Purdue-NCKU program

Lecture 3
Hypothesis Testing

Dr. Qifan Song

Beyond quantitative inferences

- Point/Interval estimations give a precise numerical argument about the parameters
- In many cases, instead of knowing the exact values, we want to know the trend, especially in a preliminary study.
- Is it better? vs how much better?
- A researcher thinks that if knee surgery patients go to physical therapy twice a week (instead of 3 times), their recovery period will be longer. Average recovery times for knee surgery patients (if they go to therapy 3 times a week) is 8.2 weeks.
- (mean of recovery times $\leq$ 8.2 weeks) versus (mean of recovery times $>$ 8.2 weeks)

Hypothesis

Hypothesis: A hypothesis is a statement about the *true distribution* or eqivalently, a statement about the *true parameter*.

Math form of a hypothesis: $\theta \in \Theta_0$ where $\Theta_0 \subset \Theta$.
Example

- Normal(μ, σ^2) modeling. The mean of the distribution is greater than 2: $\Theta_0 = (2, \infty) \otimes (0, \infty)$
- Bernoulli(p) modeling. The variance of the distribution is smaller or equal to 0.04: $\Theta_0 = \{p : p(1-p) \leq 0.04, 0 \leq p \leq 1\}$
- Exponential(λ) modeling. The probability of the distribution being greater than 10 is smaller than 0.01: $\Theta_0 = \{\lambda : \exp(-10\lambda) < 0.01\}$

Hypothesis Testing

Given a data set, we decide whehter $\theta \in \Theta_0$ or not?

Let $\Theta_1 = \Theta_0^c$, then it is equivalent to $\theta \in \Theta_0$ versus $\theta \in \Theta_1$.

Hypothesis Testing: Null vs Alternative Hypothesis

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

We need to design a decision making process (accept H_0 or accept H_1) based on the observations.

Reject Region

Decision making process can be view as a mapping from data to $\{0, 1\}$, i.e. $\psi : \mathcal{X}^n \to \{0, 1\}$

- Reject Region, a subset of $\mathcal{X}^n$, $\mathcal{R} = \{\text{data} : \psi(\text{data}) = 1\}$. All the possible data values that lead to the acceptance of H_1.

- Example: if $\mathcal{R} = R^n$, then we always accept H_1.

- There are, of course, infinite choices of $\mathcal{R}$. The question will be, how to evaluate a given $\mathcal{R}$?

- A straightforward way is to examine whether $\mathcal{R}$ can give you a *correct decision*

Power function

- Given any $\theta \in \Theta$, we define power function $\beta(\theta) := Pr(\psi(\text{data}) = 1) = Pr((X_1, \ldots, X_n) \in \mathcal{R})$ $= \int_{\mathcal{R}} \prod_{i=1}^{n} f_\theta(x_i) dx_1 \ldots dx_n$

- The chance to make a correction decision should be high, thus

i When $\theta \in \Theta_0$, we want a small $\beta(\theta)$. Small chance of *Type I error*

ii When $\theta \in \Theta_1$, we want a large $\beta(\theta)$. Small chance of *Type II error*

Trade-off

- small $\beta(\theta)$ for $\theta \in \Theta_0$ implicitly wants a small set $\mathcal{R}$
- big $\beta(\theta)$ for $\theta \in \Theta_1$ implicitly wants a large set $\mathcal{R}$
- There is a trade-off between two goals and we need a strategy to make the balance

- The common strategy of statistical hypothesis testing
  - For any $\theta \in \Theta_0$, $\beta(\theta) \leq \alpha$ for some fixed small α i.e., $\max_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.
  - While the probability of committing type I error bounded, we try to minimize the probability of type II error.

Meaning of small α

- $Pr((X_1 \ldots, X_n) \in \mathcal{R})) \leq \alpha$ means that $\mathcal{R}$ represents the set of *rare* or *extreme* cases under $\theta \in \Theta_0$
- We reject H_0, only when the data we observed is a rare case for $\theta \in \Theta_0$. That is, there looks like a strong contradiction between observations and null hypothesis.
- Small α means our strategy is: we are reluctant to reject H_0 unless data are not compatible with null hypothesis
- Alternative interpretation: H_0 is our prior belief, if unnecessary, we will continue believing in it.
- In practice, we put default or previous knowledge as null hypothesis.

Meaning of large β values over Θ_1

A good test tries maximize $\beta(\theta)$ over Θ_1. If we indeed make it, then

- $Pr((X_1 \ldots, X_n) \in \mathcal{R}))$ is non-small means that $\mathcal{R}$ represents the set of *possible or common* cases under $\theta \in \Theta_1$
- When we reject H_0, the data looks like a regular case for $\theta \in \Theta_1$. That is, data are compatible with alternative hypothesis

In conclusion, a good test rejects H_0 when data are clearly not compatible with null hypothesis, but reasonably compatible with alternative hypothesis

Examples

- A bad test:

  We want to test the biological sex of a person, male vs female.

  reject region: the person has natural green hair.

- A good test:

  A fair criminal adjudication

  A presumption of innocence, or the suspect is innocent until proven guilty.

How to find a good $\mathcal{R}$?

- Intuitively, we can examine the density $\prod_{i=1}^{n} f_\theta(x_i)$. $\mathcal{R}$ should somehow include $(x_1, \ldots, x_n)$'s that have high density under alternative but low density under null.
- This choice will lead to the optimal test (largest β) under setting $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$ (Neyman–Pearson lemma)
- It is not convenience to work on the n-dimensional space. (For example, n-dim integral is needed to justify α requirement.) Therefore, instead of working on original data, we work on summary statistics.

Test Statistic

A summary statistic $T(X_1, \ldots, X_n) \in R$, such that we define reject region as $\mathcal{R} = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) \in \mathcal{C}\}$ for some set $\mathcal{C} \subset R$

- Let g_θ be the density of the test statistic T, then

  $\beta(\theta) = \int_{\mathcal{C}} g_\theta(t)dt$

- We want a set $\mathcal{C}$, such that
  - when $\theta \in \Theta_0$, $\int_{\mathcal{C}^c} g_\theta(t)dt \geq 1 - \alpha$, i.e., $\mathcal{C}^c$ is a high density region of T
  - when $\theta \in \Theta_1$, $\int_{\mathcal{C}} g_\theta(t)dt$ is large, i.e., $\mathcal{C}$ is a high density region of T

- A good T has different behavior under null and under alternative hypotheses.

Designing test statistics

1. For simplicity, we want a T such that,

   (i) if $\theta \in \Theta_0$, T tends to be small; if $\theta \in \Theta_1$, T tends to be large. Then $\mathcal{C} = (c, \infty)$

   (ii) if $\theta \in \Theta_0$, T tends to around some fixed value; if $\theta \in \Theta_1$, T tends to be larger or smaller than that fixed value. Then $\mathcal{C} = [c_1, c_2]^c$

2. In order to fulfill $\max_{\theta \in \Theta_0} \int_{\mathcal{C}} g_\theta(t)dt \leq \alpha$, T must have a tractable distribution under null hypothesis. (We can borrow some idea from pivotal quantity)

p-value

Definition: Given a data set $x_1, \ldots, x_n$ and observed test statistic value $t = T(x_1, \ldots, x_n)$,

p-value $= \max_{\theta \in \Theta_0} Pr(T$ is more or equally rare than $t|\theta$ is true parameter$)$

- We need to define a region $\mathcal{C}_t$ of "more rare than t"

$$p\text{-value} = \max_{\theta \in \Theta_0} \int_{\mathcal{C}_t} g_\theta(t)dt$$

- Let $\mathcal{C}_\alpha$ be the rejection region under level α, then

$$\alpha = \max_{\theta \in \Theta_0} \int_{\mathcal{C}_\alpha} g_\theta(t)dt$$

- We match $\mathcal{C}_t$ with $\mathcal{C}_\alpha$, i.e. define $\mathcal{C}_t$ as $\mathcal{C}_\alpha$ for some α such that $\mathcal{C}_\alpha$ barely contains t (i.e., t is on the boundary of $\mathcal{C}_\alpha$).

- p-value $< \alpha \Leftrightarrow t$ is inside $\mathcal{C}_\alpha \Leftrightarrow$ Reject H_0

One side z test

Observe $X_1, \ldots, X_n$. Assume they come from a $norm(\mu, \sigma_0^2)$ with known σ_0^2 and unknown μ.

$\Theta_0 = (-\infty, \mu_0]$ vs $\Theta_1 = (\mu_0, \infty)$

- Test statistics: $\bar{X} - \mu_0$ (or $\sqrt{n}(\bar{X} - \mu_0)/\sigma_0$)
- T tends to small under null, and tends to be large under alternative.
- $\mathcal{C} = (z_{\text{critical}}, \infty)$
- $\bar{X} - \mu_0 \sim N(\mu - \mu_0, \sigma_0^2/n)$
- $\alpha = \max_{\mu \leq \mu_0} Pr(N(\mu - \mu_0, \sigma_0^2/n) > z_{\text{critical}})$

- $\alpha = \max_{\mu \le \mu_0} Pr(N(0, \sigma_0^2/n) > z_{\text{critical}} - \mu + \mu_0)$

- The maximum occurs when μ takes its largest possible value, i.e., $\alpha = Pr(N(0, \sigma_0^2/n) > z_{\text{critical}})$

- $z_{\text{critical}} = \sigma_0 z_{1-\alpha}/\sqrt{n}$

- Rejection region: $\bar{X} - \mu_0 > \sigma_0 z_{1-\alpha}/\sqrt{n}$ or $\sqrt{n}(\bar{X} - \mu_0)/\sigma_0 > z_{1-\alpha}$

- p-value $= \max_{\mu \le \mu_0} Pr(\bar{X} - \mu_0 \ge \bar{x} - \mu_0) = Pr(N(0, \sigma_0^2/n) \ge \bar{x} - \mu_0)$

  $= Pr(N(0, 1^2) \ge \sqrt{n}(\bar{x} - \mu_0)/\sigma_0)$

Two side z test

Observe $X_1, \ldots, X_n$. Assume they come from a $norm(\mu, \sigma_0^2)$ with known σ_0^2 and unknown μ.

$\Theta_0 = \{\mu_0\}$ vs $\Theta_1 = (-\infty, \mu_0) \cup (\mu_0, \infty)$

- Test statistics: $\bar{X} - \mu_0$ (or $\sqrt{n}(\bar{X} - \mu_0)/\sigma_0$)
- T tends to be around 0 under null, and tends to away from 0 under alternative.
- $\mathcal{C} = [z_{\text{crit1}}, z_{\text{crit2}}]^c$
- $\bar{X} - \mu_0 \sim N(\mu - \mu_0, \sigma_0^2/n)$
- $\alpha = Pr(N(0, \sigma_0^2/n) > z_{\text{crit2}}) + Pr(N(0, \sigma_0^2/n) < z_{\text{crit1}})$

- A convenient choice $\alpha/2 = Pr(N(0, \sigma_0^2/n) > z_{\text{crit2}})$
  $\alpha/2 = Pr(N(0, \sigma_0^2/n) < z_{\text{crit1}})$

- $z_{\text{crit1}} = \sigma_0 z_{\alpha/2}/\sqrt{n}$ and $z_{\text{crit2}} = \sigma_0 z_{1-\alpha/2}/\sqrt{n}$

- Rejection region: $|\bar{X} - \mu_0| > \sigma_0 z_{1-\alpha/2}/\sqrt{n}$ or $\sqrt{n}|\bar{X} - \mu_0|/\sigma_0 > z_{1-\alpha/2}$

- $\mathcal{C}_\alpha$ is of the form $[-a, a]^c$

- p-value $= Pr(\bar{X} - \mu_0 \geq |\bar{x} - \mu_0|) + Pr(\bar{X} - \mu_0 \leq -|\bar{x} - \mu_0|)$

  $= 2Pr(N(0, 1^2) \geq \sqrt{n}(\bar{x} - \mu_0)/\sigma_0)$

Two side test and confidence interval

A general two side test $\Theta_0 = \{\theta_0\}$ vs $\Theta_1 = (-\infty, \theta_0) \cup (\theta_0, \infty)$

- If null hypothesis is true, $Pr(\theta_0 \notin \text{C.I.}) \leq \alpha$
- Denote $\mathcal{R} = \{(x_1, \ldots, x_n)$: which yields a C.I. that contains $\theta_0\}$
- $Pr(\theta_0 \notin \text{C.I.} | H_0 \text{ is true}) \leq \alpha \Leftrightarrow Pr((X_1, \ldots, X_n) \in \mathcal{R} | H_0 \text{ is true}) \leq \alpha$
- $\mathcal{R}$ serves a valid reject region, although we have no guarantee that is power function is large over Θ_1.

- If we have a good C.I., then we can reject null value if the null value is inside the C.I.

One side t test

Observe $X_1, \ldots, X_n$. Assume they come from a $norm(\mu, \sigma^2)$ with unknown σ^2 and unknown μ.

$\Theta_0 = \{\mu \leq \mu_0\}$ vs $\Theta_1 = \{\mu > \mu_0\}$

- Test statistics: $\bar{X} - \mu_0$
- however, its distribution is not tractable due to unknown σ^2

- Alternative choice $\sqrt{n}(\bar{X} - \mu_0)/\sqrt{S^2}$
- it tends to small under null, and tends to be large under alternative.
- $\mathcal{C} = (t_{\text{critical}}, \infty)$

- rewrite $X_i = \mu + \sigma Z_i$ where Z_i are iid standard normal r.v.'s, then the test statistic can be represented as $\sqrt{n}(\sigma\bar{Z} + \mu - \mu_0)/\sqrt{\sigma^2 S_Z^2}$, where S_Z^2 denotes the sample variance of Z_i's

- $\alpha = \max_{\mu \le \mu_0} Pr(\sqrt{n}(\sigma\bar{Z} + \mu - \mu_0)/\sqrt{\sigma^2 S_Z^2} > t_{\text{critical}})$

- The maximum occurs when μ takes its largest possible value, i.e., $\alpha = Pr(\sqrt{n}\bar{Z}/\sqrt{S_Z^2} > t_{\text{critical}})$

- We can show that $\sqrt{n}\bar{Z}/\sqrt{S_Z^2}$ follows a t_{n-1} distribution and $t_{\text{critical}} = t_{n-1,1-\alpha}$

- Rejection region: $\sqrt{n}(\bar{X} - \mu_0)/S > t_{n-1,1-\alpha}$

- p-value $= \max_{\mu \le \mu_0} Pr(\sqrt{n}(\bar{X} - \mu_0)/S \ge \sqrt{n}(\bar{x} - \mu_0)/s) = Pr(t_{n-1} \ge \sqrt{n}(\bar{x} - \mu_0)/s)$

- Two-side t-test can be derived similarly.

two side χ^2 test

Observe $X_1, \ldots, X_n$. Assume they come from a $norm(\mu, \sigma^2)$ with unknown σ^2 and unknown μ.

$\Theta_0 = \{\sigma^2 = \sigma_0^2\}$ vs $\Theta_1 = \{\sigma^2 \neq \sigma_0^2\}$

- Test statistics: S^2
- Because S^2 is consistent, T tends to be around σ_0^2 under null, and tends to away from σ_0^2 under alternative
- $\mathcal{C} = [\chi^2_{\text{crit1}}, \chi^2_{\text{crit2}}]^c$
- $S^2 \sim \sigma^2 \chi^2_{n-1}/(n-1)$
- $\alpha = Pr(\sigma_0^2 \chi^2_{n-1}/(n-1) > \chi^2_{\text{crit2}}) + Pr(\sigma_0^2 \chi^2_{n-1}/(n-1) < \chi^2_{\text{crit1}})$

- A convenient choice $\alpha/2 = Pr(\sigma_0^2\chi^2_{n-1}/(n-1) > \chi^2_{\text{crit2}})$ $\alpha/2 = Pr(\sigma_0^2\chi^2_{n-1}/(n-1) < \chi^2_{\text{crit1}})$

- $\chi^2_{\text{crit1}} = \sigma_0^2\chi^2_{n-1,\alpha/2}/(n-1)$ and $\chi^2_{\text{crit2}} = \sigma_0^2\chi^2_{n-1,1-\alpha/2}/(n-1)$

- Non-symmetric rejection region: $(n-1)S^2/\sigma_0^2 < \chi^2_{n-1,\alpha/2}$ or $> \chi^2_{n-1,1-\alpha/2}$

- p-value $= 2 * Pr(\chi^2_{n-1} \geq (n-1)s^2/\sigma_0^2)$ or $2 * Pr(\chi^2_{n-1} \leq (n-1)s^2/\sigma_0^2)$ depending which one is smaller.

- One-side version can be derived similarly.

Two sample t test

Observe $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$ Assume they come from $norm(\mu_1, \sigma^2)$ and $norm(\mu_2, \sigma^2)$ respectively, with unknown means and variances.

$\Theta_0 = \{\mu_1 = \mu_2\}$ vs $\Theta_1 = \{\mu_1 \neq \mu_2\}$

- Test statistics: $\bar{X} - \bar{Y}$
- studentize it:

$$\frac{(\bar{X} - \bar{Y})/\sqrt{1/n + 1/m}}{\sqrt{[(n-1)S_X^2 + (m-1)S_Y^2]/(n+m-2)}}$$

- it tends to be around 0, and tends to away from 0 under alternative.
- $\mathcal{C} = [t_{\text{crit1}}, t_{\text{crit2}}]^c$

- Under null hypothesis, the studentized statistic follows a t_{n+m-2} distribution.
- $\alpha = Pr(t_{n+m-2} > t_{\text{crit2}}) + Pr(t_{n+m-2} < t_{\text{crit1}})$
- $t_{\text{crit1}} = t_{n+m-2,\alpha/2}$ and $t_{\text{crit2}} = t_{n+m-2,1-\alpha/2}$
- Rejection region: t-statistic $> t_{n-1,1-\alpha/2}$

- p-value $= Pr(t_{n+m-2} \geq t$-statistics of the observations)
- Two side t-test can be derived similarly.

Two sample F test

Observe $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$ Assume they come from $norm(\mu_1, \sigma_1^2)$ and $norm(\mu_2, \sigma_2^2)$ respectively, with unknown means and variances.

$\Theta_0 = \{\sigma_1^2 \leq \sigma_2^2\}$ vs $\Theta_1 = \{\sigma_1^2 > \sigma_2^2\}$

- Test statistics: S_X^2 / S_Y^2
- it tends to be smaller than 1, and tends to be larger than 1 under alternative.
- $\mathcal{C} = (F_{\text{crit1}}, \infty)$
- $(S_X^2/\sigma_1^2)/(S_Y^2/\sigma_2^2) \sim F_{n-1,m-1}$

- $\alpha = \max_{\Theta_0} Pr(S_X^2/S_Y^2 > F_{\text{critical}}) = \max_{\Theta_0} Pr((\sigma_1^2/\sigma_2^2)F_{n-1,m-1} > F_{\text{critical}})$

- The maximum occurs when (σ_1^2/σ_2^2) takes its largest possible value, i.e., $\alpha = Pr(F_{n-1,m-1} > F_{\text{critical}})$

- $F_{\text{critical}} = F_{n-1,m-1,1-\alpha}$

- Rejection region: $S_X^2/S_Y^2 > F_{n-1,m-1,1-\alpha}$

- p-value $= \max_{\Theta_0} Pr(S_X^2/S_Y^2 \geq s_X^2/s_Y^2) = Pr(F_{n-1,m-1} \geq s_X^2/s_Y^2)$

- Two-side F-test can be derived similarly.

Approximate Testing

In many complicated problem, it is difficult to find an exactly tractable test statistic. Thus, certain approximation can be used, e.g., via CLT type approximation

Approximate z-test

Observe $X_1, \ldots, X_n \in \{0, 1\}$. Assume they come from a Bernoulli(p) distribution

$\Theta_0 = \{p \leq p_0\}$ vs $\Theta_1 = \{p > p_0\}$

- test statistic: $\bar{X}$
- T tends to small under null, and tends to be large under alternative.
- $\mathcal{C} = (z_{\text{critical}}, \infty)$

- $\bar{X} \sim Bin(n,p)/n$
- $\alpha = \max_{p \le p_0} Pr(Bin(n,p) > nz_{\text{critical}})$
- The maximum occurs when p takes its largest possible value (why?), i.e., $\alpha = Pr(Bin(n,p_0) > nz_{\text{critical}})$
- $Bin(n,p_0) \approx N(np_0, np_0(1-p_0))$ by CLT, thus $z_{\text{critical}} = p_0 + z_{1-\alpha}\sqrt{p_0(1-p_0)/n}$
- Rejection region: $\sqrt{n}(\bar{X} - p_0)/\sqrt{p_0(1-p_0)} > z_{1-\alpha}$

- p-value $= \max_{p \le p_0} Pr(\bar{X} \ge \bar{x}) = Pr(N(np_0, np_0(1-p_0)) \ge n\bar{x})$

  $= Pr(N(0,1^2) \ge \sqrt{n}(\bar{x} - p_0)/\sqrt{p_0(1-p_0)})$

Wald Test

Observe $X_1, \ldots, X_n$. Assume they come from a distribution f_θ with unknown θ.

$\Theta_0 = \{\theta = \theta_0\}$ vs $\Theta_1 = \{\theta \neq \theta_0\}$

- Let $\widehat{\theta}$ be the MLE estimation
- $\sqrt{n}(\widehat{\theta} - \theta^*) \approx N(0, \tau^2(\theta^*))$

- Test statistic: $\sqrt{n}(\widehat{\theta} - \theta_0)/\tau(\theta_0)$
- $\mathcal{C} = [z_{\text{crit1}}, z_{\text{crit2}}]^c$ with $z_{\text{crit1}} = z_{\alpha/2}$ and $z_{\text{crit2}} = z_{1-\alpha/2}$
- Rejection region: $\sqrt{n}|\widehat{\theta} - \theta_0|/\tau(\theta_0)| > z_{1-\alpha/2}$

- p-value $= 2Pr(N(0, 1^2) \geq \sqrt{n}(\widehat{\theta} - \theta_0)/\tau(\theta_0))$

The goodness of fit χ^2 test

Assume that $Pr(\text{Outcome } i) = p_i$, $i = 1, ..., k$ with unknown $p_i \geq 0$ and $\sum p_i = 1$.

We observe n experiments, and the ith outcome occurs O_i times.

$\Theta_0 = \{p_i = p_{i,0} \text{ for all } i\}$, $\Theta_1 = \Theta_0^c$

- Test statistic $\sum_{i=1}^{k}[(O_i - E_i)^2/E_i]$ where $E_i = np_{i,0}$
- T tends to small under null (since $O_i \approx E_i$), and tends to be large under alternative.
- $\mathcal{C} = (\chi^2_{\text{critical}}, \infty)$
- Under H_0, $T \approx \chi^2_{k-1}$. Thus reject region is $\chi^2_{\text{critical}} = \chi^2_{k-1,1-\alpha}$
- p-value $= Pr(\chi^2_{k-1} \geq \text{observed test statistic})$

The likelihood ratio test

Observe $X_1, \ldots, X_n$. Assume they come from a distribution f_θ with unknown θ.

- Test statistics $-2\log(\max_{\theta\in\Theta_0}\prod f_\theta(X_i)/\max_{\theta\in\Theta}\prod f_\theta(X_i))$
- T tends to small under null , and tends to be large under alternative.
- $\mathcal{C} = (\chi^2_{\text{critical}}, \infty)$
- For any $\theta \in \Theta_0$, $T \sim \chi^2_d$, where d is the dimension difference between Θ_0 and Θ.
- reject region is $\chi^2_{\text{critical}} = \chi^2_{d,1-\alpha}$
- p-value $= Pr(\chi^2_d \geq$ observed test statistic)

Sample Size Determination

- We already pick a test procedure and an α.
- We want to khow how large the sample size needs to be, such that we a reasonably large β over the *whole* Θ_1.
- This is actually impossible when Θ is a continuous connected set.
- Large enough n, such that $\theta(\theta)$ is larger than a desired level (say 0.8) over $\tilde{\Theta}_1 \subset \Theta_1$, where $\tilde{\Theta}_1 \subset \Theta_1$ represents the parameter values that are *practically significant*
- Let us try an example for one side z test

Controversy

Consider the following statistical modeling with $\Theta = \{0, 1\}$ with a discrete data generation:

PMF	X=1	X=2	X=3
$\theta = 0$	0.95	0.04	0.01
$\theta = 1$	0.099	0.9	0.001

Given a data point, and we want to test $\Theta_0 = \{0\}$ vs $\Theta_1 = \{1\}$ under $\alpha = 0.05$

There are only 8 possible reject region, and the best one is $\mathcal{R} = \{2, 3\}$ (why?)

Therefore, if we observe $X = 3$, we reject null hypothesis. However, when $X = 3$, $\theta = 0$ is more likely than $\theta = 1$, i.e., (0.01 vs 0.001)

Chapter Review

- Hypothesis Testing
- Idea and formulation
- Power function
- Test statistics