rej	Acc	Acc	Acc	Acc	Acc
$T = 2.96$	Rej	Rej	Rej	Rej	Acc	Acc
$T = 6.20$	Rej	Rej	Rej	Rej	Rej	Rej

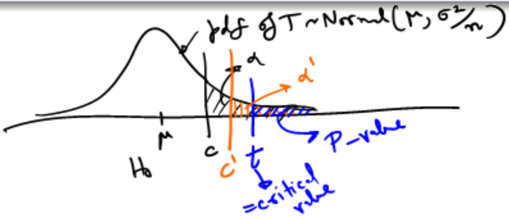


Testing for mean: Results and P-value

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001
$T = 0.23$	Rej	Acc	Acc	Acc	Acc	Acc
$T = 2.96$	Rej	Rej	Rej	Rej	Acc	Acc
$T = 6.20$	Rej	Rej	Rej	Rej	Rej	Rej

Definition (P-value)

Suppose the test statistic $T = t$ in one sampling. The lowest significance level α at which the null will be rejected for $T = t$ is said to be the P-value of the sampling.



$t > c \Rightarrow$ reject H_0 at α

$t > c' \Rightarrow$ reject H_0 at $\alpha' < \alpha$

critical value $= t \Rightarrow$ lowest significance level at which we reject H_0

Testing for mean: Results and P-value

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001
$T = 0.23$	Rej	Acc	Acc	Acc	Acc	Acc
$T = 2.96$	Rej	Rej	Rej	Rej	Acc	Acc
$T = 6.20$	Rej	Rej	Rej	Rej	Rej	Rej

Definition (P-value)

Suppose the test statistic $T = t$ in one sampling. The lowest significance level α at which the null will be rejected for $T = t$ is said to be the P-value of the sampling.

- Finding P-value for $T = t$: Put $c = t$ in computation of α

T	-0.14	0.23	2.96	6.20
P-value	0.544	0.428	0.00964	4.755e-07

If P-value is low enough, we reject H_0

What to choose? Significance level or P-value

- Samples are given, and there is some hypothesis that needs to be tested
- Step 1: Decide on the null and alternative hypotheses H_0 and H_A
- Step 2: Decide on the test statistic T
- Step 3: "Philosophy" of testing
  - Choice 1: Pick a significance level first
    - ★ Probability of Type I error can be fixed in some applications. In those cases, significance level is easy to fix
    - ★ Historically, in many applications, 0.05 or 0.01 is accepted as a common significance level
    - ★ Find rejection region (find the *critical value* c and reject H_0 if $T > c$, for example)
  - Choice 2: Use P-value
    - ★ Report the P-value
    - ★ If P-value is *low enough*, choose to reject H_0; otherwise, accept H_A
    - ★ How low is low enough? Depends on applications and other information

Section 6

z-test problems

Problem 1

Suppose $X \sim \text{Normal}(\mu, 9)$. For $n = 16$ iid samples of X, the observed sample mean is 10.2. What conclusion would a z-test reach if the null hypothesis assumes $\mu = 9.5$ (against an alternative hypothesis $\mu > 9.5$) at a signifiance level of $\alpha = 0.05$? What if the null hypothesis assumes $\mu = 8.5$ (against an alternative hypothesis $\mu > 8.5$)?

$H_0: \mu = 9.5$, $H_A: \mu > 9.5$, $\bar{X} \sim N(\mu, 9/16)$, $\bar{X} = 10.2$

$\sigma^2 = 9/16$

Test: Reject H_0 if $\bar{X} > c$

$$\alpha = P(\bar{X} > c \mid \mu = 9.5) = 1 - F_Z\left(\frac{c - 9.5}{(3/4)}\right) = 0.05$$

$$\frac{\bar{X} - 9.5}{\underbrace{(3/4)}_{\sigma}} \sim Z$$

$$c = 9.5 + \frac{3}{4} F_Z^{-1}(0.95) = 10.73$$

z-test @ $\alpha = 0.05$: Accept H_0

$H_0: \mu = 8.5$, $H_A: \mu > 8.5$

$$c = 8.5 + \frac{3}{4} F_Z^{-1}(0.95) = 9.73$$

z-test @ $\alpha = 0.05$: Reject H_0.

Problem 2

Suppose an app is desired to make an accurate identification of faces in photographs more than 90% of the time in the long run. For a random sample of 500 such photos, the app makes the correct identification 462 times - a 92.4% success rate. What does a z-test say about a null hypothesis that the app is only 90% accurate (compared to an alternative hypothesis that the app is more than 90% accurate with a significance level of $\alpha = 0.05$)?

Samples: $X_1, \ldots, X_{500} \overset{iid}{\sim} \text{Bernoulli}(p)$ $\quad X_i = \begin{cases} 1 & \text{if app correctly identifies} \\ 0 & \text{else} \end{cases}$

$$\bar{X} = \frac{462}{500} = 0.924$$

$H_0: p = 0.9$, $H_A: p > 0.9$ $\quad$ Test: Reject H_0 if $\bar{X} > c$

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{500}\right) \quad \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{500}}} \sim Z$$

$$\alpha = P(\bar{X} > c \mid p = 0.9) = 1 - F_Z\left(\frac{c - 0.9}{\sqrt{0.09/500}}\right) = 0.05$$

$$c = 0.9 + \sqrt{\frac{0.09}{500}}\, F_Z^{-1}(0.95) = 0.922$$

Since $\bar{X} > c$, z-test @ $\alpha = 0.05$: Reject H_0

Problem 3

Suppose $X \sim \text{Normal}(\mu, 36)$. For $n = 25$ iid samples of X, the observed sample mean is 6.2. What conclusion would a z-test reach if the null hypothesis assumes $\mu = 4$ (against an alternative hypothesis $\mu \neq 4$) at a signifiance level of $\alpha = 0.05$? What if the null hypothesis assumes $\mu = 8$ (against an alternative hypothesis $\mu < 8$)?

$H_0: \mu = 4$, $H_A: \mu \neq 4$, Test: Reject H_0, if $|\bar{X} - 4| > c$

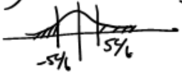
$\Updownarrow$

$\bar{X} > 4 + c$ (or) $\bar{X} < 4 - c$

$\bar{X} \sim N\left(\mu, \frac{36}{25}\right)$ $\quad \frac{\bar{X} - \mu}{(6/5)} \sim Z$

$\alpha = P(|\bar{X} - 4| > c \mid \mu = 4) = P\left(\left|\underbrace{\frac{\bar{X}-4}{6/5}}_{\sim |Z|}\right| > \frac{c}{(6/5)}\right) = P\left(|Z| > \frac{c}{6/5}\right) = 2 F_Z\left(\frac{-5c}{6}\right)$

$c = -\frac{6}{5} F_Z^{-1}\left(\frac{0.05}{2}\right) = 2.352$



Since $|6.2 - 4| = 2.2 < 2.352$, Accept H_0.

$H_0: \mu = 8$, $H_A: \mu < 8$, Test: Reject H_0 if $\bar{X} < c$

$\alpha = P(\bar{X} < c \mid \mu = 8) = F_Z\left(\frac{c - 8}{(6/5)}\right) = 0.05 \Rightarrow c = 8 + \frac{6}{5} F_Z^{-1}(0.05) = 6.02$

Since $\bar{X} = 6.2 > 6.02$, z-test @ $\alpha = 0.05$: Accept H_0

Section 7

More problems on z-test

Problem 1a (with binomial distribution)

Vaccine hesitancy (percentage of people who are unwilling to vaccinate) in a town has been reported to be 20%. To test whether the fraction is 20%, you call a randomly selected group of 10 people and find out that 3 of them are vaccine hesitant. What is the null hypothesis? Will you accept or reject null at a significance level of $\alpha = 0.05$ against an alternative that the fraction is above 20%? What is power against an alternative that the fraction is 30%?

Samples: $X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} \text{Bernoulli}(p)$ $\quad X_i = \begin{cases} 1 & \text{if } i\text{-th person is hesitant} \\ 0 & \text{else} \end{cases}$

$T = X_1 + \cdots + X_{10}$

$H_0: p = 0.2$, $H_A: p > 0.2$

Test: Reject H_0 if $T > c$

$T \sim \text{Binomial}(10, p)$ $\quad \alpha = P(T > c \mid p = 0.2)$

$$0.05 = 1 - \sum_{k=0}^{c} \binom{10}{k} p^k (1-p)^{10-k}$$

$\Rightarrow c = 4$

c	α
0	0.893
1	0.624
2	0.322
3	0.121
4	0.033
5	0.0064
6	0.00086

Type II error:

$\beta = P(\text{Accept } H_0 \mid p = 0.3) = P(T \le c \mid p = 0.3) = \underline{0.85}$

Problem 1b (with normal approximation)

Vaccine hesitancy (percentage of people who are unwilling to vaccinate) in a town has been reported to be 20%. To test whether the fraction is 20%, you call a randomly selected group of 100 people and find out that 28 of them are vaccine hesitant. What is the null hypothesis? Will you accept or reject null at a significance level of $\alpha = 0.05$ against an alternative that the fraction is above 20%? What is power against an alternative that the fraction is 30%?

$$T = \text{Binomial}(100, p) \approx \text{Normal}(100p, 100p(1-p))$$

$$\alpha = P(T > c \mid p = 0.2) = 1 - F_Z\left(\frac{c - 100 \times 0.2}{\underbrace{\sqrt{100 \times 0.2 \times 0.8}}_{4}}\right) = 0.05$$

$$c = 20 + 4 F_Z^{-1}(0.95) = 26.58$$

Since $T = 28 > c = 26.58$, Z-test @ $\alpha = 0.05$: Reject H_0

$$\beta = P(T \le c \mid p = 0.3) = F_Z\left(\frac{26.58 - 100 \times 0.3}{\sqrt{100 \times 0.3 \times 0.7}}\right) = 0.23$$

Problem 2a

The current-carrying capacity of a resistor manufactured at your company is supposed to be 3.0 A (A stands for Amperes). Because of a recent change in the manufacturing process, you suspect that the current-carrying capacity might actually be lesser than 3.0 A. You decide to test by measuring current-carrying capacities of 10 resistors with a test measurement has a standard deviation of 0.05 A. If the sample mean of the test measurements $T_{10} < 2.95$, you will conclude that the manufacturing process is faulty.

1. What is the null hypothesis? What is the alternative hypothesis? What are the samples?
2. What is the significance level α of the test?
3. If the current-carrying capacity falls to 2.9 A, there could be serious safety issues. Against the alternative hypothesis of 2.9 A, what is the power $(1 - \beta)$ of the test?

Samples: $X_1, X_2, \ldots, X_{10} \sim N(\mu, 0.05^2)$ X_i = measured capacity of i-th test resistor

$H_0: \mu = 3$ A, $H_A: \mu < 3$ A.

Working

Test: Reject H_0 if $\bar{X} < 2.95$

$\bar{X} \sim N\left(\mu, \frac{0.05^2}{10}\right)$

$$\alpha = P(\bar{X} < 2.95 \mid \mu = 3) = F_Z\left(\frac{2.95-3}{0.05/\sqrt{10}}\right) = 0.00078\ldots$$

$$\beta = P(\bar{X} \geq 2.95 \mid \mu = 2.9) = 1 - F_Z\left(\frac{2.95-2.9}{0.05/\sqrt{10}}\right) = 0.00078\ldots$$

Problem 2b

In the same problem, suppose you can test n resistors. If the sample mean $T_n < c$, you conclude that the manufacturing process is faulty. You need to determine suitable values for n and c under the following conditions:

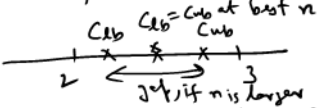
1. Significance level or probability of Type I error, $\alpha \leq 10^{-6}$
2. Probability of Type II error, $\beta \leq 10^{-12}$ (against an alternative of 2.9 A)

$\bar{X} \sim N\left(\mu, \frac{0.05^2}{n}\right)$ $\quad \alpha = P(\bar{X} < c \mid \mu = 3) = F_Z\left(\frac{c-3}{0.05/\sqrt{n}}\right) \leq 10^{-6}$

$$\frac{c-3}{0.05/\sqrt{n}} \leq F_Z^{-1}(10^{-6})$$

$$c \leq \overbrace{3 + \frac{0.05}{\sqrt{n}}(-4.753)}^{c_{ub}} \qquad \text{—— (1)}$$

$$\beta = P(\bar{X} \geq c \mid \mu = 2.9) = 1 - F_Z\left(\frac{c-2.9}{0.05/\sqrt{n}}\right) \leq 10^{-12} \Rightarrow F_Z\left(\frac{c-2.9}{0.05/\sqrt{n}}\right) \geq 1 - 10^{-12}$$



$$\frac{c-2.9}{0.05/\sqrt{n}} \geq 7.034$$

$$c \geq \underbrace{2.9 + \frac{0.05}{\sqrt{n}}(7.034)}_{c_{lb}} \qquad \text{—— (2)}$$

Best choice of n: $\quad 0.1 = \frac{0.05}{\sqrt{n}}(4.753 + 7.034) \Rightarrow n = 35$

Problem 3

The average CGPA of students in a college is reported to be 8.0 with a standard deviation of 1. You suspect that the average may be lower, possibly 7.5, and decide to sample students to find their CGPA. What sample size do you need for a test at a significance level of 0.05 and power of 0.95? How will the sample size change if you suspect the CGPA to be 7.0?

$X_1, \ldots, X_n \sim N(\mu, 1)$ $\quad H_0 : \mu = 8 \quad H_A : \mu < 8$

$\bar{X} \sim N(\mu, 1/n)$ $\quad$ Test: Reject H_0 if $\bar{X} < c$

$$\alpha = P(\bar{X} < c \mid \mu = 8) = F_Z\left(\frac{c-8}{1/\sqrt{n}}\right) = 0.05$$

$$c = 8 - \frac{1.645}{\sqrt{n}} \quad \text{——} ①$$

Case 1: $$\beta = P(\bar{X} \geq c \mid \mu = 7.5) = 1 - F_Z\left(\frac{c-7.5}{1/\sqrt{n}}\right) = 0.05$$

$$c = 7.5 + \frac{1.645}{\sqrt{n}} \quad \text{——} ②$$

$$① \,\&\, ② : \quad 0.5 = \frac{2 \times 1.645}{\sqrt{n}} \Rightarrow n \approx 44$$

Case 2: $$\beta = 1 - F_Z\left(\frac{c-7}{1/\sqrt{n}}\right) = 0.05 \Rightarrow c = 7 + \frac{1.645}{\sqrt{n}} \quad \text{——} ③$$

$$① \,\&\, ③ : \quad 1 = \frac{2 \times 1.645}{\sqrt{n}} \Rightarrow n \approx 11$$