STAT 346/446 Lecture 10

Methods of evaluating tests

CB Section 8.3.1, DS Section 9.1

Evaluating Hypothesis tests

Hypotheses: $H_0 : \theta \in \Theta_0$ vs. $H_1 : \Theta \in \Theta_0^c$

Test procedure: Reject H_0 if $\mathbf{X} \in R$

$P_\theta(X \in R)$ = prob of rejecting H_0

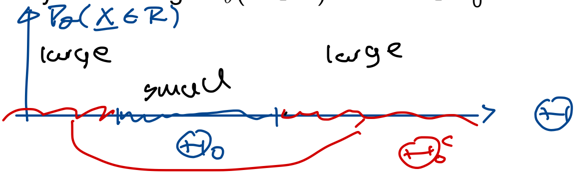
depends on the unknown θ

- Can make two types of mistakes
  - Reject $H_0 : \theta \in \Theta_0$ when in fact $\theta \in \Theta_0$ (Type I)
  - Don't reject $H_0 : \theta \in \Theta_0$ when in fact $\theta \notin \Theta_0$ (Type II)
- Want the *probability* of mistakes to be small
- Recall LRT: Reject H_0 if $\lambda(\mathbf{X}) \leq c$
  - c determines the probability of these mistakes

The challenge of finding a good test procedure

Choose R (i.e. c in the LRT procedure) such that :

- $P_\theta(\mathbf{X} \in R)$ is small for all $\theta \in \Theta_0$
    - Low probability of Type I error
    - Don't want to reject H_0 when it is true
- and $P_\theta(\mathbf{X} \in R^c)$ is small for all $\theta \in \Theta_0^c$
    - Low probability of Type II error
    - Don't want to accept H_0 when it is false
    - Equivalently: Want large $P_\theta(\mathbf{X} \in R)$ for all $\theta \in \Theta_0^c$



Power function

Describes the properties of a test procedure

Hypotheses: $H_0 : \theta \in \Theta_0$ vs. $H_1 : \Theta \in \Theta_0^c$

Test procedure: Reject H_0 if $\mathbf{X} \in R$

Def: Power function

The **power function** of a test procedure is the probability of rejecting H_0 (function of θ)

depends on the unknown parameter θ

$$\beta(\theta) = P_\theta(\mathbf{X} \in R) \quad \text{for all } \theta \in \Theta$$

$\Rightarrow$ Is a function of θ.

where R is the *rejection region* of the test

- Ideal power function:

$$\beta(\theta) = 1 \qquad \text{for } \theta \in \Theta_0^c$$

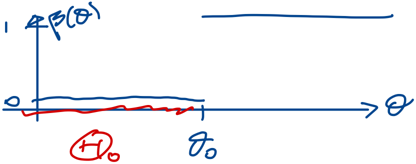
want to reject

$$\beta(\theta) = 0 \qquad \text{for } \theta \in \Theta_0$$

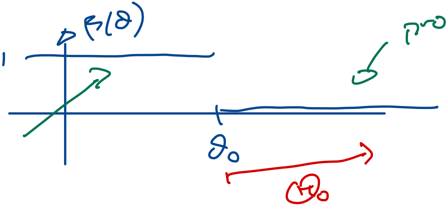
don't want to reject.

Ideal power function

$H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$



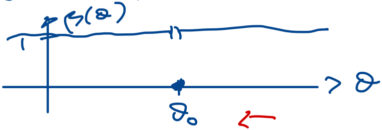
$H_0: \theta \ge \theta_0$ vs $H_1: \theta < \theta_0$



prob of type I error

1-prob of type II error.

$H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$



Type I and Type II errors

"confusion matrix"

- **Type I error**: Wrongly deciding to reject H_0
  - Rejecting $H_0 : \theta \in \Theta_0$ when in fact $\theta \in \Theta_0$
- **Type II error**: Wrongly deciding not to reject H_0
  - Don't reject $H_0 : \theta \in \Theta_0$ when in fact $\theta \notin \Theta_0$

Truth (true state)	Decision: Choose H_0	Choose H_1
H_0 is true ($\theta \in \Theta_0$)	no error	Type I error
H_1 is true ($\theta \in \Theta_0^c$)	Type II error	no error

Relation to power function:

- If $\theta \in \Theta_0$: $\beta(\theta) =$ probability of type I error
- If $\theta \in \Theta_0^c$: $1 - \beta(\theta) =$ probability of type II error

$\beta(\theta)$ = power of test when $\theta \in \Theta_0^c$

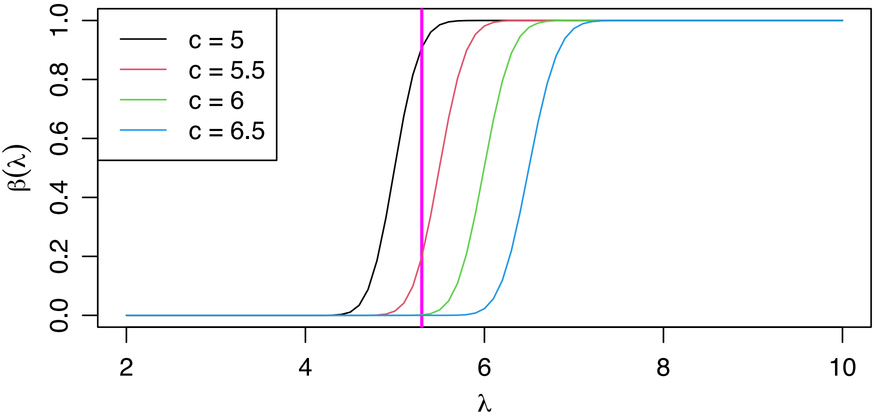
Example: Power function for a Poisson likelihood

- Let $X_1, \ldots, X_n$ be a random sample from $\text{Poisson}(\lambda)$
- We want to test the hypotheses

$$H_0 : \lambda \leq 5.3 \qquad \text{versus} \qquad H_1 : \lambda > 5.3$$

- We will reject H_0 if $\overline{X} \geq 6$
- What is the power function for this test?
- How does the power function change if instead we reject if $\overline{X} \geq 6.5$? The function shifts to the right but keeps the same shape.
- How does the power function change if sample size is increased? The function becomes steeper and closer to the ideal form of 0 in Θ_0 and 1 in Θ_0^c

Power function for Poisson Example



Example: After-School

- The Chapel Hill & Carrboro City Schools (CHCCS) operate After-School programs in all the Elementary schools in the district.
- Each month they send invoices to families that have children enrolled in these programs for both the monthly fee and any extra charges, such as care during teacher workdays (eg. election day)
- Customers are expected to use their own envelopes and stamps to return their payments
- Currently, the time it takes to pay bills has a mean of 24 days and a standard deviation of 6 days.

Example: After-School, continued

- Suppose the chief financial officer (CFO) believes that including a stamped self-addressed envelope would decrease the amount of time it takes to pay the bills.
- She calculates that the improved cash flow from a 2-day decrease in the payment period would pay for the costs for envelopes and stamps. Further decrease would generate profit.
- To test this she randomly selects 220 customers and includes a stamped self-addressed envelope with their invoice
- She assumes that the time to pay a bill follows the normal distribution $N(\mu, \sigma^2)$, where σ^2 is known.

Using the data from her experiment, how can she conclude whether this plan will be profitable?

Power function for after-school

- We have $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$ and we are interested in testing

  24-2

  $$H_0 : \ \mu \geq 22 \qquad \text{and} \qquad H_1 : \ \mu < 22$$

  We assume that σ^2 is known, $\sigma^2 = 6^2$

- Consider the LRT procedure:
  - Reject H_0 if

    $$\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq -\sqrt{-2\log(c)} \equiv c^*$$

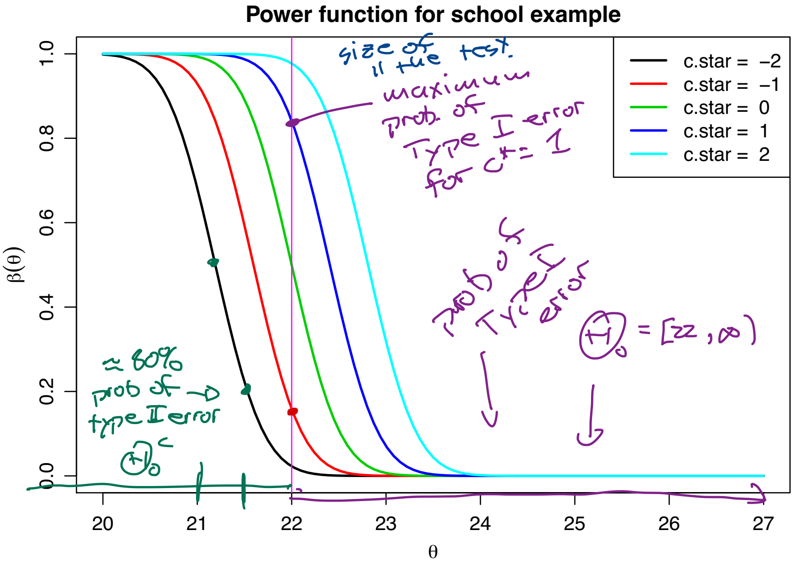
    that is if

    $$\frac{\overline{X}_n - 22}{6/\sqrt{220}} \leq c^*$$

- What is the power function for this test?

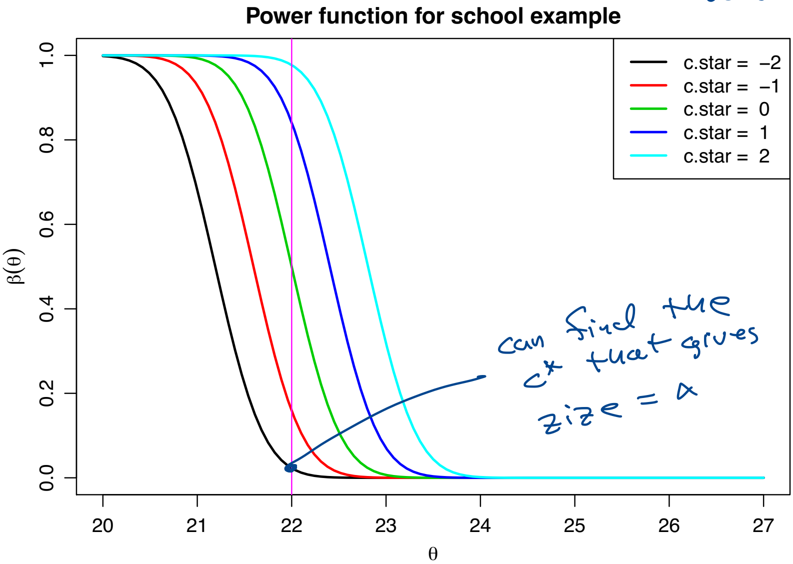
Power function for after-school

$n = 220$ $\mu_0 = 22$
$\sigma = \sqrt{6}$



Power function for after-school

$\sum_{i=0}^{n-1}$ same for many c values



Level and size of tests

↳ Refer to power function only on Θ_0

Definition: Level and size

A test with power function $\beta(\theta)$

- is a **size** α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$
- is a **level** α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$

$0 \leq \alpha \leq 1$

- size of a test $\leq$ level of a test
- Sometimes no distinction made between size and level.
- Choose test procedure that have a certain size (e.g. 0.1, 0.05, 0.01, 0.001)
  - Not unique, have a class $\mathcal{C}$ of level α tests
- For discrete distributions, exact size not always attainable.
  - Randomized tests

Level and size of tests

- Want $\beta(\theta)$ to be small for $\theta \in \Theta_0$ and large for $\theta \in \Theta_0^c$
- Generally there is a trade-off between these probabilities
- A common approach: Choose a number α and pick a procedure (e.g. c^*) such that

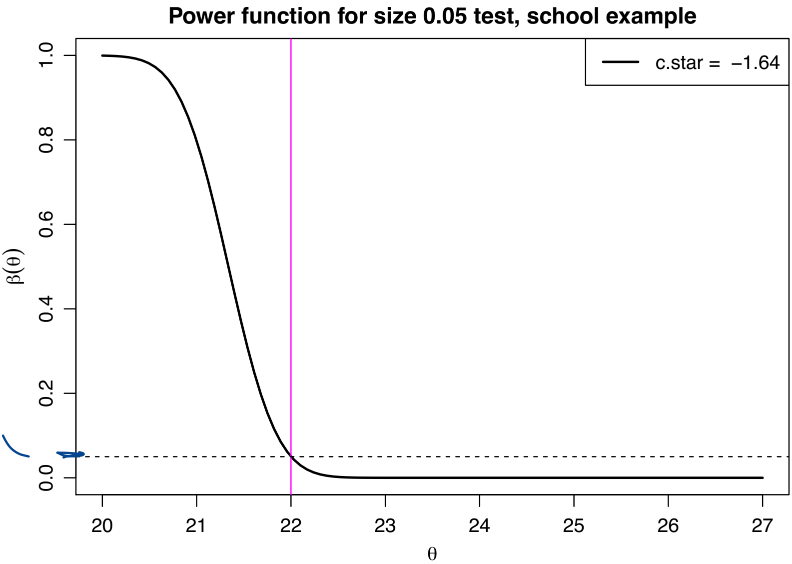
$$\beta(\theta) \leq \alpha \qquad \text{for } \theta \in \Theta_0$$

  That is, we put an upper bound on the probability of type I error.
- The test is then a level α test or we say that the test has **significance level α**

Example: Find the c^* value that makes the schools-example test a level 0.05 test. $c^* = -z_{0.05} = -1.64$

Power function for a size 0.05 test for after-school

size: $\alpha = 0.05$



Upper quantile notation

- A standard notation for upper quantiles.
- z_α is the value where

$$P(Z > z_\alpha) = \alpha \qquad \text{where} \quad Z \sim \mathrm{N}(0, 1)$$

- $t_{n,\alpha}$ is the value where

$$P(T > t_{n,\alpha}) = \alpha \qquad \text{where} \quad T \sim t_n$$

- $\chi^2_{n,\alpha}$ is the value where

$$P(X > \chi^2_{n,\alpha}) = \alpha \qquad \text{where} \quad X \sim \chi^2_n$$

Size / Level for the Poisson Example

- The set-up:
  - $X_1, \ldots, X_n$ i.i.d. $\mathrm{Poisson}(\lambda)$
  - $H_0 : \lambda \leq 5.3$ versus $H_1 : \lambda > 5.3$
  - Reject H_0 if $\overline{X} \geq c$ for some c
- Power function:

want to find the c (or $c^* = \lceil nc-1 \rceil$) such that $\alpha = \sup_{\lambda \in [-\infty, 5.3]} \beta(\lambda) = \beta(5.3)$

$$\beta(\lambda) = P_\lambda(T \geq nc) \qquad \text{where } T = \sum_{i=1}^{n} X_i$$

$$= 1 - \sum_{t=0}^{\lceil nc-1 \rceil} \frac{e^{n\lambda}(n\lambda)^t}{t!}$$

$\frac{d}{d\lambda} \geq 0 \ \forall \lambda$

A bit complicated function of λ but we saw in the RMarkdown file that it is a monotone increasing function of λ - *Need to justify this!*

Size / Level for the Poisson Example

- Since the power function $\beta(\lambda)$ is a monotone increasing function of λ:

Find this

$$\sup_{\lambda \leq 5.3} \beta(\lambda) = \beta(5.3) = 1 - \sum_{t=0}^{\lceil nc-1 \rceil} \frac{e^{n5.3}(n5.3)^t}{t!}$$

- Want to find a c such that

$$1 - \sum_{t=0}^{\lceil nc-1 \rceil} \frac{e^{n5.3}(n5.3)^t}{t!} = \alpha$$

Can't "solve for c" so we try a few values - see RMarkdown file

Size / Level for the Poisson Example

*: Based on the fact that $\beta(\lambda)$ is a monotone increasing function of λ

Justification for our statement that the power function

$$\beta(\lambda) = 1 - \sum_{t=0}^{\lceil nc-1 \rceil} \frac{e^{n\lambda}(n\lambda)^t}{t!}$$

is a monotone increasing function of λ

- One way: Take arbitrary λ_1 and λ_2 and assume that $\lambda_1 < \lambda_2$. Then show that $\beta(\lambda_1) \leq \beta(\lambda_2)$
- Another way: Show that

$$\frac{d}{d\lambda}\beta(\lambda) > 0 \qquad \text{for all } \lambda > 0$$

Unbiased tests

Definition

A test with power function $\beta(\theta)$ is called **unbiased** if

$$\beta(\theta) \leq \beta(\theta') \qquad \forall \theta \in \Theta_0, \theta' \in \Theta_0^c$$

- For any θ in Θ_0 we want the probability of rejecting H_0 to be smaller than for any θ in Θ_0^c

not unbiased.