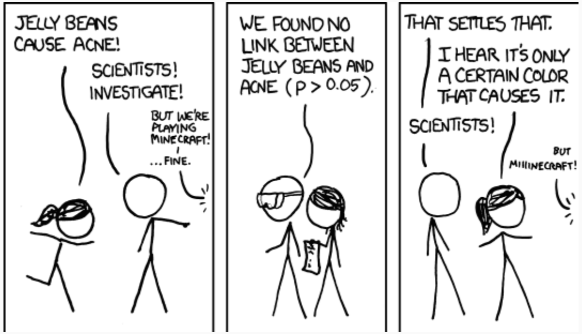
Mathematical Statistics II

STA2212H S LEC9101

Week 3

January 27 2021



Recap

- Quasi-Newton; E-M algorithm — optimization

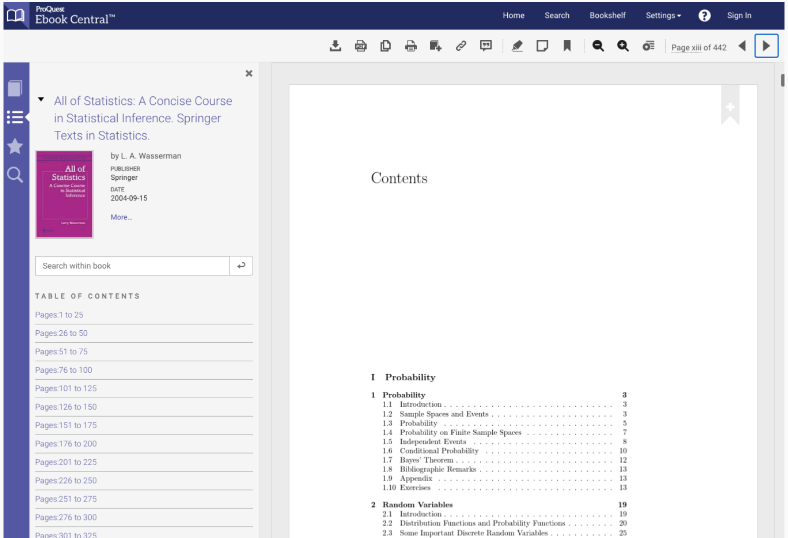
- Formal theory of testing: H_0, H_1, R, T, Type I, Type II error, β, α — hypothesis testing

- *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



… Recap

$\theta = \Theta_0 \leftarrow$ simple H_0

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

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

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

$\Theta_0 \subset \Theta$ composite

??

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

c to be chosen

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

	rej H_0	don't
H_0	FP α	TN "no disc" ✓
H_1	✓ power β	FN type 2

- Power and Size: $\beta(\overset{\delta}{\theta}) = \Pr_\theta(X \in R)$ $\quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

large

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

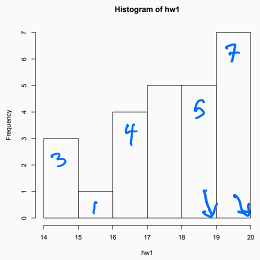
level-α means size $\le \alpha$

$\alpha = 0.05$ maximize power

Today

Start Recording

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

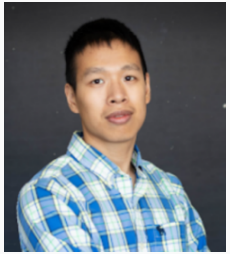


- February 1 3.00 – 4.00 Joanna Mills Flemming
- “Statistical Tools for Better Understanding Marine Life” Link
- January 28 1.00 pm EST Linbo Wang

  CANSSI National Seminar Series

  “The promises of multiple outcomes”

Data Science and Applied Research Series



Choosing test statistics

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

$T = t(\underset{\sim}{X})$, R ~~s.t.~~. $\{x ; T(x) > c\}$

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

among all f's $T = t(\underset{\sim}{X})$ find the one s.t.

$Pr_{H_1}\{T > c\}$ 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 $R = \left\{ x : \dfrac{f(\underset{\sim}{x};\theta_1)}{f(\underset{\sim}{x};\theta_0)} > k \right\}$

By the time we impose add'l 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 it's properties

known dist^{n} i) find c_α s.t. $P_{H_0}\{T > c_\alpha\} = \alpha$

2) see if it's ~~diff~~ better than T'

Pragmatic theory of tests: use likelihood f^{n}

Choosing test statistics

1. Context

2. Optimal choice – Neyman-Pearson Lemma

Define a "test fn"

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \underset{\sim}{x} \in R \\ 0 & \underset{\sim}{x} \notin R \end{cases}$$

Choosing $\phi(\cdot)$ s.t. $\max \int \phi(x) f(x; \theta_1) dx$

s.t. $\int \phi(x) f(x; \theta_0) dx \leq \underset{=}{\alpha}$

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

find k

Choosing test statistics

1. Context $H_0: \theta = \theta_0$ $H_1: \theta = \theta_1$ or $H_0: X \sim f_0(\cdot)$, $H_1: X \sim f_1(\cdot)$ $R = \left\{ \frac{f_1(\underline{x})}{f_0(\underline{x})} > k \right\}$
2. Optimal choice – Neyman-Pearson Lemma
3. Pragmatic choice – likelihood-based statistics

$[\theta_1 < \theta_0]$

Lack: MP test of $H_0: \theta = \theta_0$ vs $\underline{\underline{H_1: \theta = \theta_1}}$, $\theta_1 > \theta_0$

$(\theta_1 < \theta_0)$

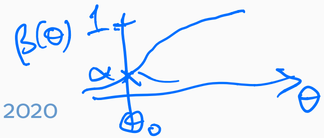
same, e.g. $\forall\, \theta_1 > \theta_0$

$\underline{\underline{\text{expl families}}}$

(only case)

Unif. MP test of $H_0: \theta = \theta_0$ against $\underline{\underline{H_0: \theta > \theta_0}}$

not $H_0: \theta \neq \theta_0$



Choosing test statistics

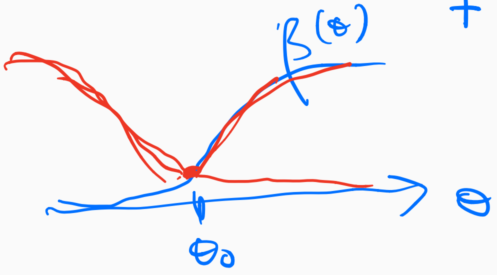
Choice 1: use Lik. Rat. (NP lemma) & hope you're lucky (UMP), eg $\theta > \theta_0$,

if not

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

+ want our test (rej^n region) to never have power $< \alpha$ (unbiased)

[~~added~~ made class of δ's $\phi(x)$ smaller]



$H_1: \theta > \theta_0$

$H_1: \theta < \theta_0$

Choosing test statistics

hypotheses

$H_0: \theta = \theta_0$ $\quad H_1: \theta = \theta_1$ $\quad$ simple vs simple

$H_0: \theta = \theta_0$ $\quad H_1: \theta \cancel{\in} \theta_0$ $\quad$ simple $\quad$ composite

"~~fake~~ comp." $\quad H_0: \theta \leq \theta_0$ $\quad H_1: \theta > \theta_0$ $\quad$ Composite vs comp.

size $\leq \alpha$, then the test of $\theta = \theta_0$ is the best; gets its size up to α.

$\theta = (\psi, \lambda)$ $\quad H_0: \psi = \psi_0$ vs $H_1: \begin{rcases} \psi > \psi_0 \\ \psi \neq \psi_0 \\ \psi < \psi_0 \end{rcases}$ true comp. H_0

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

Lehmann & Romano

Testing Stat. Hyp.

Example: Likelihood inference

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

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

$\theta = \theta_0$

$W^{(\theta)} = 2\{\ell(\hat{\theta}) - \ell(\theta)\} \sim \chi^2_k$ under $f(\underline{x};\theta)$

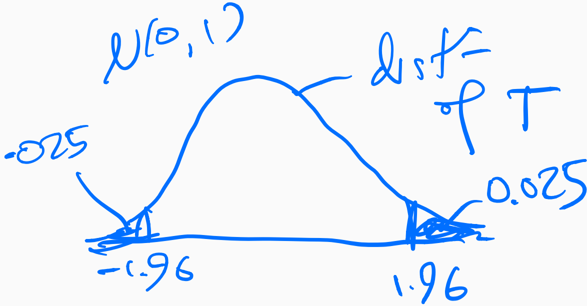
$= 2\log \dfrac{L(\hat{\theta}; \underline{X})}{L(\theta_0; \underline{X})} \xrightarrow{d} \chi^2_k$

Under H_0: $W(\theta_0) \sim \chi^2_k$

Example: Likelihood inference

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

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



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

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

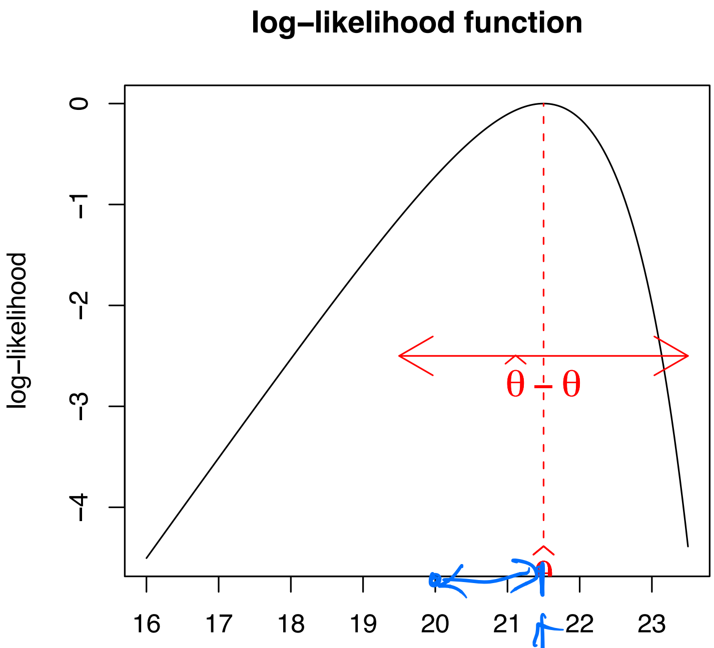
The critical region will be $\{\boldsymbol{x} : |W(\boldsymbol{x})| > z_{\alpha/2}\}$, i.e. "reject" H_0 when $|W| \geq z_{\alpha/2}$

This test has approximate size α:

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

Power? See Figure 10.1 and Theorem 10.6

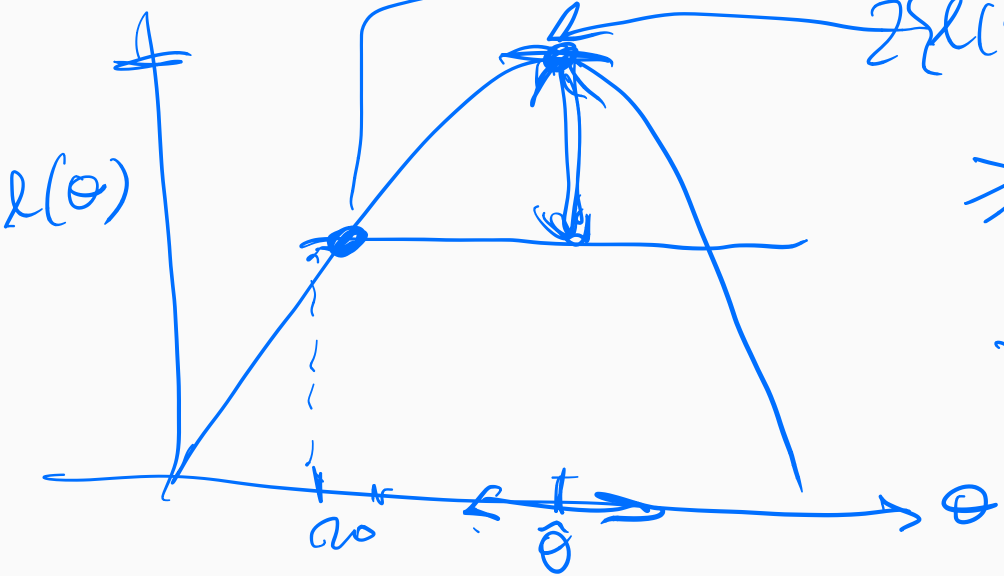
textbook calls any $\dfrac{\tilde{\theta} - \theta}{\widehat{se}} \dot\sim N(0,1)$ a Wald statistic



$N(0,1)$ to $\boxed{\dfrac{21.5-20}{\hat{se}}}$

$\hat{\theta}$ $\dfrac{20 - 21.5}{\hat{se}} \le 1.96$?

< -1.96



$$2\{\ell(\hat{\theta}) - \ell(\theta_0)\}$$

$$\geq \chi^2_1 \text{ critical value}$$

$$\geq 3.84$$

Likelihood ratio tests

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

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

$\psi \in \mathbb{R}$ or $2\{\ell_p(\hat\psi) - \ell_p(\psi_0)\}$

$\sim \chi^2_1$

SM Ch 4 = Lik. Theory

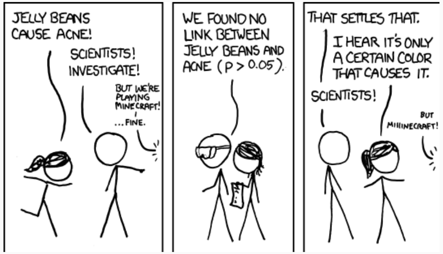
Composite null $H_0: \theta \in \Theta_0$ $H_1: \theta \notin \Theta_0$

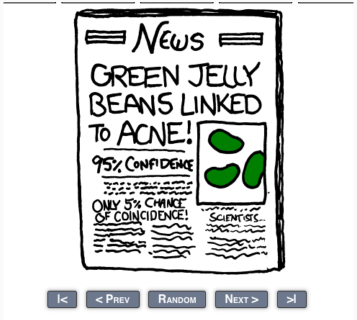
NP l.
$\downarrow$
$\frac{L(\theta_1; \underset{\sim}{x})}{L(\theta_0; \underset{\sim}{x})}$

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

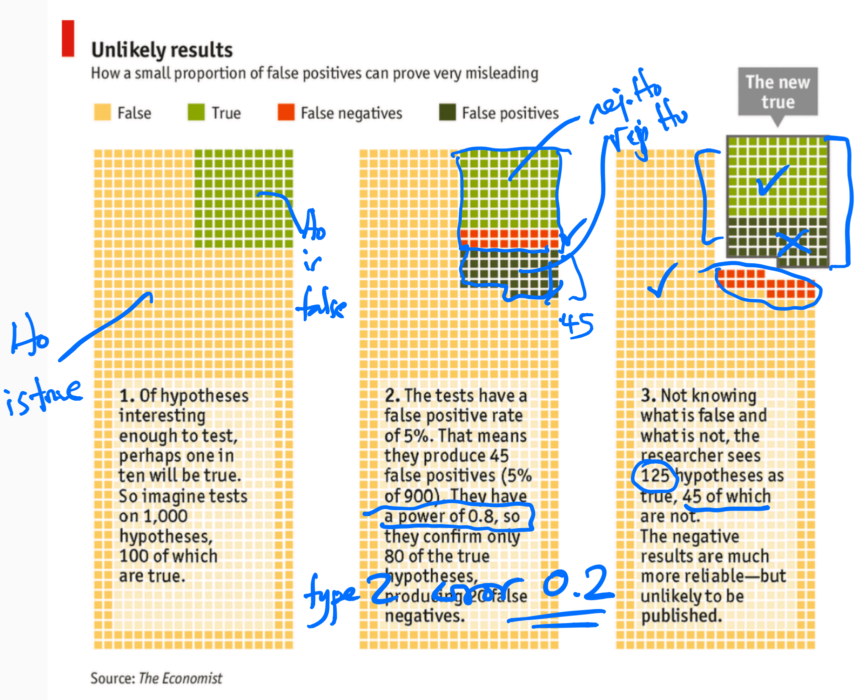
- defines a test stat.
- know $2\log W(\underset{\sim}{X}) \xrightarrow[H_0]{d} \chi^2_{\dim(\Theta_0)}$

“p-hacking”





link

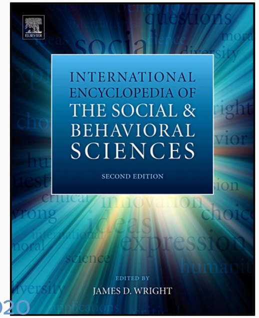


Encyclopedia article

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.



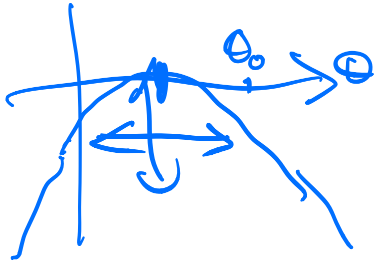
Significance Tests

N.B. LRT & deviance in R

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

$$p^{obs}(\underset{\sim}{x}) = \Pr_{H_0}\{T \geq t^{obs}\}$$

observed



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

e.g. $T = +2 \log \dfrac{L(\hat{\theta})}{L(\hat{\theta}_0)}$ $\quad \Pr_{\theta_0}\{T \geq t^{obs}\}$

"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. $t^{obs} \to p^{obs}$

$X_1, \ldots, X_n$ iid Ber(θ) $\theta = \Pr(X_i = 1)$

$\sum X_i = T$

data: 10 coin flips

H T T H H T ...

t^{obs} = # of H

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

$$p^{obs} = \Pr_{\theta=\frac{1}{2}}\Big(\sum X_i \geq \underset{7}{t^{obs}}\Big) = \sum_{x=t^{obs}}^{n} \left(\tfrac{1}{2}\right)^x$$

how (in)consistent obs'd data is w/ H_0