National Cancer Institute

Insights into p-values and Bayes Factors through False-Positive and False-Negative Bayes Factors

Hormuzd Katki

Division of Cancer Epidemiology and Genetics, NCI, NIH, DHHS

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

Introduction

- Bayes Factor and p-values often lead to different conclusions about a null hypothesis
  - Example: GWAS of prostate cancer
  - Bayes Factors may be more theoretically-justified, but yield no insight about false positive or negative results
- I introduce the
  - False Positive and False Negative Bayes Factors
  - These are functions of the p-value and power vs. an alternative hypothesis
  - Their ratio is approximately the Bayes Factor
  - They help reveal the 2 SNP mutations declared positive that, with later data, are now known to be negative

Bayes Factor

$$BF = \frac{P(T = t_{obs} | H_0)}{P(T = t_{obs} | H_1)}$$

- Likelihood of observing the study data under the null to the alternative
- Based on $Posterior = P(H_0 | T = t_{obs})$
- Uses all the study information
- Has ancient history
  - Jeffreys, 1939
  - Lindley, 1957
  - Cox and Hinkley, 1974
  - Lindley and Scott, 1984

p-values

$$P(T > t_{obs} \mid H_0) \quad vs. \quad P(|T| > t_{obs} \mid H_0)$$

- Two major differences between Bayes Factors and p-values
  - P-values don't account for alternative hypotheses H_1
    - P-value may be small, but if the alternatives are even less likely, the p-value does not provide evidence against the null
  - P-values use tail probabilities, not the point of data observed
    - P-values consider more extreme values of the test statistic that are not observed. Seems irrelevant for drawing inferences from the value of the observed test statistic.

False Positive Report Probability

$$FPRP = P(H_0 | T > t_{obs})$$

Sholom Wacholder

- P-value spirit: use tail probability (1- or 2- sided)
  - ♦ Only uses the information:“Is the study significant?”
    - Doesn’t use all the study data
  - ♦ Has long history:
    - Jeffreys (1939) – Bayesian interpretation of p-values
    - Morton (1955) -- LOD scores in linkage studies
    - Peto et al (1976) -- Interpret p-values from clinical trials
    - Lee and Zelen (2000) -- Determine sample size
    - Storey (2003) – Bayesian interpretation of False Discovery Rates
    - Wacholder et al (2004) -- Interpret p-values in data analyses

Katki, JSM Proceedings 2007; Katki, Am J Epidemiol 2008

False Positive Bayes Factor

$$FPBF = \frac{P\left(T > t_{obs} \mid H_0\right)}{P\left(T > t_{obs} \mid H_1\right)} = \frac{p}{p_1}$$

- FPBF is the p-value divided by the “power”
  - “If I declare my study significant, how much evidence is there against the null hypothesis?”
    - FPBF <= 1
  - “power” is computed after observing data; not unconditional frequentist power

Katki, JSM Proceedings 2007; Katki, Am J Epidemiol 2008; Katki, Biostatistics, submit

False Negative Report Probability

$$FNRP = P(H_1 | T < t_{obs})$$

- Two decisions/mistakes are possible
  - Only uses the information:"Is the study not significant?"
  - Decision to declare significance is merely conventional, so both possibilities must be considered
  - Need to trade off with FPRP; complements FPRP
  - The Bayes Factor doesn't have a complement since

$$Posterior = P(H_1 | T = t_{obs}) = 1 - P(H_0 | T = t_{obs})$$

Katki, JSM Proceedings 2007; Katki, Biostatistics, submitted

False Negative Bayes Factor

$$FNBF = \frac{P(T < t_{obs} | H_1)}{P(T < t_{obs} | H_0)} = \frac{1 - p_1}{1 - p}$$

- FNBF
  - Complement of power divided by complement of p-value
- How to declare 'significant' vs 'insignificant'?
  - If both FPBF and FNBF are low, how to choose?
  - If FPBF = FNBF (happens when one-sided $p+p_1=1$; BF=1)
    - Indecision region
    - If you declare significance, study favors H_0 over H_1 with same strength as that if you declared insignificance the study favors H_1 over H_0
  - Small FPBF requires large FNBF to constitute evidence against the null

Katki, JSM Proceedings 2007; Katki, Biostatistics, submitted

Reconcile BF with FPBF and FNBF

- If FPBF = FNBF:
  - When H_0 and H_1 are close, so $p \approx p_1$
  - For hypothesis-driven alternative: one-sided $p+p_1=1$
- In both cases, BF=1 as well
- Thus reconcile BF with FPBF and FNBF:

$$S = \frac{FPBF}{FNBF} = BF$$

Katki, JSM Proceedings 2007; Katki, Biostatistics, submitted

Reconcile BF with FPBF and FNBF

$$P(Z>z)\approx \frac{\exp(-z^2/2)/\sqrt{2\pi}}{z+0.8\exp(-0.4z)}$$

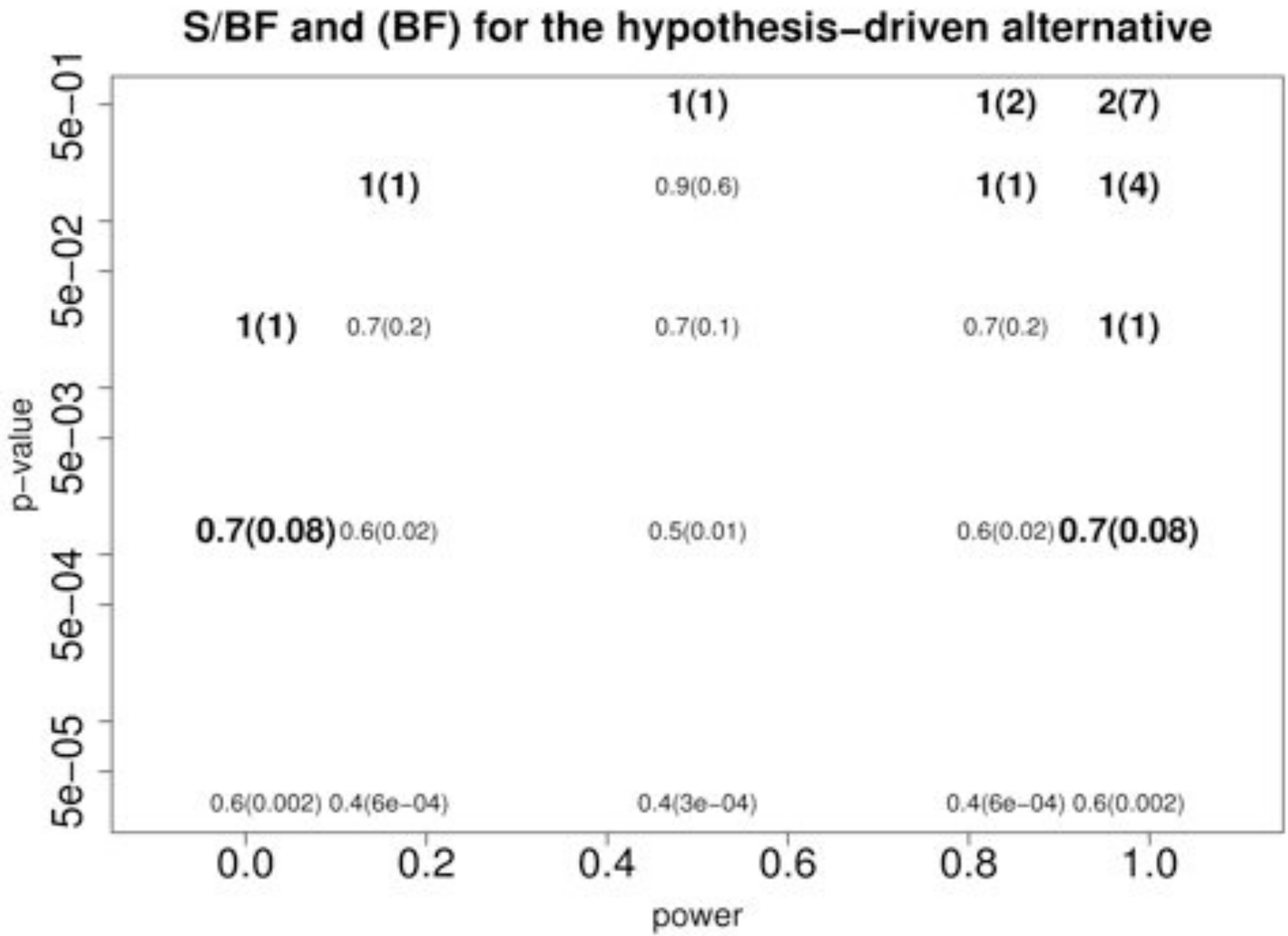
- Under Hart's approximation,

$$S=\frac{FPBF}{FNBF}\approx\frac{\exp(-z_0^2/2)}{\exp(-z_1^2/2)}\frac{|z_1|+0.8(\exp(-0.4|z_1|))}{|z_0|+0.8(\exp(-0.4|z_0|))}$$

$$=BF\times\frac{\sigma_0}{\sigma_1}\frac{|z_1|+0.8(\exp(-0.4|z_1|))}{|z_0|+0.8(\exp(-0.4|z_0|))}$$

$$(z_0=\hat{\beta}/\sigma_0; \quad z_1=(\hat{\beta}-\beta_1)/\sigma_1)$$

Katki, Biostatistics, submitted


S/BF and (BF) for the hypothesis−driven alternative
p−value
5e−01
5e−02
5e−03
5e−04
5e−05
power
0.0
0.2
0.4
0.6
0.8
1.0
1(1)
1(2)
2(7)
1(1)
0.9(0.6)
1(1)
1(4)
1(1)
0.7(0.2)
0.7(0.1)
0.7(0.2)
1(1)
0.7(0.08)
0.6(0.02)
0.5(0.01)
0.6(0.02)
0.7(0.08)
0.6(0.002)
0.4(6e−04)
0.4(3e−04)
0.4(6e−04)
0.6(0.002)

Wilcox's Example

	OR	CI	p-value
Factor A	2.0	(1.2 - 3.3)	0.0067
Factor B	4.0	(1.3 - 12)	0.013

- Set Alternative OR=3
- Assume test statistics under null and alternative are
  - Centered at OR=1 and OR=3
  - Normally distributed with equal variances
  - Back the variance out of the CI

Wilcox's Example: What Happened?

	OR	p-value	FPBF	BF	FNBF	S
Factor A	2.0	0.0067	1/278	1/11	1/17	1/16
Factor B	4.0	0.013	1/45	1/19	1/1.4	1/32

- Both FPBF and FNBF favor Factor A
  - 1/17 is between 1/8 ('fairly strong') and 1/32 ('quite strong') evidence.
  - Warns us that the BF and the p-value will disagree
  - The BF and S are close, justifying using FPBF & FNBF
- Bayes Factors do not yield insight about why they disagree with p-values
  - But FPBF and FNBF show us this insight

CGEMS: Cancer GEnetic Markers of Susceptibility

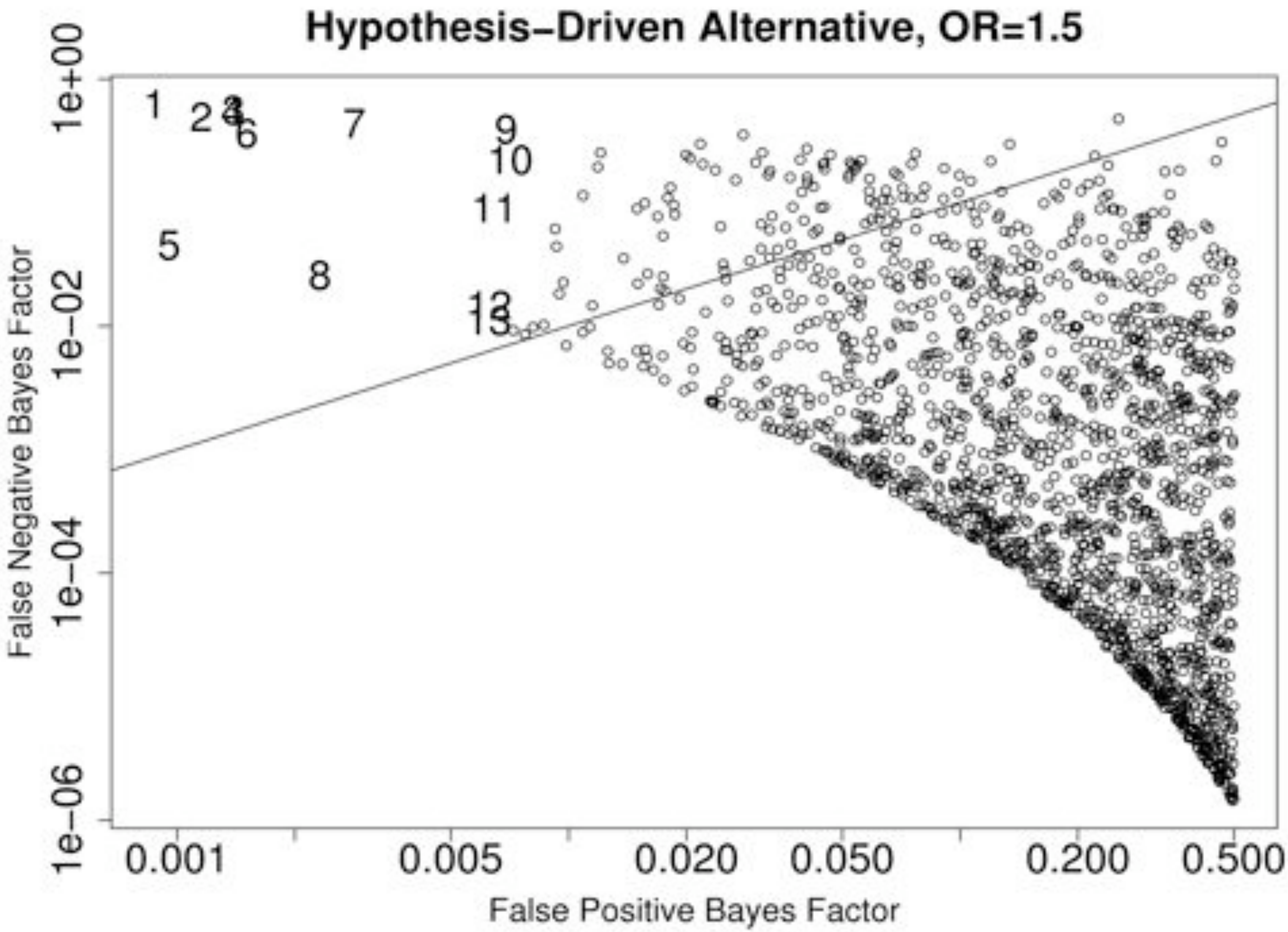
- Over 500,000 SNP genetic mutations
  - Single Nucelotide Polymorphism (SNP): mutation at a single base-pair
  - Can be 4 possibilities (A,C,G,T), but almost always binary
- Focus on Chromosome 8 in 8q24
  - PLCO: ~1200 prostate cancers, ~1200 controls per SNP
- Analysis
  - Odds ratios (OR) associating each SNP with presence of prostate cancer (binary)

CGEMS: Cancer GEnetic Markers of Susceptibility

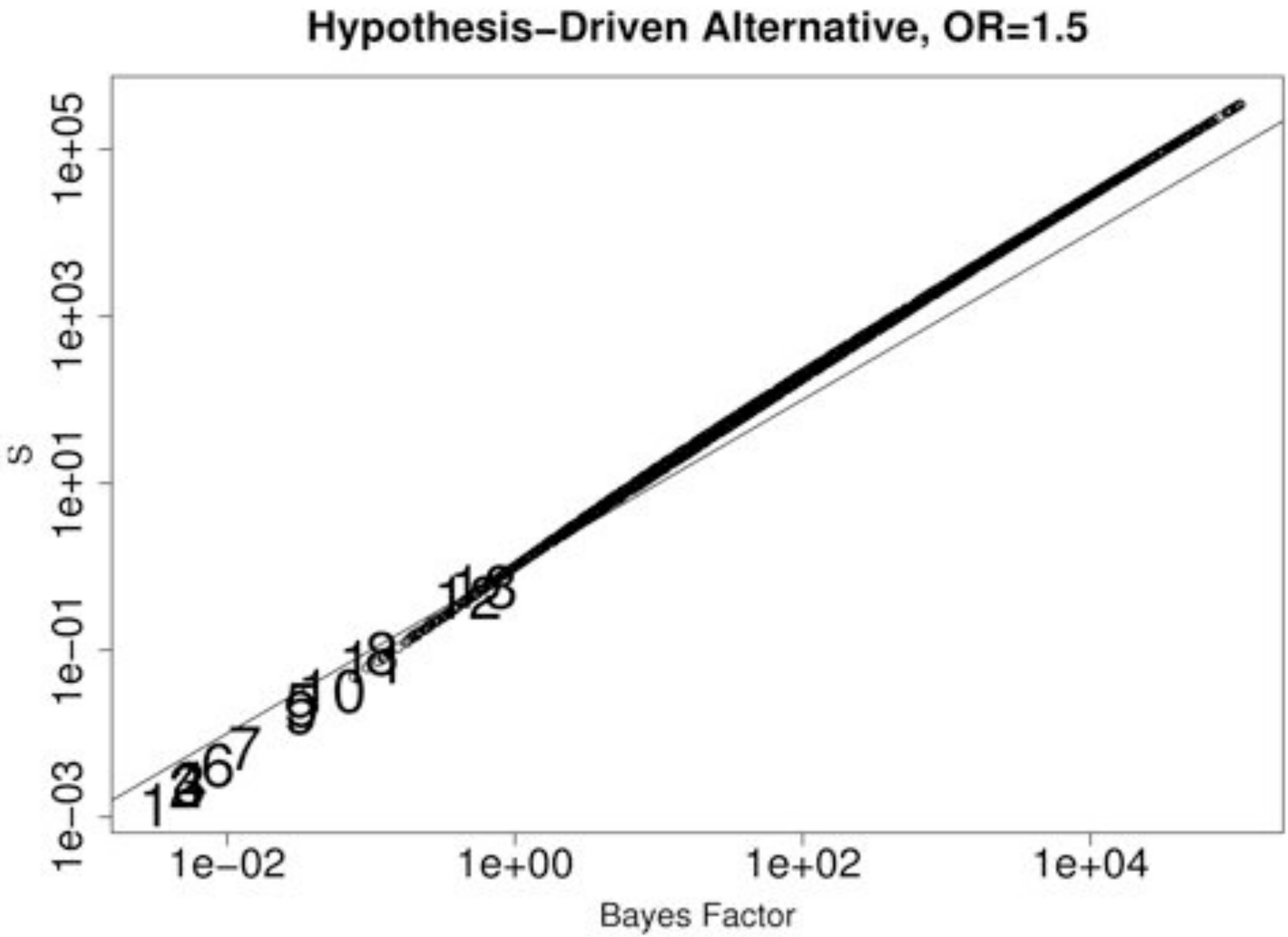
#	OR	stderr logOR	p-value	Bayes Factor
1	1.56	0.13	0.00066	0.0031
2	1.49	0.12	0.0012	0.0052
3	1.52	0.13	0.0012	0.0055
4	1.52	0.13	0.0013	0.0055
5	1.30	0.08	0.0018	0.0329
6	1.43	0.11	0.0019	0.0087
7	1.46	0.13	0.0032	0.0133
8	1.27	0.08	0.0045	0.11
9	1.44	0.14	0.0085	0.0325
10	1.37	0.12	0.011	0.0534
11	1.30	0.10	0.012	
12	1.24	0.09	0.012	
13	1.24	0.08	0.012	

CGEMS Alternative Hypotheses

- Fix variance of null to be that observed for that SNP
- Two hypothesis-driven alternatives
  - Minimally-important effect sizes we want to test against
  - Set β_1 =log(OR) = log(1.5)
  - Set β_1 =log(OR) = log(1.3)
- Two hypothesis-generating alternatives
  - Presume 95% of findings should be in (1/1.5,1.5): W=0.04
  - Presume 95% of findings should be in (1/1.1,1.1): W=0.002

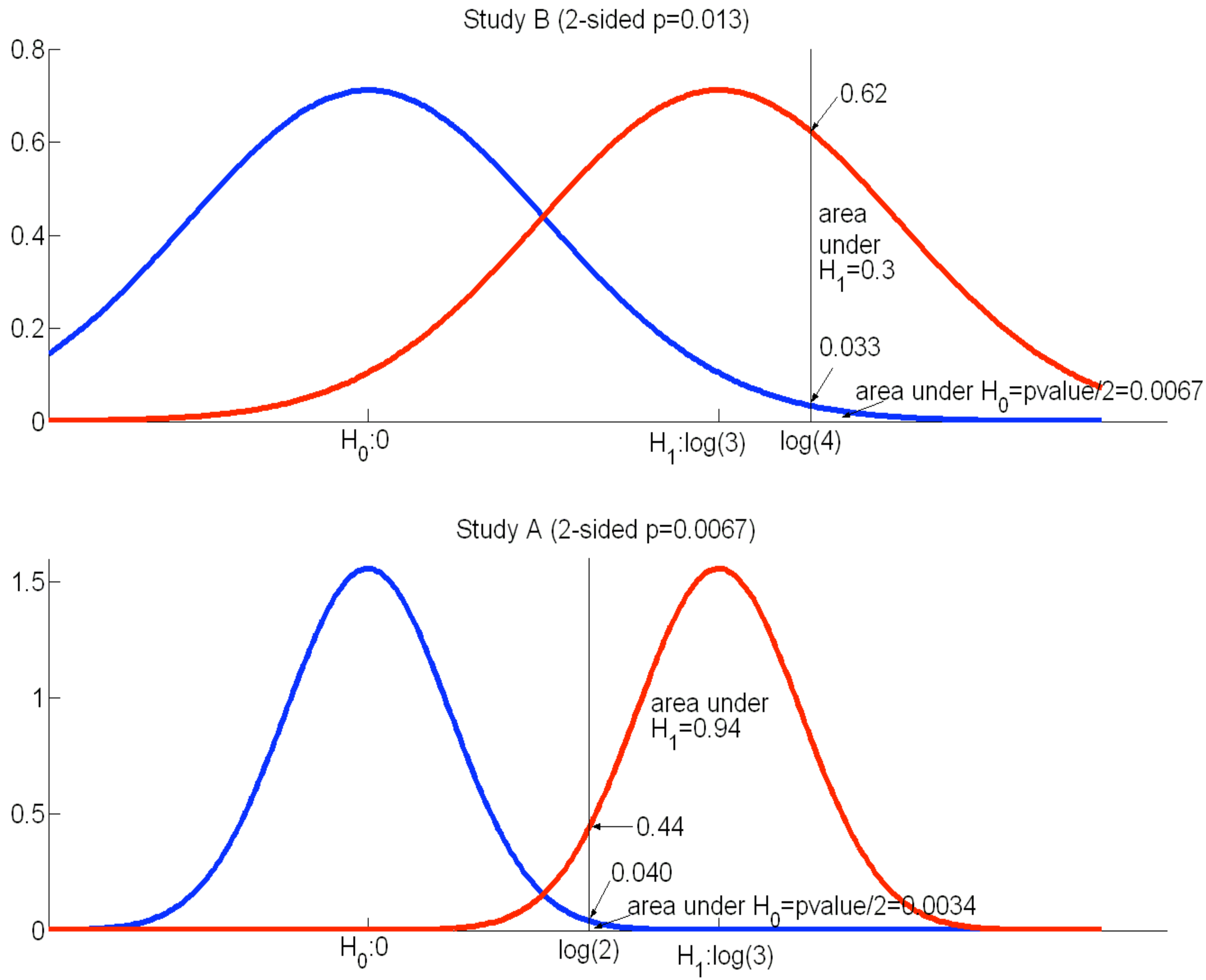


Hypothesis–Driven Alternative, OR=1.5
False Negative Bayes Factor
1e+00
1e−02
1e−04
1e−06
0.001
0.005
0.020
0.050
0.200
0.500
False Positive Bayes Factor
1
2
3
4
5
6
7
8
9
10
11
12
13


Hypothesis−Driven Alternative, OR=1.5
S
1e−03
1e−01
1e+01
1e+03
1e+05
1e−02
1e+00
1e+02
1e+04
Bayes Factor

Recap

- False Positive/Negative Bayes Factors
  - Are approximately the two components of the Bayes Factor
  - Explain disagreements between p-values and Bayes Factors
  - Complement p-values and Bayes Factors
- FPBF and FNBF require dichotomization
  - Scientists often dichotomize study results anyway
    - “If I declare my findings positive, what is the chance that I'm wrong?”
    - “If I declare my findings negative, what is the chance that I'm wrong?”
  - Relationship to Bayes Factor alleviates concern
  - Need to consider FPBF jointly with FNBF
- P-values and Bayes Factors can't be reconciled
  - But the *evidential interpretation of the p-value via False Positive/Negative Bayes Factors* can be approximately reconciled with the Bayes Factor


Study B (2-sided p=0.013)
0.8
0.6
0.4
0.2
0
0.62
area
under
H1=0.3
0.033
area under H0=pvalue/2=0.0067
H0:0
H1:log(3)
log(4)
Study A (2-sided p=0.0067)
1.5
1
0.5
0
area under
H1=0.94
0.44
0.040
area under H0=pvalue/2=0.0034
H0:0
log(2)
H1:log(3)

Wilcox's Example: What Happened?

	OR	CI	p-value	FPBF	BF
Factor A	2.0	(1.2 - 3.3)	0.0067	1/278	1/11
Factor B	4.0	(1.3 - 12)	0.013	1/45	1/19

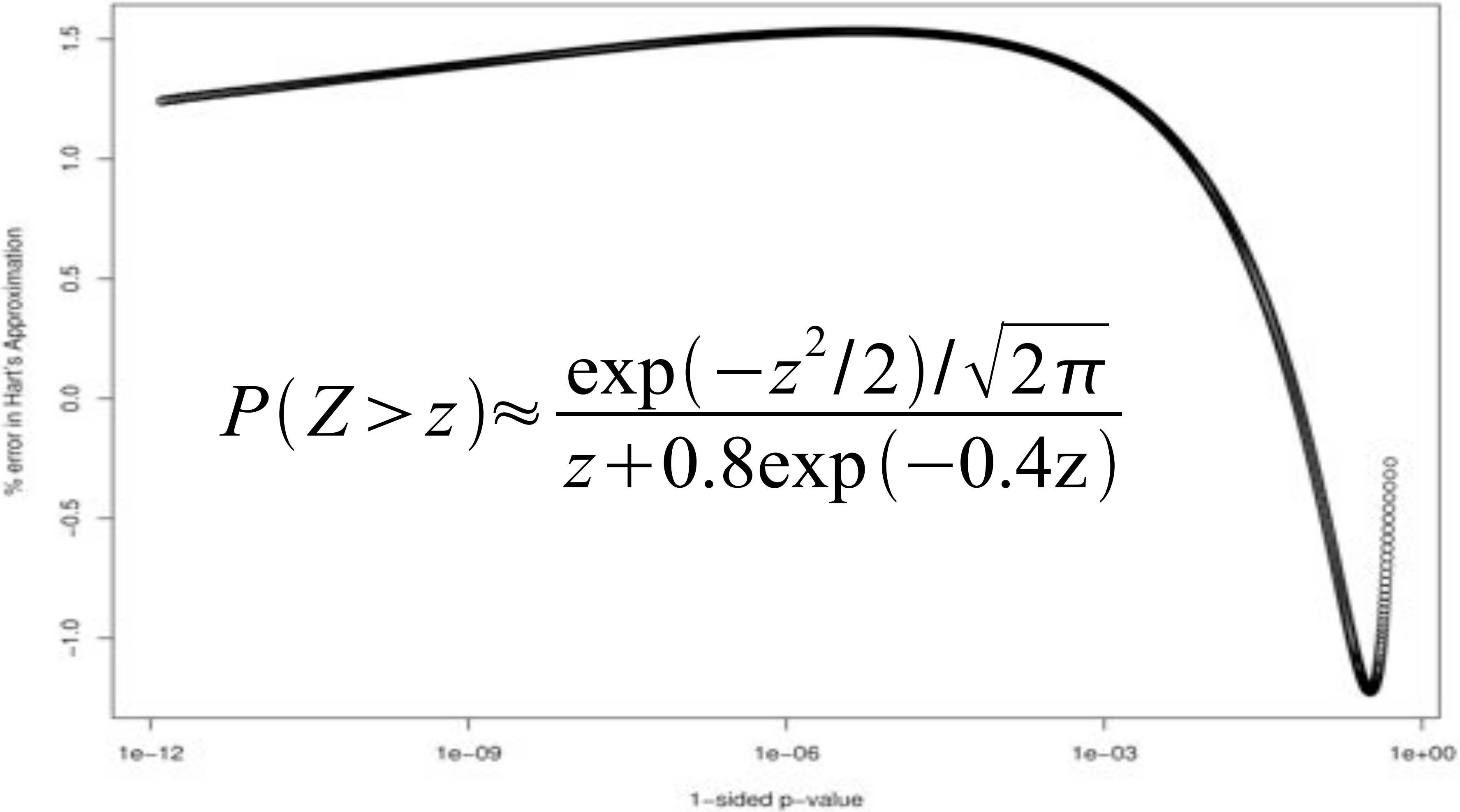
- FPBF: "You could've seen something huge!
- BF: "Both hypotheses are unlikely!"
- Both agree that alternative is favored
  - Disagree on how much more it's favored
- Famous unresolved paradox:
  - Royall, American Statistician (1986)

Katki, JSM Proceedings 2007

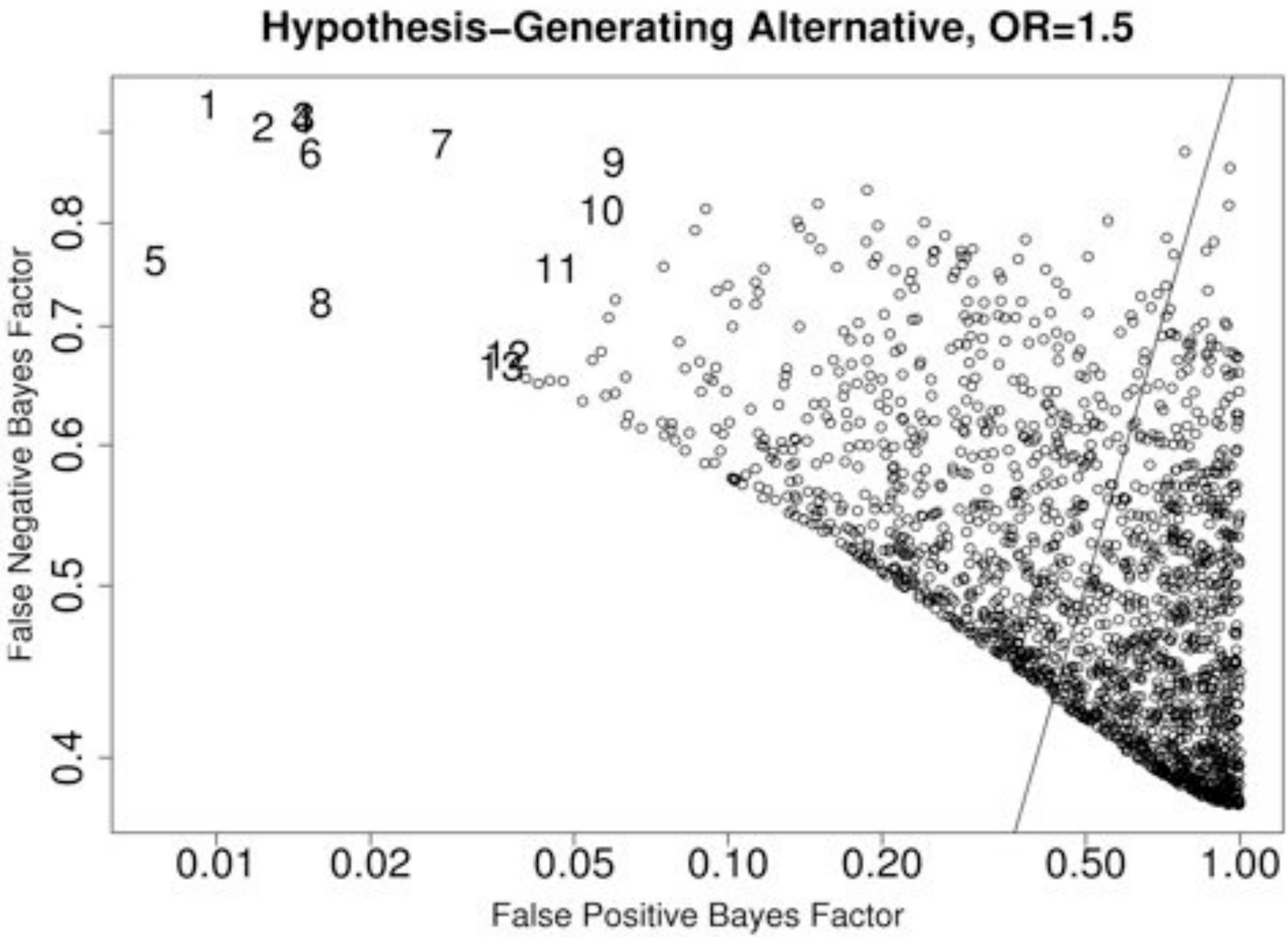
Future Work

- Choice of alternative hypothesis
  - For Wilcox example, if chose alternatives centered at log(1.5) or log(5), there are no problems
  - Critical choice of W in hypothesis-generating alternative
  - Empirical Bayes: Estimate null and alternative hypotheses from data (Efron 2004)
- Relationship to False Discovery Rates (FDR)
  - The FPRP estimates the same estimand as the FDR, only forcing you to propose a specific alternative
    - FPBF is akin to an evidential measure implicitly used by FDR
  - Similarly for FNRP and FNBF for the False Non-Discovery Rate

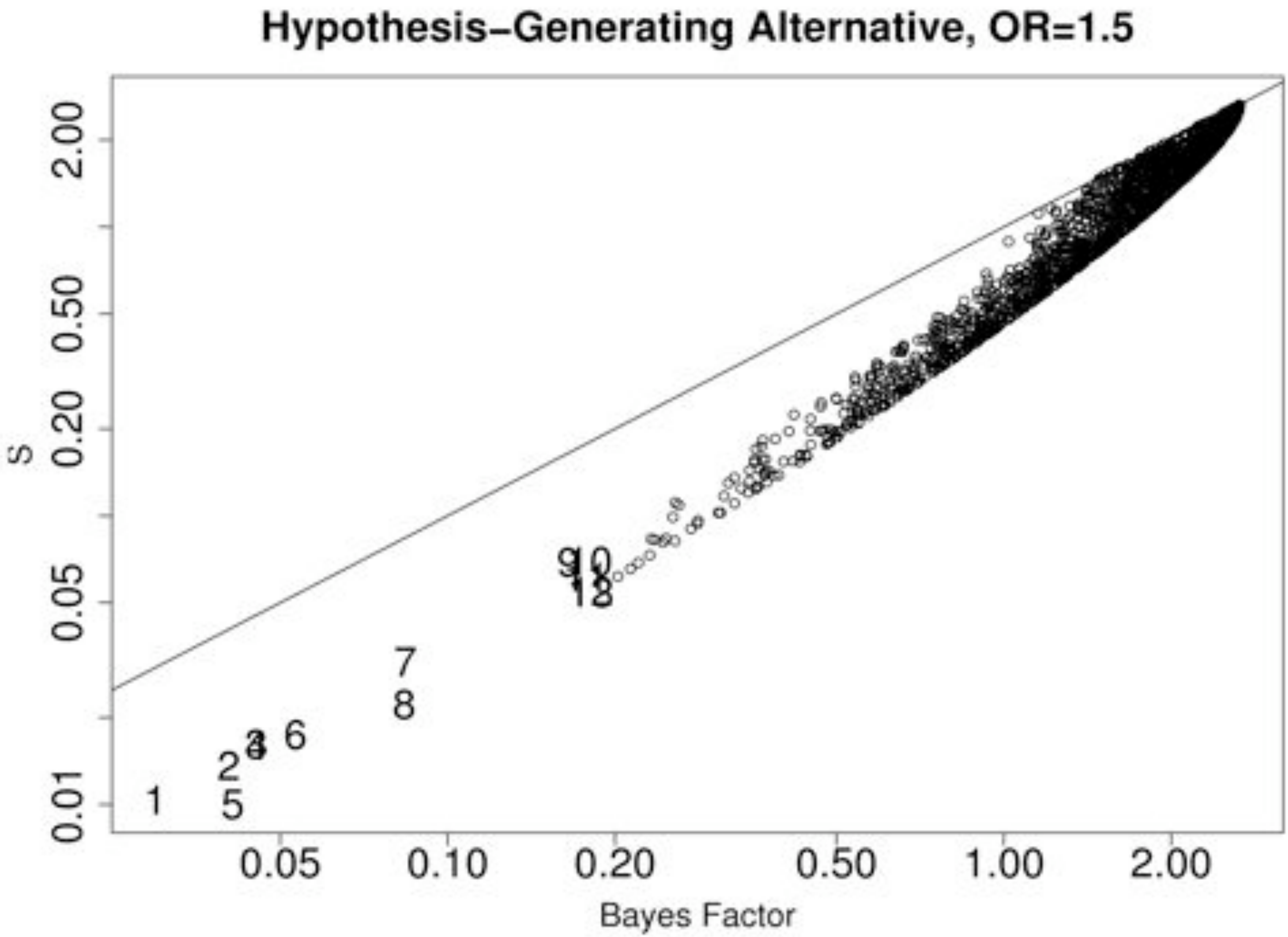
Reconcile BF with FPBF and FNBF: Hart's Approximation (1957)



$$P(Z > z) \approx \frac{\exp(-z^2/2)/\sqrt{2\pi}}{z + 0.8\exp(-0.4z)}$$



Hypothesis–Generating Alternative, OR=1.5
False Negative Bayes Factor
False Positive Bayes Factor
0.4
0.5
0.6
0.7
0.8
0.01
0.02
0.05
0.10
0.20
0.50
1.00
1
2
3
4
5
6
7
8
9
10
11
12
13



Hypothesis–Generating Alternative, OR=1.5
S
2.00
0.50
0.20
0.05
0.01
0.05
0.10
0.20
0.50
1.00
2.00
Bayes Factor
1
2
3
4
5
6
7
8
9
10
11
12
13

Reconcile BF with FPBF and FNBF

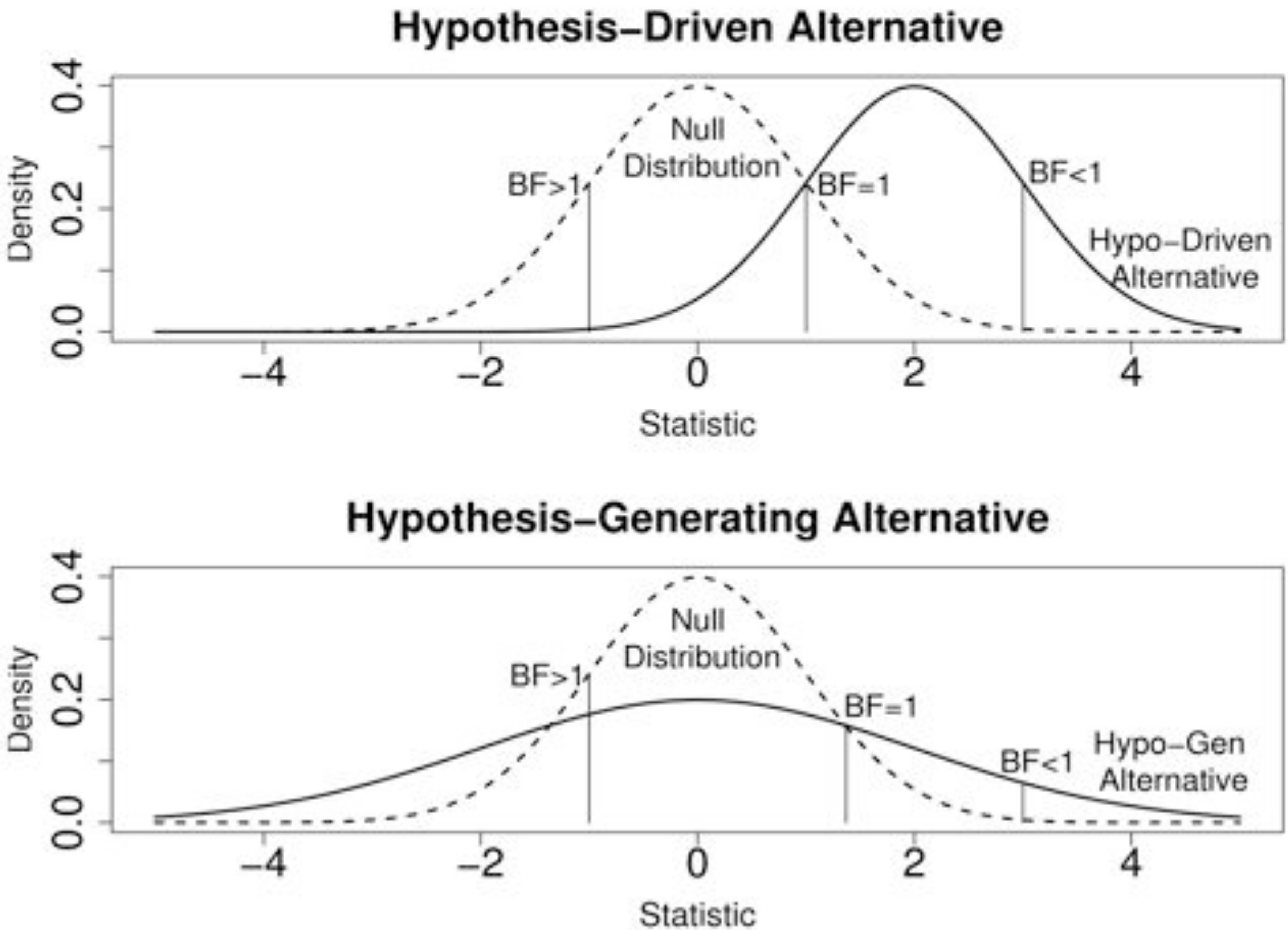
$$P(Z>z)\approx\frac{\exp(-z^2/2)/\sqrt{2\pi}}{z+0.8\exp(-0.4z)}$$

- S is approximately a ratio of tail probabilities
  - $p<p_1<0.5$: If both p,p_1 are small then $S\approx FPBF=p/p_1$ for both alternatives
  - $p<0.5$, $p_1>0.5$: If p is small, p_1 is large, then $S\approx p/(1-p_1)$ for hypothesis-driven alternative
  - $p>0.5$, $p_1<0.5$: Impossible since $p<p_1$ for both alternatives
  - $p_1>p>0.5$: If both p,p_1 are large then $S\approx 1/FPBF=(1-p)/(1-p_1)$ for both alternatives

Katki, Biostatistics, submitted

Two types of Alternative Hypotheses

- Hypothesis-Driven Alternative
  - $H_{1d}: \hat{\beta} \sim N(\beta_{1,} \sigma_1^2), \quad WLOG \quad \beta_1 > 0; \quad \sigma_0 = \sigma_1$
  - Specification of β_1 suggests the investigator has an *a priori* hypothesis or a minimally-interesting effect size
- Hypothesis-Generating Alternative
  - $H_{1g}: \hat{\beta} \sim N(0, \sigma_1^2 = \sigma_0^2 + W)$
  - Marginal distribution of: $\hat{\beta} \sim N(\beta_{1,} \sigma_0^2), \beta_1 \sim N(0, W)$
  - Often used in bioinformatic situations where there are no *a priori* hypotheses about which factors will have an effect, but the vast majority of factors are expected to have no effect



Hypothesis–Driven Alternative
Density
Null
Distribution
BF>1
BF=1
BF<1
Hypo–Driven
Alternative
Statistic
Hypothesis–Generating Alternative
Density
Null
Distribution
BF>1
BF=1
BF<1
Hypo–Gen
Alternative
Statistic

Two types of Alternative Hypotheses

- Hypothesis-Driven Alternative Bayes Factor

$$BF_d(\hat{\beta}) = \frac{\exp(-0.5((\hat{\beta}-0)^2/\sigma_0^2))}{\exp(-0.5((\hat{\beta}-\beta_1)^2/\sigma_1^2))}$$

$$= \exp(-0.5(2\hat{\beta}\beta_1 - \beta_1^2)/\sigma_0^2)$$

- Hypothesis-Generating Alternative Bayes Factor

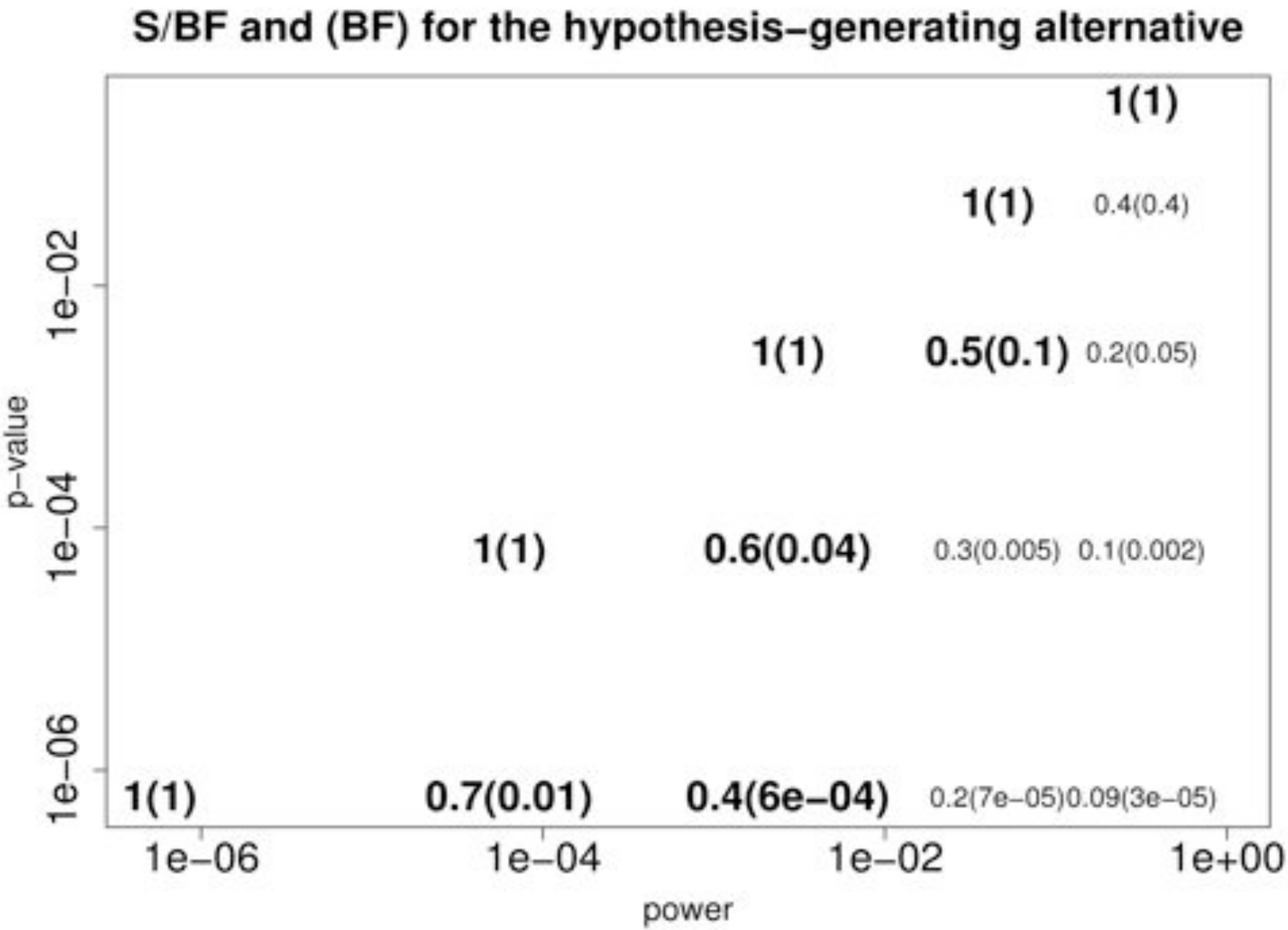
$$BF_d(\hat{\beta}) = \frac{\exp(-0.5((\hat{\beta}-0)^2/\sigma_0^2))}{\exp(-0.5((\hat{\beta}-0)^2/\sigma_1^2))}$$

$$= \sigma_0^{-1}\sqrt{\sigma_0^2 + W}\exp(-0.5(\hat{\beta}^2/\sigma_0^2 - \hat{\beta}^2/(\sigma_0^2 + W)))$$

Use of FPBF to interpret p-values

Richard Royall

- Consider betahat where densities cross (BF=1)
  - By BF, neither null nor alternative can be preferred
  - However, p-value can be very small: contradiction
  - Under either alternative, FPBF < 1: alternatives preferred
  - As p-value -> 0, FPBF << 1: alternatives strongly preferred
  - Easy to construct such contradictions: Royall (1986)
  - Basis for much argument that Bayes Factors and p-values draw different conclusions


S/BF and (BF) for the hypothesis−generating alternative
1(1)
1(1)
0.4(0.4)
1(1)
0.5(0.1)
0.2(0.05)
1(1)
0.6(0.04)
0.3(0.005)
0.1(0.002)
1(1)
0.7(0.01)
0.4(6e−04)
0.2(7e−05)
0.09(3e−05)
1e−02
1e−04
1e−06
p−value
1e−06
1e−04
1e−02
1e+00
power