Mathematical Statistics

MAS 713

Chapter 8

Previous lecture:

1. Bayesian Inference
2. Decision theory
3. Bayesian Vs. Frequentist
4. Loss functions
5. Conjugate priors

Any questions?

This lecture

8. Hypothesis Testing

- 8.1 Tests of Hypotheses
- 8.2 Confidence Intervals vs. Hypothesis Tests
- 8.3 Mathematical Formulation of Tests
- 8.4 Likelihood Ratio Test
- 8.5 Examples

Additional reading : Chapter 8

Hypotheses testing : Introduction

In the previous lectures we showed how a parameter of a population can be estimated from sample data, using either a point estimate or an interval of "likely" values called a confidence interval

This can be divided into two major areas :

- **estimation**, including point estimation and interval estimation
- **tests of hypotheses**

Today we discuss different **tests of hypotheses**.

Hypotheses testing : Introduction

There are many situations in which we must decide whether a statement concerning a parameter is true or false, that is, we must **test a hypothesis about a parameter**

Hypotheses testing : Introduction

For instance, suppose that a customer protection agency wants to test a paint manufacturer's claim that the average drying time of his new 'fast-drying' paint is 20 minutes

It instructs a member of its staff to paint each of 36 boards using a different can of the paint : the observed average drying time for this sample is 20.75 minutes

$\rightsquigarrow$ does that really contradict the manufacturer's claim ?

This type of question can be answered using a statistical inference technique called **hypothesis testing**

Statistical hypotheses

Many problems require that we decide which of two competing statements about some parameter of a population is true

$\rightsquigarrow$ the statements are called hypotheses

In the previous situation, we might express the hypothesis to test as

$$\mathcal{H}_0 : \mu = 20$$

where μ is the 'true' unknown mean drying time for this type of paint. This statement is called **null hypothesis**, and is usually denoted $\mathcal{H}_0$

Statistical hypotheses

The statement to be favoured in case we come to the conclusion that $\mathcal{H}_0$ is (most probably) not true is called the **alternative hypothesis**, and is usually denoted $\mathcal{H}_a$ (or sometimes $\mathcal{H}_1$ in some references)

The choice of $\mathcal{H}_a$ depends mostly on the problem and what we hope to be able to show

In our example, it could be either $\mathcal{H}_a : \mu \neq 20$ or $\mathcal{H}_a : \mu > 20$

Null hypothesis

The value μ_0 of the population parameter specified in the null hypothesis

$$\mathcal{H}_0 : \mu = \mu_0$$

is usually determined in one of three ways :

- it may result from past experience or knowledge of the process, or from previous tests or experiments

$\leadsto$ determine whether the parameter value has changed

- it may be determined from some theory or some model

$\leadsto$ check whether the theory or the model is valid

- it may result from external considerations, such as engineering specifications, or from contractual obligations

$\leadsto$ conformance testing

Null hypothesis

Note : in some instances, a null hypothesis of the form $\mathcal{H}_0 : \mu \geq \mu_0$ or $\mathcal{H}_0 : \mu \leq \mu_0$ may seem appropriate. However, the test procedure for such an $\mathcal{H}_0$ always exactly amounts to that for $\mathcal{H}_0 : \mu = \mu_0$
$\rightsquigarrow$ we always state a null hypothesis as an equality

Alternative hypothesis

The alternative hypothesis can essentially be of two types

We talk about **two-sided alternative** when $\mathcal{H}_a$ is of the form

$$\mathcal{H}_a : \mu \neq \mu_0$$

This is the exact negation of the null hypothesis $\mathcal{H}_0 : \mu = \mu_0$, μ_0 being the only value of some interest in the problem

However, in many situations, we may wish to favour a given direction for the alternative :

$$\mathcal{H}_a : \mu < \mu_0 \qquad \text{or} \qquad \mathcal{H}_a : \mu > \mu_0$$

We then talk about **one-sided alternative**

Alternative hypothesis

For instance, in the previous example, the customer protection agency may only wish to highlight that the average drying time of the paint is actually longer than the advertised 20 minutes (no criticisms if this time is even shorter)

The considered alternative might change the conclusion of a hypothesis test, and should be carefully formulated !

Hypothesis testing

A procedure leading to a decision about a particular hypothesis $\mathcal{H}_0$ is called a test of hypothesis

Such procedures rely on using the information contained in a random sample from the population of interest

$\rightsquigarrow$ if this information is consistent with the hypothesis $\mathcal{H}_0$, we will not reject $\mathcal{H}_0$; however, if this information is inconsistent with $\mathcal{H}_0$, we will conclude that $\mathcal{H}_0$ is false and we will reject it

Hypothesis testing

Fact

The truth or the falsity of a particular hypothesis can never be known with certainty, unless we can examine the entire population

The decision we make depends on a random sample, so is a kind of ‘random’ object

$\leadsto$ a hypothesis test should be developed with the probability of reaching a wrong conclusion in mind

To illustrate the general concepts, consider again the mean drying time problem. Imagine that we wish to test

$$\mathcal{H}_0 : \mu = 20 \qquad \text{against} \qquad \mathcal{H}_a : \mu \neq 20$$

We have a sample of $n = 36$ specimens and the sample mean $\bar{x}$ is observed

As the sample mean is a 'good' estimate of μ, we expect $\bar{x}$ to be reasonably close to μ

$\leadsto$ if $\bar{x}$ falls 'close' to 20 min, no clear contradiction with $\mathcal{H}_0$, we do not reject it

$\leadsto$ if $\bar{x}$ is considerably 'distant' from 20 min, evidence in support of $\mathcal{H}_a$, we reject $\mathcal{H}_0$

The numerical value which is computed from the sample and used to decide between $\mathcal{H}_0$ and $\mathcal{H}_a$, here the (possibly standardised) sample mean, is called the

test statistic

Hypothesis testing

Suppose that we decide to reject $\mathcal{H}_0$ if $\bar{x}$ is smaller than 19.33 or larger than 20.67 (arbitrary criterion for illustrative purpose only)

$\rightsquigarrow$ if $\bar{x} \in [19.33, 20.67]$, we do not reject $\mathcal{H}_0 : \mu = 20$

The values for which we reject $\mathcal{H}_0$, that is here values less than 19.33 and greater than 20.67, are called **critical region** for the test, the limiting values (here 19.33 and 20.67) being the **critical values**

Hypothesis testing

This provides a clear-cut criterion for the decision, however it is not guaranteed :

a) even if the true mean $\mu = 20$, there is a possibility that the sample mean $\bar{x}$ may be outside $[19.33, 20.67]$, due to bad luck

b) even if the true mean $\mu \neq 20$, say $\mu = 21$, there is a possibility that the sample mean may be in $[19.33, 20.67]$

Types of errors

There are essentially two possible wrong conclusions :

a) rejecting $\mathcal{H}_0$ when it is true : this is defined as a **type I error**

b) failing to reject $\mathcal{H}_0$ when it is false : this is defined as **type II error**

Because the decision is based on random variables, **probabilities** can be associated with these errors

The probability of type I error is usually denoted α

$$\mathrm{P}r(\text{type I error}) = \mathrm{P}r(\text{reject } \mathcal{H}_0 \text{ when it is true}) = \alpha$$

The probability of type II error is usually denoted β

$$\mathrm{P}r(\text{type II error}) = \mathrm{P}r(\text{fail to reject } \mathcal{H}_0 \text{ when it is false}) = \beta$$

$1 - \beta$, that is $\mathrm{P}r(\text{reject } \mathcal{H}_0 \text{ when it is false})$, is also called the **power** of the test

Note that β actually depends on the true (unknown) value of μ

		In Reality	
		H_0 True	H_0 False
Decision	Reject H_0	Type I Error	Correct Decision
	Fail to Reject H_0	Correct Decision	Type II Error

Assume in our running example that it is known from past experience that the drying time is normally distributed with known standard deviation $\sigma = 2$ min

As we know that $Z = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$, we have :

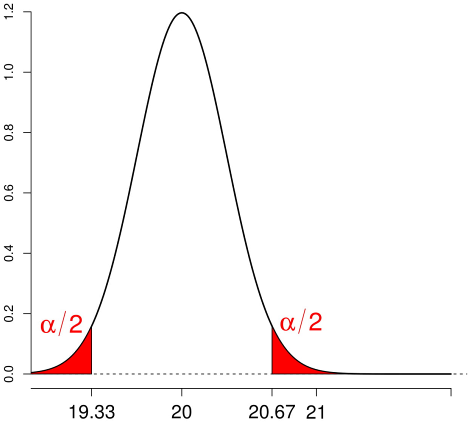
$$
\begin{aligned}
\Pr(\text{type I error}) &= \Pr((\bar{X} < 19.33) \cup (\bar{X} > 20.67) \text{ when } \mu = 20) \\
&= \Pr\left(Z < \sqrt{36}\,\frac{19.33-20}{2}\right) + \Pr\left(Z > \sqrt{36}\,\frac{20.67-20}{2}\right) \\
&= \Pr(Z < -2.01) + \Pr(Z > 2.01) = 0.044 = \alpha \qquad \text{(tables)}
\end{aligned}
$$

If $\mathcal{H}_0$ is true, we have a 4.4% chance of rejecting it (with our rule)

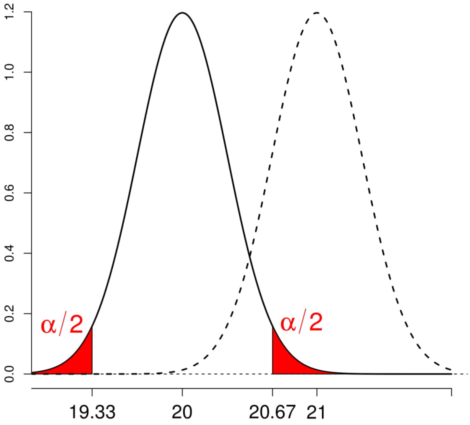
Suppose now that $\mu = 21$ (so $\mathcal{H}_0$ is not true!).
We have :

$$\begin{aligned} Pr(\text{type II error}) &= Pr(\bar{X} \in [19.33, 20.67] \text{ when } \mu = 21) \\ &= Pr\left(\sqrt{36}\,\frac{19.33 - 21}{2} \le Z \le \sqrt{36}\,\frac{20.67 - 21}{2}\right) \\ &= Pr(-5.01 \le Z \le -0.99) = 0.16 = \beta \qquad \text{(tables)} \end{aligned}$$

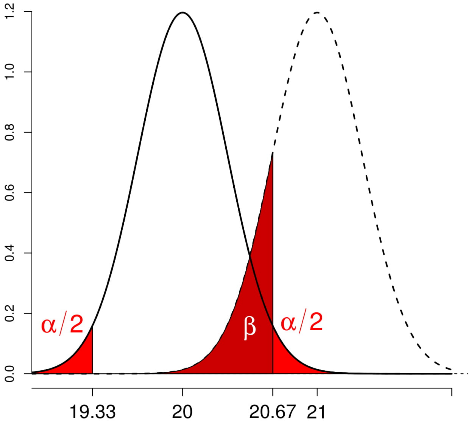
With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.33, 20.67]$



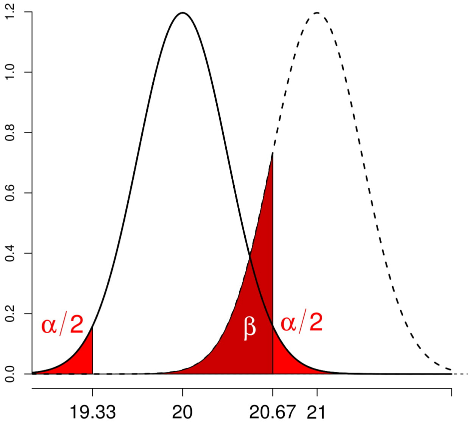
With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.33, 20.67]$



With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.33, 20.67]$



With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.33, 20.67]$



$\rightsquigarrow \alpha = 0.044$,
$\beta = 0.16$ if $\mu = 21$

See that β would rapidly increase as μ approached the hypothesized value μ_0

Suppose that you want to reduce the error I probability α

$\leadsto$ widen the acceptance region, for instance say

$$\text{reject } \mathcal{H}_0 \text{ if } \bar{x} \notin [19.2, 20.8]$$

(again for illustrative purpose only) Then,

$$\alpha = \ldots = 0.016 \quad (< 0.044)$$

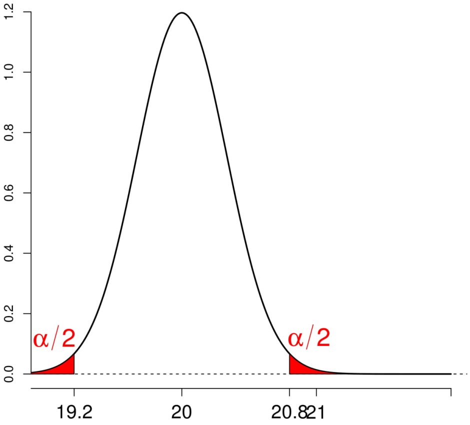
but

$$\beta = \ldots = 0.27 \quad (> 0.16) \qquad \text{when } \mu = 21$$

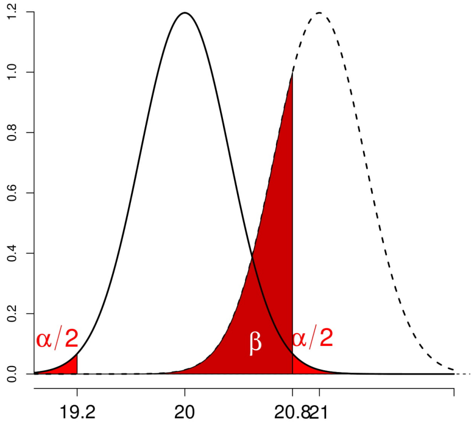
$\leadsto$ if α decreases, β must increase and vice-versa !

Impossible to make both types of error as small as possible simultaneously

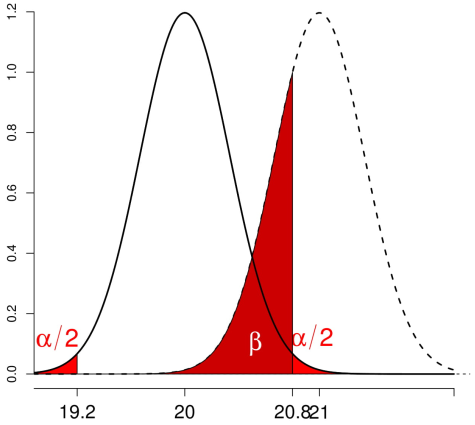
With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.2, 20.8]$



With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.2, 20.8]$



With the decision rule : reject $\mathcal{H}_0$ if $\bar{x} \notin [19.2, 20.8]$



$\rightsquigarrow \alpha = 0.016,$
$\beta = 0.27$ if $\mu = 21$

Usually, one decides to set α to a predetermined level, called the **significance level** of the test

This is because hypotheses testing methods were actually originally inspired by jury trials

In a trial, defendants are primarily **assumed innocent** ($\mathcal{H}_0$). Then,

- if strong evidence is found to the contrary, then they are declared to be guilty (reject $\mathcal{H}_0$)
- if there is insufficient evidence, they are declared nonguilty (no reject of $\mathcal{H}_0$) $\rightsquigarrow$ not the same as proving the defendant innocent (accept $\mathcal{H}_0$) !

If the jury is wrong, either an innocent person is convicted (type I error) or a culprit is let free (type II error)

$\rightsquigarrow$ one thought (and probably still thinks) that convicting an innocent person was more a serious problem than the contrary

$\rightsquigarrow$ controlling α is more important, could β be whatever it is

Errors : analogy to criminal trials

		In Reality	
		Innocent	Guilty
Decision	Convict	Type I Error	Correct Decision
	Acquit	Correct Decision	Type II Error

Significance level and decision rule

Jury must be "convinced beyond a reasonable doubt" to convict

$\rightsquigarrow$ usually, the significance level α is set to 0.10, 0.05 or 0.01 (it all depends on the consequences), and the decision rule is fixed accordingly

Definition

The ***p*-value** is the smallest level of significance that would lead to rejection of $\mathcal{H}_0$ with the observed sample

Concretely, the *p*-value is the probability that the test statistic will take on a value that is at least as extreme as the observed value when $\mathcal{H}_0$ is true ('extreme' to be understood in the direction of the alternative)

It provides a **measure of the credibility of $\mathcal{H}_0$**

It might be interpreted as the chance of being wrong if we reject $\mathcal{H}_0$

One-sided alternatives

One-sided alternatives

The whole development, yet very similar, must be slightly adapted when **one-sided alternatives** are concerned

First, *it might occasionally be difficult to choose the appropriate formulation of the alternative*

In our running example, suppose now that we would like to highlight that the average drying time is actually longer than the advertised 20 min. Would we test for

- $\mathcal{H}_0 : \mu \leq 20$ against $\mathcal{H}_a : \mu > 20$, hoping to reject $\mathcal{H}_0$, or
- $\mathcal{H}_0 : \mu \geq 20$ against $\mathcal{H}_a : \mu < 20$, hoping not to reject $\mathcal{H}_0$?

One-sided alternatives

Remember that rejecting $\mathcal{H}_0$ is a strong conclusion (we have enough evidence to do it),
unlike not rejecting (we do not have enough evidence to conclude, the decision is dictated by risk aversion, not by facts $\leadsto$ weak conclusion)

$\leadsto$ always put what we want to prove in the alternative hypothesis

Here, we should test $\mathcal{H}_0 : \mu \leq 20$ against $\mathcal{H}_a : \mu > 20$

One-sided alternatives

In a two-sided test, i.e. with alternative $\mathcal{H}_a : \mu \neq \mu_0$, an observed value $\bar{x}$ of $\bar{X}$ much smaller than μ_0 **or** much larger than μ_0 is evidence in direction of $\mathcal{H}_a$

However, if the alternative is $\mathcal{H}_a : \mu > \mu_0$, a small value of $\bar{X}$ is not evidence against $\mathcal{H}_0 : \mu = \mu_0$ in favour of $\mathcal{H}_a$ ($\mathcal{H}_0$ is more likely than $\mathcal{H}_a$ if $\bar{X}$ takes a small value!)

$\rightsquigarrow$ we must only seek evidence against $\mathcal{H}_0$ in the direction of $\mathcal{H}_a$!

One-sided alternatives

1. Thus, in testing

$$\mathcal{H}_0 : \mu = \mu_0 \qquad \text{against } \mathcal{H}_a : \mu > \mu_0$$

   we should reject $\mathcal{H}_0$ if $\bar{X}$ is much greater than μ_0

2. Similarly, in testing

$$\mathcal{H}_0 : \mu = \mu_0 \qquad \text{against } \mathcal{H}_a : \mu < \mu_0$$

   we should reject $\mathcal{H}_0$ if $\bar{X}$ is much smaller than μ_0

The critical region is again determined by the significance level α

Confidence Intervals vs. Hypothesis Tests

Hypothesis tests and confidence intervals

You might have noticed that the critical values for the (two-sided test) decision rule

$$\text{reject } \mathcal{H}_0 \text{ if } \bar{x} \notin \left[\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

look like the limits of the (two-sided) confidence interval for μ (see Chapter 4, Slide 39)

$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

As the interval widths are the same, the confidence interval (centred at $\bar{x}$) cannot contain μ_0 if the 'no-rejection area' (centred at μ_0) does not contain $\bar{x}$ (and vice-versa)

Hypothesis tests and confidence intervals

Generally speaking, there is always a **close relationship between the test of hypothesis about any parameter, say θ, and the confidence interval for θ** :

If $[l, u]$ is a $100 \times (1-\alpha)\%$ confidence interval for a parameter θ, then the hypothesis test for

$$\mathcal{H}_0 : \theta = \theta_0 \qquad \text{against} \qquad \mathcal{H}_a : \theta \neq \theta_0$$

will **reject $\mathcal{H}_0$ at significance level α if and only if θ_0 is not in $[l, u]$**

Hypothesis tests and confidence intervals

$\rightsquigarrow$ hypothesis tests and CIs are more or less equivalent, however each provides somewhat different insights :

- CIs provide a range of likely values for θ
- tests easily display the risk levels, such as p-values, associated with a specific decision

Note : the same analogy exists between one-sided tests and one-sided confidence intervals

Mathematical Formulation of Tests

- Let $X_1, \ldots, X_n$ be a random sample.
- Let $S_\theta \equiv \Theta$ be the Parameter space of unknown parameter

Hypothesis formulation: Let $\Theta_0 \subseteq \Theta$, $\Theta_a \subseteq \Theta$ so that $\Theta_0 \cap \Theta_a = \emptyset$. Then

$$\text{Null hypothesis } \mathcal{H}_0 : \theta \in \Theta_0$$
$$\text{Alternative hypothesis } \mathcal{H}_a : \theta \in \Theta_a$$

Definition: A hypothesis is called **simple** if

$$\Theta_0 = \{\theta_0\}, \qquad \Theta_a = \{\theta_a\}$$

Definition: A test is a map $\tau : \mathbb{R}^n \to \{0, 1\}$ where

- $\tau(x_1, \ldots, x_n) = 0$ means one does not reject the null hypothesis $\mathcal{H}_0$
- $\tau(x_1, \ldots, x_n) = 1$ means one rejects the null hypothesis (hence $\mathcal{H}_1$)

- $T := t(X_1, \ldots, X_n)$ is called **test statistic**
- If $K \subseteq \mathbb{R}$ denotes the critical region, the decision can be formulated as

$$\tau(x_1, \ldots, x_n) := \mathbf{1}_{\{t(x_1,\ldots,x_n) \in K\}}$$

$\Longrightarrow$ one rejects $\mathcal{H}_0$ if and only if for the observed value $x_1, \ldots, x_n$ one has that the observed value of the data satisfies $t(x_1, \ldots, x_n) \in K$.

Type I error: Whenever $\theta \in \Theta_0$ and $T \in K$.

Therefore, $\mathbb{P}_\theta(T \in K) \equiv \mathbb{P}(T \in K|\theta)$ for $\theta \in \Theta_0$ is the called the probability of having Type I error.

Type II error: Whenever $\theta \in \Theta_a$ and $T \notin K$.

Therefore, $\mathbb{P}_\theta(T \notin K) \equiv \mathbb{P}(T \notin K|\theta)$ for $\theta \in \Theta_a$ is the called the probability of having Type II error.

Definition:

The function

$$\text{pow} : \Theta_a \to [0, 1], \quad \theta \mapsto \text{pow}(\theta) := \mathbb{P}_\theta(T \in K) \equiv \mathbb{P}(T \in K|\theta)$$

is called the **power** of the test.

Typically one has that:

- Θ_0 and Θ_a form a partition of Θ
- The map $\theta \mapsto \mathbb{P}_\theta[T \in K]$ is continuous

$\Longrightarrow$ one cannot simultaneously minimize $\mathbb{P}_\theta(T \in K)$ on Θ_0 and maximize on Θ_a

Therefore, proceed as follows to construct T, K:

1) Fix significance level α and make sure that

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T \in K) \leq \alpha$$

2) Afterwards, maximize the power $\text{pow}(\theta) = \mathbb{P}_\theta(T \in K)$ for $\theta \in \Theta_a$

- Let $\Theta \subseteq \mathbb{R}$.

Definition:

A test is called **upper-tailed** if

$$\mathcal{H}_0 : \theta = \theta_0, \qquad \mathcal{H}_a : \theta > \theta_0 \qquad (\Longrightarrow K = (c_u, \infty))$$

A test is called **lower-tailed** if

$$\mathcal{H}_0 : \theta = \theta_0, \qquad \mathcal{H}_a : \theta < \theta_0 \qquad (\Longrightarrow K = (-\infty, c_l))$$

A test is called **two-tailed** if

$$\mathcal{H}_0 : \theta = \theta_0, \qquad \mathcal{H}_a : \theta \neq \theta_0 \qquad (\Longrightarrow K = (-\infty, c) \cup (C, \infty))$$

z-test

- random sample $X_1, \ldots, X_n$ i.i.d. $\sim N(\mu, \sigma)$ where μ is unknown but σ known.

- $\theta = \mu \quad (\Theta \subseteq \mathbb{R})$
- $\mathcal{H}_0 : \mu = \mu_0$
- **Test statistic**

$$T = T(X_1, \ldots, X_n) := \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \qquad \Longrightarrow T \sim N(0,1) \text{ under } \mathbb{P}_{\mu_0}.$$

z-test: How to find critical region

- Given significance level α

upper-tailed test (i.e. $\mathcal{H}_1 : \mu > \mu_0$):

$K = (c_u, \infty)$ where one finds c_u by

$$\alpha = \mathbb{P}_{\mu_0}(T \in K) = \mathbb{P}_{\mu_0}(T > c_u) = 1 - \Phi(c_u)$$

$\Longrightarrow c_u = z_{1-\alpha}$ i.e. the $1 - \alpha$ quantile of $N(0, 1)$-distribution

lower-tailed test (i.e. $\mathcal{H}_1 : \mu < \mu_0$):

$K = (-\infty, c_l)$ where one finds c_l by

$$\alpha = \mathbb{P}_{\mu_0}(T \in K) = \mathbb{P}_{\mu_0}(T < c_l) = \Phi(c_l)$$

$\Longrightarrow c_l = z_\alpha = -z_{1-\alpha}$ i.e. the α quantile of $N(0, 1)$-distribution

two-tailed test (i.e. $\mathcal{H}_1 : \mu \neq \mu_0$):

- $K = (-\infty, c) \cup (C, \infty)$
- by symmetry of $N(0,1)$ choose $c = -C \leq 0$ where one finds c by

$$\alpha = \mathbb{P}_{\mu_0}(T < c) + \mathbb{P}_{\mu_0}(T > -c) = \Phi(c) + 1 - \Phi(-c) = 2\Phi(c)$$

$$\Longrightarrow c = z_{\frac{\alpha}{2}}, \qquad C = -c = -z_{\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}$$

t-test

- random sample $X_1, \ldots, X_n$ i.i.d. $\sim N(\mu, \sigma)$ where **both** μ and σ^2 unknown.

- $\theta = (\mu, \sigma^2) \quad (\Theta \subseteq \mathbb{R} \times (0, \infty))$
- $\mathcal{H}_0 : \mu = \mu_0$ More precisely $\Theta_0 = \{\mu_0\} \times (0, \infty)$
- **Test statistic**

$$T = T(X_1, \ldots, X_n) := \sqrt{n}\frac{\bar{X}_n - \mu_0}{S} \qquad \Longrightarrow T \sim t_{n-1}$$

where:

- $S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$
- t_{n-1} denotes t-student distribution with $n-1$-degree of freedom

t-test: How to find critical region

- Given significance level α

upper-tailed test (i.e. $\mathcal{H}_1 : \mu > \mu_0$):

$K = (c_u, \infty)$ where one finds c_u by

$$\alpha = \mathbb{P}_{\mu_0}(T \in K) = \mathbb{P}_{\mu_0}(T > c_u) = 1 - F_{t_{n-1}}(c_u)$$

$\Longrightarrow c_u = z_{1-\alpha}$ i.e. the $1 - \alpha$ quantile of t_{n-1}-distribution

lower-tailed test (i.e. $\mathcal{H}_1 : \mu < \mu_0$):

$K = (-\infty, c_l)$ where one finds c_l by

$$\alpha = \mathbb{P}_{\mu_0}(T \in K) = \mathbb{P}_{\mu_0}(T < c_l) = F_{t_{n-1}}(c_l)$$

$\Longrightarrow c_l = z_\alpha = -z_{1-\alpha}$ i.e. the α quantile of t_{n-1}-distribution

two-tailed test (i.e. $\mathcal{H}_1 : \mu \neq \mu_0$):

- $K = (-\infty, c) \cup (C, \infty)$
- by symmetry of t_{n-1}-distr. choose $c = -C \leq 0$ where one finds c by

$$\alpha = \mathbb{P}_{\mu_0}(T < c) + \mathbb{P}_{\mu_0}(T > -c) = F_{t_{n-1}}(c) + 1 - F_{t_{n-1}}(-c) = 2F_{t_{n-1}}(c)$$

$$\Longrightarrow c = z_{\frac{\alpha}{2}}, \qquad C = -c = -z_{\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}$$

What happens if random sample not normal?

- Let $X_1, \ldots, X_n$ random sample, not necessarily normal
- Assume sample size is large (i.e. n large)

Case 1: μ is unknown but σ known.

- Let $\mathcal{H}_0 : \mu = \mu_0$
- **Test statistic**

$$T = T(X_1, \ldots, X_n) := \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \qquad \Longrightarrow T \overset{a}{\sim} N(0,1) \text{ under } \mathbb{P}_{\mu_0}.$$

Case 2: Both μ and σ unknown.

- Let $\mathcal{H}_0 : \mu = \mu_0$
- **Test statistic**

$$T = T(X_1, \ldots, X_n) := \sqrt{n}\frac{\bar{X}_n - \mu_0}{S} \qquad \Longrightarrow T \overset{a}{\sim} N(0,1) \text{ under } \mathbb{P}_{\mu_0}.$$

$\Longrightarrow$ **Continue as for the *z*-test or *t*-test.**

p-value

Motivation: We would like to see how "extreme" under $\mathcal{H}_0$ is the observed value $t(x_1, \ldots, x_n)$ given data $x_1, \ldots, x_n$.

- Let $\Theta_0 = \{\theta_0\}$

Definition: The *p*-value of a test statistic $T = t(X_1, \ldots, X_n)$ given observed values $x_1, \ldots, x_n$ is the probability under the null-hypothesis to obtain an at least as extreme value as the one observed.
- Hereby, the alternative hypothesis defines what is more extreme.

Mathematical Definition of p-value:

p-value of $T = t(X_1, \ldots, X_n)$ given observation $x_1, \ldots, x_n$ is defined:

Upper-tailed test:

$$p\text{-value} := \mathbb{P}_{\theta_0}\big(T \geq t(x_1, \ldots, x_n)\big)$$

Lower-tailed test:

$$p\text{-value} := \mathbb{P}_{\theta_0}\big(T \leq t(x_1, \ldots, x_n)\big)$$

Two-tailed test:

$$p\text{-value} := 2\min\Big\{\mathbb{P}_{\theta_0}\big(T \geq t(x_1, \ldots, x_n)\big), \mathbb{P}_{\theta_0}\big(T \leq t(x_1, \ldots, x_n)\big)\Big\}$$

Consequence: We reject $\mathcal{H}_0$ (and take $\mathcal{H}_a$) $\iff$ p-value $< \alpha$

Example coin toss: $n = 10, X_1, \dots, X_{10}$ i.i.d. $\sim$ Ber(θ) under P_θ
Test statistic: $T = t(X_1, \dots, X_{10}) = \sum_{i=1}^{10} X_i \sim \text{Bin}(10, \theta)$ under P_θ
Data: $x_1, \dots, x_{10}$ with $t(x_1, \dots, x_n) = \sum_{i=1}^{10} x_i = 7$.

Case 1: $\mathcal{H}_0 : \theta = 0.5, \qquad \mathcal{H}_a : \theta > 0.5.$

$$\text{p-value } = \mathbb{P}_{0.5}(T \geq 7) = 0.17$$

Case 2: $\mathcal{H}_0 : \theta = 0.5, \qquad \mathcal{H}_a : \theta < 0.5.$

$$\text{p-value } = \mathbb{P}_{0.5}(T \leq 7) = 0.94$$

Case 3: $\mathcal{H}_0 : \theta = 0.5, \qquad \mathcal{H}_a : \theta \neq 0.5.$

$$\text{p-value } = 2\,\mathbb{P}_{0.5}(T \geq 7) = 0.34$$

Conclusion: For significance level $\alpha = 0.05$,
$\mathcal{H}_0$ will in all cases not be rejected.

Likelihood Ratio Test

- Let $X_1, \ldots, X_n$ be random sample
- Given observations $\mathbf{x} := (x_1, \ldots, x_n)$
- Let $\theta \mapsto L(\theta|\mathbf{x}) := p(\mathbf{x}|\theta)$ be the Likelihood function
- Given observations $x_1, \ldots, x_n$, consider the ratio

$$\Lambda(x_1, \ldots, x_n) := \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Theta_a} L(\theta|\mathbf{x})}$$

Intuition:

If $\Lambda(x_1, \ldots, x_n)$ is small one can interpret that more likely $\mathcal{H}_a$ is true.

Definition: For any given $c \geq 0$ the **Likelihood Ratio Test (LRT)** with respect to c is defined as

$$T(X_1, \ldots, X_n) := \Lambda(X_1, \ldots, X_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{X})}{\sup_{\theta \in \Theta_a} L(\theta|\mathbf{X})}$$

with corresponding critical region $K := \{\mathbf{x} : \Lambda(\mathbf{x}) < c\}$

- There is slight alternative definition of the LRT test:

Definition: For any given $c \in (0,1)$ the alternative **Likelihood Ratio Test (LRT)** with respect to c is defined as

$$\widetilde{\Lambda}(X_1,\ldots,X_n) = \frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{X})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{X})}$$

with corresponding critical region $K := \{\mathbf{x} : \Lambda(\mathbf{x}) < c\}$

In which is sense is LRT Test optimal?

Neyman-Pearson Lemma

Assume that:

- $\Theta_0 = \{\theta_0\}, \quad \Theta_a = \{\theta_a\}$
- $T(X_1, \ldots, X_n) = \Lambda(X_1, \ldots, X_n)$ (i.e. LRT) with critical region $K = [0, c)$
- Let $\alpha^* := \mathbb{P}_{\theta_0}(T \in K)$.

Then, for any other test (T', K') with $\mathbb{P}_{\theta_0}(T' \in K') \leq \alpha^*$ we have that

$$\mathbb{P}_{\theta_a}(T \in K) \geq \mathbb{P}_{\theta_a}(T' \in K')$$

Intepretation:

Under the above assumptions, LRT is the most **powerful** test with significance level α^*

(also called **uniformly most powerful UMP** test of all level α^* tests)

5 Steps to follow when doing a statistical test

1) Choose your model (i.e. sample distribution)

2) Define $\mathcal{H}_0$ and $\mathcal{H}_a$

3) Define statistical test $T = t(X_1, \ldots, X_n)$

4) Fix significance level α. This characterizes K

5) Compute $t(x_1, \ldots, x_n)$ given observation $(x_1, \ldots, x_n)$ and conclude by the decision rule

$$\mathbf{1}_{\{t(x_1,\ldots,x_n)\in K\}}$$

Examples

Example

Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from $N(\mu, \sigma)$ random variables with σ known, but μ unknown.

1. Apply the Neyman-Pearson lemma to construct the best test of the hypotheses

$$\mathcal{H}_0 : \mu = \mu_0$$
$$\mathcal{H}_a : \mu = \mu_1$$

   where $\mu_1 > \mu_0$.
2. Is this a simple test?
3. What is the test statistic?
4. What is the critical region?
5. What is the power of the test?

Question:
Is this a simple test?

Solution:
Yes, because both $\mathcal{H}_0 = \{\mu_0\}$ and $\mathcal{H}_a = \{\mu_1\}$ are singletons.

Question:
Is this a simple test?

Solution:
Yes, because both $\mathcal{H}_0 = \{\mu_0\}$ and $\mathcal{H}_a = \{\mu_1\}$ are singletons.

- The LR is given by

$$\Lambda(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)}$$

$$p(x_1,\ldots,x_n|\mu_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)$$

$$p(x_1,\ldots,x_n|\mu_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)$$

- We use the Neyman-Pearson lemma and write

$$\Lambda(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_1)^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\mu_1)+\mu_0^2-\mu_1^2\right)\right)$$

- We use the Neyman-Pearson lemma and write

$$\Lambda(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_1)^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\mu_1)+\mu_0^2-\mu_1^2\right)\right)$$

- We use the Neyman-Pearson lemma and write

$$\Lambda(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_1)^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\mu_1)+\mu_0^2-\mu_1^2\right)\right)$$

- We use the Neyman-Pearson lemma and write

$$\Lambda(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)}$$
$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)}$$
$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_1)^2\right)}$$
$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right)\right)$$
$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\mu_1)+\mu_0^2-\mu_1^2\right)\right)$$

- We use the Neyman-Pearson lemma and write

$$\begin{aligned}
\Lambda(x_1,\ldots,x_n) &:= \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\mu_1)} \\
&= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_1)^2\right)} \\
&= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_1)^2\right)} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right)\right) \\
&= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\mu_1)+\mu_0^2-\mu_1^2\right)\right)
\end{aligned}$$

- the test statistic is for given c is:

$$\Lambda\left(x_1,\ldots,x_n\right) := \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}\left(\mu_0-\mu_1\right)+\mu_0^2-\mu_1^2\right)\right) \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} c$$

- We try to separate between random variables and constants

$\Longrightarrow$ We take the log transform

$$-\frac{n}{2\sigma^2}\left(-2\bar{x}\left(\mu_0-\mu_1\right)+\mu_0^2-\mu_1^2\right) \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} \log c$$

which is equal to

$$\bar{x} \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} \frac{\sigma^2}{n}\frac{\log c}{\mu_0-\mu_1}+\frac{\mu_0+\mu_1}{2}$$

- The critical region is

$$C = \left\{ x : \bar{x} < \frac{\sigma^2}{n} \frac{\log c}{\mu_0 - \mu_1} + \frac{\mu_0 + \mu_1}{2} \right\}$$

What is the power of the test?

Solution:

$$\mathbb{P}_{\mu_1}\left(\Lambda\left(X_1,\ldots,X_n\right) < c\right)$$

$$\mathbb{P}_{\mu_1}\left(\bar{X} < \frac{\sigma^2}{n}\frac{\log c}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2}\right)$$

- We need to find the distribution of $\bar{X}$ under the alternative:

$$\bar{X}|\mathcal{H}_a \sim N\left(\mu_1, \frac{\sigma}{\sqrt{n}}\right)$$

$$\mathbb{P}_{\mu_1}\left(\frac{\bar{X}-\mu_1}{\sigma}\sqrt{n} < \left(\frac{\sigma^2}{n}\frac{\log c}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2} - \mu_1\right)\frac{\sqrt{n}}{\sigma}\right)$$

which is for $Z \sim N(0,1)$ also equal

$$\mathbb{P}\left(Z < \left(\frac{\sigma^2}{n}\frac{\log c}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2} - \mu_1\right)\frac{\sqrt{n}}{\sigma}\right)$$

which can be easily found (using standard normal-table).

- In the same way you can find the significance level α^* (so that Neyman-Pearson Lemma holds)

$$\alpha^* := \mathbb{P}_{\mu_0}\left(\Lambda\left(X_1, \ldots, X_n\right) < c\right)$$

Example

Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from $N(\mu, \sigma)$ random variables with σ known.

1. Apply the alternative LRT $\widetilde{\Lambda}(X_1, \ldots, X_n)$ for some $c \in (0, 1)$ with the hypothesis

$$\mathcal{H}_0 : \mu = \mu_0$$
$$\mathcal{H}_a : \mu \neq \mu_0$$

2. What is the critical region?

Solution:

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1, \ldots, x_n) := \frac{p(x_1, \ldots, x_n | \mu_0)}{\sup_{\mu} p(x_1, \ldots, x_n | \mu)} \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} c$$

$$p(x_1, \ldots, x_n | \mu_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i - \mu_0)^2\right)$$

$$p(x_1, \ldots, x_n | \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i - \mu)^2\right)$$

- Since μ is unknown we estimate it from the data by finding the unrestricted **MLE**:

$$\widehat{\mu} = \bar{x}$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\widehat{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\widehat{\mu})^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\widehat{\mu})^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\widehat{\mu})^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\bar{x}) + \mu_0^2 - \bar{x}^2\right)\right)$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\widehat{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\widehat{\mu})^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\widehat{\mu})^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\widehat{\mu})^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\bar{x}) + \mu_0^2 - \bar{x}^2\right)\right)$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\widehat{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\widehat{\mu})^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\widehat{\mu})^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\widehat{\mu})^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\bar{x}) + \mu_0^2 - \bar{x}^2\right)\right)$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\widehat{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\widehat{\mu})^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\widehat{\mu})^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\widehat{\mu})^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\bar{x}) + \mu_0^2 - \bar{x}^2\right)\right)$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1,\ldots,x_n) := \frac{p(x_1,\ldots,x_n|\mu_0)}{p(x_1,\ldots,x_n|\widehat{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(x_i-\widehat{\mu})^2\right)}$$

$$= \frac{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\widehat{\mu})^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\widehat{\mu})^2\right)\right)$$

$$= \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}(\mu_0-\bar{x}) + \mu_0^2 - \bar{x}^2\right)\right)$$

- The alternative LRT is given by

$$\widetilde{\Lambda}(x_1, \ldots, x_n) := \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}\mu_0 + \bar{x}^2 + \mu_0^2\right)\right)$$

And the test is:

$$\widetilde{\Lambda}(x_1, \ldots, x_n) := \exp\left(-\frac{n}{2\sigma^2}\left(-2\bar{x}\mu_0 + \bar{x}^2 + \mu_0^2\right)\right) \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} c$$

- We take the log transform

$$-\frac{n}{2\sigma^2}\left(-2\bar{x}\mu_0 + \bar{x}^2 + \mu_0^2\right) \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} \log c$$

which is equal to

$$-\bar{x}^2 + 2\bar{x}\mu_0 \underset{\mathcal{H}_a}{\overset{\mathcal{H}_0}{\gtrless}} \frac{2\sigma^2}{n}\log c + \mu_0^2$$

- The critical region is

$$C = \left\{ x : -\bar{x}^2 + 2\bar{x}\mu_0 < \frac{2\sigma^2}{n} \log c + \mu_0^2 \right\}$$

$$= \left\{ x : -(\bar{x} - \mu_0)^2 < \frac{2\sigma^2}{n} \log c \right\}$$

$$= \left\{ x : |\bar{x} - \mu_0| > \sqrt{\left|\frac{2\sigma^2}{n} \log c\right|} \right\}$$

Example

Suppose that the distribution of lifetimes of TV tubes can be adequately modelled by an exponential distribution with mean θ so

$$p(x|\theta) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) & , x \geq 0 \\ 0, & \text{Otherwise} \end{cases}$$

Under usual production conditions, the mean lifetime is 2000 hours but if a fault occurs in the process, the mean lifetime drops to 1000 hours. A random sample of 20 tube lifetimes is to taken in order to test the hypotheses

$$\mathcal{H}_0 : \theta = 2000 = \theta_0$$
$$\mathcal{H}_a : \theta = 1000 = \theta_a$$

1. Use the Neyman-Pearson lemma to find the most powerful test with significance level α^*.
2. Find the Type I error.

Solution:

- We have that the likelihood is

$$p(x_1,\dots,x_{20}|\theta) = \prod_{i=1}^{20} p(x_i|\theta) = \prod_{i=1}^{20} \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$$
$$= \frac{1}{\theta^{20}}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{20} x_i\right)$$

- Therefore

$$\Lambda(x_1,\dots,x_{20}) = \frac{p(x_1,\dots,x_{20}|\theta=2000)}{p(x_1,\dots,x_{20}|\theta=1000)}$$
$$= \left(\frac{1000}{2000}\right)^{20}\exp\left(\frac{20\bar{x}}{1000}-\frac{20\bar{x}}{2000}\right)$$
$$= 2^{-20}\exp\left(\frac{\bar{x}}{100}\right)$$

- Using the Neyman-Pearson lemma, the most powerful test of significance α has critical region

$$\begin{aligned} C &= \left\{ x_1, \ldots, x_{20} : 2^{-20} \exp\left(\frac{\bar{x}}{100}\right) < c \right\} \\ &= \left\{ x_1, \ldots, x_{20} : \exp\left(\frac{\bar{x}}{100}\right) < 2^{20} c \right\} \\ &= \{ x_1, \ldots, x_{20} : \bar{x} < 100 \left(\log c + 20 \log 2\right) \} \end{aligned}$$

- **The type I error is**

$$
\begin{aligned}
&\mathbb{P}_{\theta_0}\left(\bar{X} < 100\left(\log c + 20\log 2\right)\right)\\
=&\mathbb{P}_{\theta_0}\left(\bar{X} < 100\left(\log c + 20\log 2\right)|\theta = 2000\right)\\
=&\mathbb{P}_{\theta_0}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \leq 100\left(\log c + 20\log 2\right)\right)\\
=&\mathbb{P}_{\theta_0}\left(\sum_{i=1}^{20} X_i \leq 20 \times 100\left(\log c + 20\log 2\right)\right)\\
=&\mathbb{P}_{\theta_0}\left(Y \leq 20 \times 100\left(\log c + 20\log 2\right)\right)
\end{aligned}
$$

where $Y = \sum_{i=1}^{20} X_i$

- It turns out that the sum of n independent exponentially distribution R.V.s with parameter $\frac{1}{\theta}$ follows the Gamma distribution with parameters n and $\frac{1}{\theta}$

$$Y \sim \Gamma\left(n, \frac{1}{\theta}\right)$$

$\Longrightarrow$ So we can calculate (using matlab)

$$\mathbb{P}_{\theta_0}\left(Y \leq 20 \times 100\left(\log \lambda - 20 \log 2\right)\right)$$

Objectives

Now you should be able to :

- understand the concepts of significance level, power, error of type I and of type II □
- explain and use the relationship between confidence intervals and hypothesis tests □
- construct the LRT with nuisance parameters □

Put yourself to the test ! $\rightsquigarrow$ Q3 p.402, Q6 p.403, Q7 p.403, Q15 p.404, Q17 p.405