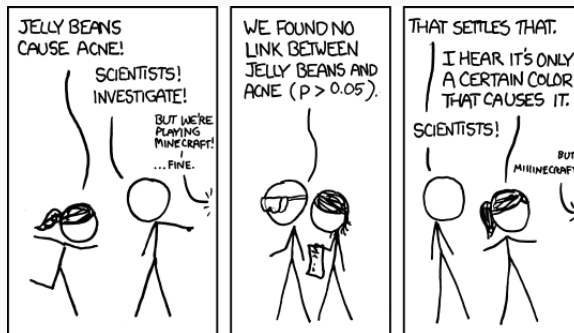


Mathematical Statistics II

STA2212H S LEC9101

Week 3

January 27 2021



Recap

- Quasi-Newton; E-M algorithm optimization
- Formal theory of testing: H_0, H_1, R, T , Type I, Type II error, β, α hypothesis testing
- p -values significance testing
- Wald test, likelihood ratio test more on this today

Recap

- Quasi-Newton; E-M algorithm

optimization

- Formal theory of testing: H_0 , H_1 , R , T , Type I, Type II error, β , α

hypothesis testing

- p -values

- Wald test, likelihood ratio test

- Chapter numbering AoS link

The screenshot displays the iDOL eBook Central interface. The top navigation bar includes links for Home, Search, Bookshelf, Settings, and Sign In. The main content area shows the book 'All of Statistics: A Concise Course in Statistical Inference' by L. A. Wasserman, published by Springer in 2004. Below the book cover, there is a search bar and a 'TABLE OF CONTENTS' section. The table lists chapters and their corresponding page ranges. The 'Contents' page is also visible, showing the detailed table of contents for the book.

TABLE OF CONTENTS	
Pages:1 to 25	
Pages:26 to 50	
Pages:51 to 75	
Pages:76 to 100	
Pages:101 to 125	
Pages:126 to 150	
Pages:151 to 175	
Pages:176 to 200	
Pages:201 to 225	
Pages:226 to 250	
Pages:251 to 275	
Pages:276 to 300	
Pages:301 to 325	

Contents	
I Probability	3
1 Probability	3
1.1 Introduction	3
1.2 Sample Spaces and Events	3
1.3 Probability	5
1.4 Probability on Finite Sample Spaces	7
1.5 Independent Events	8
1.6 Conditional Probability	10
1.7 Bayes' Theorem	12
1.8 Bibliographic Remarks	13
1.9 Appendix	13
1.10 Exercises	13
2 Random Variables	19
2.1 Introduction	19
2.2 Distribution Functions and Probability Functions	20
2.3 Some Important Discrete Random Variables	25

... Recap

- Null and alternative hypothesis: $H_0: \theta \in \Theta_0$; $H_1: \theta \in \Theta_1$, $\Theta_0 \cup \Theta_1 = \Theta$
 $\theta = \theta_0 \leftarrow$ simple H_0

- Rejection region: $R \subset \mathcal{X}$; if $\mathbf{x} \in R$ "reject" H_0

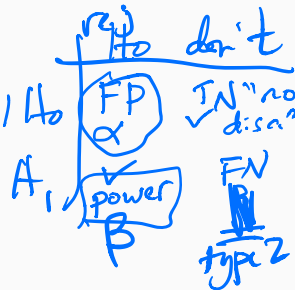
$$X \sim f(x; \theta), \theta \in \Theta$$

$$\Theta_0 \subset \Theta \text{ composite}$$

c to be chosen

- Test statistic and critical value: $R = \{x \in \mathcal{X} : t(x) > c\}$

- Type I and Type II error: $\Pr\{t(X) > c \mid \theta \in \Theta_0\}$, $\Pr\{t(X) \leq c \mid \theta \in \Theta_1\}$



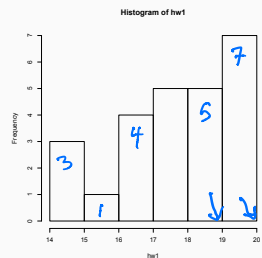
- Power and Size: $\beta(\theta) = \Pr_{\theta}(X \in R)$ $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$
 $\beta(\theta) \stackrel{5}{=} \text{large}$

- Optimal tests: among all level- α tests, find that with the highest power under H_1

level- α means size $\leq \alpha$

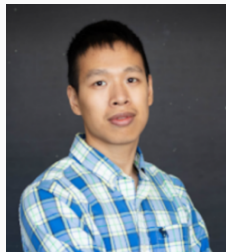
$$\alpha = 0.05 \quad \text{maximize power}$$

1. Homework
2. Choosing test statistics – NP lemma; score, Wald, LRT
3. Significance testing p-values
4. Goodness-of-fit tests



- February 1 3.00 – 4.00 Joanna Mills Flemming
 - “Statistical Tools for Better Understanding Marine Life” [Link](#)
 - January 28 1.00 pm EST Linbo Wang
- CANSSI National Seminar Series
- “The promises of multiple outcomes”

Data Science and Applied Research Series



Choosing test statistics

1. Context $\underline{X} \sim f(\underline{x}; \theta) \quad \theta \in \Theta$ $H_0: \theta \in \Theta_0, H_1: \theta \in \Theta_1$

$$T = t(\underline{X}), \quad R = \{ \underline{x}; T(\underline{x}) > c \}$$

$$\text{s.t. } P_{H_0} \{ T > c \} \leq \alpha$$

Among all f 's $T = t(\underline{X})$ find the one s.t.
 $P_{H_1} \{ T > c \}$ is max'd Most Powerful

NP Lemma If H_0 is simple ($\theta = \theta_0$) and H_1 is simple ($\theta = \theta_1$)
then MP test has $R = \{ \underline{x}: \frac{f(\underline{x}; \theta_1)}{f(\underline{x}; \theta_0)} > k \}$

By the time we impose all'd constraints,
we're really "re-discovering" existing tests
e.g. t-test. "N.P lemma isn't
useful" A's

Choice of test statistics

history 1) : educated guess at $T = t(x)$
then study its properties

known dist- 1) find c_α s.t. $P_{H_0}\{T > c_\alpha\} = \alpha$
2) see if it's ~~off~~ better than T'

Pragmatic theory of tests: use likelihood f-

Choosing test statistics

1. Context

2. Optimal choice – Neyman-Pearson Lemma

Define a "test ϕ "

$$\phi(\underline{x}) = \begin{cases} 1 & \underline{x} \in R \\ 0 & \underline{x} \in R^c \end{cases}$$

Choosing $\phi(\cdot)$ s.t. $\max \int \phi(\underline{x}) f(\underline{x}; \theta_1) d\underline{x}$
s.t. $\int \phi(\underline{x}) f(\underline{x}; \theta_0) d\underline{x} \leq \alpha$

$$R = \left\{ \underline{x}; \frac{f(\underline{x}; \theta_1)}{f(\underline{x}; \theta_0)} > k \right\}$$

find k

Choosing test statistics

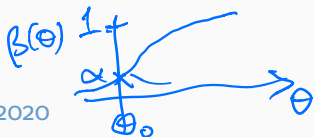
- Context $H_0: \theta = \theta_0$ $H_1: \theta = \theta_1$ or $H_0: X \sim f_0(\cdot)$
 $H_1: X \sim f_1(\cdot)$ $\rightarrow \left\{ \frac{f_1(\underline{x})}{f_0(\underline{x})} > k \right\}$
- Optimal choice – Neyman-Pearson Lemma

- Pragmatic choice – likelihood-based statistics

Luck: MP test of $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$, $\theta_1 > \theta_0$ [$\theta_1 < \theta_0$]
 same, e.g. $\forall \theta_1 > \theta_0$ ($\theta_1 < \theta_0$)

expl. families
 (only case)

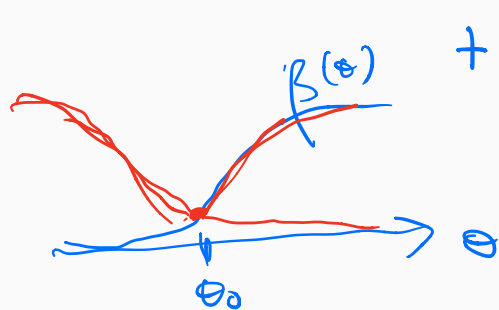
Unif. MP test of $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$
 not $H_1: \theta \neq \theta_0$



Choosing test statistics

Choice 1: use Lik. Rat. (NP lemma) & hope
you're lucky (UMP), eg $\theta > \theta_0$

if not
Choice 2: put some more constraints max'g power
s.t. size $\leq \alpha$



+ want our test (rejⁿ region)

to ~~never~~ have power $< \alpha$ (unbiased)

[~~added~~ made class of $\int$ -s $\phi(x)$
smaller]

$$H_1: \theta > \theta_0$$

$$H_0: \theta \leq \theta_0$$

Choosing test statistics

hypotheses

$$H_0: \theta = \theta_0$$

$$H_1: \theta = \theta_1$$

single vs single

$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$

single composite

"fake comp."

$$H_0: \theta \leq \theta_0$$

$$H_1: \theta \geq \theta_0$$

Composite vs comp.

size $\leq \alpha$, then the test of $\theta = \theta_0$

is the best; gets it size up to α .

$$\theta = (\psi, \lambda)$$

$$H_0: \psi = \psi_0$$

vs $H_1:$

$$\left. \begin{array}{l} \psi > \psi_0 \\ \psi \neq \psi_0 \\ \psi < \psi_0 \end{array} \right\}$$

true comp. H_0

[invariant] impose yet more constraints on $\{\phi(x)\}$

Lehmann & Romano
Testing Stat. Hyp.

Example: Likelihood inference

$$\hat{\theta} \sim N(\theta, I_n^{-1}(\hat{\theta})) \quad \text{know this}$$

$$H_0: \theta = \theta_0 \quad t(x) = \frac{(\hat{\theta} - \theta_0)}{\sqrt{I_n^{-1}(\hat{\theta})}} \underset{H_0}{\sim} N(0, 1) \quad \theta = \theta_0$$

$$W^{(\theta)} = 2 \{ \ell(\hat{\theta}) - \ell(\theta_0) \} \sim \chi_k^2 \quad \text{under } f(x; \theta)$$

$$= 2 \log \frac{L(\hat{\theta}; \underline{x})}{L(\theta_0; \underline{x})} \xrightarrow{d} \chi_k^2$$

$$\text{under } H_0: W(\theta_0) \sim \chi_k^2$$

Example: Likelihood inference

$X_1, \dots, X_n$ i.i.d. $f(x; \theta)$; $\hat{\theta}(X_n)$ is maximum likelihood estimate. From last week:

$$(\hat{\theta} - \theta) / \widehat{se} \sim N(0, 1)$$

To test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ we could use

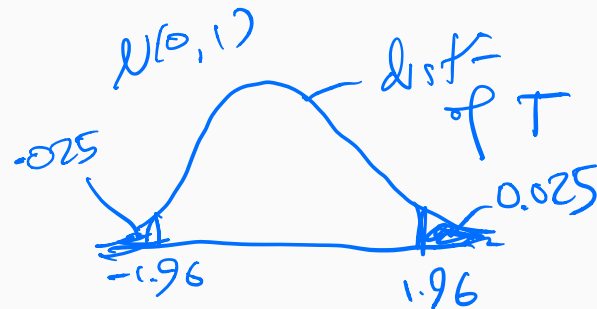
$$W = \widehat{T}(X_n) = (\hat{\theta} - \theta_0) / \widehat{se},$$

The critical region will be $\{\mathbf{x} : |\widehat{T}(\mathbf{x})| > z_{\alpha/2}\}$, i.e. “reject” H_0 when $|\widehat{T}| \geq z_{\alpha/2}$

This test has approximate size α :

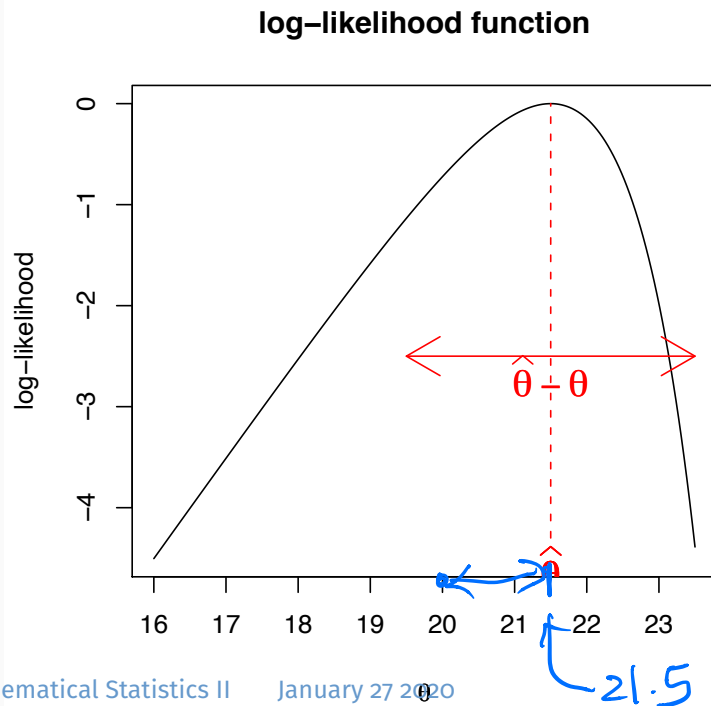
$$\Pr(|\widehat{T}| > z_{\alpha/2}) \doteq \alpha.$$

Power? See Figure 10.1 and Theorem 10.6

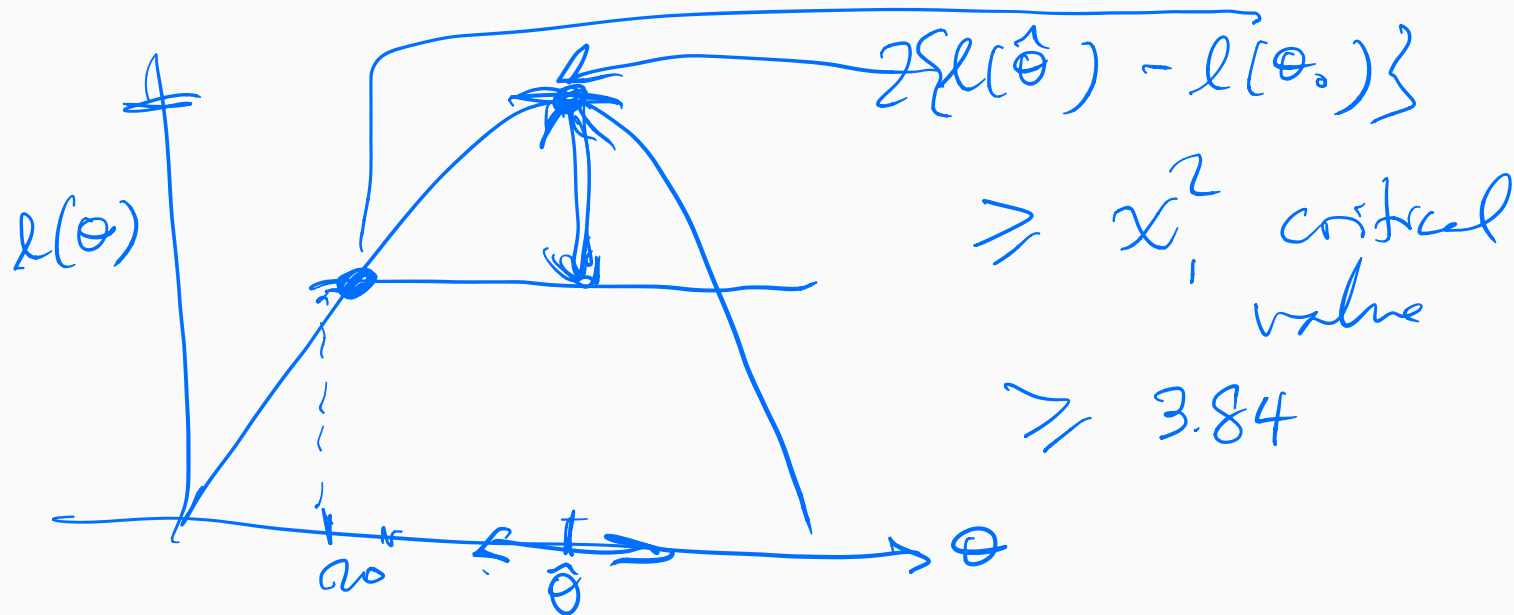


textbook calls
any $\frac{\hat{\theta} - \theta}{\widehat{se}} \sim N(0, 1)$
a Wald statistic

... likelihood inference



$$\mathcal{N}(0,1) \text{ to } \left| \frac{21.5 - 20}{\hat{se}} \right|$$
$$\frac{20 - 21.5}{\hat{se}} \approx 1.96$$
$$< \underline{\underline{-1.96}}$$



can be used in composite $H_0: \psi = \psi_0$

$$\text{bec. } l_p(\hat{\psi}) \rightarrow \frac{\hat{\psi} - \psi_0}{\widehat{se}} \sim N(0, 1)$$

$$\psi \in \mathbb{R} \quad \text{or} \quad 2\{l_p(\hat{\psi}) - l_p(\psi_0)\}$$

$$\text{SM Ch 4} = \underline{\text{Lik. Theory}} \quad \sim \chi^2_1$$

Composite null $H_0: \theta \in \Theta_0$ $H_1: \theta \notin \Theta_0$

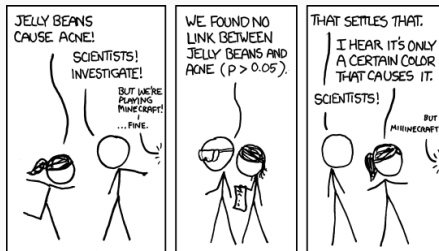
NP I.
 $\downarrow$
 $\frac{L(\theta_1; \underline{x})}{L(\theta_0; \underline{x})}$

GLRT $\triangleq$ $W(\underline{x}) = \frac{\sup_{\Theta} L(\theta; \underline{x})}{\sup_{\Theta_0} L(\theta; \underline{x})} \left[\frac{L(\hat{\theta}; \underline{x})}{L(\hat{\theta}_{0j}; \underline{x})} \right]$

- defines a test stat.

- know $2 \log W(\underline{x}) \xrightarrow[H_0]{d} \chi^2_{\dim(\Theta_0)}$

"p-hacking"



link

Unlikely results

How a small proportion of false positives can prove very misleading

False

True

False negatives

False positives

H_0 is true

1. Of hypotheses interesting enough to test, perhaps one in ten will be true. So imagine tests on 1,000 hypotheses, 100 of which are true.

type 2 error 0.2

2. The tests have a false positive rate of 5%. That means they produce 45 false positives (5% of 900). They have a power of 0.8, so they confirm only 80 of the true hypotheses, producing 20 false negatives.

3. Not knowing what is false and what is not, the researcher sees 125 hypotheses as true, 45 of which are not. The negative results are much more reliable—but unlikely to be published.

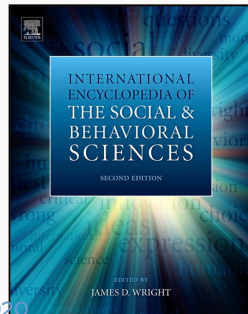
Source: *The Economist*

The new true

Author's personal copy

**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This article was originally published in the *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.

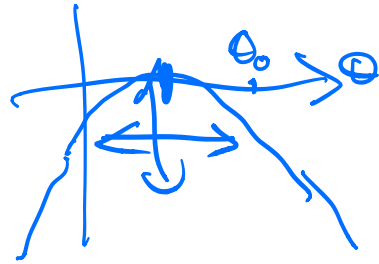


Significance Tests

N.B. LRT & deviance in R

p-value for testing H_0 using statistic $T = t(X)$

$$p^{obs}(x) = P_{H_0} \{ T \geq t^{obs} \}$$



observed

e.g. $T = \frac{|\hat{\theta} - \theta_0|}{\hat{se}}$

e.g. $T = +2 \log \frac{L(\hat{\theta})}{L(\hat{\theta}_0)}$

$$P_{\theta_0} \{ T \geq t^{obs} \}$$

"Probability of observing a result
as or more extreme than we have"
 $\{T \geq t^{obs}\}$

extreme = stronger evidence against H_0

1. H_0 2. T 3. $t^{obs} \rightarrow p^{obs}$

$X_1, \dots, X_n$ iid $\text{Ber}(\theta)$ $\theta = P_n(X_i=1)$

~~$\sum$~~ $\sum X_i = T$

data: 10 coin flips

H T T H H T ...

$t^{obs} = \underline{\underline{\# \text{ of } H}}$

$H_0: \theta = \frac{1}{2}$ "coin is fair"

$p^{obs} \stackrel{\text{def}}{=} P_{\theta=\frac{1}{2}}(\sum_{i=1}^n X_i \geq \underset{7}{t^{obs}}) = \sum_{x=t^{obs}}^n (\frac{1}{2})^x$

how (in)consistent ^{obs'd} data is w H_0