Lecture 9

Shuo Jiang

The Wang Yanan Institute for Studies in Economics, Xiamen University

Autumn, 2023

Introduction

All knowledge is, in the final analysis, history. All sciences are, in the abstract, mathematics. All judgements are, in their rational, statistics.

—-C. R. Rao

Introduction

- The modern society is essentially data-driven.
- In science, medicine, public policy, judicial practice and business, most decisions are data-driven.
- Example:
  - Is a coin really fair?
  - Does a nutrition help the growth of a plant?
  - Is the COVID medication really effective?
  - Should the government launch an education program?
  - Should Microsoft change the headline of ads shown on search engine Bing?
  - ...

Introduction

Having data itself is not enough. We need a systematic and rigorous method for processing data and arrive at a conclusion.

Example: Is a coin fair? That is to say, if X denotes a random variable that equals 1 if head faces up and 0 otherwise. Do we have $X \sim Bernoulli(0.5)$?

We could try to possibly answer this question with some data. Let $\widehat{p} = \frac{\sum_i X_i}{100}$.

Suppose a student toss the coin 100 times and get the following data: $\widehat{p} = 0.52$

Another student toss the coin 100 times and get the following data $\widehat{p} = 0.45$

A third student toss the coin 100 times and get $\widehat{p} = 0.2$

Introduction

Shall the first, second or third student conclude that the coin is fair?
If we decide the coin is fair in any of the above cases, are we sure we've made the correct choice? If not, how likely we are wrong?

Introduction

Example: In signal processing, we might want to see if noises are centered at zero. Suppose Y is some quantity of interest. Define noise as $X = Y_{received} - Y_{send}$. To do this we could try to collect a data set $X_1, ..., X_{50}$ which follows i.i.d. $N(\mu, \sigma^2)$, estimate the sample mean and compare it to zero. Suppose one engineer estimates $\bar{X} = 0.6$ based on a sample he collected.
Another engineer estimates it to be $\bar{X} = 1.2$ based on his sample.
Which engineer should conclude that the noise is centered at zero?
If we decide the center is zero, what's the probability that we got it wrong?

Introduction

Definition 1.1 (Hypothesis)

A hypothesis is a statement about a population parameter.

Definition 1.2 (Null and Alternative Hypothesis)

The two complementary hypothesis in a hypothesis testing problem are called the null hypothesis and the alternative hypothesis. They are denoted as $\mathbb{H}_0$ and $\mathbb{H}_1$, respectively.

Definition 1.3 (Test Statistic)

A test statistic is a function of the sample denoted by $T(\mathbf{X})$, which is used to make a decision on a hypothesis based on $\mathbf{X}$.

Introduction

If θ denotes a population parameter, the general format of the null and alternative hypothesis is $\mathbb{H}_0 : \theta \in \Theta_0$ and $\mathbb{H}_1 : \theta \in \Theta_0^c$, where Θ_0 is some subset of the parameter space and Θ_0^c is its complement.

For example,

$$\mathbb{H}_0 : \theta = 0, \text{ versus } \mathbb{H}_1 : \theta \neq 0$$

.

Definition 1.4

For a given sample, a hypothesis test makes the decision on whether we should

- not reject the null hypothesis $\mathbb{H}_0$, or
- reject the null hypothesis and accept the alternative hypothesis.

Exact Test

We start by constructing a test for the second example. Suppose that iid random variables $X_1, ..., X_n$ follow a Normal distribution with mean unknown μ and known variance $\sigma^2 < \infty$. We want to test the hypothesis

$$\mathbb{H}_0 : \mu = 0 \quad \text{versus} \quad \mathbb{H}_1 : \mu \neq 0.$$

We consider the sample average of $X_1, ..., X_n$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Exact Test

- Intuitively, we should reject the null if $\bar{X}_n$ deviates from 0 too much in either direction.
- Thus intuitively our test statistic better be a strictly increasing function of $|\bar{X}_n|$ that leads to rejection when the test statistic is large.

Exact Test

Also, since $X_i \sim N(\mu, \sigma^2)$, for $i = 1, 2, ..., n$. We have

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Therefore,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

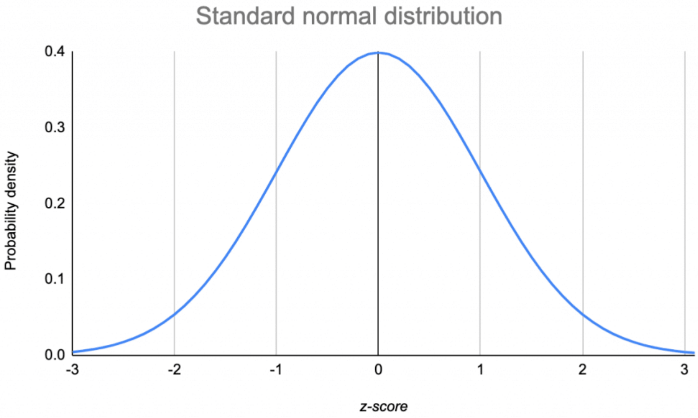
To test $\mathbb{H}_0 : \mu = 0$, we establish the test statistic

$$t = |\frac{\bar{X}_n}{\sigma/\sqrt{n}}| = |Z|,$$

where the object inside the absolute value is called the Z-score, and $Z \sim N(0, 1)$.

Exact Test

If $\mathbb{H}_0$ is correct, then $t \sim |Z|$. That is, for most of the time, the test statistic will be located around 0, and with a very small probability the test statistic will be very far awary from 0. The following graph shows the pdf of the Z-score:

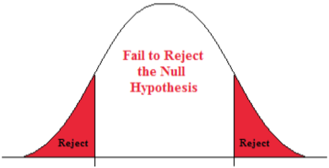


Exact Test

$$\mathbb{H}_0 : \mu = 0 \quad \text{versus} \quad \mathbb{H}_1 : \mu \neq 0.$$

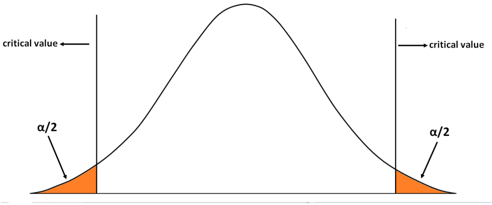
However, if $\mathbb{H}_0$ is incorrect, i.e., $\mathbb{H}_1$ is correct, then $\bar{X}_n$ will in general be far away from 0 with higher probability, such that the test statistic will be more likely to be far away from 0.

Therefore, **we reject the null hypothesis when t is sufficiently large, or equivalently when the Z-score is sufficiently large or sufficiently small(see the picture below)**.



Exact Test

Then, how large should we reject the null hypothesis? Where are the boundaries?



We set a *significance level* α, which is the probability of incorrectly reject $\mathbb{H}_0$ when $\mathbb{H}_0$ is correct. The boundary c is called the *critical value* for test statistic t, and we reject $\mathbb{H}_0$ when $t > c$. Because $t > c$ if and only if $Z > c$ or $Z < -c$, we could equivalently use Z as the test statistic and set the critical value for the Z-score (as shown in the figure above) as $-c$ and c.

Constructing tests using likelihood ratio

- In this example above, we have proposed a test statistic based on intuition.
- The form of the test statistic could also be derived using likelihood ratio with vs without restriction.

Constructing tests using likelihood ratio

Definition 1.5 (Likelihood ratio test)

The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\underset{\Theta_0}{sup} L(\theta|\mathbf{x})}{\underset{\Theta}{sup} L(\theta|\mathbf{x})} = \frac{L(\widehat{\theta}_0|\mathbf{x})}{L(\widehat{\theta}|\mathbf{x})}.$$

A likelihood ratio test(LRT) is any test that has a rejection region of the form $\{\mathbf{x} : \lambda(\mathbf{x}) \le c\}$ where $c \in [0, 1]$.

This is a general method we could use to construct tests for fully parametric models.

Constructing tests using likelihood ratio

Example: Let $X_1, ..., X_n$ be a random sample from $N(\theta, 1)$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

$$\begin{aligned}\lambda(\mathbf{x}) &= \frac{(2\pi)^{-n/2} exp(-\sum_{i=1}^{n}(x_i - \theta_0)^2/2)}{(2\pi)^{-n/2} exp(-\sum_{i=1}^{n}(x_i - \bar{x})^2/2)} \\ &= exp[-\sum_{i=1}^{n}(x_i - \theta_0)^2/2 - \sum_{i=1}^{n}(x_i - \bar{x})^2/2],\end{aligned}$$

Simplify the expression gives:

$$\lambda(\mathbf{x}) = exp[-n(\bar{x} - \theta_0)^2/2]$$

The rejection region $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ where $c \in [0, 1]$ could be written as $\{\mathbf{x} : |\bar{x} - \theta_0| \geq \sqrt{-2(logc)/n}\}$. Thus, all tests that reject the null if the sample mean deviate too much from the hypothesized value are equivalent to LRT up to a constant in the critical value.

Exact Test

- Now forget about the likelihood ratio and return to the original scenario of testing $H_0 : \mu = 0$ from a normal population.
- How should we proceed with the testing problem if variance is unknown?
- What if our goal is to decide if μ is smaller than 0?

Exact Test

- Note that in this process we did everything under the null: deriving the exact distribution of the test statistic, select the critical value and make the decision assuming that H_0 is true.
- The philosophy of statistical testing is to decide whether there is sufficient evidence from the data to conclude that the alternative hypothesis should be believed.
- A similar philosophy is also adopted in the modern court practice, sciences or businesses.

Exact Test

- For example, hypothesis testing is very similar to a court practice or criminal trial, in which the judge (or jury) must use evidence to decide which of 2 possible truths: innocent (H_0) or guilty (H_1), is to be believed.
- Just as the court is instructed to assume that the defendant is innocent unless proven otherwise, during investigation the forensic investigator should assume there is no association unless there is strong evidence to the contrary.

Exact Test

- In court practice, evidence that qualify as proof for conviction should be "beyond a reasonable doubt." That implies, evidence should be very unlikely given that the suspect is innocent.
- For hypothesis testing, we do the same.
- We select the level of significance (usually a small value) for the test that results in a critical value beyond which we reject the null.
- The critical value in our example is relatively large, and all values of t lower than this value are considered still reasonable under H_0. Only values of t greater than the critical value are considered beyond the reasonable range if H_0 were true.

Exact Test

- Since it is still possible that t goes beyond the critical value even if H_0 is true, we could make a mistake in this decision.
- The probability of rejecting H_0 when H_0 is true is exactly α. The standard value chosen for level of significance is $\alpha = 5\%$
- This standard means that even if H_0 is false, we are willing to accept a 1 in 20 chance of a false conclusion.

Two Types of Errors

The subset of the sample space for which $\mathbb{H}_0$ will be rejected is called the **rejection region** or **critical region**. The complement of the rejection region is called the **acceptance region**.

Note that they are subsets of the sample space of $\mathbf{X}$.
For any given θ, we could define the following two types of error probabilities:

- If $\theta \in \Theta_0$ but we falsely reject $\mathbb{H}_0$, it is called the **Type I error**;
- If $\theta \in \Theta_0^c$ but we fail to reject $\mathbb{H}_0$, it is called the **Type II error**.

Two Types of Errors

Table 8.3.1. *Two types of errors in hypothesis testing*

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct decision	Type I Error
	H_1	Type II Error	Correct decision

Power Function

The probability for these errors could be characterized using the rejection region R. $P_\theta(\mathbf{X} \in R)$ as a function of the true θ essentially contains all information about a certain test.

- $P_\theta(\mathbf{X} \in R)$ gives the probability of Type I error if $\theta \in \Theta_0$.
- $P_\theta(\mathbf{X} \in R)$ gives the probability of 1 minus Type II error if $\theta \in \Theta_0^c$.

Introduction

Definition 1.6

The power function of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.

- In other words, the power function could be viewed as the proabability of rejecting the null hypothesis.
- Ideally, we would expect the power function is 0 when $\theta \in \Theta_0$ and 1 when $\theta \in \Theta_0^c$.
- In practice, this ideal situation cannot be attained. A good test has power function near 1 for most $\theta \in \Theta_0^c$ and near 0 for most $\theta \in \Theta_0$.

Power Function

Example: Let $X \sim binomial(5, \theta)$. Consider testing $\mathbb{H}_0 : \theta \leq 1/2$ versus $\mathbb{H}_1 : \theta > 1/2$. Consider first the test that rejects $\mathbb{H}_0$ if and only if all "successes" are observed. Then, the power function of the test is

$$\beta_1(\theta) = \theta^5$$

as shown in Figure 8.3.1.

Power Function

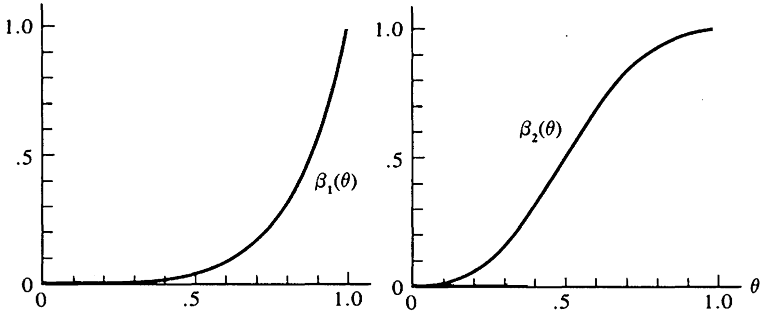


Figure 8.3.1. *Power functions for Example 8.3.2*

Power Function

- Consider the first case when $\mathbb{H}_0 : \theta \leq 1/2$ is correct: we find that the probability of making the type I error is very small (close to 0). That is $\beta_1(\theta) \leq 0.5^5 = 0.0312$.
- Consider the second case when $\mathbb{H}_1 : \theta > 1/2$ is correct: the probability of failing to reject $\mathbb{H}_0$ is $1 - \beta_1(\theta)$, which is smaller than $1/2$ only when $\theta > 0.5^{1/5} = 0.87$, so that only for a small proportion of $\theta \in \Theta_0^c$, the probability of making the Type II error is smaller than 50%.

Power Function

To achieve a smaller Type II error, we consider another test that rejects more frequently: we reject the null hypothesis if $X = 3, 4, 5$. Then, the power of the test is

$$\beta_2(\theta) = \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta) + \binom{5}{5}\theta^5.$$

We find that the probability of making the Type II error becomes smaller, however, the probability of making the Type I error increased when $\theta \leq 1/2$.

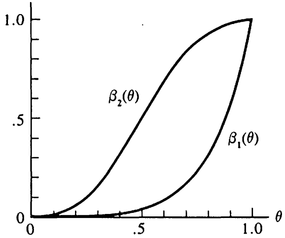


Figure 8.3.1. *Power functions for Example 8.3.2*

Power Function

Example: Let $X_1, ..., X_n$ be a random sample from a $N(\theta, \sigma^2)$ population in which σ^2 is known. We consider the hypothesis of $\mathbb{H}_0 : \theta \leq \theta_0$ versus $\mathbb{H}_1 : \theta > \theta_0$. Suppose that we establish a test that rejects the null hypothesis if

$$\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c$$

for some constant $c > 0$. Then the power function of the test is

$$\begin{aligned}\beta(\theta) =& P_\theta\left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c\right) = P_\theta\left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\ =& P_\theta\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right),\end{aligned}$$

where Z is a standard normal random variable.

Power Function

It is easy to find that

$$\lim_{\theta \to -\infty} \beta(\theta) = 0, \lim_{\theta \to \infty} \beta(\theta) = 1,$$

and

$$\beta(\theta_0) = \alpha \text{ if } P(Z > c) = \alpha.$$

The following figure shows $\beta(\theta)$ for $c = 1.28$.

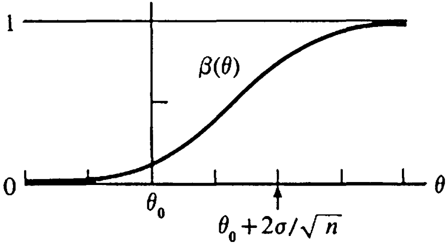


Figure 8.3.2. *Power function for Example 8.3.3*

Power Curve

- We could adjust the rejection region of the test by changing c.
- As c increases, it becomes more difficult to reject, and the power curve will be driven down.
- After the power curve has been driven down, which increases type II error for all θ but decreases type I error for all θ.

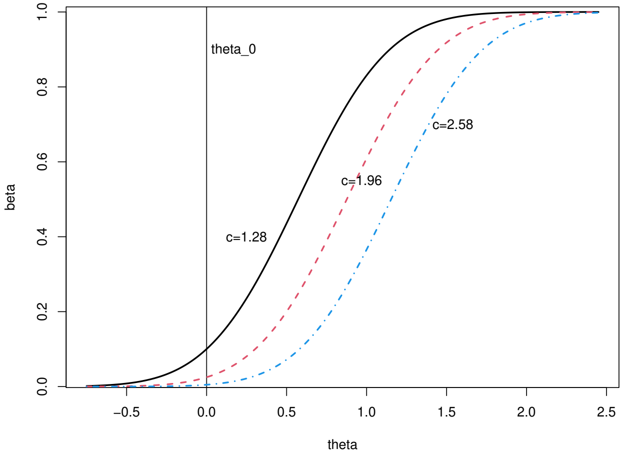


Figure 1: Power functions for different values of c (fix $n = 20$).

- For a fixed sample size, it is usually impossible to make both types of error probabilities arbitrarily small.
- Usually, we control the Type I error at a specified level, and then search for a test whose Type II error is as small as possible.

Definition 8.3.5

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.

Definition 8.3.6

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.

In many books, α is called the **size of a test**. Usually, we choose $\alpha = 0.1, 0.05$ or 0.01 in practice.
For power functions that are monotone in θ, it suffices to look at the power function at one point (in the previous two examples, this point is θ_0).

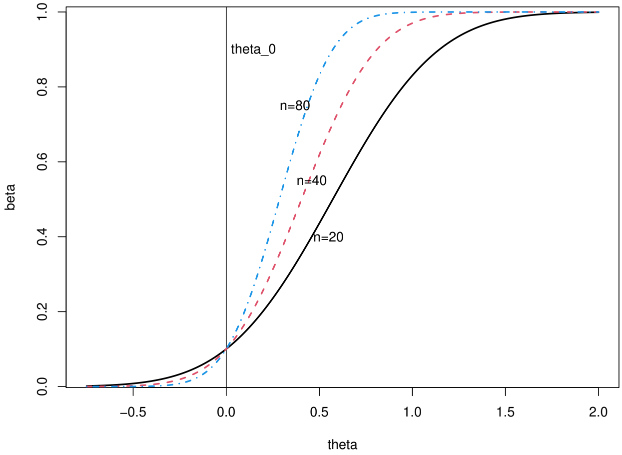


Figure 2: Power functions for different values of n (fix $c = 1.28$).

Example If we want to have a maximum type I error probability of 0.1, and have a maximum type II error probability of 0.2 for $\theta \geq \theta_0 + \sigma$, how should we choose c and n?

Because $\beta(\theta)$ is increasing in θ, to satisfy the above two requirements, we have

$$\beta(\theta_0) = 0.1 \text{ and } \beta(\theta_0 + \sigma) = 0.8.$$

By choosing $c = 1.28$, we achieve
$\beta(\theta_0) = P(Z > 1.28) = 0.1$, regardless of the value of n.

Given that $c = 1.28$, we could adjust n to meet the second requirement.

Other Concepts

- Consistency of a test

Approximate Tests: Asymptotic Approximation

- Suppose I ask you to work out the example of testing whether a coin is fair
- By a similar logic, we could propose a test statistic $T = |\frac{\hat{p}-0.5}{sd}| = |Z|$, but the problem is that now the exact sampling distribution of Z is unknown.
- In this case, we could use asymptotic approximation to conduct an approximate test.
- Similarly for all other distributions.

Approximate Tests: Bootstrap

- An alternative to doing asymptotic approximation is to do the bootstrap.
- In this approach, we approximate the exact sampling distribution by resampling our data B times with replacement, and obtain the distribution of the statistic of interest using the distribution of the B values.

Final Words

Uncertain knowledge + knowledge about the extent of uncertainty in it =Usable knowledge
—- C. R. Rao