

p -Values, Confidence Regions

10/17/2023

Outline

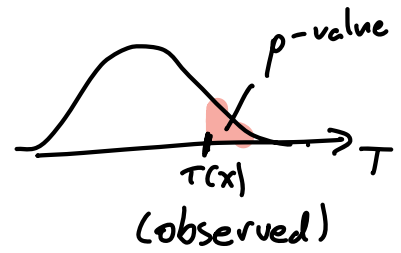
- 1) p -Values
- 2) Confidence regions
- 3) (Mis-)interpreting tests

p-Values

Informal definition: Suppose $\phi(x)$ rejects for large values of $T(x)$.

$$p(x) = \mathbb{P}_{H_0} (T(X) \geq T(x))$$

$$= \sup_{\theta \in \Theta_0} \mathbb{P}_{\theta} (T(X) \geq T(x))$$



Ex $X \sim N(\theta, 1)$ $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$

Two-sided test rejects for large $T(x) = |x|$

$$(\Leftrightarrow \phi_{\alpha}(x) = 1\{|x| > z_{\alpha/2}\})$$

The two-sided p-value is $p(x)$ where

$$\begin{aligned} p(x) &= \mathbb{P}_0 (|X| > |x|) \\ &= 2(1 - \Phi(|x|)) \end{aligned}$$

Formal definition: $\mathcal{P}, \Theta_0, \Theta$.

Assume we have a test ϕ_α for each significance level, $\sup_{\theta \in \Theta_0} \mathbb{E}_\theta \phi_\alpha(X) \leq \alpha$

(non-randomized case: $\phi_\alpha = 1\{x \in R_\alpha\}$)

Assume tests are monotone in α :

if $\alpha_1 \leq \alpha_2$ then $\phi_{\alpha_1}(x) \leq \phi_{\alpha_2}(x)$

(non-randomized: $R_{\alpha_1} \subseteq R_{\alpha_2}$)

Then $p(x) = \inf \{ \alpha : \phi_\alpha(x) = 1 \}$
(= $\inf \{ \alpha : x \in R_\alpha \}$)

(possible to define randomized p-value but not worth it)

Note $p(x) \leq \alpha \Leftrightarrow \phi_{\tilde{\alpha}}(x) = 1 \quad \forall \tilde{\alpha} > \alpha$

For $\theta \in \Theta_0$, $\mathbb{P}_\theta(p(X) \leq \alpha) \leq \inf_{\tilde{\alpha} > \alpha} \underbrace{\mathbb{P}_\theta(\phi_{\tilde{\alpha}}(X) = 1)}_{\leq \tilde{\alpha}} \leq \alpha$

\Rightarrow p-value stochastically dominates $u[0,1]$

IF ϕ_α rejects for large $T(X)$,
reduces to original definition.

Note the p -value depends on

- the model & null hyp.,
- the data, AND
- the choice of test

Ex $X \sim N_d(\theta, I_d)$ $H_0: \theta = 0$ vs $H_1: \theta \neq 0$

We can use $T_1(x) = \|x\|^2$ (χ^2 test)

or $T_2(x) = \|x\|_\infty$ (max test)
 $= \max_i |x_i|$

Very different p -values / power if d large
(choice reflects belief about whether θ is sparse)

Confidence Sets

- Accept/reject decision only so interesting:
- usually we care how big θ is
 - tiny p -value doesn't imply big θ
(big p -value doesn't imply small θ either)

Def $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$

$C(X)$ is a $1-\alpha$ confidence set for $g(\theta)$ if

$$P_{\theta}(C(X) \ni g(\theta)) \geq 1-\alpha, \quad \forall \theta \in \Theta$$

subject verb object

We say $C(X)$ covers $g(\theta)$ if $C(X) \ni g(\theta)$
 $P_{\theta}(C(X) \ni g(\theta))$ is coverage probability
 $\inf_{\theta} P_{\theta}(C \ni g(\theta))$ is conf. level

- Notes
- $C(X)$ is random, not $g(\theta)$
 - Often misinterpreted as Bayesian guarantee
 - Say "C(X) has a 95% chance of covering"
NOT "g(θ) has a 95% chance of being in C"
NEVER "95% chance $g(\theta) \in [0.5, 1.5]$ " (e.g.)

Duality of Testing & Confidence Sets

Suppose we have a level- α test $\phi(x; a)$
of $H_0: g(\theta) = a$ vs. $H_1: g(\theta) \neq a$, $\forall a \in g(\Theta)$

We can use it to make a confidence set for $g(\theta)$:

$$\text{Let } C(X) = \{a : \phi(x; a) < 1\}$$

= "all non-rejected values of θ "

$$\text{Then } \mathbb{P}_\theta(C(X) \not\ni g(\theta)) = \mathbb{P}_\theta(\phi(x; g(\theta)) = 1) \\ \leq \alpha \quad \forall \theta$$

Alternatively, suppose $C(X)$ is a $1-\alpha$ confidence set for $g(\theta)$.

We can use C to construct a test $\phi(x)$ of
 $H_0: g(\theta) = a$ vs. $H_1: g(\theta) \neq a$

$$\phi(x) = \mathbb{1}\{a \notin C(x)\}$$

For θ s.t. $g(\theta) = a$:

$$\mathbb{E}_\theta \phi(x) = \mathbb{P}_\theta(C(X) \not\ni g(\theta)) \leq \alpha$$

This is called inverting the test

Confidence Intervals / Bounds

If $C(X) = [C_1(X), C_2(X)]$ we say
 $C(X)$ is a confidence interval (CI)

$C(X) = [C_1(X), \infty)$: lower conf. bd. (LCB)

$C(X) = (-\infty, C_2(X)]$: upper conf. bd. (UCB)

We usually get LCB / UCB by inverting
a one-sided test in appropriate direction

Called uniformly most accurate (UMA) if test UMP

Get CI by inverting a two-sided test

Called UMAU if test is UMPU

Ex $X \sim \text{Exp}(\theta) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0, \theta > 0$

CDF $P_{\theta}(X \leq x) = 1 - e^{-x/\theta}$

LCB: Invert test for $H_0: \theta \leq \theta_0$

Solve $\alpha = P_{\theta_0}(X > c(\theta_0)) = e^{-c(\theta_0)/\theta_0}$

$c(\theta_0) = \theta_0 \log(1/\alpha) (> 0)$

$X \leq c(\theta_0) \Rightarrow \theta_0 \geq \frac{X}{-\log \alpha}$

$C(X) = \left[\frac{X}{-\log \alpha}, \infty \right)$

UCB: Similar, $C(X) = \left(-\infty, \frac{X}{-\log(1-\alpha)} \right]$

Equal-tailed CI:

Invert equal-tailed test of $H_0: \theta = \theta_0$

$\underbrace{\phi_{\alpha}^{\text{ET}}(X)}_{\text{equal-tailed } H_0: \theta = \theta_0} = \underbrace{\phi_{\alpha/2}^{\geq \theta_0}(X)}_{H_0: \theta \geq \theta_0} + \underbrace{\phi_{\alpha/2}^{\leq \theta_0}(X)}_{H_0: \theta \leq \theta_0}$

$\Rightarrow C(X) = \left[\frac{X}{-\log \alpha/2}, \infty \right) \cap \left(-\infty, \frac{X}{-\log(1-\alpha/2)} \right]$

$= \left[\frac{X}{-\log \alpha/2}, \frac{X}{-\log(1-\alpha/2)} \right]$

Similar for UMPU 2-sided test

(Mis-) Interpreting Hypothesis Tests

Hypothesis tests ubiquitous in science

Common misinterpretations:

1) $p < 0.05$ therefore "there is an effect"
or "the effect size = the estimate"

2) $p > 0.05$ therefore "there is no effect"

3) $p = 10^{-6}$ therefore "the effect is huge"

4) $p = 10^{-6}$ therefore "the data are signif."
and everything about our model
is correct in most naive interp.

5) Effect CI for men is $[0.2, 3.2]$,
for women is $[-0.2, 2.8]$ therefore
"there is an effect for men and not
for women."

Dichotomous test doesn't eliminate uncertainty
(CIs usually less misleading to novices)

How to interpret testing

Learning about the world from data is not easy or automatic!

Hypothesis tests let us ask specific questions about specific data sets under specific modeling assumptions, using specific testing method.

All of these choices bear on the interpretation.

Top-tier medical journals let people publish claims, reporting p -values without saying what model was used or what test was employed

Pretty bad when you think about it!

Hyp. tests can be a good companion to

critical thinking, never a substitute

"All models are wrong, some are useful" but need experience and theory to understand when assumptions do or don't cause real trouble

Conceptual Objections

Q1: Why should I test $H_0: \theta = 0$? No θ is ever exactly 0.

A1: a) Test $H_0: |\theta| \leq \delta$ if you want
If $\text{se.}(\hat{\theta}) \gg \delta$, not much difference.

b) Most two-sided tests justify directional inference:

"If $T > c_\alpha$ declare $\theta > 0$, if $T < c_\alpha$,
declare $\theta < 0$ " with $P(\text{false claim}) \leq \alpha$

c) Harder to answer in non-parametric problems,
e.g. $H_0: P=Q$ vs $H_1: P \neq Q$ for
perm. test, but alternative frameworks like
Bayes force very strong assumptions on us.

Q2: People only like frequentist results like
 p -values, CIs because they mistake them
for Bayesian results.

95% chance $C(X) \ni \theta$ is misinterpreted as
a claim about $p(\theta | X)$.

A2: True, but subjective Bayesian results often
misinterpreted as "the posterior dist. of θ "
when really should be "my posterior opinion about θ "