Choosing Between Hypotheses

Dick Startz

Pearson quote

[the] idea which has formed the basis of all the…researches of Neyman and myself…is the simple suggestion that the only valid reason for rejecting a statistical hypothesis is that some alternative hypothesis explains the events with a greater degree of probability.

Arrow(1)

[statistical significance] is not useless, merely grossly unbalanced if one does not speak also of the power of the test.

- (according to McCloskey)

Arrow(2)

It is very remarkable that the rapid development of decision theory has had so little effect on statistical practice. Ever since the classic work of Neyman and Pearson (1933), it has been apparent that, in the choice of a test for a hypothesis, the power of a test should play a role coordinate with the level of significance.

– writing in honor of Hotelling.

Five take-aways

1. The traditional use of classical hypothesis testing to choose between hypotheses leads to misleading results. As a practical matter, standard practice can be very, *very* misleading. It is entirely possible to strongly reject the null in cases where the null is more likely than the alternative, and vice versa.

Five take-aways

2. Choosing between hypotheses requires invoking Bayes theorem. For the most common empirical applications at least, those where the estimated coefficients are approximately normal, applying Bayes theorem is very easy.

Five take-aways

3. Something has to be said about the likelihood of particular values of the parameter of interest under the alternative. Use of Bayes theorem does require specifying some prior beliefs. Sometimes this can be done in a way in which the specified priors take a neutral stance between null and alternative; sometimes a completely neutral stance is more difficult.

Five take-aways

4. The notion that frequentist procedures specify a null and then take a neutral stance with regard to parameter values under the alternative is wrong. Frequentist decision rules are equivalent to adopting an implicit prior. The implicit prior is often decidedly non-neutral.

Five take-aways

5. Economic hypotheses are usually best distinguished by some parameter being small or large, rather than some parameter being exactly zero versus non-zero. The calculations required for choosing between non-sharp hypotheses are straightforward.

Bayes theorem

- Basis for traditional hypothesis testing

$$\Pr\left(\hat{\theta} \middle| H_0\right)$$

- What we're actually interested in

$$\Pr\left(H_0 \middle| \hat{\theta}\right)$$

Connection: Bayes theorem

$$\Pr(H_0|\hat{\theta}) = \Pr\left(\hat{\theta} \middle| H_0\right) \times \frac{\pi(H_0)}{\Pr\left(\hat{\theta}\right)}$$

Posterior odds

$$\underbrace{PO_{0A}}_{\text{posterior odds}} = \frac{\Pr\left(H_0 \middle| \hat{\theta}\right)}{\Pr\left(H_A \middle| \hat{\theta}\right)} = \underbrace{\frac{\Pr\left(\hat{\theta} \middle| H_0\right)}{\Pr\left(\hat{\theta} \middle| H_A\right)}}_{\text{Bayes factor}} \times \underbrace{\frac{\pi(H_0)}{\pi(H_\mathrm{A})}}_{\text{prior odds}}$$

- For brevity, let $p_{H_0} \equiv \Pr\left(H_0 \middle| \hat{\theta}\right)$, $p_{H_A} \equiv \Pr\left(H_\mathrm{A} \middle| \hat{\theta}\right)$

$$p_{H_0} \equiv \Pr\left(H_0 \middle| \hat{\theta}\right) = \frac{\Pr\left(\hat{\theta} \middle| H_0\right) \cdot \pi(H_0)}{\Pr\left(\hat{\theta} \middle| H_0\right) \cdot \pi(H_0) + \Pr\left(\hat{\theta} \middle| H_A\right) \cdot \left(1 - \pi(H_0)\right)}$$

Arguably neutral priors

$$p_{H_0} = \frac{\Pr(\hat{\theta}|H_0) \cdot \pi(H_0)}{\Pr(\hat{\theta}|H_0) \cdot \pi(H_0) + \Pr(\hat{\theta}|H_A) \cdot (1 - \pi(H_0))}$$

If $\pi(H_0) = \pi(H_A) = {}^1/_2$

$$p_{H_0} = \frac{\Pr(\hat{\theta}|H_0)}{\Pr(\hat{\theta}|H_0) + \Pr(\hat{\theta}|H_A)}$$

Coin toss example

Probability of a head is θ.

- Null is fair coin, $\theta = \theta_0 = 1/2$.
- Alternative is $\theta = \theta_A$.
  - We'll play with $\theta_A = 0.80$.

Suppose 21 heads out of 32 tosses. $\hat{\theta} = 0.656$

$$t = 1.77$$

Suppose 21 heads out of 32 tosses

traditional

- *p*-value 0.039 from normal approximation. (0.025 from exact binomial).
- ➢ Strongly reject null in favor of the alternative

Bayes theorem

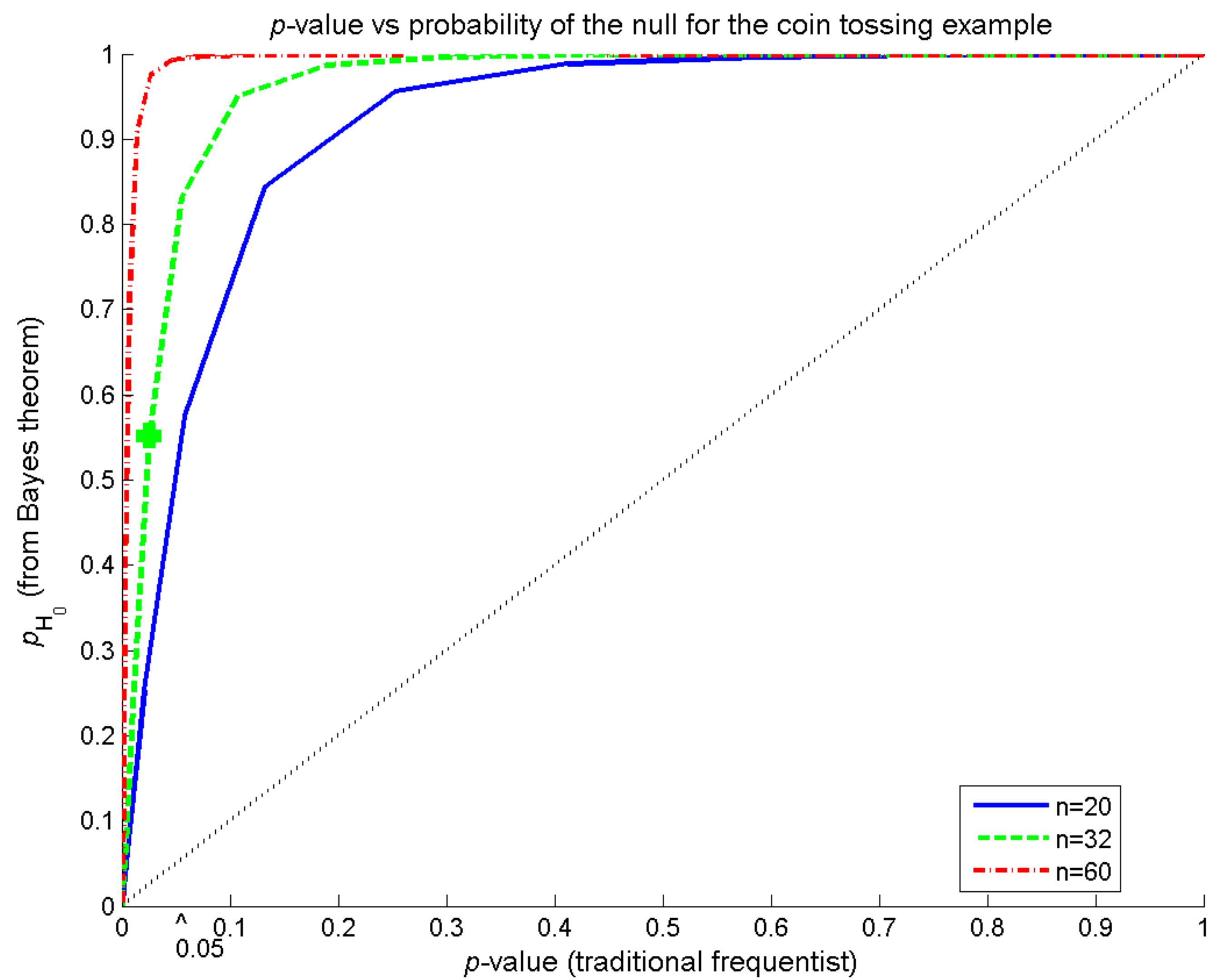
- $\Pr\left(\hat{\theta}\middle|H_0\right) = 0.030$, $\Pr\left(\hat{\theta}\middle|H_A\right) = 0.024$
- $p_{H_0} = \frac{\Pr(\hat{\theta}|H_0)}{\Pr(\hat{\theta}|H_0)+\Pr(\hat{\theta}|H_A)}$ $= 0.55$
- ➢ Evidence is mildly *in favor* of the null

p-value vs p_{H_0}

p-value $= 0.025$ vs $p_{H_0} = 0.55$

Because the tail area… is a probability, there seems to be a natural tendency for a scientist, at least one who is not a trained statistician, to interpret the value … as being closely related to, if not identical to, the probability that the hypothesis H is true.

– DeGroot


p-value vs probability of the null for the coin tossing example
p_H0 (from Bayes theorem)
p-value (traditional frequentist)
^
0.05
n=20
n=32
n=60
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Central lesson

If you start off indifferent between the null and alternative hypothesis, then using statistical significance as a guide as to *which* hypothesis is more likely can be wrong and thinking of the *p*-value as a guide as to *how much* faith you should put in one hypothesis compared to the other can be very wrong.

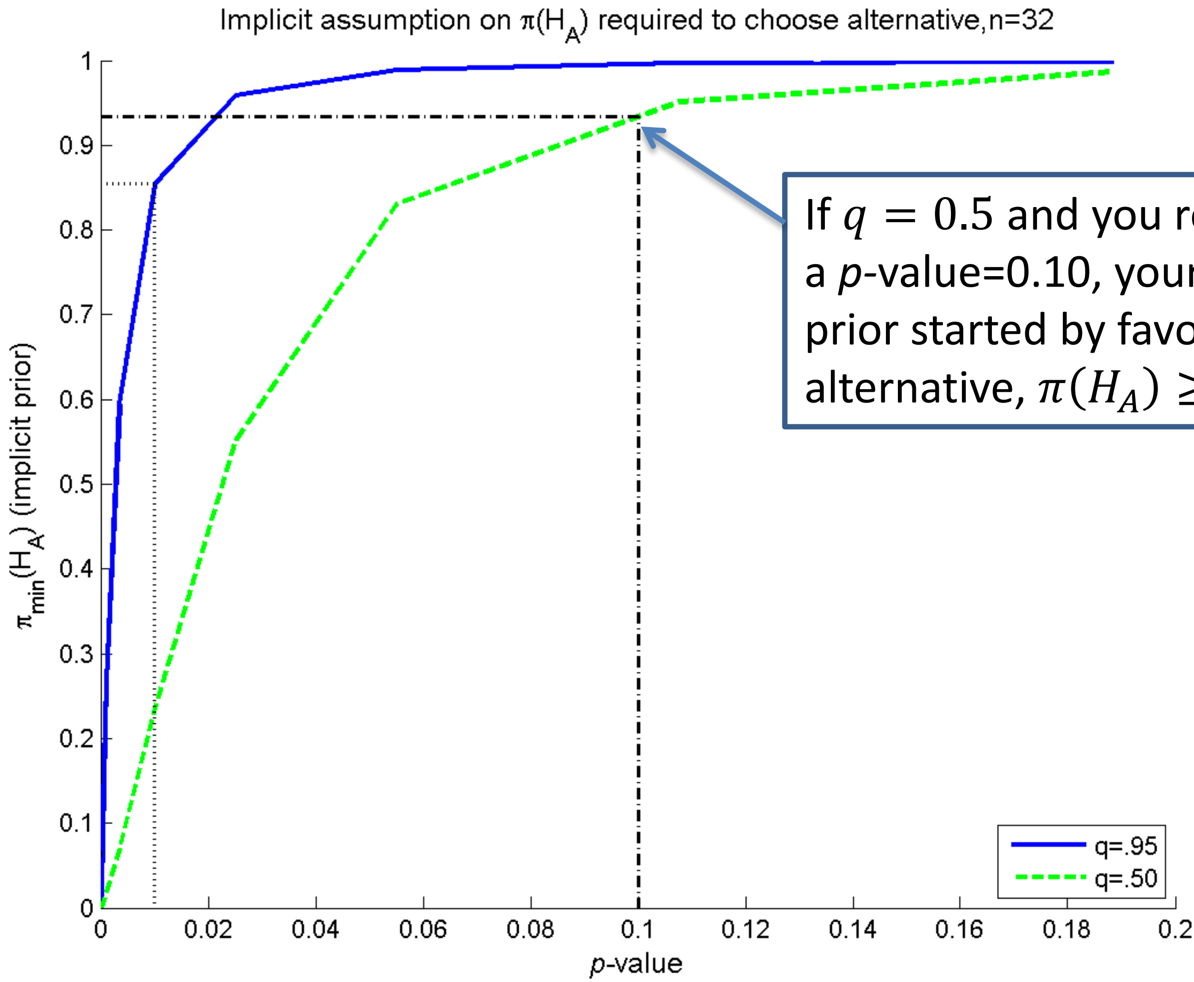
Implicit prior

- Suppose we want to reject null in favor of alternative when $p_{H_A} > q$ (equivalently $p_{H_0} < 1 - q$).
  - $q = 0.95$?
- And we base the decision on the *p*-value.

$$p_{H_0} = \frac{\Pr\left(\hat{\theta}\middle|H_0\right) \cdot \pi(H_0)}{\Pr\left(\hat{\theta}\middle|H_0\right) \cdot \pi(H_0) + \Pr\left(\hat{\theta}\middle|H_A\right) \cdot \left(1 - \pi(H_0)\right)}$$

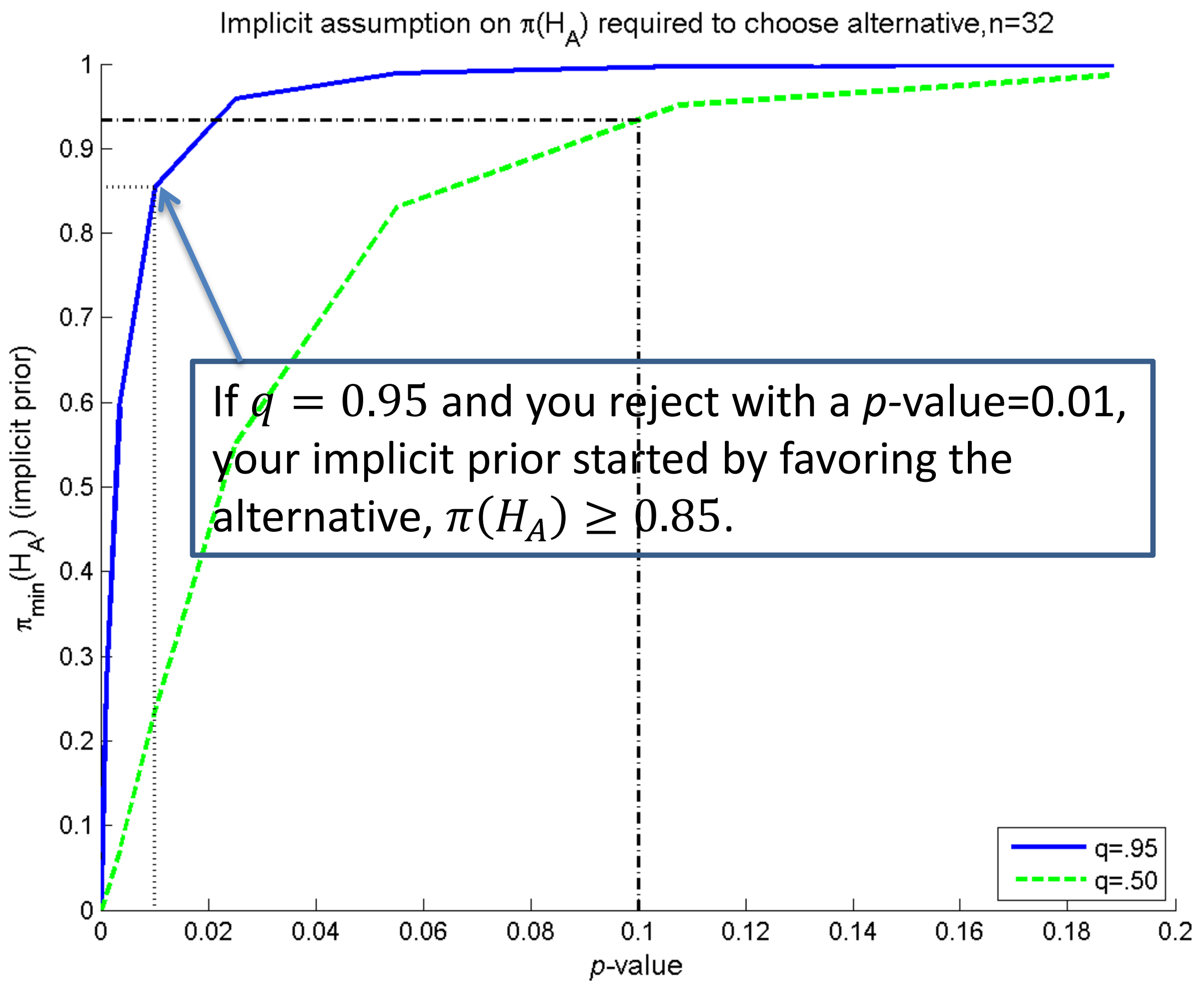
$$\pi_{min}(H_A) = \frac{q \cdot \Pr\left(\hat{\theta}\middle|H_0\right)}{q \cdot \Pr\left(\hat{\theta}\middle|H_0\right) + (1 - q) \cdot \Pr\left(\hat{\theta}\middle|H_A\right)}$$

Implicit prior



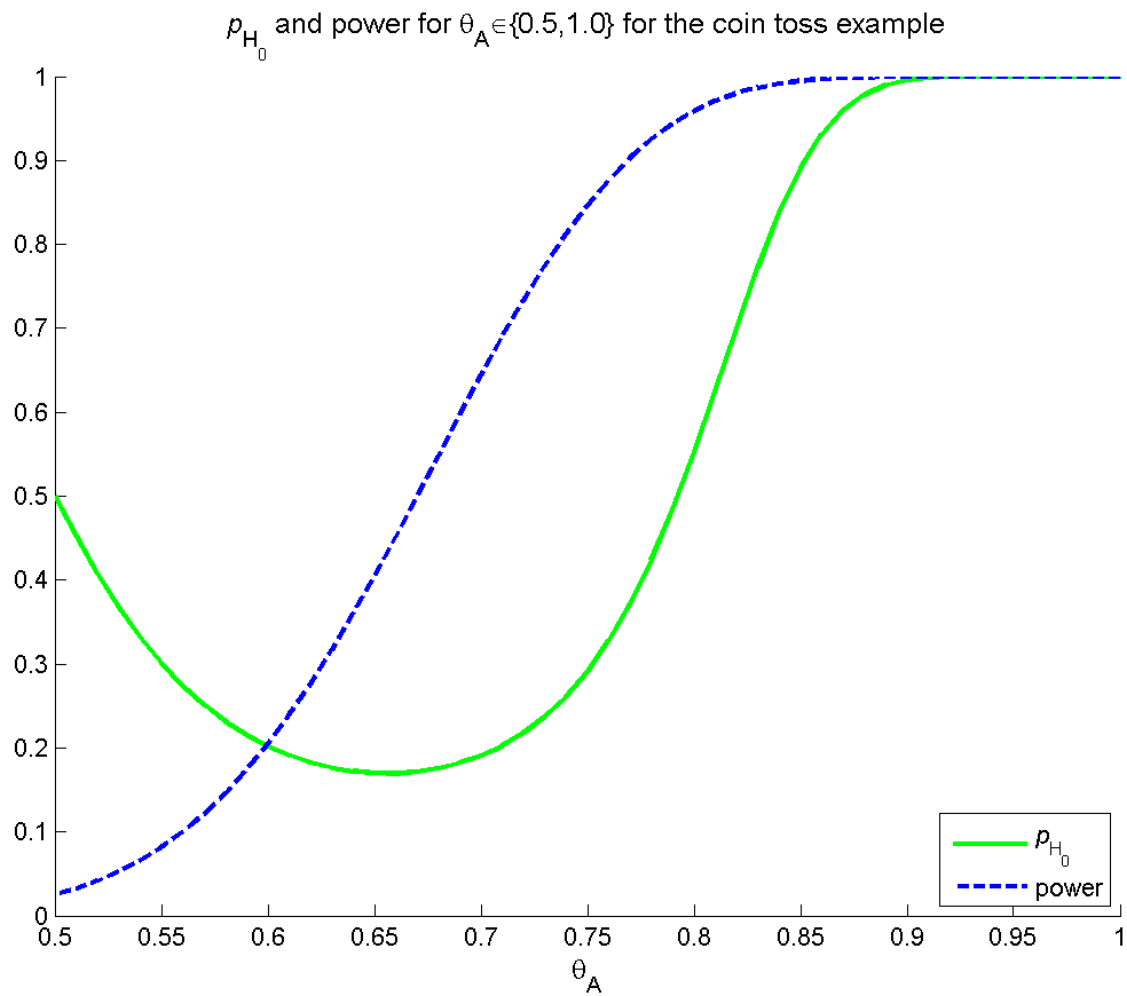
If $q = 0.5$ and you reject with a *p*-value=0.10, your implicit prior started by favoring the alternative, $\pi(H_A) \geq 0.93$.

Implicit prior



If $q = 0.95$ and you reject with a *p*-value=0.01, your implicit prior started by favoring the alternative, $\pi(H_A) \geq 0.85$.

The alternative matters



Hierarchical prior

$$\pi(\theta_A, H_A) = \pi(\theta_A | H_A)\pi(H_A)$$

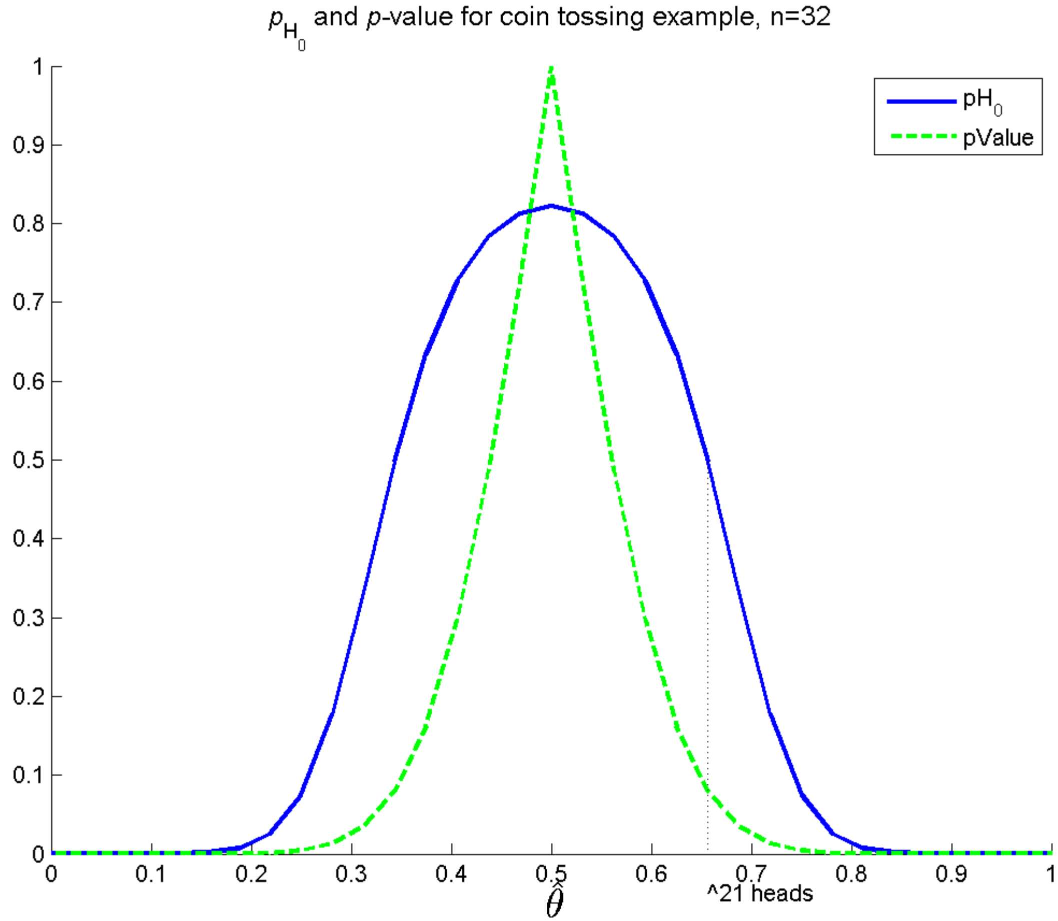
$$p_{H_0} = \frac{\Pr\left(\hat{\theta}\middle|H_0\right)}{\Pr\left(\hat{\theta}\middle|H_0\right) + \int_{-\infty}^{\infty} \Pr\left(\hat{\theta}\middle|\theta_A\right)\pi(\theta_A|H_A)d\theta_A},$$

if $\pi(\mathrm{H}_0) = .5$

Example for coin toss:

$$\theta_A|H_A \sim U(0,1)$$

p_{H_0} under $U(0,1)$ alternative and two-sided p-value, for different outcomes



Approximate p_{H_0} for normal $\hat{\theta}$

$$p_{H_0} \approx \frac{\phi(t)}{\phi(t) + \left[{}^{\mathrm{c}}/_{\sigma_{\hat{\theta}}}\right]^{-1}}$$

$$\theta | H_A \sim U(\theta_0 - c/2\,, \theta_o + c/2)$$

$$p_{H_0} \approx \frac{\phi(t)}{\phi(t) + \left[{}^{\sigma_A}/_{\sigma_{\hat{\theta}}}\right]^{-1} \phi(0)}$$

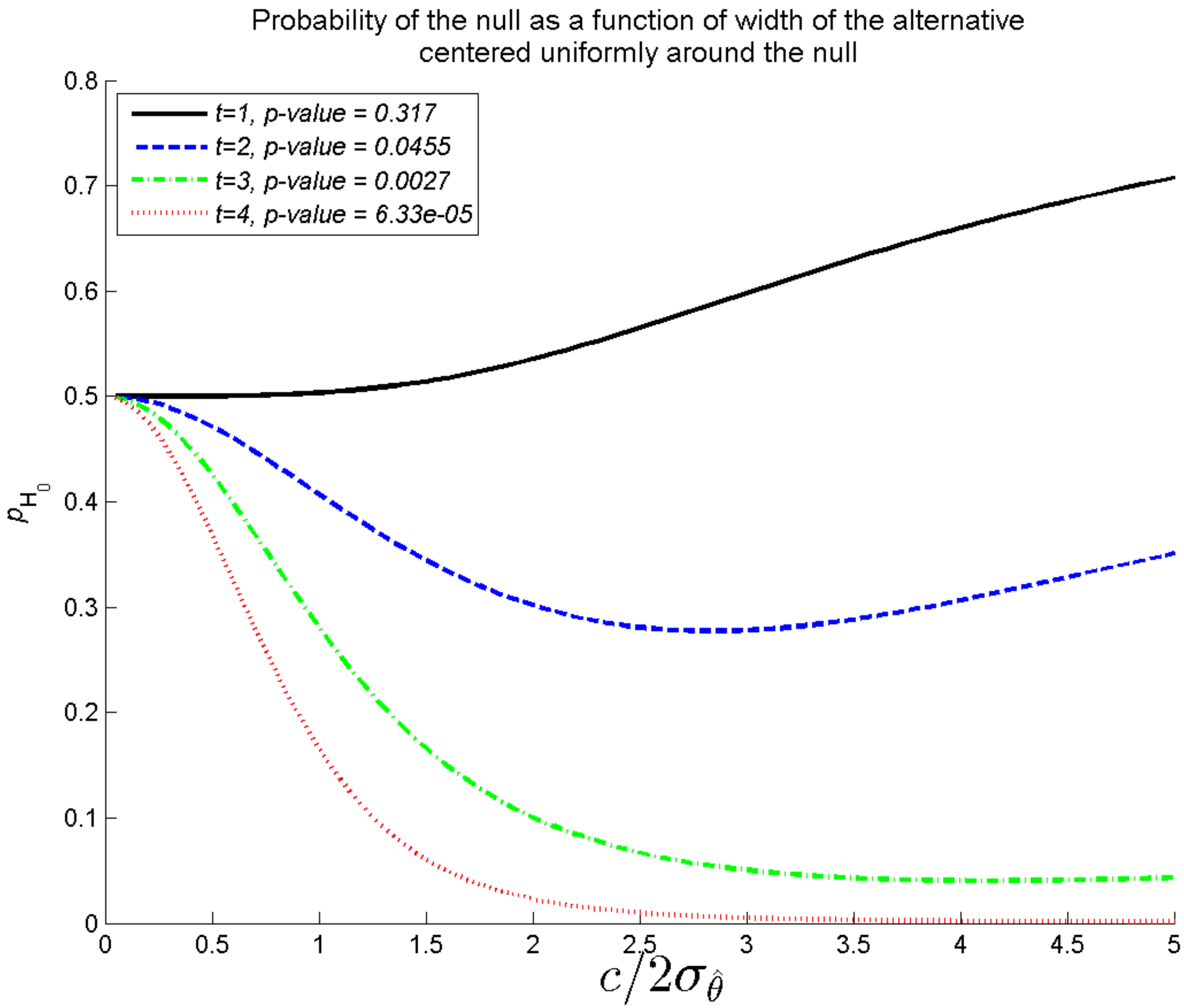
$$\theta | H_A \sim N(\theta_0, \sigma_A^2)$$

p_{H_0} for $\theta | H_A \sim U\left(\theta_0 - \frac{c}{2}, \theta_o + \frac{c}{2}\right)$

$$p_{H_0} \approx \frac{\phi(t)}{\phi(t) + \left[{}^{\mathrm{c}}/_{\sigma_{\hat{\theta}}}\right]^{-1}}$$

Exact:

$$p_{H_0} = \frac{\phi(t)}{\phi(t) + \left[{}^{\mathrm{c}}/_{\sigma_{\hat{\theta}}}\right]^{-1}\left[\Phi\left(t + \frac{c}{2\sigma_{\hat{\theta}}}\right) - \Phi\left(t - \frac{c}{2\sigma_{\hat{\theta}}}\right)\right]}$$



Probability of the null as a function of width of the alternative centered uniformly around the null
t=1, p-value = 0.317
t=2, p-value = 0.0455
t=3, p-value = 0.0027
t=4, p-value = 6.33e-05
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
p_{H_0}
$c/2\sigma_{\hat{\theta}}$

Prior width matters

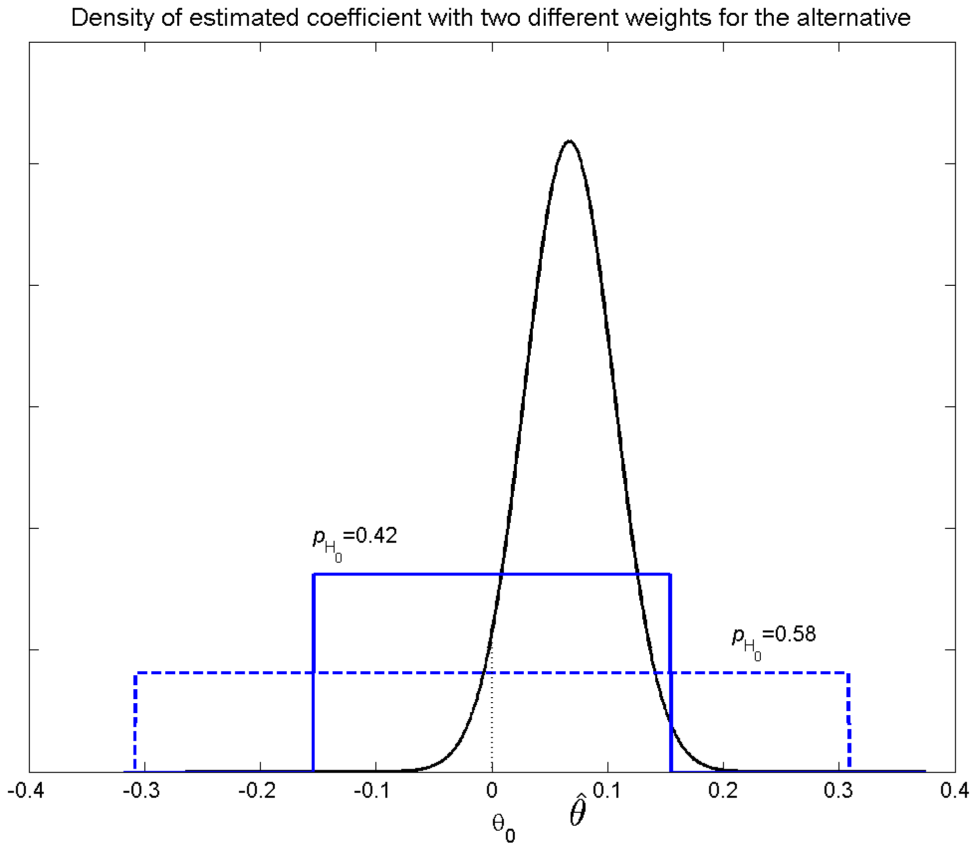
$$p_{H_0} = \frac{\phi(t)}{\phi(t) + \left[{}^{\mathrm{c}}/_{\sigma_{\hat{\theta}}}\right]^{-1} \left[\Phi\left(t + \frac{c}{2\sigma_{\hat{\theta}}}\right) - \Phi\left(t - \frac{c}{2\sigma_{\hat{\theta}}}\right)\right]}$$

$$\lim_{c \to 0} p_{H_0} = 1/2$$

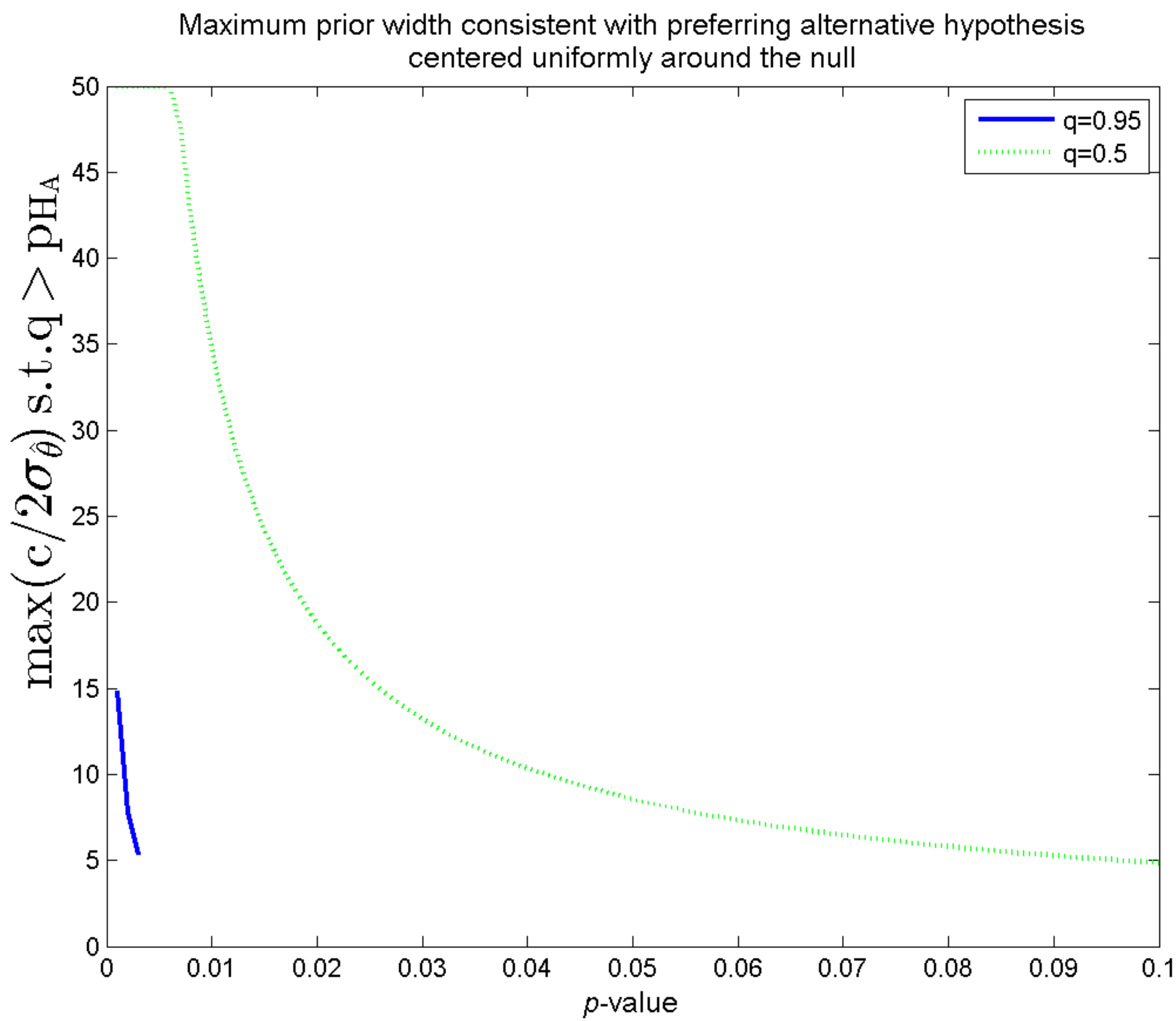
$$\lim_{c \to \infty} p_{H_0} = 1.0$$

Lindley "paradox"

Lindley "paradox"



$$p_{H_0} = \frac{\Pr(\hat{\theta} \mid H_0)}{\Pr(\hat{\theta} \mid H_0) + \int_{-\infty}^{\infty} \Pr(\hat{\theta} \mid \theta_A)\pi(\theta_A \mid H_A)\,d\theta_A}$$



Maximum prior width consistent with preferring alternative hypothesis
centered uniformly around the null
50
45
40
35
30
25
20
15
10
5
0
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
q=0.95
q=0.5
p-value
max(c/2σ_θ̂) s.t. q > pH_A

Weak-form efficiency

- r_t is return on stock market (S&P500)

$$r_t = \alpha + \theta r_{t-1} + \varepsilon_t$$

- Under weak form efficiency,

$$H_0\colon \theta = 0$$

Stock returns are not predictable from lagged returns.

Weak-form efficiency

Monthly

Dependent Variable: RET
Method: Least Squares
Date: 09/02/14 Time: 14:27
Sample: 1957M03 2012M08
Included observations: 666

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
C	5.765684	2.026401	2.845283	0.0046
RET(-1)	0.067381	0.038695	1.741363	0.0821

R-squared	0.004546	Mean dependent var	6.176808
Adjusted R-squared	0.003047	S.D. dependent var	52.01844
S.E. of regression	51.93914	Akaike info criterion	10.74102
Sum squared resid	1791256.	Schwarz criterion	10.75454
Log likelihood	-3574.760	Hannan-Quinn criter.	10.74626
F-statistic	3.032345	Durbin-Watson stat	1.993456
Prob(F-statistic)	0.082083		

Daily

Dependent Variable: RET
Method: Least Squares
Date: 09/02/14 Time: 14:26
Sample: 1/04/1957 8/30/2012
Included observations: 14014

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
C	0.000236	8.50E-05	2.782920	0.0054
RET(-1)	0.026091	0.008445	3.089474	0.0020

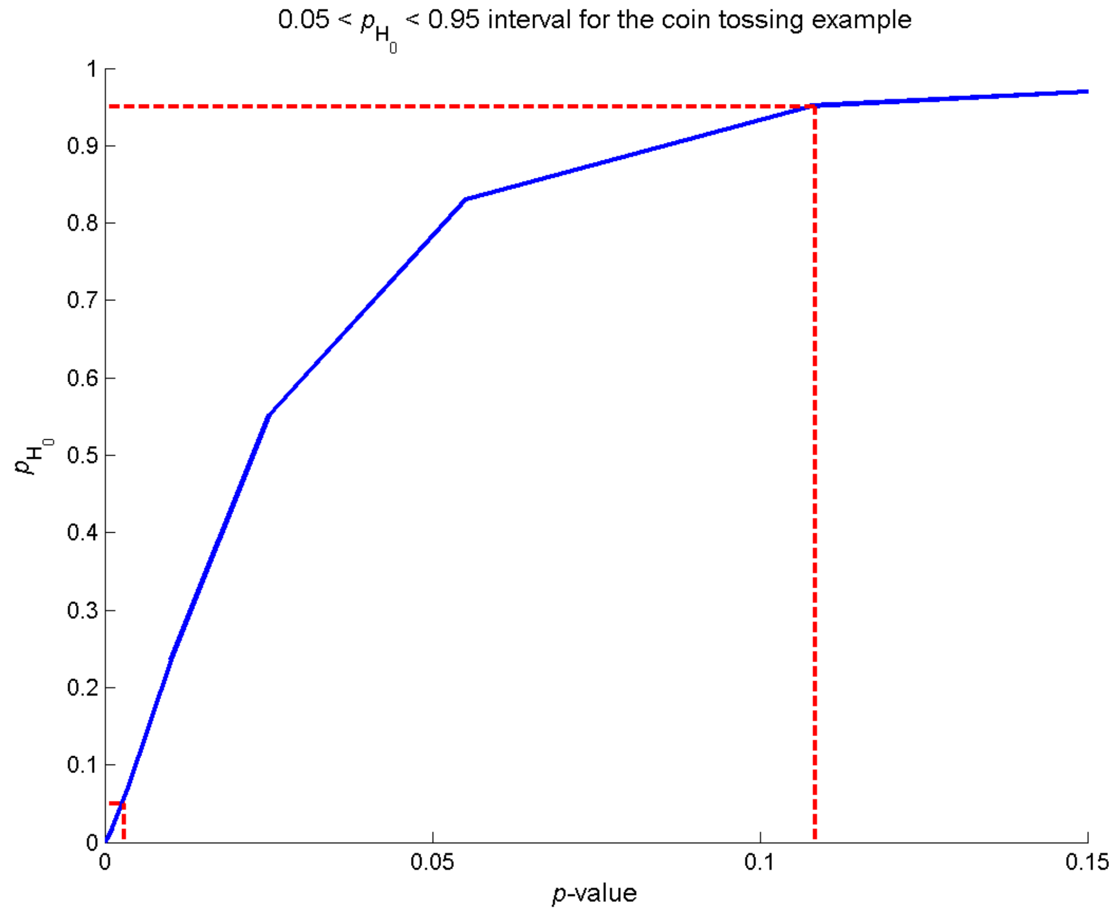
R-squared	0.000681	Mean dependent var	0.000243
Adjusted R-squared	0.000609	S.D. dependent var	0.010057
S.E. of regression	0.010054	Akaike info criterion	-6.361638
Sum squared resid	1.416248	Schwarz criterion	-6.360561
Log likelihood	44578.00	Hannan-Quinn criter.	-6.361280
F-statistic	9.544851	Durbin-Watson stat	1.997881
Prob(F-statistic)	0.002009		

	Data	S&P 500 returns, monthly 1957M03 - 2012M08	S&P 500 returns, daily 1/04/1957 - 8/30/2012
(1)	observations	666	14,014
(2)	Coefficient on lagged return $(\hat{\theta})$ (std. error) [*t*-statistic]	0.067 (0.039) [1.74]	0.026 (0.008) [3.09]
(3)	*p*-value	0.082	0.002
	Probability of weak form efficiency, i.e. $\theta = 0$, single-point null		
(4)	Prior on alternative for lag coefficient $U[-.15, .15]$	0.42	0.11
(5)	Prior on alternative for lag coefficient $U[-.31, .31]$	0.58	0.20
(6)	Prior on alternative for lag coefficient $U[-1,1]$	0.82	0.44
(7)	BIC approximation	0.85	0.50
	Implicit prior to reject weak form efficiency		
(8)	Reject with probability>0.5	$\theta \sim U[-0.22, 0.22]$	$\theta \sim U[-1.25, 1.25]$
(9)	Reject with probability>0.95	$\emptyset$	$\theta \sim U[-0.07, 0.07]$

“Accept” vs Don’t Reject

- $p_{H_0} > 0.95$ (or whatever)
  - Accept
- $p_{H_0} < 0.05$
  - Reject
- $0.05 < p_{H_0} < 0.95$
  - Need another grant

Example from the original coin toss model



Nonsharp null

	Data	S&P 500 returns, monthly 1957M03 2012M08	S&P 500 returns, daily 1/04/1957 8/30/2012
	Probability of weak form efficiency, finite null		
(1)	Null $\sim U[-0.0001, 0.0001]$, alternative $\sim U[-.15, .15]$ excluding $U[-0.0001, 0.0001]$	0.42	0.11
(2)	Null $\sim U[-0.02, 0.02]$, alternative $\sim U[-.15, .15]$ excluding $U[-0.02, 0.02]$	0.43	0.68

Bayesian Information Criterion (BIC)

$$B = -t^2 + \log n$$

$$p_{H_0} = \frac{\exp(.5 \cdot B)}{1 + \exp(.5 \cdot B)}$$

- Disadvantage:
  - Seems to “overly” favor the null
- Advantage
  - Investigator can’t monkey with prior

Decision-theoretic problem

- *Loss function over action* A*,* $L(A,\theta)$

$$\min_A \int L(A,\theta)\,p\left(\theta \middle| \hat{\theta}\right) d\theta$$

- Diffuse prior just fine.
- Not a short cut method for hypothesis testing.

"That is the case as it appears to the police, and improbable as it is, all explanations are more improbable still."

-Sherlock Holmes

Five take-aways

1. Standard practice can be very, *very* misleading.
2. Choosing between hypotheses requires invoking Bayes theorem.
3. Sometimes neutral priors are easy to specify; sometimes not.
4. Frequentist decision rules are equivalent to adopting an implicit prior. The implicit prior is often decidedly non-neutral.
5. The calculations required for choosing between non-sharp hypotheses are straightforward.

One take-away

$$\cancel{p - \text{value} < 0.05}$$

$$p_{H_0} \approx \frac{\phi(t)}{\phi(t) + \left[{}^{\mathrm{c}}/_{\sigma_{\hat{\theta}}}\right]^{-1}}$$