

Uniformly and Restricted Most Powerful Bayesian Tests

Valen E. Johnson and Scott Goddard
Texas A&M University

June 6, 2014

Overview of Null Hypothesis Significance Testing (NHST)

- ▶ Central to scientific publication
- ▶ Tens of thousands of hypothesis tests are reported every year
- ▶ Prior beliefs of investigators, editors, other scientists, patients, general public, etc. vary regarding alternative hypotheses implicit to each hypothesis test
- ▶ Ditto for loss functions
- ▶ Scientific hypothesis tests are normally conducted under frequentist paradigm using standard threshold for significance (i.e, 0.05).

Towards routine Bayesian testing

For Bayesian testing procedures to find routine application in NHST, practitioners need

- ▶ Standard evidence thresholds for rejection of null hypothesis, *based on Bayes factors*
- ▶ Default alternatives hypotheses
- ▶ Simple expressions/programs to compute Bayes factors

Further requirements for routine Bayesian testing

For specified evidence threshold,

- ▶ Scientists want default alternatives that provide highest probability for rejection of null hypothesis

Further requirements for routine Bayesian testing

For specified evidence threshold,

- ▶ Scientists want default alternatives that provide highest probability for rejection of null hypothesis

Uniformly most powerful Bayesian tests provide highest probability of rejection in one-parameter exponential family models

Further requirements for routine Bayesian testing

For specified evidence threshold,

- ▶ Scientists want default alternatives that provide highest probability for rejection of null hypothesis

Uniformly most powerful Bayesian tests provide highest probability of rejection in one-parameter exponential family models

Restricted most powerful Bayesian tests provide high probability of rejection when UMPBTs do not exist, and they often have simple expressions

Notation and assumptions

- ▶ H_0, H_1 denote models/hypotheses
- ▶ $f(\mathbf{y} | \boldsymbol{\theta})$ denotes the sampling density under all models
- ▶ $m_i(\mathbf{y})$ denotes the marginal density of data under model i
- ▶ Θ denotes parameter space
- ▶ $\pi_i(\boldsymbol{\theta})$ denotes the prior density for $\boldsymbol{\theta} \in \Theta$ under model i
- ▶ $BF_{10}(\mathbf{y})$ denotes the Bayes factor between H_1 and H_0
- ▶ γ denotes evidence threshold, i.e., reject H_0 if $BF_{10} > \gamma$

Definition

A **uniformly most powerful Bayesian test** for a given evidence threshold γ , in favor of an alternative hypothesis H_1 against a fixed null hypothesis H_0 is a Bayesian hypothesis test in which the Bayes factor for the test satisfies the following inequality

$$\mathbf{P}_{\theta_t} [BF_{10}(\mathbf{y}) > \gamma] \geq \mathbf{P}_{\theta_t} [BF_{20}(\mathbf{y}) > \gamma] \quad (1)$$

for any $\theta_t \in \Theta$ and for all alternative hypotheses $H_2 : \theta \sim \pi_2(\theta)$:

Binomial data

- ▶ Suppose $y \sim \text{Binom}(n, \pi)$
- ▶ $H_0 : \pi = \pi_0$
- ▶ Suppose under alternative $H_1 : \pi = \pi_1 > \pi_0$
- ▶ Let π_t denote true (i.e., data-generating) value of π

Binomial data (cont)

- For a point alternative hypothesis, log of Bayes factor is

$$\log[BF_{10}(y)] = y \log \left[\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \right] + n \log \left[\frac{1 - \pi_1}{1 - \pi_0} \right]$$

Binomial data (cont)

- For a point alternative hypothesis, log of Bayes factor is

$$\log[BF_{10}(y)] = y \log \left[\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \right] + n \log \left[\frac{1 - \pi_1}{1 - \pi_0} \right]$$

- $\mathbf{P}_{\pi_t}[BF_{10}(y) > \gamma]$ can be expressed

$$\mathbf{P}_{\pi_t} \left[y > \frac{\log(\gamma) - n \log[(1 - \pi_1)/(1 - \pi_0)]}{\log[\pi_1(1 - \pi_0)]/[(1 - \pi_1)\pi_0]} \right]$$

Binomial data (cont)

- For a point alternative hypothesis, log of Bayes factor is

$$\log[BF_{10}(y)] = y \log \left[\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \right] + n \log \left[\frac{1 - \pi_1}{1 - \pi_0} \right]$$

- $\mathbf{P}_{\pi_t}[BF_{10}(y) > \gamma]$ can be expressed

$$\mathbf{P}_{\pi_t} \left[y > \frac{\log(\gamma) - n \log[(1 - \pi_1)/(1 - \pi_0)]}{\log[\pi_1(1 - \pi_0)]/[(1 - \pi_1)\pi_0]} \right]$$

Minimizing the RHS of inequality maximizes $\mathbf{P}[BF_{10}(y) > \gamma]$, regardless of the value of π_t .

One parameter exponential family models

- Suppose $\mathbf{x} = \{x_1, \dots, x_n\}$ are iid with joint density function

$$f(\mathbf{x}) = \exp \left[-\eta(\theta) \sum_{i=1}^n T(x_i) - nA(\theta) \right] \prod_{i=1}^n h(x_i),$$

where $\eta(\theta)$ is strictly monotonic

- Consider a one-sided test of a point null hypothesis that $H_0 : \theta = \theta_0$ against an arbitrary alternative hypothesis.

UMPBT(γ) for one parameter exponential family models

Theorem

Define

$$g_{\gamma}(\theta, \theta_0) = \frac{\log(\gamma) + n[A(\theta) - A(\theta_0)]}{\eta(\theta) - \eta(\theta_0)},$$

and define $u = \pm 1$ according to whether $\eta(\theta)$ is monotonically increasing or decreasing, and define $v = \pm 1$ according to whether the alternative hypothesis requires θ to be greater than or less than θ_0 , respectively.

Then a UMPBT(γ) can be obtained by restricting the support of $\pi_1(\theta)$ to values of θ that belong to the set

$$\arg \min_{\theta} uv g_{\gamma}(\theta, \theta_0).$$

Implications

- ▶ Like classical uniformly most powerful tests, UMPBTs exist for all common 1PEFs
- ▶ Unique UMPBTs are often defined by simple alternative hypotheses; exceptions occur when several values of parameter define the same rejection region
- ▶ Rejection regions for UMPBTs in exponential family models can generally be matched to rejection regions of UMPTs by appropriate choice of γ and Type I error

Implications

- ▶ Like classical uniformly most powerful tests, UMPBTs exist for all common 1PEFs
 - ▶ Unique UMPBTs are often defined by simple alternative hypotheses; exceptions occur when several values of parameter define the same rejection region
 - ▶ Rejection regions for UMPBTs in exponential family models can generally be matched to rejection regions of UMPTs by appropriate choice of γ and Type I error
- ⇒ This property establishes a connection between BFs and p-values

Asymptotic Properties

Theorem

Let θ^* denote a value of θ that defines a $UMPBT(\gamma)$ test. Then under certain regularity conditions, the following are true.

1. For some $t \in (\theta_0, \theta^*)$,

$$|\theta^* - \theta_0| = \sqrt{\frac{2 \log(\gamma)}{nA''(t)}}. \quad (2)$$

2. Under the null hypothesis,

$$\log(BF_{10}) \rightarrow N(-\log(\gamma), 2 \log(\gamma)) \quad \text{as } n \rightarrow \infty. \quad (3)$$

Asymptotics

- ▶ For fixed γ , as $n \rightarrow \infty$ the $\text{UMPBT}(\gamma)$ alternative converges to null hypothesis.
- ▶ In practice, very large samples are collected for hypothesis tests when either
 1. A very small effect size is being tested, or
 2. Very strong evidence against H_0 is required
- ▶ Asymptotic properties of UMPBTs seem consistent with actual statistical practice
- ▶ Evidence in favor of true null is probabilistically bounded by $\log(\gamma)$
- ▶ Rates at which to increase γ with n are topic for additional research

Examples

Binomial data

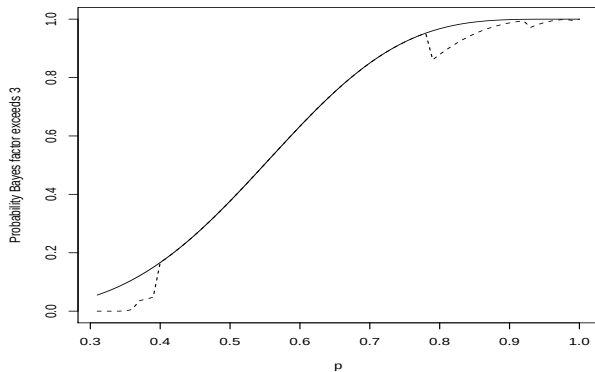
- ▶ Suppose $y \sim \text{Binom}(n, \pi)$
- ▶ $H_0 : \pi = 0.3, n = 10, \gamma = 3; H_1 : \pi > 0.3$

Binomial data

- ▶ Suppose $y \sim \text{Binom}(n, \pi)$
- ▶ $H_0 : \pi = 0.3$, $n = 10$, $\gamma = 3$; $H_1 : \pi > 0.3$
- ▶ UMPBT(γ) value of π_1 satisfies

$$\begin{aligned}\pi_1 &= \arg \min_{\pi} \frac{\log(\gamma) - n[\log(1 - \pi) - \log(1 - \pi_0)]}{\log[\pi/(1 - \pi)] - \log[\pi_0/(1 - \pi_0)]} \\ &= 0.525\end{aligned}$$

$P[BF_{10} > 3]$ vs data-generating parameter



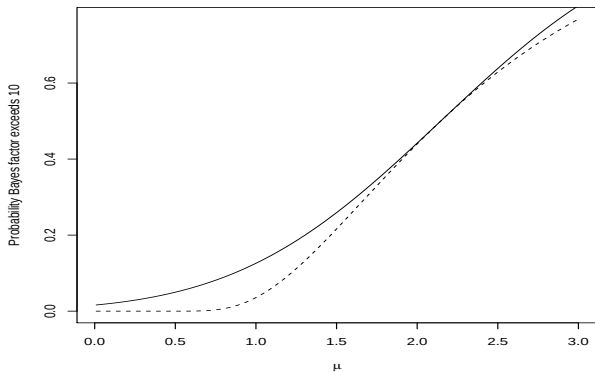
Normal data

- ▶ Suppose $x_1, \dots, x_n$ iid $N(\mu, \sigma^2)$, σ^2 known
- ▶ UMPBT(γ) test of $H_0 : \mu = \mu_0$ is given by

$$\mu_1 = \mu_0 \pm \sigma \sqrt{\frac{2 \log \gamma}{n}},$$

depending on whether $\mu_1 > \mu_0$ or $\mu_1 < \mu_0$.

$P[BF_{10} > 10]$ vs data-generating parameter for $\sigma^2, n = 1$



STATISTICS
TEXAS A&M UNIVERSITY

Comparison to classical UMPT of normal mean

- ▶ Classical one-sided test's rejection region is

$$\bar{x} \geq \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

- ▶ Equating the rejection regions for the UMPBT(γ) test and the UMPT of size α leads to

$$\gamma = \exp(z_{\alpha}^2/2)$$

- ▶ UMPBT places μ_1 on boundary of classical UMPT rejection region

Comparison to classical UMPT of normal mean

- ▶ Classical one-sided test's rejection region is

$$\bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

- ▶ Equating the rejection regions for the UMPBT(γ) test and the UMPT of size α leads to

$$\gamma = \exp(z_\alpha^2/2)$$

- ▶ UMPBT places μ_1 on boundary of classical UMPT rejection region
- ▶ UMPBT the most “subjective” of objective Bayesian hypothesis tests?

Other Exact UMPBTs

UMPBTs exist for

- ▶ Simple tests regarding coefficients in linear models with known observational variances
- ▶ Chi-squared tests on one degree of freedom when $H_0 : \lambda = 0$ and $H_1 : \lambda > 0$, λ the non-centrality parameter
- ▶ Two-sided tests in 1PEF, under constraint of symmetric alternative.

Approximate UMPBTs

- ▶ Approximate UMPBTs can be obtained in normal model hypothesis tests with unknown variances (require data dependent alternative hypotheses).
 1. T-tests (one-sample, paired, two-sample)
 2. Simple tests of linear regression coefficients with unknown observational variance

T-tests

- ▶ For one-sample t-test, $P(BF_{10} > \gamma)$ can be expressed as

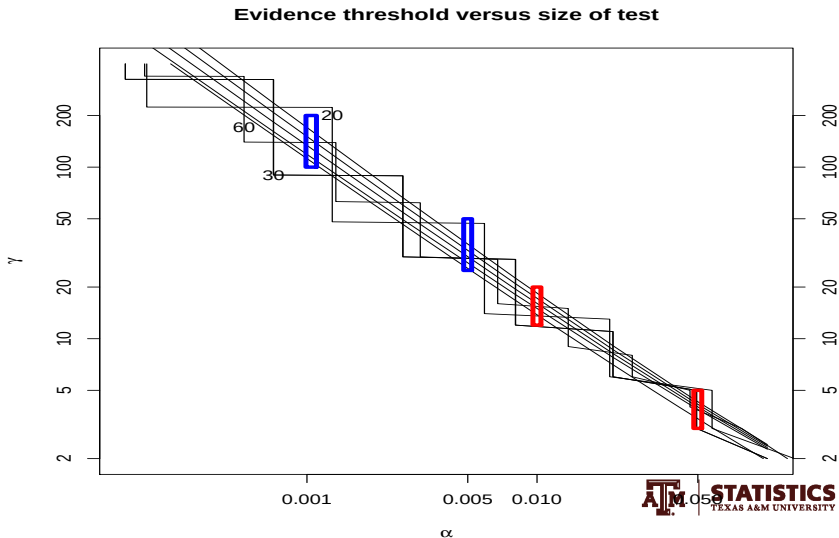
$$P_{\mu_t} [a < \bar{x} < b]$$

- ▶ For two-sample t-test, $P(BF_{10} > \gamma)$ can be expressed as

$$P_{\mu_1 - \mu_0} [c < \bar{x}_1 - \bar{x}_2 < d]$$

- ▶ The parameters (a, b, c, d) depend on n , γ , and s^2 .
- ▶ Upper bounds $b, d \rightarrow \infty$ with n
- ▶ Ignoring upper bound, data-dependent (s^2) approximate UMPBT can be obtained by minimizing a or c .

Bayes evidence thresholds versus test size



Bayes evidence thresholds versus test size

Under assumption of equipoise (i.e., $P(H_0) = P(H_1)$),

- ▶ $p = 0.05 \Rightarrow \gamma \in (3, 5) \Rightarrow P(H_0 | \mathbf{x}) \in (.17, .25)$
- ▶ $p = 0.01 \Rightarrow \gamma \in (12, 20) \Rightarrow P(H_0 | \mathbf{x}) \in (.05, .08)$

- ▶ $p = 0.005 \Rightarrow \gamma \in (25, 50) \Rightarrow P(H_0 | \mathbf{x}) \in (.02, .04)$
- ▶ $p = 0.001 \Rightarrow \gamma \in (100, 200) \Rightarrow P(H_0 | \mathbf{x}) \in (.005, .001)$

Bayes evidence thresholds versus test size

- ▶ Standard definitions of "significant" and "highly significant" results correspond to only weak evidence against null hypotheses.
- ▶ Definition of "significant" or "highly significant" should require evidence of $> 25 : 1$ or $> 100 : 1$ against the null $\Rightarrow$ p -values of 0.005 or 0.001

π -Restricted Most Powerful Bayesian Tests

A π -restricted most powerful Bayesian test for evidence threshold $\gamma > 0$ in favor of the alternative hypothesis $H_1 : \theta \sim \pi(\theta|\psi_1)$ against a fixed null hypothesis H_0 is a Bayesian hypothesis test in which the Bayes factor for the test satisfies

$$\mathbf{P}_{\theta_t}[BF_{10}(\mathbf{x}) > \gamma] \geq \mathbf{P}_{\theta_t}[BF_{20}(\mathbf{x}) > \gamma],$$

for any $\theta_t \in \Theta$ and for all alternative hypotheses $H_2 : \theta \sim \pi(\theta|\psi_2)$, where π is a density function parameterized by ψ , and $\psi_1, \psi_2 \in \Psi$.

g-prior-Restricted Most Powerful Tests

- Suppose

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}), \quad H_0 : \boldsymbol{\beta} = 0, \quad H_1 : \boldsymbol{\beta} \sim N[0, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$$

- With non-informative prior on σ^2 , value of g that maximizes $\mathbf{P}[BF_{10} > \gamma]$ is

$$\underset{g}{\operatorname{argmin}} \frac{(g+1)}{g} \left[1 - (g+1)^{-p/n} \gamma^{-2/n} \right]$$

g-prior-Restricted Most Powerful Tests

Restricted most powerful Bayesian tests have applications in

- ▶ ANOVA, where they provide correspondence to F tests
- ▶ One and two sample t-tests
- ▶ Bayesian variable selection, where γ can be set according to p and n
- ▶ To get corresponding size α test, set

$$g = F - 1,$$

where F is $1 - \alpha$ quantile from corresponding F distribution.

Bayes factors of RMPBT's in linear models

► ANOVA:

$$BF_{10}(\hat{F}) = F^{(n-J)/2} \left[\frac{F + \frac{J-1}{n-J} \hat{F}}{1 + \frac{J-1}{n-J} \hat{F}} \right]^{-(n-1)/2}$$

► Two-sample t :

$$BF_{10}(\hat{t}) = t^{n-2} \left[\frac{t^2 + \frac{1}{n-2} \hat{t}^2}{1 + \frac{1}{n-2} \hat{t}^2} \right]^{-(n-1)/2}$$

► One-sample t :

$$BF_{10}(\hat{t}) = t^{n-1} \left[\frac{t^2 + \frac{1}{n-1} \hat{t}^2}{1 + \frac{1}{n-1} \hat{t}^2} \right]^{-n/2}$$



STATISTICS
TEXAS A&M UNIVERSITY

Summary

- ▶ UMPBTs provide default objective Bayes factors for the most common of statistical hypothesis tests
- ▶ Large sample behavior is reasonable
- ▶ Approximately mimic the subjective alternative hypothesis implicit to classical tests (for matched γ and Type I error)
- ▶ Correspondence between UMPBTs and UMPTs provide guidance on appropriate definition of significant and highly significant findings, and insight into the non-reproducibility of scientific studies
- ▶ Restricted most powerful Bayesian tests can provide default settings for hyperparameters for parametric alternative hypotheses

The End

Emanuel Parzen
1819 Laura Lane
College Station, TX 77840