Lecture 6: The Elements of Hypothesis Testing

STA261 – Probability & Statistics II

Ofir Harari

Department of Statistical Sciences

University of Toronto

Outline

Introduction

The Statistical Hypothesis Testing Framework

- Basic Definitions

Simple Hypotheses

- Significance Level and Power
- Likelihood Ratio Tests and the Neyman–Pearson Lemma

UNIVERSITY OF TORONTO

Introduction

- In order to receive the FDA's approval, new drugs must go through Clinical Research, or *Clinical Trials* (specifically: *Treatment Trials*).

- Clinical trials are conducted in phases:

- *Phase* I *and* II *trials* (20–300 patients): evaluating safety, identifying side effects and determining effectiveness.

- *Phase* III *trials* (1,000–3,000 patients): confirming effectiveness. We shall focus on this phase.

- Typically there will be (at least) two groups: treatment and control.

- Each group consists of patients (volunteers) with the disease/condition, who have met the selection criteria, blindly and randomly allocated to avoid *research bias*. i.e. woman more prone to disease..

- In "placebo-controlled" trials, patients in the treatment group receive the drug under investigation, while patients in the control group receive a placebo.

double blind: both patient and doctor does not know which group peps are in

Introduction

- How do we determine the effectiveness of a treatment?

- Denote:

$$\begin{cases} n_{\text{tr}} \text{ -- Number of patients in the treatment group} \\ n_{\text{pl}} \text{ -- Number of patients in the placebo group} \\ p_{\text{tr}} \text{ -- the probability of being } \underline{\text{cured}} \text{ among patients in the treatment group} \\ p_{\text{pl}} \text{ -- the probability of being cured among patients in the placebo group} \\ X_{\text{tr}} \text{ -- Number of patients in the treatment group who were cured} \\ X_{\text{pl}} \text{ -- Number of patients in the placebo group who were cured} \end{cases}$$

  Then $X_{\text{tr}} \sim \text{Binom}(n_{\text{tr}}, p_{\text{tr}})$ and $X_{\text{pl}} \sim \text{Binom}(n_{\text{pl}}, p_{\text{pl}})$

- To prove the effectiveness of the drug under investigation, the pharmaceutical company must **use the data** to show that

$$p_{\Delta} := p_{\text{tr}} - p_{\text{pl}} > 0$$

  beyond a reasonable doubt.

UNIVERSITY OF TORONTO

Hypotheses and types of errors

- The data: $X_1, \dots, X_n \sim f_\theta$
- The parameter: θ (in our example $p_\Delta := p_{\text{tr}} - p_{\text{pl}}$)

publication bias: discard result that does not support alternative hythesis

- The *Null Hypothesis* $\mathcal{H}_0$ – usually the conservative view ("no effect")
  - In our example: $\mathcal{H}_0 : p_\Delta \leq 0$ (i.e. the new drug is no better than placebo)
- The *Alternative Hypothesis* $\mathcal{H}_1$ – usually represents change of reality
  - In our example: $\mathcal{H}_1 : p_\Delta > 0$ (i.e. the new drug IS better than placebo)

rejection rule

- The decision: to reject $\mathcal{H}_0$ or not to reject $\mathcal{H}_0$ there is no accept H_1
- *Type* I *Error*: incorrectly rejecting $\mathcal{H}_0$ (i.e. a *false discovery*)
  - In our example: falsely declaring the medication effective
- *Type* II *Error*: incorrectly retaining $\mathcal{H}_0$
  - In our example: failing to approve an authentically effective medication

⋆ Which type of error would you consider to be more serious?

type 1: false discovery is more serious than sparing the world an authentic discovery

UNIVERSITY OF TORONTO

Formulating the hypotheses

- The parameter space: Θ – all possible values of θ
  - In our example: $\Theta = [-1, 1]$ (all possible values of $p_{\Delta} := p_{\text{tr}} - p_{\text{pl}}$)

- The competing hypotheses are typically of the form

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases}$$

  - In our example: theta_0 and theta_1 definitely dont overlap

$$\begin{cases} \mathcal{H}_0 : p_{\Delta} \in [-1, 0] \\ \mathcal{H}_1 : p_{\Delta} \in (0, 1] \end{cases}$$

- When $\Theta = \{\theta_0, \theta_1\}$, the hypotheses

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{cases}$$

  are called *simple hypotheses*.

- A statistical test: a data driven, probabilistic decision rule with regard to $\mathcal{H}_0$ (reject/not reject). input: sample ==> output: hypothesis

Significance Level and Power

Definition

Suppose that we test the simple hypotheses

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0, \\ \mathcal{H}_1 : \theta = \theta_1. \end{cases}$$

1. The *significance level* of the test is the probability of a type I error,

$$\alpha = \mathbb{P}\left(\begin{matrix} \text{rejecting} \\ \mathcal{H}_0 \end{matrix} \middle| \theta = \theta_0 \right).$$

not conditional probability, just with given param

2. The *power* of a statistical test is the probability of NOT making a type II error, $\pi = 1 - \beta$, where

$$\beta = \mathbb{P}\left(\begin{matrix} \text{not rejecting} \\ \mathcal{H}_0 \end{matrix} \middle| \theta = \theta_1 \right).$$

rejecting H_0 when we should reject it

a good test-> small beta and high power

UNIVERSITY OF TORONTO

Example

Example

You just purchased a matchbox. The company states on its website that 10% of their matches are defective, but you have that growing feeling that the true proportion is 50%. The company's customer relations representative proposes that you sample two matches at random, and if at least one turns out to be defective – your claim will be accepted.

- Denote: X – the number of defective matches in a random sample of size 2;
- $X \sim \text{Binom}(2, p)$ specifies probability distribution completely => hence simple hypotheses
- The (simple) hypotheses in this case are –

$$\begin{cases} \mathcal{H}_0 : p = 0.1 \\ \mathcal{H}_1 : p = 0.5 \end{cases}$$

null: 10% defective
alternative: 50% defective

- The test: reject $\mathcal{H}_0$ if $X \geq 1$

UNIVERSITY OF TORONTO

Example (cont.)

- $\begin{cases} \mathcal{H}_0 : p = 0.1 \\ \mathcal{H}_1 : p = 0.5 \end{cases}$, reject $\mathcal{H}_0$ if $X \geq 1$
- The significance level of the test is – rejecting H_0 when H_0 is true

$$\alpha = \mathbb{P}\left(\begin{array}{c} \text{rejecting} \\ \mathcal{H}_0 \end{array} \middle| p = 0.1 \right) = \mathbb{P}(X \geq 1 | p = 0.1)$$

$$= 1 - \underline{\mathbb{P}(X = 0 | p = 0.1)} = 1 - 0.9^2 = 0.18 0.19$$

just binom

- The probability of a type II error is – not rejecting H_0 when H_1 is true

$$\beta = \mathbb{P}\left(\begin{array}{c} \text{not rejecting} \\ \mathcal{H}_0 \end{array} \middle| p = 0.5 \right) = \mathbb{P}(X = 0 | p = 0.5)$$

$$= 0.5^2 = 0.25,$$

note here alpha and beta can be calculated before observations, with given decision rule, test statistic, and its underlying distribution

and the power of the test is $\pi = 1 - \beta = 0.75$.

UNIVERSITY OF TORONTO

The α–β tradeoff

Decision \ Truth	$\mathcal{H}_0$ is correct	$\mathcal{H}_0$ is incorrect
Not rejecting $\mathcal{H}_0$	Correct decision $(1-\alpha)$	Type I error (Significance level α) beta:false negative
Rejecting $\mathcal{H}_0$	Type II error (β)	Correct decision (Power $\pi = 1-\beta$)

alpha: false positive

- Ideally, we would like both α and β to be as small as possible
- Unfortunately, they have conflicting agendas...
- Suppose that we wish to test

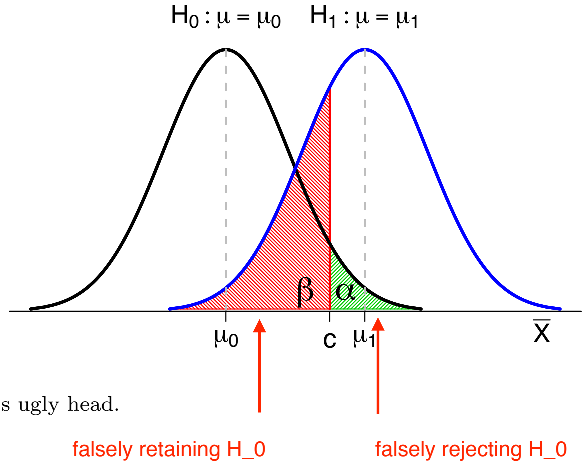
$$\begin{cases} \mathcal{H}_0 : \mu = \mu_0 \\ \mathcal{H}_1 : \mu = \mu_1 \end{cases},$$

  where $\mu_1 > \mu_0$, based on a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known.
- We reject $\mathcal{H}_0$ if $\overline{X} \geq c$ for some threshold c, and recall that $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

test statistic —> mean

The α–β tradeoff (cont.)

- As α decreases...



β rears its ugly head.

The α–β tradeoff (cont.)

- The general convention is that type I errors are more dangerous than type II errors, ergo, controlling α is prioritized over controlling β.

- The acceptable practice is thus to focus on tests with significance level α (where α is a small number, typically 0.05 or less), and among those to search for the test with the smallest β.
  - also known as the *most powerful test*.

- But can we search for such a test in a principled way?

- Yes! And (perhaps not so) surprisingly, it involves the likelihood function.

UNIVERSITY OF TORONTO

Test Statistics

- Any statistical test is based on a *test statistic* $\mathrm{T}(\underline{\mathrm{X}})$: a sample statistic whose distribution under $\mathcal{H}_0$ is known.
- In a previous example wished to test

$$\begin{cases} \mathcal{H}_0 : \mu = \mu_0 \\ \mathcal{H}_1 : \mu = \mu_1 \end{cases},$$

where $\mu_1 > \mu_0$, based on a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 was known.

- The proposed test was: reject $\mathcal{H}_0$ if $\overline{X} \geq c$, for some c.
- Here $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, hence $\overline{X} \overset{\mathcal{H}_0}{\sim} \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$.

known by H_0

- In this example $\overline{X}$ is our test statistic.

UNIVERSITY OF TORONTO

Rejection Regions

$$\begin{cases} \mathcal{H}_0 : \mu = \mu_0 \\ \mathcal{H}_1 : \mu = \mu_1 \ \ (> \mu_0) \end{cases}, \quad \text{reject } \mathcal{H}_0 \text{ if } \overline{X} \geq c$$

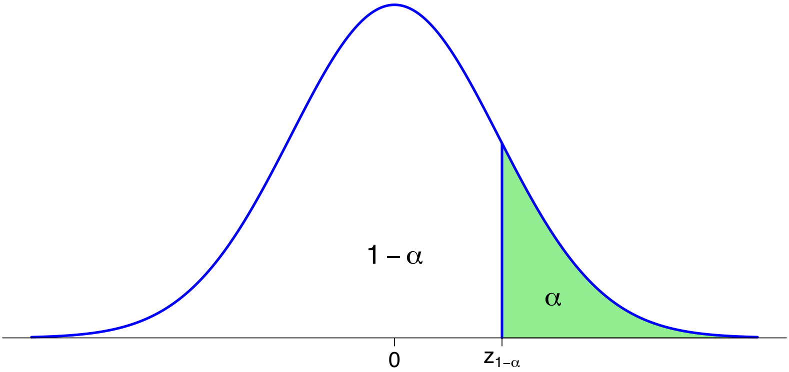
- Remember that we restrict ourselves to tests with significance level α
- Can we find a value of c for which the probability of a type I error will be α?
- Recall that under $\mathcal{H}_0$ (i.e. assuming $\mu = \mu_0$), $\overline{X} \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$, thus

$$\alpha = \mathbb{P}\left(\begin{array}{c}\text{rejecting} \\ \mathcal{H}_0\end{array} \middle| \mu = \mu_0\right) = \mathbb{P}\left(\overline{X} \geq c \middle| \mu = \mu_0\right) = 1 - \mathbb{P}\left(\overline{X} \leq c \middle| \mu = \mu_0\right)$$

$$= 1 - \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{c - \mu_0}{\sigma/\sqrt{n}} \middle| \mu = \mu_0\right) = 1 - \Phi\left(\frac{c - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\Longrightarrow \Phi\left(\frac{c - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \alpha \Longrightarrow \frac{c - \mu_0}{\sigma/\sqrt{n}} = z_{1-\alpha} \Longrightarrow \boxed{c = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}}$$

Rejection Regions (cont.)



UNIVERSITY OF TORONTO

Rejection Regions (cont.)

- We calculated $c = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$, hence the test –

$$\text{“reject } \mathcal{H}_0 \text{ if } \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\text{”}$$

  has significance level α.

- In general, any statistical test that is based on test statistic $T(\underline{X})$ will be of the form

$$\text{“reject } \mathcal{H}_0 \text{ if } T(\underline{x}) \in \mathcal{C}\text{,” for some region } \mathcal{C} \subset \mathbb{R}^n.$$

  We call $\mathcal{C}$ the *Rejection Region*.

- In our example the rejection region was

$$\mathcal{C} = \left\{ \underline{x} \in \mathbb{R}^n \, : \, \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}.$$

reject H_0 if x \in rejection region —> gives confidence level of alpha

Rejection Regions (cont.)

- For example, suppose that $n = 16$, $\sigma^2 = 25$, and we wish to test

$$\begin{cases} \mathcal{H}_0 : \mu = 175 \text{ vs.} \\ \mathcal{H}_0 : \mu = 180 \end{cases}$$

  at a 5% level.

- Here the rejection region is

$$\mathcal{C} = \left\{ \underline{x} : \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\} = \left\{ \underline{x} : \overline{x} \geq 175 + \frac{5}{4} \underbrace{z_{0.95}}_{1.645} \right\}$$

$$= \{ \underline{x} : \overline{x} \geq 177.06 \}.$$

- Congratulations! You just designed your first statistical test. But is it the optimal (i.e. most powerful) test?

UNIVERSITY OF TORONTO

The Likelihood Ratio statistic

- Consider again the problem of testing the simple hypotheses

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{cases}$$

based on $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$.

- The *Likelihood Ratio* is

$$\lambda(\underline{x}) := \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{f(x_1, \ldots, x_n | \theta_1)}{f(x_1, \ldots, x_n | \theta_0)}.$$

- Loosely speaking, $\lambda(\underline{x})$ measures how likely $\mathcal{H}_1$ is to be true compared to $\mathcal{H}_0$, with large values supporting the case for rejecting $\mathcal{H}_0$.

- It is thus reasonable to consider tests with rejection regions of the form

$$\mathcal{C} = \{\underline{x} \in \mathbb{R}^n : \lambda(\underline{x}) \geq c\}.$$

Likelihood Ratio Tests

Definition

A statistical test based on the rejection region

$$\mathcal{C} = \left\{ \underline{x} \in \mathbb{R}^n \ : \ \lambda(\underline{x}) = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \geq c \right\},$$

for c satisfying $\mathbb{P}\left(\lambda(\underline{x}) \geq c \middle| \theta = \theta_0\right) = \alpha$, is called a *likelihood ratio test* (LRT) at (significance) level α.

- Back to our example:

$$\begin{cases} \mathcal{H}_0 : \mu = \mu_0 \\ \mathcal{H}_1 : \mu = \mu_1 \ \ (> \mu_0) \end{cases}, \quad X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

- Here

$$\lambda(\underline{x}) = \frac{\mathcal{L}(\mu_1)}{\mathcal{L}(\mu_0)} = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_1)^2 \right\}}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_0)^2 \right\}}$$

Likelihood Ratio Tests (cont.)

$$\lambda(\underline{x}) = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_1)^2\right\}}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right\}} = \frac{\exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{n\mu_1}{\sigma^2}\overline{x} - \frac{n\mu_1^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{n\mu_0}{\sigma^2}\overline{x} - \frac{n\mu_0^2}{2\sigma^2}\right\}}$$

$$= \exp\left\{\frac{n(\mu_1-\mu_0)}{\sigma^2}\overline{x} - \frac{n(\mu_1^2-\mu_0^2)}{\underset{2}{\cancel{s}}\sigma^2}\right\},$$

therefore

a constant

$$\lambda(\underline{x}) = \geq c \iff \frac{n(\mu_1-\mu_0)}{\sigma^2}\overline{x} - \frac{n(\mu_1^2-\mu_0^2)}{\underset{2}{\cancel{s}}\sigma^2} \geq c_1$$

$$\iff \frac{n(\mu_1-\mu_0)}{\sigma^2}\overline{x} \geq c_2 \iff \overline{x} \geq c_3, \ (\text{since } \mu_1 > \mu_0)$$

hence

positive constant

$$\boxed{\mathcal{C} = \{\underline{x} \in \mathbb{R}^n : \lambda(\underline{x}) \geq c\} = \{\underline{x} \in \mathbb{R}^n : \overline{x} \geq c_3\}.}$$

this is the test we did earlier on normal distribution samples

UNIVERSITY OF TORONTO

Likelihood Ratio Tests (cont.)

$$\mathcal{C} = \{\underline{x} \in \mathbb{R}^n : \lambda(\underline{x}) \geq c\} = \{\underline{x} \in \mathbb{R}^n : \overline{x} \geq c_3\}$$

- It now remains to find c_3 such that

$$\alpha = \mathbb{P}\begin{pmatrix} \text{Type I} \\ \text{error} \end{pmatrix} = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = \mu_0\right) = \mathbb{P}\left(\overline{X} \geq c_3 \middle| \mu = \mu_0\right)$$

in rejection region

- Wait a minute – we've already calculated it! It was

$$c_3 = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

- It turns out that the test we proposed was a likelihood ratio test all along.
- But is it the *most powerful* (MP) test at level α?

The Neyman–Pearson Lemma

Definition

Suppose that we observe $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$, and consider the simple hypotheses $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. We say that a test is a *most powerful* (MP) test at level α if

1. the significance level of the test is α, and
2. no other test at level α has greater power.

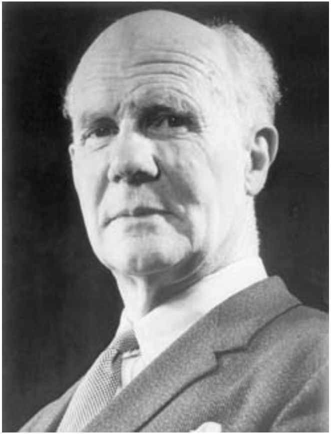
Lemma (Neyman–Pearson Lemma)

The likelihood ratio test, based on the rejection region

$$\mathcal{C} = \left\{ \underline{x} \in \mathbb{R}^n \, : \, \lambda(\underline{x}) = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \geq c \right\},$$

with c satisfying $\mathbb{P}\left(\lambda(\underline{X}) \geq c\right) = 1 - \alpha$, \alpha is an MP test at level α.

The Neyman–Pearson Lemma (cont.)



Egon Pearson, 1895-1980

Source: swlearning.com

UNIVERSITY OF TORONTO

The Neyman–Pearson Lemma (cont.)

Proof for a continuous f_θ:

Denote the rejection region of the LRT by $\mathcal{C}$. Note that

$$\alpha = \mathbb{P}(\underline{X} \in \mathcal{C} \big| \theta = \theta_0) = \int_{\mathcal{C}} f(\underline{x} \big| \theta_0) \mathrm{d}\underline{x} = \int_{\mathcal{C}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x}.$$

rejecting null given null correct

Consider now another test at level α, with rejection region $\mathcal{D}$. The same calculations would yield $\alpha = \int_{\mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x}$, hence

prove that C yields higher power when compared to any other region D (i.e. MP)

$$\int_{\mathcal{C}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x} = \int_{\mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x}.$$

since both equal to alpha

Write $\mathcal{C} = (\mathcal{C} \cap \mathcal{D}) \cup \left(\mathcal{C} \cap \overline{\mathcal{D}}\right)$ (a disjoint union), we have

$$\int_{\mathcal{C}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x} = \int_{\mathcal{C} \cap \mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x} + \int_{\mathcal{C} \cap \overline{\mathcal{D}}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x}.$$

by one property of integration

Likewise, $\displaystyle\int_{\mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x} = \int_{\mathcal{C} \cap \mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x} + \int_{\overline{\mathcal{C}} \cap \mathcal{D}} \mathcal{L}(\theta_0) \mathrm{d}\underline{x}.$

The Neyman–Pearson Lemma (cont.)

Proof (cont.):

We have shown so far

$$\int_{\mathcal{C}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} + \int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} = \int_{\mathcal{C}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} + \int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x},$$

or simply

$$\boxed{\int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} = \int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x}.}$$

It is now time to recall that for any $\underline{x} \in \mathcal{C}$ we have $\lambda(\underline{x}) = \dfrac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \geq c$, and for every $\underline{x} \in \overline{\mathcal{C}}$ we have $\dfrac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} < c$, hence

by definition of region C

$$c\int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} \leq \int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} \quad \text{and} \quad c\int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} \geq \int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x}.$$

Example: Binomial distribution

UNIVERSITY OF TORONTO

The Neyman–Pearson Lemma (cont.)

Proof (cont.):

Thus far we have

by equality of alpha for region C and D

by definition of region C

$$\int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} \le c\int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} = c\int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_0)\mathrm{d}\underline{x} \le \int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x}.$$

Let π be the power of the LRT, and let π' be the power of the other test. We have

power = reject null when H_1 is true

$$\pi = \mathbb{P}\left(\underline{X}\in\mathcal{C}\middle|\theta=\theta_1\right) = \int_{\mathcal{C}} f(\underline{x}|\theta_1)\mathrm{d}\underline{x} = \int_{\mathcal{C}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x}$$

$$= \int_{\mathcal{C}\cap\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} + \int_{\mathcal{C}\cap\overline{\mathcal{D}}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} \ge \int_{\mathcal{C}\cap\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} + \int_{\overline{\mathcal{C}}\cap\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x}$$

$$= \int_{\mathcal{D}} \mathcal{L}(\theta_1)\mathrm{d}\underline{x} = \int_{\mathcal{D}} f(\underline{x}|\theta_1)\mathrm{d}\underline{x} = \mathbb{P}\left(\underline{X}\in\mathcal{D}\middle|\theta=\theta_1\right) = \pi'.$$

prove that power of LRT > power of any other test

Example: Bernoulli data

Example

Suppose that we wish to test $\mathcal{H}_0 : p = p_0$ vs. $\mathcal{H}_1 : p = p_1$ (for $p_1 > p_0$), based on sequence $X_1, \ldots, X_n$ of Bernoulli trials with an unknown probability of success p. Find an <u>MP test</u> at level α. should be does not exceed alpha.

Solution: just use LRT

Here

note do not use pdf of binomial distribution, just product of bernoulli pdf

$$\lambda(\underline{x}) = \frac{\mathcal{L}(p_1)}{\mathcal{L}(p_0)} = \frac{p_1^{\sum x_i}(1-p_1)^{n-\sum x_i}}{p_0^{\sum x_i}(1-p_0)^{n-\sum x_i}} = \left(\frac{p_1}{p_0}\right)^{\sum x_i}\left(\frac{1-p_1}{1-p_0}\right)^{n-\sum x_i}$$

$$= \left(\underbrace{\frac{p_1}{p_0}}_{>1} \cdot \underbrace{\frac{1-\overset{p\_0}{\cancel{p_1}}}{1-\underset{p\_1}{\cancel{p_0}}}}_{>1}\right)^{\sum x_i} \underset{<1}{\underline{\left(\frac{1-p_1}{1-p_0}\right)^n}} = a^{\sum x_i} b,$$

where $a > 1$ and $b > 0$.

Example: Bernoulli data (cont.)

Solution (cont.):

$$\lambda(\underline{x}) = a^{\sum x_i} b, \text{ where } a > 1 \text{ and } b > 0.$$

monotonic increasing function

Now,

$$\lambda(\underline{x}) \geq c \iff a^{\sum x_i} b \geq c \iff a^{\sum x_i} \geq c_1 \iff \sum x_i \geq c_2,$$

hence, the rejection region of the LRT is

$$\boxed{\mathcal{C} = \{\underline{x} : \lambda(\underline{x}) \geq c\} = \left\{\underline{x} : \sum x_i \geq c_2\right\}.}$$

To find c_2, recall that $\sum_{i=1}^n X_i \sim \text{Binom}(n, p)$, hence

note here how alpha is the upper bound on possible size of test

note this is a problem for discrete distribution, where can cant really find size of test to be exact alpha

$$\alpha \geq \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| p = p_0\right) = \mathbb{P}\left(\sum_{i=1}^n X_i \geq c_2 \middle| p = p_0\right)$$

rejecting null when null is true

$$= 1 - \mathbb{P}\left(\sum_{i=1}^n X_i < c_2 \middle| p = p_0\right) = 1 - \boxed{\sum_{k=0}^{\lfloor c_2 \rfloor - 1} \binom{n}{k} p_0^k (1 - p_0)^{n-k}.}$$

sum < c SAME AS sum <= c - 1

cdf of binomial

UNIVERSITY OF TORONTO

Example: Bernoulli data (cont.)

Solution (cont.): take smallest c_2 such that the output <= alpha

So c_2 is the smallest integer satisfying

$$\sum_{k=0}^{c_2-1} \binom{n}{k} p_0^k (1-p_0)^{n-k} \geq 1-\alpha.$$

For example, to test $\mathcal{H}_0 : p = 0.2$ vs. $\mathcal{H}_1 : p = 0.3$ at a 5% level with $n = 50$, we need to look for the smallest c_2 satisfying

$$\sum_{k=0}^{c_2-1} \binom{50}{k} 0.2^k 0.8^{50-k} \geq 0.95.$$

```
> probs <- pbinom(c(0:50), size=50, prob=.2)
> (c2 <- min(which(probs > .95)))
[1] 16
```

assume that c_2 is integer, because sum of int is int, no point in setting a threshhold that is not int

and the rejection region in this case is $\mathcal{C} = \left\{ \underline{x} : \sum_{i=1}^{50} x_i \geq 16 \right\}$.

so reject null if sum of x is >= 16.
gives significant level of alpha = 0.05 with smallest beta

Example: Bernoulli data (cont.)

Solution (cont.):

a conservative test; not exactly 5% because pdf is discrete

⋆ CAUTION: the actual significance level of this test is actually not 5%, but

$$\text{alpha} = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| p = 0.2\right) = \mathbb{P}\left(\sum_{i=1}^{50} X_i \geq 16 \middle| p = 0.2\right) = 0.031$$

(but if we chose the cutoff to be 15 instead, it would be 0.061).

⋆ The Neyman-Pearson Lemma states that no test with greater power exists at the 0.031 level, but there could be a test with a significance level that is closer to 5% and a greater power.

because pearson lemma states MP test at exactly alpha

- For a large n, $\text{Binom}(n,p) \overset{\text{CLT}}{\approx} \mathcal{N}\left(np, np(1-p)\right)$. We can then calculate

$$\alpha = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| p = p_0\right) = \mathbb{P}\left(\sum_{i=1}^{n} X_i \geq c_2 \middle| p = p_0\right) \approx 1 - \Phi\left(\frac{c_2 - np_0}{\sqrt{np_0(1-p_0)}}\right)$$

$$\implies c_2 = np_0 + z_{1-\alpha}\sqrt{np_0(1-p_0)}$$

UNIVERSITY OF TORONTO

Example: Bernoulli data (cont.)

Solution (cont.):

- The rejection region based on the large sample approximation is

$$\mathcal{C}' = \left\{ \underline{x} \,:\, \sum_{i=1}^{n} x_i \geq np_0 + z_{1-\alpha}\sqrt{np_0(1-p_0)} \right\}$$

- For $n = 50$, $\alpha = 0.05$ and $p_0 = 0.2$ we get

$$\mathcal{C}' = \left\{ \underline{x} \,:\, \sum_{i=1}^{50} x_i \geq 50 \times 0.2 + 1.645\sqrt{50 \times 0.2 \times 0.8} \right\}$$

$$= \left\{ \underline{x} \,:\, \sum_{i=1}^{50} x_i \geq 14.65 \right\} = \left\{ \underline{x} \,:\, \sum_{i=1}^{50} x_i \geq 15 \right\}.$$

in this case, normal approximation is continuous hence able to find c such that P(Type I error) = \alpha. But still since binomial is discrete, c gives to a natural number, whose actual significance level is not 5%, in fact in this case it is 6.1%