

Now hypothesis testing. Recall the Sequential hypothesis testing
AKA "significance testing"

$H_0: \theta = 0.5$ "pt. hyp. test", "sharp null hypothesis"

$H_A: \theta \neq 0.5$

$\alpha = 5\%$, $n = 100$

create rejection region

binomial or $\hat{\theta} = \hat{p} \sim N(\theta, \frac{\theta(1-\theta)}{n})^2$

$$\text{Res Reg} = \left[\underset{0.5}{\theta} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\theta(1-\theta)}{n}} \right] = \left[0.5 \pm 2 \sqrt{\frac{0.5 \cdot 0.5}{100}} \right] = [0.4, 0.6]$$

$\hat{p} = 0.35 \notin \text{Res Reg} \Rightarrow \text{Reject } H_0 / \text{Accept } H_A$

$\hat{p} = 0.45 \in \text{Res Reg} \Rightarrow \text{Reject } H_0$

"Significance testing"

$$p\text{-val} := P(\text{seeing this data or more extreme} \mid H_0 \text{ is true}) = \alpha \cdot \mathbb{I}_{\{\hat{\theta} \in \text{Res Reg}\}}$$

Talbot: "How could a significant finding really
hold up: other diseases?? this is not
objective..."

Bayesian Hypothesis Testing for Sharp Null \Rightarrow Loss of discreteness

$P(H_0: \theta = \theta_0 \mid x)$ can be computed directly now!

But if $P(\theta)$ is continuous, $P(\theta = \theta_0 \mid x) = 0 \forall \theta_0$
Since it's a cont. distr.

So what do we do? \approx Res Reg

Calc CR. If $\theta_0 \in \text{CR} \Rightarrow \text{Reject } H_0$

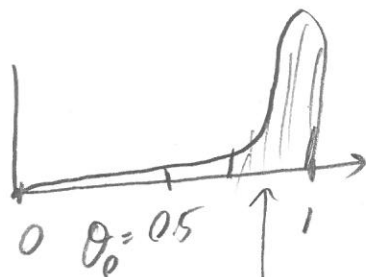
(the qualitative version) $\theta_0 \notin \text{CR} \Rightarrow \text{Reject } H_0 / \text{Accept } H_A$

$$p_{\text{val}} := P(\theta \text{ is more "extreme" than } \theta_0 | X) \\ \approx 2 \min \{ P(\theta > \theta_0 | X), P(\theta < \theta_0 | X) \}$$

Why?
Non-symm
post. distrib.

e.g.

$$H_0: \theta = 0.5, H_1: \theta \neq 0.5$$



95% CR

$$\theta_0 \notin \text{CR} \Rightarrow \text{Reject } H_0$$

$$p_{\text{val}} = 2 P(\theta < 0.5 | X) \approx \text{small}$$

Bayesian Hypothesis Test for "One-Sided Hypothesis"

$$H_0: \theta \leq \theta_0 \text{ e.g. } \theta < 0.5 \text{ coin skewed towards Tails}$$

$$H_1: \theta > \theta_0 \text{ e.g. } \theta > 0.5 \text{ " " " Heads}$$

Matters? No... continuous functions don't matter.

Look at what we have now:

$$p_{\text{val}} := P(H_0 | X)$$

prob. of null hyp. is same given the data:
EXACTLY WHAT WE WANT!!!

$$= P(\theta \leq \theta_0 | X)$$

if $p_{\text{val}} < \alpha \Rightarrow \text{Reject } H_0 / \text{Accept } H_1$

if $p_{\text{val}} \geq \alpha \Rightarrow \text{Accept } H_0 / \text{Reject } H_1$ Why??

Why not pick hypothesis with greater prob i.e. 50%?
 Ockham's Razor... H_0 should be picked to be the
 simple explanation which should only be rejected
 if you have sufficient evidence.

proof by contradiction
 modulo randomness

/ proof by
 low prob.

reduction to
 absurdum

General case of:

$$H_0: \theta \in (H)_0 \subset (H) \quad \text{s.t.} \quad P((H)_0) \in (0,1)$$

$$H_1: \theta \in (H)_0^c \subset (H)$$

$$p_{\text{val}} := P(H_0 | X)$$

In a two sided test

$$H_0: \theta = \theta_0 \quad P(\theta_0) = 0 \quad \text{if } P(\theta) \text{ is continuous}$$

$$H_1: \theta \neq \theta_0 \quad \text{etc... doesn't work}$$

But in the real world $\theta = \theta_0$ is absurd.

E.g.

$\theta = 0.5$ i.e. coin is fair

but you know the coin isn't fair (flip out coin)!

In reality $\theta = 0.5001$ due to neural on one side.

But $0.5001 = 0.5$ for all practical interest & purposes.

Thus in reality, you only care about:

$H_0: \theta \in [\theta_0 \pm \epsilon]$ where ϵ is a m.o.e. you are indifferent to

$H_1: \theta \in [\theta_0 \pm \epsilon]^c$

e.g.

$H_0: \theta \in [0.49, 0.51]$ coin fair

$H_1: \theta \in [0, 0.49) \cup (0.51, 1]$ coin unfair

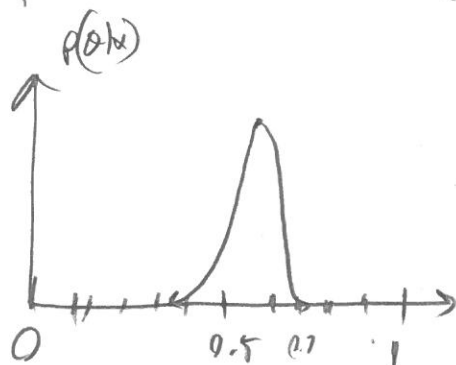
$p_{\text{rel}} := P(H_0 | X)$, Without a lot of data or a coherent prior, you will be falsely ignoring H_0 a lot. Why?
Be careful in specifying ϵ !

Let's do an example. Inference for $\theta_i = P(\text{Heads})$.

$\theta \sim U(0,1) = \text{Bern}(\alpha=1, \beta=1)$ Pick this before everything!

$n=100$, $\#H=54$

$$\theta | X \sim \text{Beta}(\alpha+x, \beta+n-x) = \text{Beta}(1+54, 1+100-54) = \text{Beta}(55, 47)$$



95% CR = HDR due to unimodal

$$= [\text{qbern}(0.025, 55, 47), \text{qbern}(0.975, 55, 47)]$$

$$= [.472, .635]$$

$$\alpha = 5\%$$

(5)

$$H_0: \theta = 0.5, H_1: \theta \neq 0.5 \Rightarrow \text{FTR } H_0 \text{ since } \theta = 0.5 \in CR$$

$$H_0: \theta \in [0.49, 0.51]$$

$$p_{\text{val}} = 2 P(\theta < 0.5) = 2 \cdot 0.213 = .426$$

$$H_1: 0/6$$

Clearly not stat. sign. ~~5%~~

"p-hat"

$$P(H_0 | X) = F(0.51; \alpha = 55, \beta = 47) - F(0.49; \alpha = 55, \beta = 47)$$

$$= 0.117 \notin 5\% \Rightarrow \text{Reject } H_0$$

$$P(H_1 | X) = 0.883 \text{ Isn't that more likely than } H_0 | X?$$

prob con =
infr

change frequency

$$\hat{p} = 0.54$$

$$CI_{\theta, 95\%} = [\hat{p} \pm 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}] = [0.440, 0.640] \not\subset CR_{\theta, 95\%}$$

$$H_0: \theta = 0.5, H_1: \theta \neq 0.5$$

$$p_{\text{val}} := 2 P(\hat{p} > \beta | \hat{p} \sim N(0.5, \sqrt{\frac{0.5 \cdot 0.5}{100}})) = 0.24 \notin 5\%$$

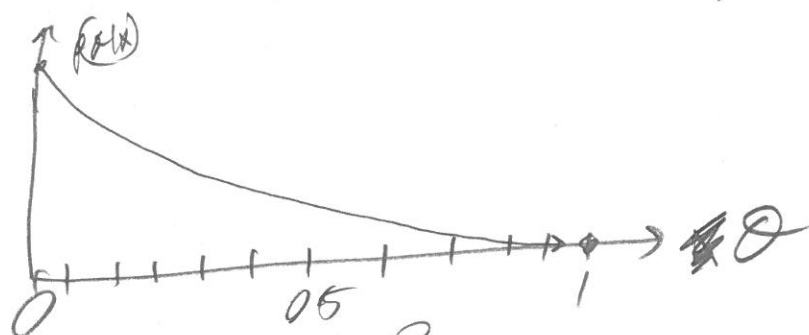
about de same answers...

How about...

$$\theta \sim U(0, 1) = \text{Bern}(\alpha = 1, \beta = 1)$$

$$X = (0, 0, 0)$$

$$\theta|x \sim \text{Beta}(\alpha+x, \beta+n-x) = \text{Beta}(1, 4)$$



Why so low?

$$CR = [0.006, 0.602]$$

$$H_0: \theta = 0.5, \theta \neq 0.5 \Rightarrow 0.5 \in CR \Rightarrow \text{FTR}$$

$$p_{\text{val}} = 2P(\theta > 0.5|x) = .13 \neq 5\%$$

$$H_0: \theta \in [0.49, 0.51]$$

$$H_a: \theta \neq 0.5 = P(\theta \in [0.49, 0.51] | x=0,0,0)$$

$$p_{\text{val}} = 0.01 < 5\% \Rightarrow \text{Reject } H_0$$

Is this a fair $\epsilon = 0.01$ for $n=3$?

$$H_0: \theta \leq 0.1 \text{ severely weighted in favor of tails}$$

$$H_a: \theta > 0.1$$

$$p_{\text{val}} = P(\theta \leq 0.1 | x=0,0,0) = .344 \neq 5\% \quad \text{FTR / accept } H_0$$

\Rightarrow coin is severely weighted in favor of Tails

Frequentist methods?

No chance! $\hat{p}=0$, $SE=0$ & Normal approx begins at $n=3$.

Another way to look at significance testing

- Bayes Factors: ... ~~usually used for~~ ~~Bayesian~~ ~~Model~~ $\Rightarrow P(X|Z)$

$M_1 \quad H_0: \theta = \theta_0 \stackrel{eg}{=} 0.5$

$M_2 \quad H_a: \theta \sim U(0,1) \quad i.e. \quad \theta \neq \theta_0 \quad \text{since} \quad P(\theta_0) = 0 \quad \forall \theta$

Bayes Factor

$$K = B = \frac{P(X|M_1)}{P(X|M_2)} = \frac{\int_{\Theta_{M_1}} P(X|\theta, m_1) P(\theta|m_1) d\theta}{\int_{\Theta_{M_2}} P(X|\theta, m_2) P(\theta|m_2) d\theta}$$

Jeffreys Note

binomial likelihood
degenerate prior

Share of
Numerator
under M_1, M_2

In our case

$P(X|\theta, m_1)$ is the binomial lik.

$P(\theta|m_1)$ is degenerate at $\theta = 0.5$ & $\Theta_{M_1} = \{0.5\}$

$P(X|\theta, m_2)$ is the binomial lik.

$P(\theta|m_2)$ is $U(0,1)$ & $\Theta_{M_2} = [0,1]$

$$= \frac{p(x|\theta=0.5)}{\int_0^1 p(x|\theta) (1) d\theta}$$

$$= \frac{\binom{n}{x} (0.5)^x (0.5)^{n-x}}{\int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta} = B(x+1, n-x+1)$$

with $x=58, n=100$

$$= \frac{(0.5)^{100}}{B(59, 42)} = 2.66 \in [1, 3] \Rightarrow \text{bias much remaining}$$

$$\frac{p(x|H_1)}{p(x|H_0)} = (2.66)^{-1} < 1 \Rightarrow \text{log-odds evidence against coin is unfair}$$

Good example $\alpha=5\%$

$x=61, n=100$

Frequentist pval:

$$pval = 2P\left(\frac{\hat{p} - .5}{\sqrt{\frac{.5 \cdot .5}{100}}} > \frac{.61 - .5}{\sqrt{\frac{.5 \cdot .5}{100}}}\right) = 2P(Z > 2.2) = .027 < \alpha$$

$\Rightarrow \text{Reject } H_0 !!$
Win is unfair

Use Bayes Factors $H_0: \theta = 0.5$
 $H_1: \theta \sim U(0,1)$

19

$$K = B = \frac{P(X|H_1)}{P(X|H_0)} = \frac{B(62, 40)}{0.5^{61} 0.5^{39}} = 1.39 \in (1, 3)$$

Badly work
reasoning !!

Freq concl \neq Bayesian concl.

Why? $\theta \sim U(0,1)!!$

Back to beta-binomial prob

$$P(X|\theta) = \text{bin}(n, \theta), \quad P(\theta) = \text{Beta}(\alpha, \beta)$$

$$P(\theta|X) = \text{Beta}(\alpha + x, n - x + \beta)$$

hyperparameters
 they are "fixed"
 just like the
 separate beta &
 no fixed

If we believe...

$$E(\theta) = 0.9 \Rightarrow \frac{\alpha}{\alpha + \beta} = 0.9$$

$$SE(\theta) = 0.02 \Rightarrow \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} = 0.02$$

& Solve...

— OR —

If we believe we see this coin flip H α times
 and tails β times we get them this way...

Bayesian distributions