36-705 09/07/2020

Youtube lecture 19/10/16

- Homework hints are provided at start of lecture.

Lecture Notes 10 - Hypothesis Testing

- Within parametric framework.
- Instead of trying to estimate parameter θ using appropriate estimators, we are testing a hypothesis about the parameter θ.

- 1. Introduction

- $X_1, \dots, X_n \sim p(x;\theta)$
- Start with null, alternate hypothesis:-

$H_0: \theta = \theta_0$ (null)

$H_1: \theta \neq \theta_0$ (alternate)

- Generally; we have: $H_0: \theta \in \Theta_0$ $H_1: \theta \in \Theta_1$
- Only requirement is that parameter spaces Θ_0 and Θ_1 are different, formally $\Theta_0 \cap \Theta_1 = \emptyset$.
- Simple/composite null are terms to describe whether param space is a point, or more than one point.

Example 1

- $X_1, \dots, X_n \sim \text{Bernoulli}(p)$

$H_0: p = \frac{1}{2}$

$H_1: p \neq \frac{1}{2}$

- Context: Statistical analysis of coin flipping (via hyp. test).

~~W~~ Hypothesis testing is 'asymmetric'
- Analogy with legal cases (trial
  - presumption of innocence
  - Question is not innocent/guilty, but whether sufficient evidence to reject your innocence
- without evidence, we retain null hyp. (that you are innocent.

'world'	Decision: Retain H_0	Reject H_0
H_0 true	✓	Type I error (false positive)
H_1 true	Type II error (false negative)	✓

(*) False positive - yes drug works; but just chance finding

(*) False negative -

IW: Statistical hypothesis testing is conducted in a situation where false positives are worse than false negatives.

- Asymmetric situation
- Type II errors 'more about career' → amazing scientific discovery missed.
- Assume we are in a situation where hypothesis testing is appropriate
- We want to design a good test

(*) Set a false positive rate; context dependent (on field)

- Medicine: 5% or 1%
- Particle physics: 0.0000001%

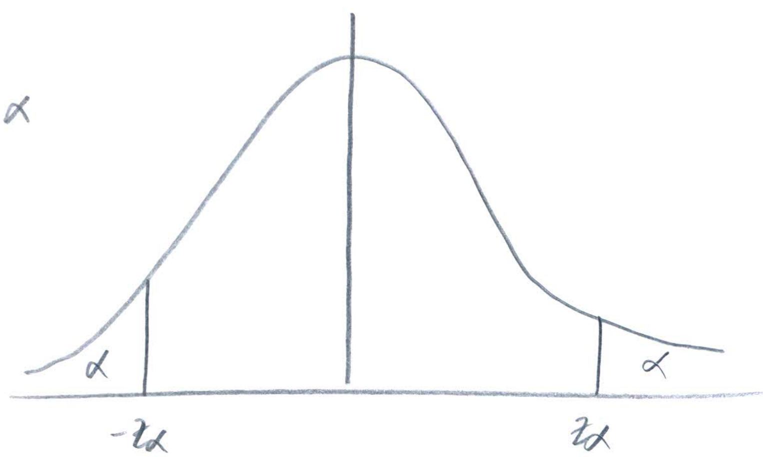
(type II)

(*) In a general sense, test design ~~to~~ involves minimising false negative rate, given an acceptable rate on false positives (type I error), controlled for first.

(*) Priority on Type I error (false positives). Type II is subsidiary

Notation: - Let Φ be CDF of a standard Normal r.v Z.

- For $0 < \alpha < 1$ let $z_\alpha = \Phi^{-1}(1-\alpha)$
- Hence $P(Z > z_\alpha) = \alpha$, $P(Z < -z_\alpha) = \alpha$



2. Constructing Tests

IW: make a rule to reject H_0 when something happens:-

1) Choose a test statistic $T_n = T_n(X_1, \dots, X_n)$
2) Choose a rejection region R
3) If $T_n \in R$ then reject H_0 otherwise retain H_0.

- Have to select T and R so that test has good statistical properties.

- will consider following tests:-
  1. Neyman-Pearson test
  2. Wald test
  3. Likelihood Ratio test
  4. Permutation test

(*) Discuss/decide properties of test, how to evaluate them
(*) can think of rejection region as a subset of the sample space

3. Error Rates and Power

- suppose we reject H_0 when $(X_1, \ldots, X_n) \in R$
- Define power function :-

$$\beta(\theta) = P_\theta(\text{reject } H_0) = P_\theta(X_1, \ldots, X_n \in R)$$

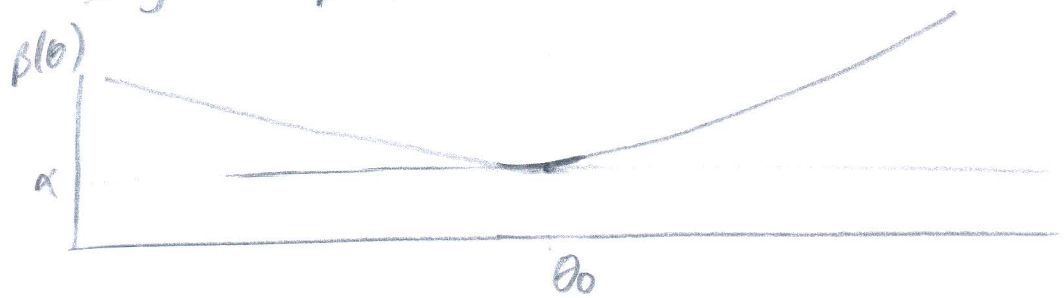
(*) suppose we fix the parameter/set of parameters at a particular value; then ask what probability of rejecting null is.
(*) That will be a function of parameter θ.
(*) we want $\beta(\theta)$ to be small when $\theta \in \Theta_0$ i.e. null hypothesis is true
$\beta(\theta)$ to be large when $\theta \in \Theta_1$ i.e. null hypothesis is false

general strategy:-

1. Fix $\alpha \in [0,1]$
2. Maximise $\beta(\theta)$ for $\theta \in \Theta_1$
   subject to $\beta(\theta) \le \alpha$ for $\theta \in \Theta_0$

IW: Fix 'amount of type I error' we are willing to tolerate.
- when null is true, there is some chance we will make a false claim
- How high are we going to tolerate?

(A) - stare at the logic of this.



(*) one parameter case, null is a single value
(*) want to make power function $\beta(\theta)$ as large as possible over Θ_1 subject to condition that it cannot be bigger than α at the i.e.

$\beta(\theta) \leq \alpha$ over $\theta \in \Theta_0$.

(*) optimisation problem.

(*) Different test statistics will have different power functions

- α is known as the size or level of the test (roughly speaking)

- A test is size α if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$$

- level-α test refers to when cannot construct exact size-α test, settle for smaller error rate.

- A test is level α if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

---

example 3

$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with σ^2 known (1 param problem).

(distinct from notes)

- $H_0: \theta \leq \theta_0$

$H_1: \theta > \theta_0$

· FDA placebo-drugs

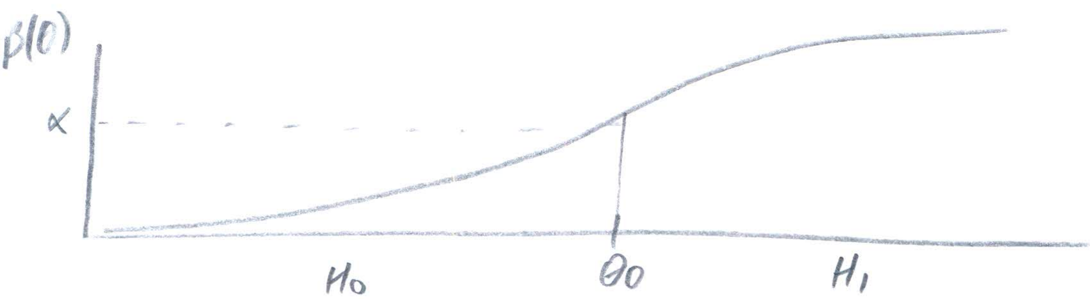
- one-tailed / one-sided alternative, composite null.
- construct test-stat, compute mean, compare to θ_0. (can be any test stat.)
- Define this test statistic T_n, standardise:-

$$T_n = \frac{\bar{X}_n - \theta_0}{\sigma / \sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma}$$

· T_n used for comp. convenience (funct. form).
- can also use $\bar{X}_n$ directly

(*) Intuitively, reject H_0 when T_n is very large (i.e. far to the right of θ_0 null)

- reject H_0 if $T_n > c$
- one to one-sided alternative, one-sided alternative:-



(*) want to be sure that for $\theta < \theta_0$ (ie. null region), $\beta(\theta)$ is below α.

- compute power function:-

$$\beta(\theta) = P(\text{reject } H_0) = P(T_n > c)$$ (i)

$$= P_\theta\left(\frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} > c\right) = P_\theta\left(\bar{X}_n > \frac{\sigma c}{\sqrt{n}} + \theta_0\right)$$

- (i) - 'incorrectly standardised' so rewrite

P_θ - as dependent on true value of θ

$$= P_\theta\left(\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

- (ii) Standardise using true, unknown θ

- Note $Z \sim N(0,1)$
$P(Z < t) = \Phi(t)$

$$= P\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

- σ fixed $\quad \theta_0$ fixed
- c fixed
- n fixed

↳ a function of θ

(*) Have to determine c

- know that when $\theta = \theta_0$, we want $\beta(\theta_0) = \alpha$
- Note that when $\theta = \theta_0$, the above formula for $\beta(\theta)$ reduces to

$$\beta(\theta_0) = 1 - \Phi(c) = \alpha$$

(ie. set this to be the case)

$$\Rightarrow \Phi(c) = 1 - \alpha$$

$$\Rightarrow c = \Phi^{-1}(1-\alpha) = z_\alpha$$

(*) In order to have a <u>size-α</u> test, to make sure that under the null H_0,

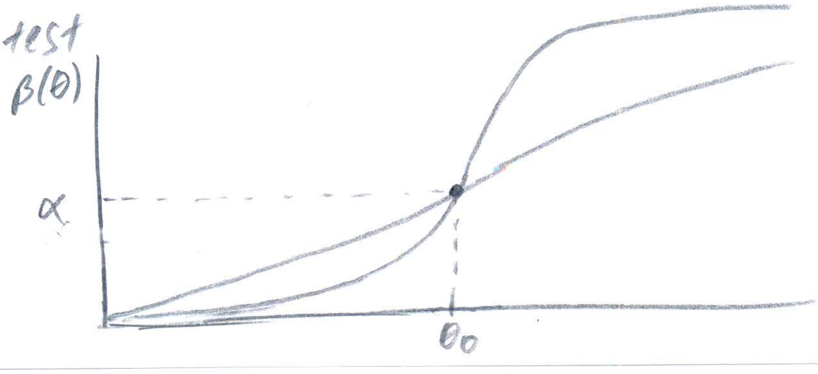
- reject H_0 when

$$T_n = \frac{\bar{x}_n - \theta_0}{\sigma/\sqrt{n}} > z_\alpha$$

(A) - Need to solidify what is going on <u>diagramatically</u>

(*) uniformly most powerful test

(*) As $n \to \infty$:-



(*) By construction $\beta(\theta_0) = \alpha$ (always)

(*) Slope of power function $\beta(\theta)$ increases with n, i.e. as you get more and more data, your ability to discover that the null is false i.e. reject the null H_0, gets higher (you get more information) (OT).

(*) At any alternative value $\theta \in \Theta_1$, or in this context, $\theta > \theta_0$, the power function $\beta(\theta) \xrightarrow{n \to \infty} 1$ over region $\theta \in \Theta_1$

Example 4

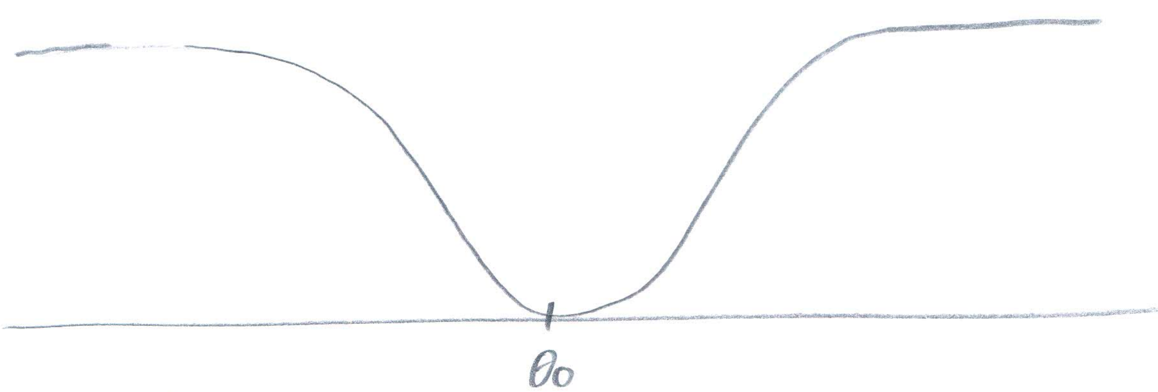
- Two-sided alternative

(A3) Review calculation here in notes

- $H_0: \theta = \theta_0$

  $H_1: \theta \neq \theta_0$

- Reject H_0 if $|T_n| > c$ where T_n is defined as above.



- Reject if $|T_n| > z_{\alpha/2}$

LW: In a two-sided alternative; reject if test statistic very large or very small

- z_α, $z_{\alpha/2}$ are critical values of the test.

(*) These are canonical tests

---

4. Neyman-Pearson test

- LW: Specific test, not used much in practice, but conceptually v. important.
- Specific

(*) simple null vs simple alternative (both specific)

$H_0: \theta = \theta_0$

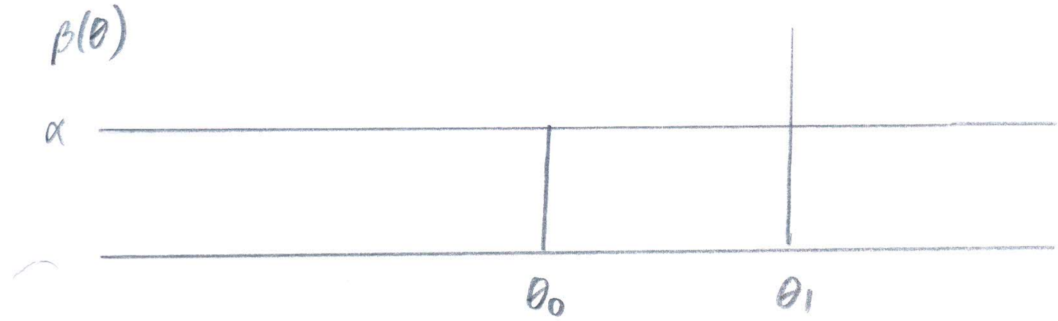
$H_1: \theta = \theta_1$

e.g. $H_0: X_1, \dots, X_n \sim N(0, \sigma^2)$

$H_1: X_1, \dots, X_n \sim N(1, \sigma^2)$

(*) LW: normally, alternate hypothesis non-specific, so above is rare situation.
theoretically, is most basic, simplest.

- consider power function — only 2 values of param.



- specific optimisation problem:-
- How can I make the probability of rejecting the null H_0 as large as possible in region where $\theta \in \Theta_1$ (i.e. $\theta = \theta_1$ in this context) while making sure it's held at α at $\theta \in \Theta_0$ (i.e. $\theta = \theta_0$ in —"—) $\beta(\theta)$

(*) Neyman-Pearson lemma gives an exact answer to this.

Theorem 5

- let $L(\theta) = p(X_1, \dots, X_n; \theta)$ (i.e. likelihood) and test stat be ratio of likelihoods of alternative and null:- — Assuming $p(x^n; \theta_i)$ are well-behaved p. densities

$$T_n = \frac{L(\theta_1)}{L(\theta_0)} = \frac{p(X_1, \dots, X_n; \theta_1)}{p(X_1, \dots, X_n; \theta_0)}$$

(*) Intuitively reject H_0 when likelihood under alternative $L(\theta_1)$ is bigger

(*) Reject H_0 if $T_n > c$

- choose c so that:- probability under null is exactly α.

$$P_{\theta_0}(T_n > c) = P_{\theta_0}(X^n \in R) = \alpha$$

(*) If we take null H_0 to be true, then T_n is a random variable and can request distribution of T_n (likelihood ratio)

(*) We can compute $P_{\theta_0}(T_n > c)$ as once we have specified null hypothesis, we have specified the complete distribution (analytically or numerically).

(*) The theorem says this is uniformly most powerful (UMP). (level-α test).

- Formally:

(*) If we have another size-α test with power function β; then

$$\beta_{NP}(\theta_1) \geq \beta'(\theta_1)$$

(*) in this simple context, NP is most powerful test.

(*) Typo in notes - $\beta(\theta) \geq \beta'(\theta) \ \forall \theta \in \Theta_1$

LW: Reinforcement:

- If null is true, probability of a false rejection cannot be more than α.
- If θ_0 is true value, have to make sure that whatever test is used the probability of rejecting H_0 is less than α. (constraint).
- Subject to above constraint, if θ_1 is the true value, we would like to reject H_0, and to make that probability as large as possible.

(*) 3 common tests are next 3 tests.

- LW: Not going to cover theory of 'optimal tests'
  - In the scalar hypothesis test valid parametric framework, will examine Wald Test
  - will rely on asymptotic Normality

(*) In hypothesis testing, asymptotic Normality often used for computational reasons

(*) In principle can do things analytically, but comp. expensive. - use asymptotic approximations.