two sided **LECTURE 6 Math 341 3/22/10**

Now, hypothesis testing. Recall the frequentist 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 retention region → Similar as $\hat{\theta} = \hat{p} \sim N\left(\theta, \sqrt{\frac{\theta(1-\theta)}{n}}^2\right)$

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

$\hat{p} = 0.35 \notin$ Ret Region $\Rightarrow$ Reject H_0 / Accept H_a

$\hat{p} = 0.45 \in$ Ret Region $\Rightarrow$ Retain H_0

"significance testing"

$$\text{Pval} := P\left(\begin{matrix}\text{seeing this data} \\ \text{or more} \\ \text{extreme}\end{matrix} \;\middle|\; H_0 \text{ is true}\right) =_{\text{rightarrow}}^{\alpha} \left\{\hat{\theta} \in \text{Ret Region}\right\}$$

Jeffreys would say: "How could a frequentist imagine seeing other datasets?? this's not objective..."

Bayesian Hypothesis Testing for Sharp Null $\Rightarrow$ Lots of disagreements

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

But if $P(\theta)$ is continuous, $P(\theta = \theta_0 \mid X) = 0 \;\forall \theta_0$
since its a cont. distr.

So what do we do?

$\approx$ Ret Region

Calc CR. If $\theta_0 \in CR \Rightarrow$ Retain H_0

(the quantile version) $\theta_0 \notin CR \Rightarrow$ Reject H_0 / Accept H_a

$$pval := P(\theta \text{ is more "extreme" than } \theta_0 \mid X)$$

$$\approx 2 \min\{ P(\theta > \theta_0 \mid X),\ P(\theta < \theta_0 \mid X) \}$$

Why? Non-symm. post. distr's.

e.g. $H_0: \theta = 0.5$, $H_a: \theta \neq 0.5$

0 $\theta_0 = 0.5$ 1

95% CR

$\theta_0 \notin CR \Rightarrow$ Reject H_0

$pval = 2P(\theta < 0.5 \mid X) \approx$ small

Bayesian Hypothesis Testing for "One-Sided Hypothesis"

α sign. level

$H_0: \theta \leq \theta_0$ eg $\theta < 0.5$ coin skewed towards Tails

$H_a: \theta > \theta_0$ eg. $\theta > 0.5$ " " " Heads

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

Look at what we have now:

$pval := P(H_0 \mid X)$ prob the null hyp. is true given the data:

EXACTLY WHAT WE WANT!!!

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

if $pval < \alpha \Rightarrow$ Reject H_0 / Accept H_a

if $pval \geq \alpha \Rightarrow$ Accept H_0 / Reject H_a ← Why??

Why not pick hypothesis w/ 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. reductio ad absurdum

General case of:

$H_0 : \theta \in \Theta_0 \subset \Theta$ s.t. $P(\Theta_0) \in (0,1)$

$H_a : \theta \in \Theta_0^c \subset \Theta$

$pval := P(H_0 | X)$

In a two sided test

$H_0 : \theta = \theta_0$ $P(\theta_0) = 0$ if $P(\Theta)$ is continuous

$H_a : \theta \neq \theta_0$ So... 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 (pull out coin)!

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

But $0.5001 = 0.5$ for all practical intents & 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

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

e.g.

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

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

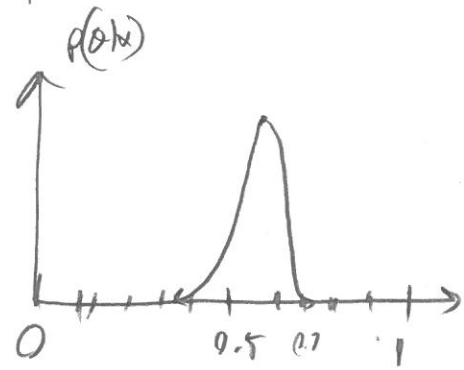
$\text{pval} := P(H_0 \mid X)$. Without a lot of data or a concentrated prior, you will be falsely rejecting H_0 a lot. Why? Be careful in specifying ϵ!

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

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

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

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



95% CR = HDR due to unimodal

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

$= [.442, .635]$

$\alpha = 5\%$

$H_0: \theta = 0.5,\ H_1: \theta \neq 0.5 \Rightarrow$ FTR H_0 s/c $\theta = 0.5 \in CR$

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

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

$H_a: o/t$

Clearly not stat. sign. $\not\leq 5\%$

"pbeta"

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

$= 0.117 \not< 5\% \Rightarrow$ Retain H_0

$P(H_a \mid X) = 0.883$ Isn't that more likely than $H_0 \mid X$?

probcon = confin

Compare frequentist

$\hat{p} = 0.54$

$CI_{\theta, 95\%} = \left[\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [.440, .640] \stackrel{?}{=} CR_{\theta, 95\%}$ ($\hat{p} = .54$, $n = 100$)

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

$p_{val} := 2P\left(\hat{P} > \hat{p} \mid \hat{P} \sim N\left(0.5, \sqrt{\frac{0.5 \cdot 0.5}{100}}\right)\right) = .24 \not< 5\%$

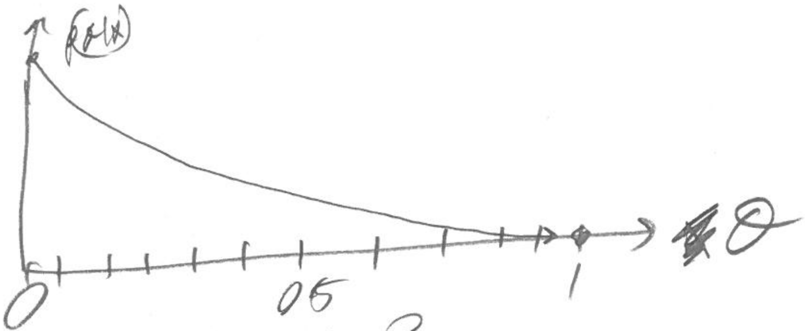
about the same answer...

How about...

$\theta \sim U(0,1) = Beta(\alpha = 1, \beta = 1)$

$X = \langle 0, 0, 0 \rangle$

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



Why so low?

$CR = [0.006, 0.602]$

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

$\text{pval} = 2P(\theta > 0.5 | x) = .13 \not< 5\%$

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

$H_a: \theta \notin$

$= P(\theta \in [0.49, 0.51] \mid x = 0,0,0)$

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

Is this a fair $\overset{=0.01}{\varepsilon}$ for $n = 3$?

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

$H_a: \theta > 0.1$

$\text{pval} = P(\theta \leq 0.1 \mid x = 0,0,0) = .344 \not< 5\%$ FTR / accept H_0

$\Rightarrow$ coin is severely weighted in favor of Tails

Frequentist methods?

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

Another way to look at significance testing

— Bayes Factors: ... ~~(previously used Beta-Binomial Model) $\Rightarrow P(\tilde{X}^* \mid \tilde{X})$~~

$M_1 \quad H_0: \theta = \theta_0 \overset{e.g.}{=} 0.5$

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

Bayes Factor

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

binomial likelihood

degenerate prior

Jeffrey's Notation

Abuse of Notation

Under M_1, M_2

conditional on model is an abuse of notation. It should really be P_m (X | theta) for likelihood and P_m (theta) for prior. The model M specifies both! It allows a general comparison between any likelihood-prior vs any other likelihood prior.

In our case

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

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

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

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

$$= \frac{P(x \mid \theta = 0.5)}{\int_0^1 P(x \mid \theta)\,(1)\, d\theta}$$

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

with $x = 54$, $n = 100$

$$= \frac{(0.5)^{100}}{B(55, 47)} = 2.66 \in [1, 3] \Rightarrow \text{barely worth mentioning}$$

$$\frac{P(X \mid H_a)}{P(X \mid H_0)} = (2.66)^{-1} < 1 \Rightarrow \text{negative evidence against coin is unfair}$$

Another example $\alpha = 5\%$ Frequentist pval:

$x = 61$, $n = 100$

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

$\Rightarrow$ Reject H_0!!

Coin is unfair

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

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

Barely worth mentioning !!

Freq concl $\neq$ Bayesian concl.

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

Back to beta-binomial prob

hyperparameters... they are "fixed" just like the frequentist believe θ is fixed

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

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

If we believe...

$E(\theta) = 0.9$

$SE(\theta) = 0.02$

$$\Rightarrow \quad \frac{\alpha}{\alpha+\beta} = 0.9 \qquad \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = 0.02$$

& Solve...

— or —

If we believe we've seen this coin flip H α times and tails β times... we set them this way...

Mixture distributions...