Statistics

MATH 324

McGill University, Montréal, Canada

Fall 2018

In this section we will discuss:

(1) Principles of hypothesis testing

(2) Neyman-Pearson Lemma

(3) Likelihood ratio statistic

(4) Common hypothesis testing problems

Introduction

Based on an iid sample $X_1, X_2, \ldots, X_n$ from a parametric distribution $f(\cdot;\theta)$ with unknown parameter θ, we discussed:

(i) Point estimator(s) of θ:

$$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$$

e.g. method of moment or maximum likelihood estimators

(ii) $100(1-\alpha)\%$ confidence interval(s) for θ:

$$P(L \leq \theta \leq U) = 1 - \alpha$$

(iii) Next, we will discuss hypothesis testing about θ, which is very different from (i)-(ii).

Hypothesis testing problem

- A statistical hypothesis test is a “decision rule” that uses the data to infer which of two mutually exclusive hypotheses, that reflect two competing hypothetical states of the nature, is correct.
- The decision rule partitions the sample space $\mathcal{X}$ into two regions that respectively reflect support for the two hypotheses.
- Note: $\mathcal{X}$ is the set of all possible values of $(x_1, x_2, \ldots, x_n)$.

Hypothesis testing problem

- A statistical hypothesis test is a “decision rule” that uses the data to infer which of two mutually exclusive hypotheses, that reflect two competing hypothetical states of the nature, is correct.
- The decision rule partitions the sample space $\mathcal{X}$ into two regions that respectively reflect support for the two hypotheses.
- Note: $\mathcal{X}$ is the set of all possible values of $(x_1, x_2, \ldots, x_n)$.

McGill

Hypothesis testing problem

- A statistical hypothesis test is a “decision rule” that uses the data to infer which of two mutually exclusive hypotheses, that reflect two competing hypothetical states of the nature, is correct.
- The decision rule partitions the sample space $\mathcal{X}$ into two regions that respectively reflect support for the two hypotheses.
- Note: $\mathcal{X}$ is the set of all possible values of $(x_1, x_2, \ldots, x_n)$.

The null and alternative hypotheses

- Two hypotheses that characterize the two possible states of the nature are:

  $\mathcal{H}_0$: null hypothesis

  $\mathcal{H}_1$: alternative hypothesis

- The alternative hypothesis $\mathcal{H}_1$ is usually the one that we seek to support on the basis of the data.

The null and alternative hypotheses

- Two hypotheses that characterize the two possible states of the nature are:

  $\mathcal{H}_0$: null hypothesis

  $\mathcal{H}_1$: alternative hypothesis

- The alternative hypothesis $\mathcal{H}_1$ is usually the one that we seek to support on the basis of the data.

Hypothesis testing problem in a parametric family

- Consider the parameter space Θ, where $\theta \in \Theta$.
- Suppose Θ is partitioned into two disjoint subsets Θ_0 and Θ_1 such that
$$\Theta = \Theta_0 \cup \Theta_1$$
- Rather than estimating θ, the goal is to decide (based on the data $X_1, X_2, \ldots, X_n$) whether the unknown θ lies in Θ_0 or in Θ_1.
- Note: the choice of Θ_0 and Θ_1 are decided by the researcher.

Hypothesis testing problem in a parametric family

- Consider the parameter space Θ, where $\theta \in \Theta$.
- Suppose Θ is partitioned into two disjoint subsets Θ_0 and Θ_1 such that

$$\Theta = \Theta_0 \cup \Theta_1$$

- Rather than estimating θ, the goal is to decide (based on the data $X_1, X_2, \ldots, X_n$) whether the unknown θ lies in Θ_0 or in Θ_1.
- Note: the choice of Θ_0 and Θ_1 are decided by the researcher.

Hypothesis testing problem in a parametric family

- Consider the parameter space Θ, where $\theta \in \Theta$.
- Suppose Θ is partitioned into two disjoint subsets Θ_0 and Θ_1 such that
$$\Theta = \Theta_0 \cup \Theta_1$$
- Rather than estimating θ, the goal is to decide (based on the data $X_1, X_2, \ldots, X_n$) whether the unknown θ lies in Θ_0 or in Θ_1.
- Note: the choice of Θ_0 and Θ_1 are decided by the researcher.

The null and alternative hypotheses in a parametric family

Defintion:

For a parametric family $\mathcal{F} = \{f(x; \theta) : \theta \in \Theta\}$, set

$\mathcal{H}_0 : \theta \in \Theta_0$

$\mathcal{H}_1 : \theta \in \Theta_1$

such that $\Theta = \Theta_0 \cup \Theta_1$ and $\Theta_0 \cap \Theta_1 = \emptyset$.

- The goal would be to test $\mathcal{H}_0$ versus $\mathcal{H}_1$ using the data.

The null and alternative hypotheses in a parametric family

Defintion:

For a parametric family $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta\}$, set

$\mathcal{H}_0 : \theta \in \Theta_0$

$\mathcal{H}_1 : \theta \in \Theta_1$

such that $\Theta = \Theta_0 \cup \Theta_1$ and $\Theta_0 \cap \Theta_1 = \emptyset$.

- The goal would be to test $\mathcal{H}_0$ versus $\mathcal{H}_1$ using the data.

Example

Suppose a political candidate, say John, claims that he will gain more than 50% of the votes in a city election and thereby he emerges as the winner. If we consider the votes as a Bernoulli random sample with unknown probability $p \in (0, 1)$, the testing problem under consideration is the following:

$$\begin{aligned} \mathcal{H}_0 : \quad & p \geq .50, \\ \mathcal{H}_1 : \quad & p < .50. \end{aligned}$$

In other words, the null hypothesis is that John will win the election, while the alternative is that he will loose.

Example

Suppose a political candidate, say John, claims that he will gain more than 50% of the votes in a city election and thereby he emerges as the winner. If we consider the votes as a Bernoulli random sample with unknown probability $p \in (0,1)$, the testing problem under consideration is the following:

$$\begin{aligned} \mathcal{H}_0 &: \quad p \geq .50, \\ \mathcal{H}_1 &: \quad p < .50. \end{aligned}$$

In other words, the null hypothesis is that John will win the election, while the alternative is that he will loose.

Statistical Test procedure

Defintion:

A statistical procedure that is used to decide whether to reject or not to reject the null hypothesis $\mathcal{H}_0$ in favour of the alternative $\mathcal{H}_1$ is called a statistical test procedure, or simply a test.

- A test defines a partition of the sample space $\mathcal{X}$ into two regions. The hypothesis $\mathcal{H}_0$ is then rejected in favour of $\mathcal{H}_1$ depending where the data $X_1, X_2, \ldots, X_n$ or a suitably chosen statistic $T(X_1, X_2, \ldots, X_n)$ fall within $\mathcal{X}$.
- The $T(X_1, X_2, \ldots, X_n)$ is called test statistic.

Statistical Test procedure

Defintion:

A statistical procedure that is used to decide whether to reject or not to reject the null hypothesis $\mathcal{H}_0$ in favour of the alternative $\mathcal{H}_1$ is called a statistical test procedure, or simply a test.

- A test defines a partition of the sample space $\mathcal{X}$ into two regions. The hypothesis $\mathcal{H}_0$ is then rejected in favour of $\mathcal{H}_1$ depending where the data $X_1, X_2, \ldots, X_n$ or a suitably chosen statistic $T(X_1, X_2, \ldots, X_n)$ fall within $\mathcal{X}$.
- The $T(X_1, X_2, \ldots, X_n)$ is called test statistic.

Statistical Test procedure

Defintion:

A statistical procedure that is used to decide whether to reject or not to reject the null hypothesis $\mathcal{H}_0$ in favour of the alternative $\mathcal{H}_1$ is called a statistical test procedure, or simply a test.

- A test defines a partition of the sample space $\mathcal{X}$ into two regions. The hypothesis $\mathcal{H}_0$ is then rejected in favour of $\mathcal{H}_1$ depending where the data $X_1, X_2, \ldots, X_n$ or a suitably chosen statistic $T(X_1, X_2, \ldots, X_n)$ fall within $\mathcal{X}$.
- The $T(X_1, X_2, \ldots, X_n)$ is called test statistic.

Critical or rejection region

Defintion:

A test of $\mathcal{H}_0$ versus $\mathcal{H}_1$ consists of partitioning $\mathcal{X}$ into two regions $\mathcal{R}$ and $\mathcal{R}^c$, and rejecting $\mathcal{H}_0$ if and only if $(X_1, X_2, \ldots, X_n) \in \mathcal{R}$ (or $T(X_1, X_2, \ldots, X_n) \in \mathcal{R}$). The region $\mathcal{R}$ is called critical region or rejection region of the test.

Errors in making decisions about the two hypotheses

- Any given statistical test can make two types of errors (mistakes):

		Decision	
		$\mathcal{H}_0$	$\mathcal{H}_1$
Truth	$\mathcal{H}_0$	✓	×
	$\mathcal{H}_1$	×	✓

- Type I error: is made if $\mathcal{H}_0$ is rejected when $\mathcal{H}_0$ is true.
- Type II error: is made if $\mathcal{H}_0$ is not rejected when $\mathcal{H}_1$ is true.

Errors in making decisions about the two hypotheses

- Any given statistical test can make two types of errors (mistakes):

		Decision	
		$\mathcal{H}_0$	$\mathcal{H}_1$
Truth	$\mathcal{H}_0$	✓	×
	$\mathcal{H}_1$	×	✓

- Type I error: is made if $\mathcal{H}_0$ is rejected when $\mathcal{H}_0$ is true.
- Type II error: is made if $\mathcal{H}_0$ is not rejected when $\mathcal{H}_1$ is true.

Power function of a test

Defintion:

Consider a statistical test, say δ, with a rejection region $\mathcal{R}$. The power function of the test is given by

$$\pi(\theta) = P\{\text{rejecting } \mathcal{H}_0 \text{ when the parameter value is } \theta \in \Theta\}$$

$$\equiv P\{\text{rejecting } \mathcal{H}_0 | \theta \in \Theta\} \equiv P_\theta\{\text{rejecting } \mathcal{H}_0\}.$$

Probability of Type I & II errors of a test δ

- Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

$$\alpha(\delta) = P(\text{Type I error}) = P\{\text{rejecting } \mathcal{H}_0 \text{ when } \theta \in \Theta_0\}$$

$$\equiv P\{\text{rejecting } \mathcal{H}_0 | \theta \in \Theta_0\} = P\{\boldsymbol{x} \in R | \theta \in \Theta_0\}$$

$$\beta(\delta) = P(\text{Type II error}) = P\{\text{not rejecting } \mathcal{H}_0 \text{ when } \theta \in \Theta_1\}$$

$$\equiv P\{\text{not rejecting } \mathcal{H}_0 | \theta \in \Theta_1\} = P(\boldsymbol{x} \in R^c | \theta \in \Theta_1).$$

Probability of Type I & II errors of a test δ

- Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

$$\alpha(\delta) = P(\text{Type I error}) = P\{\text{rejecting } \mathcal{H}_0 \text{ when } \theta \in \Theta_0\}$$

$$\equiv P\{\text{rejecting } \mathcal{H}_0 | \theta \in \Theta_0\} = P\{\boldsymbol{x} \in R | \theta \in \Theta_0\}$$

$$\beta(\delta) = P(\text{Type II error}) = P\{\text{not rejecting } \mathcal{H}_0 \text{ when } \theta \in \Theta_1\}$$

$$\equiv P\{\text{not rejecting } \mathcal{H}_0 | \theta \in \Theta_1\} = P(\boldsymbol{x} \in R^c | \theta \in \Theta_1).$$

Controlling the errors

- Ideally, given a test δ we would like to have $\alpha(\delta) = \beta(\delta) = 0$. But cannot do. It is also NOT possible to find a test for which both $\alpha(\delta)$ and $\beta(\delta)$ are arbitrarily small.
- It is, however, possible to find a testing procedure δ such that

$$a\,\alpha(\delta) + b\,\beta(\delta)$$

is minimized, for some given $a, b > 0$.

- We proceed by focusing on so-called simple hypotheses.

Controlling the errors

- Ideally, given a test δ we would like to have $\alpha(\delta) = \beta(\delta) = 0$. But cannot do. It is also NOT possible to find a test for which both $\alpha(\delta)$ and $\beta(\delta)$ are arbitrarily small.

- It is, however, possible to find a testing procedure δ such that

$$a\,\alpha(\delta) + b\,\beta(\delta)$$

  is minimized, for some given $a, b > 0$.

- We proceed by focusing on so-called simple hypotheses.

Simple hypotheses

- Consider a situation where $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, where θ_0 and θ_1 are known. Then, we have

$$\begin{aligned} \mathcal{H}_0 : &\quad \theta = \theta_0 \\ \mathcal{H}_1 : &\quad \theta = \theta_1 . \end{aligned}$$

  These are called simple hypotheses.

- The goal is to design an optimal test for the above hypothesis testing problem.

Simple hypotheses

- Consider a situation where $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, where θ_0 and θ_1 are known. Then, we have

$$\begin{aligned} \mathcal{H}_0 : &\quad \theta = \theta_0 \\ \mathcal{H}_1 : &\quad \theta = \theta_1 . \end{aligned}$$

  These are called simple hypotheses.

- The goal is to design an optimal test for the above hypothesis testing problem.

Optimal tests: (Proof of this theorem is posted in myCourses)

- Suppose that $a, b > 0$ are specified constants. Also, denote

$$f(\boldsymbol{x}; \theta) = f(x_1; \theta) \times f(x_2; \theta) \times \ldots \times f(x_n; \theta)$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Theorem: Let δ^* denote a test procedure such that $\mathcal{H}_0$ is rejected if

$$a\, f(\boldsymbol{x}; \theta_0) < b\, f(\boldsymbol{x}; \theta_1)$$

and $\mathcal{H}_0$ is not rejected if

$$a\, f(\boldsymbol{x}; \theta_0) > b\, f(\boldsymbol{x}; \theta_1).$$

If $a\, f(\boldsymbol{x}; \theta_0) = b\, f(\boldsymbol{x}; \theta_1)$, either $\mathcal{H}_0$ is rejected or not. Then for any other test procedure δ,

$$a\, \alpha(\delta^*) + b\, \beta(\delta^*) \leq a\, \alpha(\delta) + b\, \beta(\delta).$$

Optimal tests: (Proof of this theorem is posted in myCourses)

- Suppose that $a, b > 0$ are specified constants. Also, denote

$$f(\boldsymbol{x};\theta) = f(x_1;\theta) \times f(x_2;\theta) \times \ldots \times f(x_n;\theta)$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Theorem: Let δ^* denote a test procedure such that $\mathcal{H}_0$ is rejected if

$$a\, f(\boldsymbol{x};\theta_0) < b\, f(\boldsymbol{x};\theta_1)$$

and $\mathcal{H}_0$ is not rejected if

$$a\, f(\boldsymbol{x};\theta_0) > b\, f(\boldsymbol{x};\theta_1).$$

If $a\, f(\boldsymbol{x};\theta_0) = b\, f(\boldsymbol{x};\theta_1)$, either $\mathcal{H}_0$ is rejected or not. Then for any other test procedure δ,

$$a\, \alpha(\delta^*) + b\, \beta(\delta^*) \leq a\, \alpha(\delta) + b\, \beta(\delta).$$

Critical region of the optimal test δ^*

- Note that by the theorem, given $\boldsymbol{x}$, the optimal test δ^* rejects $\mathcal{H}_0$ if

$$a\, f(\boldsymbol{x};\theta_0) < b\, f(\boldsymbol{x};\theta_1) \Longleftrightarrow \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b}.$$

In other words, the critical region for such test is

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b} \right\}.$$

- The ratio $\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} = \frac{L_n(\theta_1)}{L_n(\theta_0)}$ is called the likelihood ratio of the sample.

Critical region of the optimal test δ^*

- Note that by the theorem, given $\boldsymbol{x}$, the optimal test δ^* rejects $\mathcal{H}_0$ if

$$a\, f(\boldsymbol{x};\theta_0) < b\, f(\boldsymbol{x};\theta_1) \Longleftrightarrow \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b}.$$

In other words, the critical region for such test is

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b} \right\}.$$

- The ratio $\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} = \frac{L_n(\theta_1)}{L_n(\theta_0)}$ is called the likelihood ratio of the sample.

Critical region of the optimal test δ^*

- Note that by the theorem, given $\boldsymbol{x}$, the optimal test δ^* rejects $\mathcal{H}_0$ if

$$a\, f(\boldsymbol{x};\theta_0) < b\, f(\boldsymbol{x};\theta_1) \Longleftrightarrow \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b}.$$

In other words, the critical region for such test is

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} > \frac{a}{b} \right\}.$$

- The ratio $\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} = \frac{L_n(\theta_1)}{L_n(\theta_0)}$ is called the likelihood ratio of the sample.

Minimizing type I error

- In practice, we fix a specified upper bound α for the probability of type I error, which is called significance level α. The common choices of α are $\{0.1, .0.5, 0.01\}$.
- We then try to design a test δ^* whose probability of type I error is below a specified significance level α and has the probability of an error of type II as small as possible. Neyman-Pearson Lemma comes to the rescue!

McGill

Minimizing type I error

- In practice, we fix a specified upper bound α for the probability of type I error, which is called significance level α. The common choices of α are $\{0.1, .0.5, 0.01\}$.
- We then try to design a test δ^* whose probability of type I error is below a specified significance level α and has the probability of an error of type II as small as possible. Neyman-Pearson Lemma comes to the rescue!

Neyman-Pearson Lemma

- Recall the simple hypotheses, $\begin{array}{ll} \mathcal{H}_0: & \theta = \theta_0 \\ \mathcal{H}_1: & \theta = \theta_1. \end{array}$

  Theorem 10.1: Let δ^* denote a test such that $\mathcal{H}_0$ is rejected if

  $$a\, f(\boldsymbol{x}; \theta_0) < b\, f(\boldsymbol{x}; \theta_1)$$

  and $\mathcal{H}_0$ is not rejected if

  $$a\, f(\boldsymbol{x}; \theta_0) > b\, f(\boldsymbol{x}; \theta_1).$$

  If $a\, f(\boldsymbol{x}; \theta_0) = b\, f(\boldsymbol{x}; \theta_1)$, either $\mathcal{H}_0$ is rejected or not. Then for any other test δ such that $\alpha(\delta) \leq \alpha(\delta^*)$, we have $\beta(\delta) \geq \beta(\delta^*)$.

Proof: Neyman-Pearson Lemma

Set $a = 1$ and $b = k$ in the previous theorem. Then, for any test δ

$$\alpha(\delta^*) + k\ \beta(\delta^*) \le \alpha(\delta) + k\ \beta(\delta).$$

It is easy to see that if $\alpha(\delta) \le \alpha(\delta^*)$, we must have

$$k\ \beta(\delta^*) \le k\ \beta(\delta) \Longleftrightarrow \beta(\delta^*) \le \beta(\delta).$$

Similarly, if $\alpha(\delta) < \alpha(\delta^*)$, then $\beta(\delta^*) < \beta(\delta)$.
This completes the proof.

The use of NP lemma:

- Suppose in the problem of testing a simple null hypothesis $\mathcal{H}_0 : \theta = \theta_0$ versus the simple alternative $\mathcal{H}_1 : \theta = \theta_1$, we wish the probability of Type I error to be at most α.
- Suppose we find a value k such that

$$\alpha(\delta^*) = P\left\{ f(\boldsymbol{X}; \theta_0) < k\, f(\boldsymbol{X}; \theta_1) \middle| \theta = \theta_0 \right\} \leq \alpha.$$

  where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.
- Then, the NP lemma guarantees that the probability of Type II error is the smallest possible, among all tests whose probability of Type I error is at most α.

The use of NP lemma:

- Suppose in the problem of testing a simple null hypothesis $\mathcal{H}_0 : \theta = \theta_0$ versus the simple alternative $\mathcal{H}_1 : \theta = \theta_1$, we wish the probability of Type I error to be at most α.
- Suppose we find a value k such that

$$\alpha(\delta^*) = P\Big\{ f(\boldsymbol{X};\theta_0) < k\, f(\boldsymbol{X};\theta_1) \Big| \theta = \theta_0 \Big\} \leq \alpha.$$

  where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.
- Then, the NP lemma guarantees that the probability of Type II error is the smallest possible, among all tests whose probability of Type I error is at most α.

The use of NP lemma:

- Suppose in the problem of testing a simple null hypothesis $\mathcal{H}_0 : \theta = \theta_0$ versus the simple alternative $\mathcal{H}_1 : \theta = \theta_1$, we wish the probability of Type I error to be at most α.

- Suppose we find a value k such that

$$\alpha(\delta^*) = P\left\{ f(\boldsymbol{X}; \theta_0) < k\, f(\boldsymbol{X}; \theta_1) \middle| \theta = \theta_0 \right\} \leq \alpha.$$

  where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

- Then, the NP lemma guarantees that the probability of Type II error is the smallest possible, among all tests whose probability of Type I error is at most α.

Example 1

Let $X_1, X_2, \ldots, X_n$ be iid from $N(\mu, 1)$ such that $\mu \in \{0, 1\}$. The hypothesis testing problem of interest is:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mu = 0 \\ \mathcal{H}_1 : &\quad \mu = 1. \end{aligned}$$

Find an optimal test δ^* for which $\alpha(\delta^*) \leq 0.01$; i.e. a test with smallest possible value of $\beta(\delta^*)$.

Example 1: Step 1

- First we need to construct the likelihood ratio $f(\boldsymbol{x};\theta_1)/f(\boldsymbol{x};\theta_0)$, where in our problem we have $\theta_0 = 0$ and $\theta_1 = 1$:

- The likelihood ratio is given by:

$$\frac{f(\boldsymbol{x};1)}{f(\boldsymbol{x};0)} = \frac{L_n(1)}{L_n(0)} = \frac{\frac{1}{(\sqrt{2\pi})^n}\exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-1)^2\right\}}{\frac{1}{(\sqrt{2\pi})^n}\exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}}$$

$$= \exp\{n\bar{x}_n - n/2\}.$$

Example 1: Step 1

- First we need to construct the likelihood ratio $f(\boldsymbol{x};\theta_1)/f(\boldsymbol{x};\theta_0)$, where in our problem we have $\theta_0 = 0$ and $\theta_1 = 1$:
- The likelihood ratio is given by:

$$\begin{aligned}\frac{f(\boldsymbol{x};1)}{f(\boldsymbol{x};0)} &= \frac{L_n(1)}{L_n(0)} = \frac{\frac{1}{(\sqrt{2\pi})^n}\exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2\right\}}{\frac{1}{(\sqrt{2\pi})^n}\exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}} \\ &= \exp\{n\bar{x}_n - n/2\}.\end{aligned}$$

Example 1: Step 2

- Using the NP lemma, we construct the critical region:

$$\begin{aligned}
\mathcal{R} &= \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{f(\boldsymbol{x};1)}{f(\boldsymbol{x};0)} > \frac{1}{k} \right\} \\
&= \left\{ \boldsymbol{x} \in \mathcal{X} : \exp\{n\bar{x}_n - n/2\} > \frac{1}{k} \right\} \Longrightarrow \\
\mathcal{R} &= \left\{ \boldsymbol{x} \in \mathcal{X} : \bar{x}_n > k^* \right\},
\end{aligned}$$

where

$$k^* = \frac{1}{2} - \frac{\ln k}{n} \ , \ \ \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

...

That is, given the data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we will reject the null hypothesis $\mathcal{H}_0 : \mu = 0$ in favour of the alternative $\mathcal{H}_1 : \mu = 1$, if $\bar{x}_n > k^*$.

Example 1: Step 3

- Given the significance level $\alpha = 0.01$, calculate the value of k^*:

$$\alpha(\delta^*) = P(\overline{X}_n > k^* | \mu = 0) = 0.01$$

- Note that under the null hypothesis that $\mu = 0$, we have that $\overline{X}_n \sim N(0, \frac{1}{n})$, and therefore,

$$P(\overline{X}_n > k^* | \mu = 0) = P(\sqrt{n}\overline{X}_n > \sqrt{n}k^* | \mu = 0) = 0.01$$

- Using the standard Normal table, we must have $\sqrt{n}k^* = z_{0.01} = 2.326$ which implies that $k^* = 2.326/\sqrt{n}$.

Example 1: ...

- Put together, at the significance level $\alpha = 0.01$, the optimal test δ^* rejects the null hypothesis $\mathcal{H}_0 : \mu = 0$ in favour of the alternative $\mathcal{H}_1 : \mu = 1$, if
$$\bar{x}_n > \frac{2.326}{\sqrt{n}}.$$

- Note:

  the test statistic in this example is $T(\boldsymbol{X}) = \overline{X}_n$, and $k^* = \frac{2.326}{\sqrt{n}}$ is called critical value.

Example 1: ...

- Put together, at the significance level $\alpha = 0.01$, the optimal test δ^* rejects the null hypothesis $\mathcal{H}_0 : \mu = 0$ in favour of the alternative $\mathcal{H}_1 : \mu = 1$, if
$$\bar{x}_n > \frac{2.326}{\sqrt{n}}.$$

- Note:

  the test statistic in this example is $T(\boldsymbol{X}) = \overline{X}_n$, and $k^* = \frac{2.326}{\sqrt{n}}$ is called critical value.

Example 1: calculating $\beta(\delta^*)$

- The probability of Type II error of the above test: by definition,

$$\beta(\delta^*) = P\left(\text{not rejecting } \mathcal{H}_0 \middle| \mu \in \Theta_1\right) = P\left(\overline{X}_n \le \frac{2.326}{\sqrt{n}} \middle| \mu = 1\right)$$

- Note that under the alternative hypothesis that $\mu = 1$, we have that $\overline{X}_n \sim N(1, \frac{1}{n})$, and therefore,

$$\beta(\delta^*) = P\left(\sqrt{n}(\overline{X}_n - 1) \le 2.326 - \sqrt{n} \middle| \mu = 1\right) = \Phi(2.326 - \sqrt{n}).$$

  where $\Phi(\cdot)$ is the CDF of the standard Normal distribution.

- Note that for this test, $\alpha(\delta^*) = 0.01$ but $\beta(\delta^*)$ changes with n. For example, if $n = 20$, then $\beta(\delta^*) = 0.0159$.

McGill

Example 1: calculating $\beta(\delta^*)$

- The probability of Type II error of the above test: by definition,

$$\beta(\delta^*) = P\left(\text{not rejecting } \mathcal{H}_0 \middle| \mu \in \Theta_1\right) = P\left(\overline{X}_n \le \frac{2.326}{\sqrt{n}} \middle| \mu = 1\right)$$

- Note that under the alternative hypothesis that $\mu = 1$, we have that $\overline{X}_n \sim N(1, \frac{1}{n})$, and therefore,

$$\beta(\delta^*) = P\left(\sqrt{n}(\overline{X}_n - 1) \le 2.326 - \sqrt{n} \middle| \mu = 1\right) = \Phi(2.326 - \sqrt{n}).$$

  where $\Phi(\cdot)$ is the CDF of the standard Normal distribution.

- Note that for this test, $\alpha(\delta^*) = 0.01$ but $\beta(\delta^*)$ changes with n. For example, if $n = 20$, then $\beta(\delta^*) = 0.0159$.

Example 2

Let $X_1, X_2, \ldots, X_n$ be iid from $\text{Ber}(p)$ such that $p \in \{0.2, 0.4\}$. The hypothesis testing problem of interest is:

$$\begin{aligned} \mathcal{H}_0 : &\quad p = 0.2 \\ \mathcal{H}_1 : &\quad p = 0.4. \end{aligned}$$

Find an optimal test δ^* for which $\alpha(\delta^*) \leq 0.05$; i.e. a test with smallest possible value of $\beta(\delta^*)$.

Example 2: Step 1

- First we need to construct the likelihood ratio $f(\boldsymbol{x}; p_1)/f(\boldsymbol{x}; p_0)$, where in our problem we have $p_0 = 0.2$ and $p_1 = 0.4$:
- The likelihood ratio is given by:

$$\frac{f(\boldsymbol{x}; 0.4)}{f(\boldsymbol{x}; 0.2)} = \frac{L_n(0.4)}{L_n(0.2)} = \frac{(0.4)^{\sum_{i=1}^n x_i} \times (0.6)^{n-\sum_{i=1}^n x_i}}{(0.2)^{\sum_{i=1}^n x_i} \times (0.8)^{n-\sum_{i=1}^n x_i}}$$

$$= \left(\frac{3}{4}\right)^n \left(\frac{8}{3}\right)^{\sum_{i=1}^n x_i}.$$

Example 2: Step 1

- First we need to construct the likelihood ratio $f(\boldsymbol{x}; p_1)/f(\boldsymbol{x}; p_0)$, where in our problem we have $p_0 = 0.2$ and $p_1 = 0.4$:
- The likelihood ratio is given by:

$$\frac{f(\boldsymbol{x}; 0.4)}{f(\boldsymbol{x}; 0.2)} = \frac{L_n(0.4)}{L_n(0.2)} = \frac{(0.4)^{\sum_{i=1}^n x_i} \times (0.6)^{n-\sum_{i=1}^n x_i}}{(0.2)^{\sum_{i=1}^n x_i} \times (0.8)^{n-\sum_{i=1}^n x_i}}$$

$$= \left(\frac{3}{4}\right)^n \left(\frac{8}{3}\right)^{\sum_{i=1}^n x_i}.$$

Example 2: Step 2

- Using the NP lemma, we construct the critical region:

$$\begin{aligned}
\mathcal{R} &= \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{f(\boldsymbol{x};1)}{f(\boldsymbol{x};0)} > \frac{1}{k} \right\} \\
&= \left\{ \boldsymbol{x} \in \mathcal{X} : \left(\frac{3}{4}\right)^n \left(\frac{8}{3}\right)^{\sum_{i=1}^n x_i} > \frac{1}{k} \right\} \Longrightarrow \\
\mathcal{R} &= \left\{ \boldsymbol{x} \in \mathcal{X} : \sum_{i=1}^n x_i > k^* \right\},
\end{aligned}$$

where

$$k^* = \frac{-\ln(k) + n\ln(4/3)}{\ln(8/3)}.$$

Example 2: Step 2...

That is, given the data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we will reject the null hypothesis $\mathcal{H}_0 : p = 0.2$ in favour of the alternative $\mathcal{H}_1 : p = 0.4$, if $\sum_{i=1}^{n} x_i > k^*$.

- Note:

the test statistic in this example is $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$, and k^* is called critical value.

Example 2: Step 2...

That is, given the data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we will reject the null hypothesis $\mathcal{H}_0 : p = 0.2$ in favour of the alternative $\mathcal{H}_1 : p = 0.4$, if $\sum_{i=1}^{n} x_i > k^*$.

- Note:

  the test statistic in this example is $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$, and k^* is called critical value.

Example 2: Step 3

- Given $\alpha = 0.05$, we must calculate the value of k^* such that:

$$\alpha(\delta^*) = P\left(\sum_{i=1}^{n} X_i > k^* \middle| p = 0.2\right) = 0.05 \tag{1}$$

- Note that under the null hypothesis $\mathcal{H}_0 : p = 0.2$, we have that $\sum_{i=1}^{n} X_i \sim \text{Bin}(n, 0.2)$, and therefore k^* can only be an integer and as a consequence the desired significance level may not be attainable.

Example 2: Step 3...

- For example, if $n = 10$, then under $\mathcal{H}_0$, $\sum_{i=1}^{10} X_i \sim \text{Bin}(10, 0.2)$.
  Using the Binomial table,

$$P\left(\sum_{i=1}^{10} X_i > 4 \middle| p = 0.2\right) = 0.033,$$

$$P\left(\sum_{i=1}^{10} X_i > 3 \middle| p = 0.2\right) = 0.121.$$

- Therefore, we cannot find a k^* such that (1) is satisfied !
  What do we do? (Randomized tests)

Example 2: Step 3...

- For example, if $n = 10$, then under $\mathcal{H}_0$, $\sum_{i=1}^{10} X_i \sim \text{Bin}(10, 0.2)$. Using the Binomial table,

$$P\left(\sum_{i=1}^{10} X_i > 4 \middle| p = 0.2\right) = 0.033,$$

$$P\left(\sum_{i=1}^{10} X_i > 3 \middle| p = 0.2\right) = 0.121.$$

- Therefore, we cannot find a k^* such that (1) is satisfied ! What do we do? (Randomized tests)

Example 2: Step 3...

- For example, if $n = 10$, then under $\mathcal{H}_0$, $\sum_{i=1}^{10} X_i \sim \text{Bin}(10, 0.2)$. Using the Binomial table,

$$P\left(\sum_{i=1}^{10} X_i > 4 \middle| p = 0.2\right) = 0.033,$$

$$P\left(\sum_{i=1}^{10} X_i > 3 \middle| p = 0.2\right) = 0.121.$$

- Therefore, we cannot find a k^* such that (1) is satisfied ! What do we do? (Randomized tests)

Randomized tests

- We can design a so-called randomized optimal test that will attain the desired significance level exactly.
- In Example 2 when $n = 10$, we can consider the following test:

$$\delta^* = \begin{cases} \text{reject } \mathcal{H}_0 & , \text{if } \sum_{i=}^{n} x_i > 4 \\ \text{reject } \mathcal{H}_0 \text{ with probability } 0.195 & , \text{if } \sum_{i=}^{n} x_i = 4 \\ \text{do not reject } \mathcal{H}_0 & , \text{if } \sum_{i=}^{n} x_i < 4 \end{cases}$$

in other words, when $\sum_{i=}^{n} x_i = 4$, we toss a coin to decide about $\mathcal{H}_0$. It can be shown that this test has $\alpha(\delta^*) = 0.05$.

Randomized tests

- However, randomized tests may be good in theory but they are not popular in practice. We do not want to make a decision about the null hypothesis by tossing a coin !

Remarks

- If a test rejects a null hypothesis $\mathcal{H}_0$ at a significance level α, we are certain that the probability of Type I error is no larger than α. Therefore, it is considered safe to reject the null hypothesis.

- If a test does NOT reject null hypothesis $\mathcal{H}_0$ at a significance level α, we would NOT know what the probability of Type II error might be. Therefore, it is NOT considered safe to accept the null hypothesis. We say that:

  based on the given data, we do not have enough evidence to reject the null hypothesis at the significance level α.

Remarks

- If a test rejects a null hypothesis $\mathcal{H}_0$ at a significance level α, we are certain that the probability of Type I error is no larger than α. Therefore, it is considered safe to reject the null hypothesis.

- If a test does NOT reject null hypothesis $\mathcal{H}_0$ at a significance level α, we would NOT know what the probability of Type II error might be. Therefore, it is NOT considered safe to accept the null hypothesis. We say that:

  based on the given data, we do not have enough evidence to reject the null hypothesis at the significance level α.

More complex hypothesis testing problems

- As before, let $X_1, X_2, \dots, X_n$ be iid from a parametric distribution $f(\cdot;\theta)$ with unknown parameter θ. We are interested in testing

$$\begin{aligned} \mathcal{H}_0 : &\quad \theta \in \Theta_0 \\ \mathcal{H}_1 : &\quad \theta \in \Theta_1. \end{aligned}$$

- We discussed simple versus simple hypothesis testing problems. We now consider

$$\mathcal{H}_0 : \theta \leq \theta_0, \quad \mathcal{H}_0 : \theta \geq \theta_0, \quad \mathcal{H}_0 : \theta = \theta_0$$

$$\mathcal{H}_1 : \theta > \theta_0, \quad \mathcal{H}_1 : \theta < \theta_0, \quad \mathcal{H}_1 : \theta \neq \theta_0$$

where the alternatives are called composite while the null could be simple (if $=$) or composite (if $\leq$ or $\geq$).

Likelihood ratio statistic for hypothesis testing

- This is by far the most popular method of hypothesis testing in statistics. Recall the likelihood function based on an iid sample from $f(\cdot;\theta)$,

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The likelihood ratio statistic is given by

$$\lambda_n(\boldsymbol{X}) = \frac{\max_{\theta\in\Theta_0} L_n(\theta)}{\max_{\theta\in\Theta} L_n(\theta)}$$

i.e. the ratio of the likelihoods evaluated at the maximum likelihood estimators of θ under $\mathcal{H}_0 : \theta \in \Theta_0$ and when $\theta \in \Theta$.

Likelihood ratio statistic for hypothesis testing

- This is by far the most popular method of hypothesis testing in statistics. Recall the likelihood function based on an iid sample from $f(\cdot;\theta)$,

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The likelihood ratio statistic is given by

$$\lambda_n(\boldsymbol{X}) = \frac{\max_{\theta\in\Theta_0} L_n(\theta)}{\max_{\theta\in\Theta} L_n(\theta)}$$

i.e. the ratio of the likelihoods evaluated at the maximum likelihood estimators of θ under $\mathcal{H}_0 : \theta \in \Theta_0$ and when $\theta \in \Theta$.

Likelihood ratio statistic for hypothesis testing

- This is by far the most popular method of hypothesis testing in statistics. Recall the likelihood function based on an iid sample from $f(\cdot;\theta)$,

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The likelihood ratio statistic is given by

$$\lambda_n(\boldsymbol{X}) = \frac{\max_{\theta\in\Theta_0} L_n(\theta)}{\max_{\theta\in\Theta} L_n(\theta)}$$

i.e. the ratio of the likelihoods evaluated at the maximum likelihood estimators of θ under $\mathcal{H}_0 : \theta \in \Theta_0$ and when $\theta \in \Theta$.

Likelihood ratio statistic for hypothesis testing

- A test based on the LR statistic has the following rejection region:

$$\mathcal{R} = \{\boldsymbol{x} \in \mathcal{X} : \lambda(\boldsymbol{x}) \leq C\}$$

for some $C \in [0, 1]$.

- For a given $\alpha \in (0, 1)$, if there exists a $C_\alpha \in [0, 1]$ such that

$$\max_{\theta \in \Theta_0} P\Big\{\lambda(\boldsymbol{X}) \leq C_\alpha\Big\} \leq \alpha$$

then the test is called of size α.

Likelihood ratio statistic for hypothesis testing

- A test based on the LR statistic has the following rejection region:

$$\mathcal{R} = \{\boldsymbol{x} \in \mathcal{X} : \lambda(\boldsymbol{x}) \leq C\}$$

for some $C \in [0, 1]$.

- For a given $\alpha \in (0, 1)$, if there exists a $C_\alpha \in [0, 1]$ such that

$$\max_{\theta \in \Theta_0} P\left\{\lambda(\boldsymbol{X}) \leq C_\alpha\right\} \leq \alpha$$

then the test is called of size α.

Specification of the critical value

- Under the Regularity Conditions, we have that for large n,

$$-2\ln[\lambda(\boldsymbol{X})] = 2\left[\max_{\theta\in\Theta} l_n(\theta) - \max_{\theta\in\Theta_0} l_n(\theta)\right]$$

has approximately a chi-squared distribution, $\chi^2_{(d)}$, where $d = \dim(\Theta) - \dim(\Theta_0)$.

- Note that

$$\lambda(\boldsymbol{X}) \leq C_\alpha \Longleftrightarrow -2\ln[\lambda(\boldsymbol{X})] \geq C^*_\alpha.$$

where $C^*_\alpha = -2\ln[C_\alpha]$.

Specification of the critical value

- Under the Regularity Conditions, we have that for large n,

$$-2\ln[\lambda(\boldsymbol{X})] = 2\left[\max_{\theta\in\Theta} l_n(\theta) - \max_{\theta\in\Theta_0} l_n(\theta)\right]$$

  has approximately a chi-squared distribution, $\chi^2_{(d)}$, where $d = \dim(\Theta) - \dim(\Theta_0)$.

- Note that

$$\lambda(\boldsymbol{X}) \leq C_\alpha \Longleftrightarrow -2\ln[\lambda(\boldsymbol{X})] \geq C^*_\alpha.$$

  where $C^*_\alpha = -2\ln[C_\alpha]$.

...

- Thus, for large n,

$$\max_{\theta\in\Theta_0} P\Big\{\lambda(\boldsymbol{X}) \le C_\alpha\Big\} = \max_{\theta\in\Theta_0} P\Big\{-2\ln[\lambda(\boldsymbol{X})] \ge C_\alpha^*\Big\}$$

$$\approx \max_{\theta\in\Theta_0} P\Big\{\chi^2_{(d)} \ge C_\alpha^*\Big\} \le \alpha$$

which implies that $C_\alpha^* \ge \chi^2_{d;\alpha}$, from the chi-squared table.

...

- Put together, at the significance level α, the rejection region of the LR-based test is given by:

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : -2\ln[\lambda(\boldsymbol{x})] \geq \chi^2_{d;\alpha} \right\},$$

where

$$-2\ln[\lambda(\boldsymbol{x})] = 2\left\{ \max_{\theta\in\Theta} l_n(\theta) - \max_{\theta\in\Theta_0} l_n(\theta) \right\}$$

and $l_n(\theta)$ is the log-likelihood function.

- Note that the above rejection region is constructed using an approximation, that is why we said for large n.

...

- Put together, at the significance level α, the rejection region of the LR-based test is given by:

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : -2\ln[\lambda(\boldsymbol{x})] \geq \chi^2_{d;\alpha} \right\},$$

  where

$$-2\ln[\lambda(\boldsymbol{x})] = 2\left\{ \max_{\theta \in \Theta} l_n(\theta) - \max_{\theta \in \Theta_0} l_n(\theta) \right\}$$

  and $l_n(\theta)$ is the log-likelihood function.

- Note that the above rejection region is constructed using an approximation, that is why we said for large n.

Example 3

- Let $X_1, X_2, \ldots, X_n$ be iid from $N(\mu, \sigma^2)$ and both parameters are unknown. We wish to test

$$\begin{aligned} \mathcal{H}_0 : &\quad \mu = \mu_0 \\ \mathcal{H}_1 : &\quad \mu \neq \mu_0. \end{aligned}$$

for some known μ_0.

Using the LR statistic, design a statistical test at a significance level $\alpha \in (0, 1)$.

Example 3: Step 1

- The likelihood function is given by

$$L_n(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

- We have that

$$\begin{aligned} \Theta &= \left\{ (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \\ \Theta_0 &= \left\{ (\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0 \right\} \end{aligned}$$

Note that μ_0 is known.

Example 3: Step 2

- Obtain the maximum likelihood estimates of the parameters over each of Θ and Θ_0.
- Log-likelihood function:

$$l_n(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

- We have already shown that the maximum likelihood estimates over Θ are: (note that these are the ordinary MLEs)

$$\hat{\mu}_n = \bar{x}_n \quad , \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{(n-1)s_n^2}{n}.$$

Example 3: Step 2

- Obtain the maximum likelihood estimates of the parameters over each of Θ and Θ_0.

- Log-likelihood function:

$$l_n(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

- We have already shown that the maximum likelihood estimates over Θ are: (note that these are the ordinary MLEs)

$$\hat{\mu}_n = \bar{x}_n \quad , \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{(n-1)s_n^2}{n}.$$

...

- Next, we obtain the MLE over the restricted parameter space Θ_0. Over this space, $\mu = \mu_0$ is already known.

  We only need to estimate σ^2 over Θ_0. It is straightforward to see that the maximizer of $l_n(\mu_0, \sigma^2)$ with respect to σ^2 is:

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2.$$

  which can be re-written as

$$\tilde{\sigma}_n^2 = \frac{(n-1)s_n^2}{n} + (\bar{x}_n - \mu_0)^2$$

...

- Next, we obtain the MLE over the restricted parameter space Θ_0. Over this space, $\mu = \mu_0$ is already known.

  We only need to estimate σ^2 over Θ_0. It is straightforward to see that the maximizer of $l_n(\mu_0, \sigma^2)$ with respect to σ^2 is:

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

  which can be re-written as

$$\tilde{\sigma}_n^2 = \frac{(n-1)s_n^2}{n} + (\bar{x}_n - \mu_0)^2$$

Example 3: Step 3

- The LR statistic:

$$\lambda_n(\boldsymbol{X}) = \left(\frac{\hat{\sigma}_n^2}{\tilde{\sigma}_n^2}\right)^{n/2} = \left(\frac{1}{1 + \frac{n}{n-1}\frac{(\overline{X}_n-\mu_0)^2}{S_n^2}}\right)^{n/2} = \left(\frac{1}{1 + \frac{1}{n-1}T^2}\right)^{n/2}$$

  where $T = \frac{\sqrt{n}(\overline{X}_n-\mu_0)}{S_n}$.

- Based on the LR statistic, we reject the null hypothesis $\mathcal{H}_0 : \mu = \mu_0$ if and only if

$$\lambda_n(\boldsymbol{x}) \leq C \Longleftrightarrow \left|\frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}\right| > k$$

McGill

Example 3: Step 3

- The LR statistic:

$$\lambda_n(\boldsymbol{X}) = \left(\frac{\hat{\sigma}_n^2}{\tilde{\sigma}_n^2}\right)^{n/2} = \left(\frac{1}{1 + \frac{n}{n-1}\frac{(\overline{X}_n-\mu_0)^2}{S_n^2}}\right)^{n/2} = \left(\frac{1}{1 + \frac{1}{n-1}T^2}\right)^{n/2}$$

where $T = \frac{\sqrt{n}(\overline{X}_n-\mu_0)}{S_n}$.

- Based on the LR statistic, we reject the null hypothesis $\mathcal{H}_0 : \mu = \mu_0$ if and only if

$$\lambda_n(\boldsymbol{x}) \leq C \Longleftrightarrow \left|\frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}\right| > k$$

...

- Therefore, the critical region of the LR-based test δ^* is given by

$$\mathcal{R} = \left\{\boldsymbol{x} \in \mathcal{X} : \lambda_n(\boldsymbol{x}) \le C\right\} = \left\{\boldsymbol{x} \in \mathcal{X} : \left|\frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}\right| > k\right\}$$

for some $k > 0$.

Example 3: Step 3

- Given a significance level α, calculate the value of k:

$$\alpha(\delta^*) = P(|T| > k | \mu = \mu_0) = \alpha$$

where under $\mathcal{H}_0$, $T \sim t_{(n-1)}$, i.e. a Student t distribution with $(n-1)$ degrees of freedom.
Thus, we must have $k = t(n-1; \alpha/2)$.

- At the significance level α, we reject $\mathcal{H}_0 : \mu = \mu_0$ if and only if

$$\left| \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n} \right| > t(n-1; \alpha/2).$$

Example 3: Step 3

- Given a significance level α, calculate the value of k:

$$\alpha(\delta^*) = P(|T| > k|\mu = \mu_0) = \alpha$$

where under $\mathcal{H}_0$, $T \sim t_{(n-1)}$, i.e. a Student t distribution with $(n-1)$ degrees of freedom.
Thus, we must have $k = t(n-1; \alpha/2)$.

- At the significance level α, we reject $\mathcal{H}_0 : \mu = \mu_0$ if and only if

$$\left|\frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}\right| > t(n-1; \alpha/2).$$

Remark

- Note that in Example 3 we were able to construct a LR-based test that finally lead to a test statistic that has an exact Student t distribution.

  Instead, we could also directly use the χ^2 approximation to the LR statistic and use the rejection region on page 43/58:

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : -2\ln[\lambda(\boldsymbol{x})] \geq \chi^2_{d;\alpha} \right\} = \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{n}{2} \ln\left(\frac{\tilde{\sigma}_n^2}{\hat{\sigma}_n^2} \right) \geq \chi^2_{d;\alpha} \right\}.$$

Remark

- Note that in Example 3 we were able to construct a LR-based test that finally lead to a test statistic that has an exact Student t distribution.

  Instead, we could also directly use the χ^2 approximation to the LR statistic and use the rejection region on page 43/58:

$$\mathcal{R} = \left\{ \boldsymbol{x} \in \mathcal{X} : -2\ln[\lambda(\boldsymbol{x})] \geq \chi^2_{d;\alpha} \right\} = \left\{ \boldsymbol{x} \in \mathcal{X} : \frac{n}{2} \ln\left(\frac{\tilde{\sigma}^2_n}{\hat{\sigma}^2_n}\right) \geq \chi^2_{d;\alpha} \right\}.$$

Most common hypothesis testing problems

- I have posted a list of common hypothesis testing problems on myCourses. The test procedures are obtained using Neyman-Pearson Lemma or the LR statistic. Most of these tests, as a significance level $0 < \alpha < 1$, are so-called uniformly most powerful tests for the given hypotheses.
- These tests focus on the following parameters:

  population mean (μ), population variance σ^2, population proportion (p), difference in two population means ($\mu_1 - \mu_2$), difference in two population proportions ($p_1 - p_2$).

Most common hypothesis testing problems

- I have posted a list of common hypothesis testing problems on myCourses. The test procedures are obtained using Neyman-Pearson Lemma or the LR statistic. Most of these tests, as a significance level $0 < \alpha < 1$, are so-called uniformly most powerful tests for the given hypotheses.

- These tests focus on the following parameters:

  population mean (μ), population variance σ^2, population proportion (p), difference in two population means ($\mu_1 - \mu_2$), difference in two population proportions ($p_1 - p_2$).

p-value

- So far, at a significance level α, we have been using the observed value of a test statistic T, say T_{obs}, to test the null hypothesis H_0 versus the alternative hypothesis H_1, based on a rejection region.
- We can also test H_0 versus H_1 by using a quantity called **p-value**.

p-value

- p-value measures the strength of evidence against a null hypothesis H_0.
- It gives the minimum significance level for which we could have rejected the null hypothesis H_0 based on the observed value T_{obs} of the test statistic.

p-value

- p-value is in fact a conditional probability:

$$\text{p-value} =$$

P(observing a value as extreme or more extreme than T_{obs} for the test statistic $T \,\big|\, H_0$ is true).

- Note that a p-value is NOT the probability of the null hypothesis being false.

Calculating p-value

- For a one-sided test, p-value is simply the probability or area to the right (or left) of the observed value T_{obs} of the test statistic, under the null hypothesis H_0.
- For a two-sided test, p-value is equal to twice the probability or area beyond the absolute value of the observed value T_{obs} of the test statistic, under the null hypothesis H_0.

  (These will be carefully discussed in class).

How to use p-value?

- Smaller a p-value, stronger the evidence against the null hypothesis H_0.
- Note: having a small p-value is equivalent to observing an extreme value of the test statistic T, which leads to the rejection of the null hypothesis H_0.
- Just write it down, and make no decision and give it as a piece of evidence. Leave it to the reader (or expert) decide how strong they think the evidence is in the p-value, in favor of H_0.

Examples

- In the following slides we will discuss several hypothesis testing problems.

Example 4

- Atlantic bluefin tuna is the largest and most endangered of the tuna species; the concern is that this species has been overfished and that the mean weight has decreased. Suppose a random sample of 12 Atlantic blue fin tuna was obtained from commercial fishing boats and weighted. The sample is normally distributed with $\bar{x}_n = 535.7$ and $s_n = 37.8$. Is there any evidence that the mean weight is less than 550 pounds? Use the significance level $\alpha = 0.05$.

Example 5

- Despite a sophisticated recycling system, a water park informs the city water department of their need for 1 million liter of water per day. The city water department selected a random sample of $n = 21$ days; the mean and sample standard deviation of the park?s water usage (in thousands of liter) were $\bar{x}_n = 927.43$, $s_n = 154.45$. Assuming the usage is normally distributed, is there evidence to suggest the mean water usage is different from 1 million liter per day? Use the significance level $\alpha = 0.05$.

Example 6

- A study conducted by the Florida Game and Fish Commission aims at assessing the amounts of the DDT insecticide in the brain tissue of brown pelicans. Approximately Normal and independent samples of $n = 10$ juveniles and $m = 13$ nestlings gave (in parts per million),

  $$\bar{x}_n = 0.041 \,,\; s_n = 0.017 \;\;,\;\; \bar{y}_m = 0.026 \,,\; s_m = 0.006.$$

  Test whether the mean amounts of DDT in juveniles and nestlings are the same. Use the significance level $\alpha = 0.05$.

Example 7

A company produces machine engine parts that are supposed to have a diameter variance no larger than 0.0002.

A random sample of $n = 10$ parts gave a sample variance of 0.0003. We wish to test

$$\begin{aligned} \mathcal{H}_0: &\quad \sigma^2 \leq 0.0002 \\ \mathcal{H}_1: &\quad \sigma^2 > 0.0002. \end{aligned}$$

at the significance level $\alpha = 0.05$. Assume that the random sample is iid from $N(\mu, \sigma^2)$ with both parameters unknown.

Example 8

An experimenter was convinced that the variability in his/her measuring equipment results in a standard deviation of 2; $n = 16$ measurements yielded $s_n^2 = 6.1$. Do the data disagree with his/her claim? Use the significance level $\alpha = 0.05$. Assume the measurements are normally distributed with both mean and variance unknown.

Example 9

A study published in 2004 in Current Allergy & Clinical Immunology concerns the allergy to the powder on latex gloves. Among other things, the exposure to the powder of $n = 46$ hospital employees with diagnosed latex allergy was investigated. The number of latex gloves used per week by these sampled workers is summarized as

$$\bar{x}_n = 19.3 \quad , \quad s_n = 11.9.$$

Is there evidence to conclude that the mean number of latex gloves used per week by hospital employees with latex allergy is more than 15? Use $\alpha = 0.01$.

Example 10

A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results (in seconds) are summarized as

$$\bar{x}_m = 3.6\ ,\ s_m^2 = 0.18\ \ ,\ \ \bar{x}_n = 3.8\ ,\ s_n^2 = 0.14$$

Is there evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = 0.05$.

Example 11

A machine in a factory produces 10% of defectives among a large lot of items that it produces in a day. A random sample of 100 items from the day?s production contains 15 defectives, and the supervisor says that the machine must be repaired. Is there evidence that the machine produces more than 10% of defectives on average? Use $\alpha = 0.05$.

Example 12

Lipitor is a drug that is used to control cholesterol. In a randomized clinical trial, 94 subjects were treated with Lipitor and 270 independently selected subjects were given a placebo. Among 94 treated with Lipitor, 7 developed infections, while among 270 given a placebo, 27 developed infections. Is there a difference between the infection rates for the two drugs? Use $\alpha = 0.05$