

*University of Texas, MD Anderson Cancer
Center*

UT MD Anderson Cancer Center Department of Biostatistics
Working Paper Series

Year 2012

Paper 74

Uniformly Most Powerful Bayesian Tests

Valen E. Johnson*

*Dept. of Biostatistics, UT M. D. Anderson Cancer Center, vejohanson@mdanderson.org

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/mdandersonbiostat/paper74>

Copyright ©2012 by the author.

Uniformly Most Powerful Bayesian Tests

Valen E. Johnson

Abstract

Uniformly most powerful tests are statistical hypothesis tests that provide the greatest power against a fixed null hypothesis among all possible tests of a given size. In this article, I extend the notion of uniformly most powerful tests by defining a uniformly most powerful Bayesian test to be a test which maximizes the probability that the Bayes factor in favor of the alternative hypothesis exceeds a given threshold. Like their classical counterpart, uniformly most powerful Bayesian tests are most easily defined in one-parameter exponential family models, although I demonstrate that extensions outside of this class are possible. I also show how the connection between uniformly most powerful tests and uniformly most powerful Bayesian tests can be used to provide an approximate calibration between p -values and Bayes factors. Several examples of these new Bayesian tests are provided.

Uniformly Most Powerful Bayesian Tests

Valen E. Johnson

July 20, 2012

Abstract

Uniformly most powerful tests are statistical hypothesis tests that provide the greatest power against a fixed null hypothesis among all possible tests of a given size. In this article, I extend the notion of uniformly most powerful tests by defining a uniformly most powerful Bayesian test to be a test which maximizes the probability that the Bayes factor in favor of the alternative hypothesis exceeds a given threshold. Like their classical counterpart, uniformly most powerful Bayesian tests are most easily defined in one-parameter exponential family models, although I demonstrate that extensions outside of this class are possible. I also show how the connection between uniformly most powerful tests and uniformly most powerful Bayesian tests can be used to provide an approximate calibration between p -values and Bayes factors. Several examples of these new Bayesian tests are provided.

Keywords: Bayes factor, objective Bayes, one-parameter exponential family model, Neyman-Pearson lemma, non-local prior density, uniformly most powerful test, Higgs boson.

1 Introduction

Uniformly most powerful tests (UMPTs) were proposed by Neyman and Pearson in a series of articles published nearly a century ago (e.g., 1928, 1933; see Lehman and Romano 2005 for a comprehensive review of the subsequent literature). They are defined as statistical hypothesis tests which provide the greatest power among all tests of a given size. In this article, I extend the classical notion of UMPTs to the Bayesian setting by defining a uniformly most powerful Bayesian test (UMPBT) as a test that maximizes the probability that the Bayes factor against a fixed null hypothesis exceeds a specified threshold. The class of problems for which UMPBTs exist appears to be comparable in size to the class of problems for which UMPTs exist. Indeed, the overlap between the classes is substantial, and contains primarily one parameter exponential family models. In many settings, the similarity of the two types of tests provides interesting insights into the relation between significance levels and posterior model probabilities.

In contrast to classical statistical hypothesis tests, Bayesian hypothesis tests are based on comparisons of the posterior probabilities assigned to competing hypotheses. In parametric tests, competing hypotheses are characterized by the prior densities which they impose on the parameters that define a sampling density shared by both hypotheses. Such tests comprise the focus of this article. Specifically, I assume throughout that the posterior odds between two hypotheses H_1 and H_0 can be expressed as

$$\frac{\mathbf{P}(H_1 | \mathbf{x})}{\mathbf{P}(H_0 | \mathbf{x})} = \frac{m_1(\mathbf{x})}{m_0(\mathbf{x})} \times \frac{p(H_1)}{p(H_0)}, \quad (1)$$

where $BF_{10}(\mathbf{x}) = m_1(\mathbf{x})/m_0(\mathbf{x})$ is the Bayes factor between hypotheses H_1 and H_0 ,

$$m_i(\mathbf{x}) = \int_{\Theta} f(\mathbf{x} | \boldsymbol{\theta}) \pi_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (2)$$

is the marginal density of the data under hypothesis H_i , $f(\mathbf{x} | \boldsymbol{\theta})$ is the sampling density of data \mathbf{x} given $\boldsymbol{\theta}$, $\pi_i(\boldsymbol{\theta})$ is the prior density on $\boldsymbol{\theta}$ under H_i , and $p(H_i)$ is the prior probability assigned to hypothesis H_i , for $i = 0, 1$. I denote the parameter space by Θ and the sample space by \mathcal{X} . The logarithm of the Bayes factor is called the *weight of evidence*. All densities are assumed to be defined with respect to an appropriate underlying measure (e.g., Lebesgue or counting measure).

Within this restricted testing framework, the specification of a prior density on $\boldsymbol{\theta}$ is synonymous with the specification of a statistical hypothesis. To simplify terminology, I therefore refer to hypotheses and prior densities interchangeably in the sequel.

In order to obtain UMPBTs, I also impose two additional restrictions on the class of tests that are considered. First, I restrict attention to tests in which one hypothesis—the null hypothesis H_0 —can be specified unambiguously. Typically, this means that the null hypothesis is defined by either taking $\pi_0(\boldsymbol{\theta})$ to be a point mass concentrated on a parameter value of interest, say $\boldsymbol{\theta}_0$, or (less frequently) as an evidence-based prior density derived from historical data. For example, in the context of single-arm Phase II trials, $\boldsymbol{\theta}_0$ might represent a known response rate for patients treated with a standard therapy, or $\pi_0(\boldsymbol{\theta})$ might represent a beta density centered on the proportion of patients who responded to standard therapy in a previous trial.

The second condition is less of a restriction than it is a philosophical constraint. It requires that the competing, alternative hypothesis is not defined in an unambiguous manner based on either subjective or scientific considerations. If a clear alternative hypothesis exists, then the methods described below should not be employed. Instead, equation (1) should be used to obtain the posterior odds between null and alternative hypotheses directly. The alternative hypotheses considered in this article can be interpreted as the negation of the null hypothesis.

Before describing UMPBTs, it is worthwhile to first review a property of Bayes factors that pertains when the prior density defining an alternative hypothesis is misspecified.

Letting $\pi_t(\boldsymbol{\theta})$ denote the true prior density on $\boldsymbol{\theta}$ and $m_t(\mathbf{x})$ the resulting marginal density of the data, it follows from Gibb's inequality that

$$\begin{aligned} \int_{\mathcal{X}} m_t(\mathbf{x}) \log \left[\frac{m_t(\mathbf{x})}{m_0(\mathbf{x})} \right] d\mathbf{x} &- \int_{\mathcal{X}} m_t(\mathbf{x}) \log \left[\frac{m_1(\mathbf{x})}{m_0(\mathbf{x})} \right] d\mathbf{x} \\ &= \int_{\mathcal{X}} m_t(\mathbf{x}) \log \left[\frac{m_t(\mathbf{x})}{m_1(\mathbf{x})} \right] d\mathbf{x} \\ &\geq 0. \end{aligned}$$

That is,

$$\int_{\mathcal{X}} m_t(\mathbf{x}) \log \left[\frac{m_t(\mathbf{x})}{m_0(\mathbf{x})} \right] d\mathbf{x} \geq \int_{\mathcal{X}} m_t(\mathbf{x}) \log \left[\frac{m_1(\mathbf{x})}{m_0(\mathbf{x})} \right] d\mathbf{x}, \quad (3)$$

which means that the expected weight of evidence in favor of the alternative hypothesis is always decreased when $\pi_1(\boldsymbol{\theta})$ differs from $\pi_t(\boldsymbol{\theta})$ (on a set with measure greater than 0). In general, the UMPBTs described below will thus decrease the average weight of evidence obtained in favor of a true alternative hypothesis. Although they maximize the probability that the Bayes factor will exceed a given threshold, they will also tend to underestimate the posterior probability of the "true" alternative hypothesis. *That is, the weight of evidence reported from a UMPBT will tend to underestimate the actual weight of evidence provided by an experiment in favor of a true alternative hypothesis.*

Like classical statistical hypothesis tests, the tangible consequence of a Bayesian hypothesis test is often the rejection of one hypothesis, say H_0 , in favor of the second, say H_1 , if the posterior probability of H_1 exceeds a certain threshold. Given the prior odds between the hypotheses, this is equivalent to determining a threshold, say γ , over which the Bayes factor between H_1 and H_0 must fall in order to reject H_0 in favor of H_1 . It is therefore of some practical interest to determine alternative hypotheses that maximize the probability that the Bayes factor from the test exceeds a given threshold.

With this motivation and notation in place, a UMPBT(γ) may be formally defined as follows.

Definition. A uniformly most powerful Bayesian test for evidence $\gamma > 0$ in favor of the alternative hypothesis H_1 against a fixed null hypothesis H_0 , denoted by UMPBT(γ), is a Bayesian hypothesis test in which the Bayes factor for the test satisfies the following inequality for any $\boldsymbol{\theta}_t \in \Theta$ and for all alternative hypotheses $H_2 : \boldsymbol{\theta} \sim \pi_2(\boldsymbol{\theta})$:

$$\mathbf{P}_{\boldsymbol{\theta}_t} [BF_{10}(\mathbf{x}) > \gamma] \geq \mathbf{P}_{\boldsymbol{\theta}_t} [BF_{20}(\mathbf{x}) > \gamma]. \quad (4)$$

The UMPBT(γ) is thus a Bayesian test for which the alternative hypothesis is specified so as to maximize the probability that the Bayes factor $BF_{10}(\mathbf{x})$ exceeds the evidence threshold γ for all possible values of the data generating parameter $\boldsymbol{\theta}_t$.

The remainder of this article is organized as follows. In the next section, I describe UMPBTs for one parameter exponential family models. As in the case of UMPTs, this appears to be the broadest general class for which UMPBTs exist. In applying UMPBTs to one parameter exponential family models, I expose an approximate equivalence between type I errors for UMPTs and the Bayes factors obtained from UMPBTs.

In Section 3, I apply UMPBTs in two canonical testing situations: the test of a binomial proportion, and the test of a normal mean. These two tests are perhaps the most commonly employed tests by practitioners of statistics. The binomial test is illustrated in the context of a clinical trial, while the normal mean test is applied to evaluate evidence reported in support of the Higgs boson. A proposal for replacing significance tests with UMPBTs in standard problems is described in Section 4. In Section 5, I describe several settings outside of one parameter exponential family models for which UMPBTs exist. These include cases in which the nuisance parameters under the null and alternative hypothesis can be considered to be equal (though unknown), and situations in which it is possible to marginalize over the nuisance parameters to obtain expressions for data densities that are similar to those obtained in one parameter exponential family models. Section 6 describes approximations to UMPBTs obtained by specifying alternative hypotheses that depend on data through statistics that are ancillary to the parameter of interest. Concluding comments appear in Section 7.

2 One parameter exponential family models

Assume that $\{x_1, \dots, x_n\} \equiv \mathbf{x}$ are i.i.d. with a sampling density (or probability mass function in the case of discrete data) of the form

$$f(x|\theta) = h(x) \exp[\eta(\theta)T(x) - A(\theta)], \quad (5)$$

where $T(x)$, $h(x)$, $\eta(\theta)$ and $A(\theta)$ are known functions, and $\eta(\theta)$ is monotonic. Consider a one-sided test of a point null hypothesis $H_0 : \theta = \theta_0$ against an arbitrary alternative hypothesis. The value of θ_0 is assumed to be fixed based on substantive considerations.

Lemma 1 *Assume the conditions of the previous paragraph pertain, and define $g(\theta, \theta_0)$ according to*

$$g(\theta, \theta_0) = \frac{\log(\gamma) + n[A(\theta) - A(\theta_0)]}{\eta(\theta) - \eta(\theta_0)}. \quad (6)$$

In addition, define u to be 1 or -1 according to whether $\eta(\theta)$ is monotonically increasing or decreasing, respectively, and define v to be either 1 or -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

$$\arg \min_{\theta} uv g(\theta, \theta_0). \quad (7)$$

Proof: Consider the case in which the alternative hypothesis requires θ to be greater than θ_0 and $\eta(\theta)$ is increasing (so that $uv = 1$), and let θ_t denote the true (i.e., data-generating) parameter for \mathbf{x} under (5). Consider first simple alternatives for which the prior on θ is a point mass at θ_1 . Then

$$\begin{aligned} \mathbf{P}_{\theta_t}(BF_{10} > \gamma) &= \mathbf{P}_{\theta_t}[\log(BF_{10}) > \log(\gamma)] \\ &= \mathbf{P}_{\theta_t}\left\{\sum_{i=1}^n T(x_i) > \frac{\log(\gamma) + n[A(\theta_1) - A(\theta_0)]}{\eta(\theta_1) - \eta(\theta_0)}\right\}. \end{aligned} \quad (8)$$

It follows that the probability in (8) achieves its maximum value when the right-hand side of the inequality is minimized, regardless of the distribution of $\sum T(x_i)$.

Now consider arbitrary alternative hypotheses, and define an indicator function s according to

$$s(\mathbf{x}, \theta) = \text{Ind}\left(\exp\left\{[\eta(\theta) - \eta(\theta_0)]\sum_{i=1}^n T(x_i) - n[A(\theta) - A(\theta_0)]\right\} > \gamma\right). \quad (9)$$

Let θ^* be a value that minimizes $g(\theta, \theta_0)$. Then it follows from (8) that

$$s(\mathbf{x}, \theta) \leq s(\mathbf{x}, \theta^*) \quad \text{for all } \mathbf{x}. \quad (10)$$

This implies that

$$\int_{\Theta} s(\mathbf{x}, \theta)\pi(\theta) \leq s(\mathbf{x}, \theta^*) \quad (11)$$

for all probability densities $\pi(\theta)$. It follows that

$$\mathbf{P}_{\theta_t}(BF_{10} > \gamma) = \int_{\mathcal{X}} s(\mathbf{x}, \theta)f(\mathbf{x}|\theta_t)d\theta_t \quad (12)$$

is maximized by a prior that concentrates its mass on the set for which $g(\theta, \theta_0)$ is minimized. Proofs of other cases follow by noting that the direction of the inequality in (8) changes according to the sign of $\eta(\theta_1) - \eta(\theta_0)$.

It should be noted that in some cases the values of θ that maximize $\mathbf{P}_{\theta_t}(BF_{10} > \gamma)$ are not unique. This might happen if, for instance, no value of the sufficient statistic obtained from the experiment could produce a Bayes factor that exceeded the γ threshold. For example, it would not be possible to obtain a Bayes factor of 10 against a null hypothesis that a binomial success probability was 0.5 based on a sample size of $n = 1$. In that case, the probability of exceeding the threshold is 0 for all values of the success probability, and a unique UMPBT does not exist.

3 Examples

Although not an active research topic among research statisticians, tests of simple hypotheses in one parameter exponential family models continue to be the most common statistical hypothesis tests used by practitioners. These tests play a central role in many science, technology and business applications. This section illustrates the use of UMPBT tests in two archetypical examples.

3.1 Test of binomial success probability

Suppose $x \sim \text{Bin}(n, p)$, and consider the test of $H_0 : p = p_0$ versus an alternative in which it is assumed that $p > p_0$. Assume further that an evidence threshold of γ is desired for the test; that is, the alternative hypothesis is accepted if $BF_{10} > \gamma$.

From Lemma 1, the UMPBT(γ) is defined by finding p_1 that satisfies $p_1 > p_0$ and

$$p_1 = \arg \min_p \frac{\log(\gamma) - n[\log(1-p) - \log(1-p_0)]}{\log[p/(1-p)] - \log[p_0/(1-p_0)]}. \quad (13)$$

Although this equation cannot be solved in closed form, its solution can be found easily using optimization functions available in most statistical programs.

3.1.1 Phase II clinical trials with binary outcomes

To illustrate the application of the resulting test to a real-world application, consider a one-arm Phase II trial of a new drug intended to improve the response rate to a disease from the standard-of-care rate of $p_0 = 0.3$. Suppose also that budget and time constraints limit the number of patients that can be accrued on the trial to $n = 10$, and suppose that the new drug will be pursued only if the odds that it offers an improvement over the standard of care are at least 3 : 1, based on data collected during the trial. Taking $\gamma = 3$, it follows from (13) that the UMPBT alternative is defined by taking $H_1 : p_1 = 0.525$. At this value of p_1 , the Bayes factor BF_{10} in favor of H_1 exceeds 3 whenever 6 or more of the 10 patients enrolled on the trial respond to treatment.

A plot of the probability that BF_{10} exceeds 3 as function of the true response rate p appears in Figure 1. For comparison, also plotted in this figure (dashed curve) is the probability that BF_{10} exceeds 3 when p_1 is set to the data-generating parameter; that is, when $p_1 = p$.

Figure 1 shows that the probability that BF_{10} exceeds 3 when calculated under the true alternative hypothesis is significantly smaller than it is under the UMPBT alternative for values of $p < 0.4$ and for values of $p > 0.78$. Indeed, for values of $p < .034$, there is no chance that BF_{10} will exceed 3. The decrease in the probability that the Bayes factor exceeds 3 for large values of p stems from the relatively small probability that these models assign to the observation of intermediate values of x . For example when $p = 0.8$, the

probability of observing 6 out of 10 successes is only 0.88, while the corresponding probability under H_0 is 0.37. Thus, $BF_{10} = 2.39$, and the evidence in favor of the true success probability does not exceed 3.

It is also interesting to note that the solid curve depicted in Figure 1 represents the power curve for an approximate 5% one-sided significance test of the null hypothesis that $p = 0.3$ (note that $P_{0.3}(X \geq 6) = 0.047$). This rejection region for the 5% significance test also corresponds to the region for which the Bayes factor corresponding to the UMPBT(γ) exceeds γ for all values of $\gamma \in (2.40, 6.83)$. If equal prior probabilities are assigned to H_0 and H_1 , this suggests that a p -value of 0.05 for this test corresponds roughly to the assignment of a posterior probability between $(1.0/7.83, 1.0/3.40) = (0.13, 0.29)$ to the null hypothesis. This range of values for the posterior probability of the null hypothesis is in approximate agreement with values suggested by other authors, for example Berger and Selke (1987).

This example also indicates that a UMPBT can result in large Type I errors if the threshold γ is chosen to be too small. For instance, taking $\gamma = 2$ in this example would lead to Type I errors that were larger than 0.05.

Of course, it is important to note that the UMPBT does not provide a test that maximizes the expected weight of evidence, as equation (3) demonstrates. This point is illustrated in Figure 2, which depicts the expected weight of evidence obtained in favor of H_1 by a solid curve as the data-generating success probability is varied in $(0.3, 1.0)$. For comparison, the dashed curve shows the expected weight of evidence obtained as a function of the true parameter value. As predicted by the inequality in (3), the UMPBT provides less evidence in favor of the true alternative hypothesis for all values of $p \in (0.3, 1.0)$ except $p = 0.525$, the UMPBT value.

3.2 Test of normal mean, σ^2 known

Suppose $x_i, i = 1, \dots, n$ are i.i.d. $N(\mu, \sigma^2)$ with σ^2 known. The null hypothesis is $H_0 : \mu = \mu_0$, and the alternative hypothesis is accepted if $BF_{10} > \gamma$. Assuming that the alternative hypothesis takes the form $H_1 : \mu = \mu_1$ in a one-sided test, it follows that

$$\log(BF_{10}) = \frac{n}{\sigma^2} \left[\bar{x}(\mu_1 - \mu_0) + \frac{1}{2}(\mu_0^2 - \mu_1^2) \right] \quad (14)$$

If the data-generating parameter is μ_t , the probability that BF_{10} is greater than γ can be written as

$$\mathbf{P}_{\mu_t} \left[(\mu_1 - \mu_0)\bar{x} > \frac{\sigma^2 \log(\gamma)}{n} - \frac{1}{2}(\mu_0^2 - \mu_1^2) \right] \quad (15)$$

If $\mu_1 > \mu_0$, then the UMPBT(γ) value of μ_1 satisfies

$$\arg \min_{\mu_1} \frac{\sigma^2 \log(\gamma)}{n(\mu_1 - \mu_0)} + \frac{1}{2}(\mu_0 + \mu_1) \quad (16)$$

Conversely, if $\mu_1 < \mu_0$, then optimal value of μ_1 satisfies

$$\arg \max_{\mu_1} \frac{\sigma^2 \log(\gamma)}{n(\mu_1 - \mu_0)} + \frac{1}{2}(\mu_0 + \mu_1) \quad (17)$$

It follows that the UMPBT(γ) value for μ_1 is given by

$$\mu_1 = \mu_0 \pm \sigma \sqrt{\frac{2 \log \gamma}{n}}, \quad (18)$$

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

Figure 3 depicts the probability that the Bayes factor exceeds $\gamma = 10$ when testing a null hypothesis that $\mu = 0$ based on a single, standard normal observation (i.e., $n = 1$, $\sigma^2 = 1$). In this case, the UMPBT(10) is obtained by taking $\mu_1 = 2.146$. For comparison, the probability that the Bayes factor exceeds 10 when the alternative is defined to be the data-generating parameter is depicted by the dashed curve in the plot.

UMPBTs can also be used to interpret the evidence obtained from classical UMPTs. In a classical one-sided test of a normal mean with known variance, the null hypothesis is rejected if

$$\bar{x} > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}},$$

where α is the level of the test designed to detect $\mu_1 > \mu_0$. In the UMPBT, from (14-15) it follows that the null hypothesis is rejected if

$$\bar{x} > \frac{\sigma^2 \log(\gamma)}{n(\mu_1 - \mu_0)} + \frac{1}{2}(\mu_1 + \mu_0).$$

Setting $\mu_1 = \mu_0 + \sigma \sqrt{2 \log(\gamma)/n}$ and equating the two rejection regions, it follows that the rejection regions for the two tests are identical if

$$\gamma = \exp\left(\frac{z_{1-\alpha}^2}{2}\right). \quad (19)$$

According to this correspondence, p -values of 0.05, 0.01, 0.001 values against H_0 translate into Bayes factors of 3.86, 15.0, and 118 in favor of H_1 , respectively.

3.2.1 Evaluating evidence for the Higgs boson

On July 4, 2012, scientists at CERN made the following announcement.

We observe in our data clear signs of a new particle, at the level of 5 sigma, in the mass region around 126 gigaelectronvolts (GeV).”
(<http://press.web.cern.ch/press/PressReleases/Releases2012/PR17.12E.html>”).

In very simplified terms, the 5 sigma claim resulted from fitting a model for a Poisson mean that had this approximate form:

$$\mu(x) = \exp(a_0 + a_1x + a_2x^2) + s\phi(x; m, w).$$

Here, x denotes mass in GeV, $\{a_i\}$ denote nuisance parameters that model background events, s denotes signal above background, m denotes the mass of a new particle, w denotes a convolution parameter, and $\phi(x; m, w)$ denotes a Gaussian density centered on m with standard deviation w (Prosper, 2012). Poisson events collected from a series of high energy experiments conducted in the Large Hadron Collider (LHC) at CERN provide the data to estimate the parameters in this somewhat stylized model. The background parameters $\{a_i\}$ are considered nuisance parameters. Interest, of course, focuses on testing whether $s > 0$ at a mass location m . The null hypothesis is that $s = 0$ for all m .

The accepted criterion for declaring the discovery of a new particle in the field of particle physics is the 5 sigma rule, which in this case requires that the estimate of s be 5 standard errors from 0 (<http://public.web.cern.ch/public/>).

Calculation of a Bayes factor based on the original mass spectrum data is complicated by the fact that prior distributions for the nuisance parameters $\{a_i\}$, m , and w are either not available or are not agreed upon. For this reason, it is more straightforward to compute a Bayes factor for these data based on the test statistic $z = \hat{s}/se(\hat{s})$ where \hat{s} denotes the maximum likelihood estimate of s and $se(\hat{s})$ its standard error (Johnson 2005, 2008). To perform this test, I assume that under the null hypothesis z has a standard normal distribution, and that under the alternative hypothesis z has a normal distribution with mean μ and variance 1.

The standard of evidence required by CERN for the declaration of a discovery requires that $z_{1-\alpha} = 5$ in (19). This implies that a Bayes factor of $\gamma = \exp(12.5) \approx 27,000$ is required to reject the null hypothesis of no new particle, which in turn implies that the UMPBT(γ) alternative is 5 (18). If the observed value of z was exactly 5, then the Bayes factor in favor of a new particle would be approximately 27,000. If the observed value was, say 5.1, then the Bayes factor would be $\exp(-0.5[0.1^2 - 5.1^2]) = 44,000$. These values suggest very strong evidence in favor of a new particle, but not as much evidence as might be inferred by non-statisticians by the report of a p -value of 3×10^{-7} .

There are, of course, a number of important caveats that should be considered when interpreting the outcome of this analysis. This analysis assumes that an experiment with a fixed endpoint was conducted, and that the UMPBT value of the Poisson rate at 126 GeV was of physical significance. Referring to (18) and noting that the asymptotic standard error of z decreases at rate \sqrt{n} , it follows that the UMPBT alternative hypothesis favored by this analysis is $O(n^{-1/2})$. For sufficiently large n , it is then clear that the systematic errors in the estimation of the background rate could eventually lead to the rejection of the null hypothesis in favor of the hypothesis of a new particle. This is of particular concern if the high energy experiments were continued until a 5 sigma result was obtained.

Model	Test	Objective Function
Binomial	$p_1 > p_0$	$\{\log(\gamma) - n \log[(1-p)/(1-p_0)]\} (\log\{[p(1-p_0)]/[(1-p)p_0]\})^{-1}$
Exponential	$\mu_1 > \mu_0$	$\{\log(\gamma) + n[\log(\mu_1) - \log(\mu_0)]\} [1/\mu_0 - 1/\mu_1]^{-1}$
Neg. Bin.	$p_1 > p_0$	$\{\log(\gamma) - r \log[(1-p_1)/(1-p_0)]\} [\log(p_1) - \log(p_0)]^{-1}$
Normal	$\sigma_1^2 > \sigma_0^2$	$\{2\sigma_1^2\sigma_0^2 (\log(\gamma) + \frac{n}{2}[\log(\sigma_1^2) - \log(\sigma_0^2)])\} [\sigma_1^2 - \sigma_0^2]^{-1}$
Normal	$\mu_1 > \mu_0$	$[\sigma^2 \log(\gamma)] (\mu_1 - \mu_0)^{-1} + \frac{1}{2}(\mu_0 + \mu_1)$
Poisson	$\mu_1 > \mu_0$	$[\log(\gamma) + n(\mu_1 - \mu_0)] [\log(\mu_1) - \log(\mu_0)]^{-1}$

Table 1: Common one parameter exponential family models for which UMPBT(γ) exist.

3.3 Other one-parameter exponential family models

Table 1 provides the functions that must be minimized to obtain UMPBTs for a number of common exponential family models. The objective functions listed in this table correspond to the function $g(\cdot, \cdot)$ specified in Lemma 1 and $v = 1$. The negative binomial is parameterized by the fixed number of failures r and random number of successes $x = 0, 1, \dots$ observed before termination of sampling. The other models are parameterized so that μ and p denote means and proportions, respectively, while σ^2 values refer to variances.

4 A paradigm shift for standard hypothesis tests

As pointed out at the beginning of Section 3, a vast majority of the hypothesis tests conducted by practitioners are tests of simple hypotheses in exponential family models. These tests are generally conducted at the 5% significance level, even though a p -value of 0.05 only corresponds to a Bayes factor of 3 or 4 in favor of the alternative hypothesis, as demonstrated in Section 3.1 and by other authors (e.g. Berger and Sellke, 1987).

UMPBTs provide a simple alternative to UMPTs in these settings. Instead of conducting tests at the 5% level of significance, default thresholds for Bayes factors might be set according to the categorization scheme proposed by Kass and Raftery (1995). In small experiments (i.e., $n < 30$), default thresholds for Bayes factors might be set so as to obtain positive evidence, say $\gamma = 10$, in support of the alternative hypothesis. For larger experiments, γ might be set to obtain either strong evidence (e.g., $\gamma = 50$) or very strong (e.g., $\gamma = 250$) evidence in favor of the alternative hypothesis.

The report of Bayes factors based on the UMPBT as the outcome of a hypothesis also provides a convenient link to the test's Type I error. Because the rejection regions for the UMPBT have the same form as the UMPT, calculating the Type I error corresponding to a UMPBT test is relatively straightforward. Of course, it should be noted if the threshold for the UMPBT is assumed to be too small, the resulting test might have unacceptably high Type I errors when viewed from a frequentist perspective.

5 Extensions to other models

Like UMPTs, UMPBTs are most easily defined within one parameter exponential family models. In unusual cases, UMPBTs can be defined for data modeled by vector-valued exponential family models, but in general such extensions appear to require stringent constraints on nuisance parameters. For example, if the density of an observation can be expressed as

$$f(x | \boldsymbol{\theta}) = h(x) \exp \left[\sum_{i=1}^d \eta_i(\boldsymbol{\theta}) T_i(x) - A(\boldsymbol{\theta}) \right], \quad (20)$$

then UMPBTs can be defined when all but one of the $\eta_i(\boldsymbol{\theta})$ are constrained to have identical values under both hypotheses.

More useful extensions can be obtained when it is possible to integrate out nuisance parameters in order to obtain a marginal density for the parameter of interest that falls within the class of exponential family of models. An important example of this type occurs in testing whether a regression coefficient in a linear model is zero.

5.1 Test of linear regression coefficient, σ^2 known

Suppose that

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \quad (21)$$

where σ^2 is known, \mathbf{y} is an $n \times 1$ observation vector, \mathbf{X} an $n \times p$ design matrix of full column rank, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ denotes a $p \times 1$ regression parameter. The null hypothesis is defined as $H_0 : \beta_p = 0$. For concreteness, suppose that interest focuses on testing whether $\beta_p > 0$, and that under both the null and alternative hypotheses, the prior density on the first $p - 1$ components of $\boldsymbol{\beta}$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 \boldsymbol{\Sigma}$. Then the marginal density of \mathbf{y} under H_0 is

$$m_0(\mathbf{y}) = (2\pi\sigma^2)^{-n/2} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} |\mathbf{F}|^{-\frac{1}{2}} \exp \left(-\frac{R}{2\sigma^2} \right), \quad (22)$$

where

$$\mathbf{F} = \mathbf{X}'_{-p} \mathbf{X}_{-p} + \boldsymbol{\Sigma}^{-1}, \quad \mathbf{H} = \mathbf{X}_{-p} \mathbf{F}^{-1} \mathbf{X}'_{-p}, \quad R = \mathbf{y}'(\mathbf{I}_n - \mathbf{H})\mathbf{y}, \quad (23)$$

and \mathbf{X}_{-p} is the matrix consisting of the first $p - 1$ columns of \mathbf{X} .

Let β_{p*} denote the value of β_p under the alternative hypothesis H_1 that defines the UMPBT(γ), and let \mathbf{x}_p denote the p^{th} column of \mathbf{X} . Then the marginal density of \mathbf{y} under H_1 is

$$m_1(\mathbf{y}) = m_0(\mathbf{y}) \times \exp \left\{ -\frac{1}{2\sigma^2} [\beta_{p*}^2 \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{x}_p - 2\beta_{p*} \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{y}] \right\} \quad (24)$$

It follows that the probability that the Bayes factor BF_{10} exceeds γ can be expressed

$$\mathbf{P} \left[\mathbf{x}'_p(\mathbf{I}_n - \mathbf{H})\mathbf{y} > \frac{\sigma^2 \log(\gamma)}{\beta_{p^*}} + \frac{1}{2}\beta_{p^*}\mathbf{x}'_p(\mathbf{I}_n - \mathbf{H})\mathbf{x}_p \right], \quad (25)$$

which is maximized by minimizing the right-hand side of the inequality. The UMPBT(γ) is thus obtained by taking

$$\beta_{p^*} = \sqrt{\frac{2\sigma^2 \log(\gamma)}{\mathbf{x}'_p(\mathbf{I}_n - \mathbf{H})\mathbf{x}_p}}. \quad (26)$$

The corresponding one-sided test of $\beta_p < 0$ is obtained by reversing the sign of β_{p^*} in (26).

Because this expression for the UMPBT assumes that σ^2 is known, it is not of great practical significance by itself. However, this result may guide the specification of alternative models in, for example, model selection algorithms in which the priors on regression coefficients are specified conditionally on the value of σ^2 . For example, the mode of the non-local priors described in Johnson and Rossell (2012) might be set to the UMPBT values after determining an appropriate value of γ based on both the sample size n and number of potential covariates p .

6 Approximations to UMPBTs using data-dependent alternatives

In some situations –most notably in linear models with unknown variances–data-dependent alternative hypotheses can be defined to obtain tests that are approximately uniformly most powerful in maximizing the probability that a Bayes factor exceeds a threshold. This strategy is only attractive when the statistics used to define the alternative hypothesis are ancillary to the parameter of interest.

6.1 Test of normal mean, σ^2 unknown

Suppose that x_i , $i = 1, \dots, n$ are i.i.d. $N(\mu, \sigma^2)$, that σ^2 is unknown, and that the null hypothesis is $H_0 : \mu = \mu_0$. For convenience, assume further that the prior distribution on σ^2 is an inverse gamma distribution with parameters α and λ under both the null and alternative hypotheses.

To obtain an approximate UMPBT(γ), begin by marginalizing over σ^2 in both models. Noting that $(1 + a/t)^t \rightarrow e^a$, it follows that the Bayes factor in favor of the alternative

hypothesis satisfies

$$BF_{10}(\mathbf{x}) = \left[\frac{\sum_{i=1}^n (x_i - \mu_0)^2 + 2\lambda}{\sum_{i=1}^n (x_i - \mu_1)^2 + 2\lambda} \right]^{n/2+\alpha} \quad (27)$$

$$= \left[\frac{1 + (\bar{\mathbf{x}} - \mu_0)^2/s^2}{1 + (\bar{\mathbf{x}} - \mu_1)^2/s^2} \right]^{n/2+\alpha} \quad (28)$$

$$\approx \exp \left\{ -\frac{n}{2s^2} [(\bar{\mathbf{x}} - \mu_1)^2 - (\bar{\mathbf{x}} - \mu_0)^2] \right\}, \quad (29)$$

where

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2 + 2\lambda}{n + 2\alpha}. \quad (30)$$

The expression for the Bayes factor in (29) reduces to (14) if σ^2 is replaced by s^2 . This implies that an approximate, but data-dependent, UMPBT alternative hypothesis can be specified by taking

$$\mu_1 = \mu_0 \pm s \sqrt{\frac{2 \log \gamma}{n}}, \quad (31)$$

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

Figure 4 depicts the probability that the Bayes factor exceeds $\gamma = 10$ when testing a null hypothesis that $\mu = 0$ based on an independent sample of size $n = 30$ normal observations with unit variance ($\sigma^2 = 1$) and using (31) to set the value of μ_1 under the alternative hypothesis. For comparison, the probability that the Bayes factor exceeds 10 when the alternative is defined by taking $\sigma^2 = 1$ and μ_1 to be the data-generating parameter is depicted by the dashed curve in the plot. Interestingly, the data-dependent, approximate UMPBT(10) provides a higher probability of producing a Bayes factor that exceeds 10 than do alternatives fixed at the data generating parameters.

6.2 Test of linear regression coefficient, σ^2 unknown

As final example, suppose that the sampling model of section 5.1 holds, but assume now that the observational variance σ^2 is unknown and assumed under both hypotheses to be drawn from an inverse gamma distribution with parameters α and λ . Also assume that the prior distribution for the first $p - 1$ components of $\boldsymbol{\beta}$, given σ^2 , is a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma^2 \boldsymbol{\Sigma}$. As before, assume that $H_0 : \beta_p = 0$. Our goal is to determine a value β_{p*} so that $H_1 : \beta_p = \beta_{p*}$ is the UMPBT(γ) under the constraint that $\beta_p > 0$.

Define $\mathbf{y}_1 = \mathbf{y} - \mathbf{x}_p \beta_{p*}$ and let $\mathbf{y}_0 = \mathbf{y}$. By integrating with respect to the prior densities on σ^2 and the first $p - 1$ components of $\boldsymbol{\beta}$, the marginal density of the data under hypothesis i , $i = 0, 1$ can be expressed as

$$m_i(\mathbf{y}) = 2^\alpha \pi^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \frac{\lambda^\alpha}{\Gamma(\alpha)} \Gamma(n/2 + \alpha) |\mathbf{F}|^{-1/2} R_i^{-n/2-\alpha} \quad (32)$$

where \mathbf{F} is defined in (23), and

$$R_i = \mathbf{y}'_i(\mathbf{I}_n - \mathbf{H})\mathbf{y}_i + 2\lambda. \quad (33)$$

It follows that the Bayes factor in favor of H_1 can be written as

$$BF_{10} = \left[1 + \frac{\beta_{p^*}^2 \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{x}_p - 2\beta_{p^*} \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{y}}{R_0} \right]^{-n/2-\alpha} \quad (34)$$

$$\approx \exp \left\{ -\frac{1}{2s_p^2} [\beta_{p^*}^2 \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{x}_p - 2\beta_{p^*} \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{y}] \right\}, \quad (35)$$

where

$$s_p^2 = \frac{R_0}{n + 2\alpha}. \quad (36)$$

The UMPBT(γ) is defined from (35) according to

$$\mathbf{P}(BF_{10} > \gamma) = \mathbf{P} \left[\mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{y} > \frac{s_p^2 \log(\gamma)}{\beta_{p^*}} + \frac{1}{2} \beta_{p^*} \mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{x}_p \right]. \quad (37)$$

Minimizing the right-hand side of the last inequality with respect to β_{p^*} results in

$$\beta_{p^*} = \sqrt{\frac{2s_p^2 \log(\gamma)}{\mathbf{x}'_p (\mathbf{I}_n - \mathbf{H}) \mathbf{x}_p}}. \quad (38)$$

This expression is consistent with the result obtained in the known variance case, but with s_p^2 substituted for σ^2 .

7 Discussion

The UMPBTs described in this article provide a practical Bayesian alternative to classical UMPTs. In addition to providing Bayes factors in favor of the alternative hypothesis in the most common of testing situations, they also offer consumers of these tests an opportunity to obtain an approximate calibration between p -values and posterior model probabilities. They also provide a new approach for defining objective Bayesian hypothesis tests.

Perhaps as importantly as providing a calibration between posterior model probabilities and p , the use of UMPBTs also provides practitioners with an estimate of the effect size that is implicitly being tested. In normal models, equation (18) shows that the UMPBT alternative value of μ decreases at rate $O(\sqrt{n})$. For large samples, the alternative hypothesis tested for moderate values of γ can thus be very close to the null value. Explicitly calculating the alternative value of the parameter that is being favored over null may thus provide a useful check for assessing whether the rejection of a null hypothesis is significant

from a scientific perspective. Equation (18) also suggests that investigators should expect evidence to accumulate exponentially fast against a false null hypothesis.

This article has focused on the specification of UMPBTs for one-sided alternatives. The natural extension of these tests to two-sided alternatives can be obtained by assuming that the alternative hypothesis is represented by two equally-weighted point masses located at the UMPBT values determined for one-sided tests. The Bayes factors for such tests can be written as

$$\mathbf{P} \left[\frac{0.5m_l(\mathbf{x}) + 0.5m_h(\mathbf{x})}{m_0(\mathbf{x})} > \gamma \right], \quad (39)$$

where m_l and m_h denote marginal densities corresponding to one-sided UMPBTs. Letting $m_*(\mathbf{x}) = \max(m_l(\mathbf{x}), m_h(\mathbf{x}))$ for the data actually observed, and assuming that the favored marginal dominates the other, it follows that

$$\mathbf{P} \left[\frac{0.5m_l(\mathbf{x}) + 0.5m_h(\mathbf{x})}{m_0(\mathbf{x})} > \gamma \right] \approx \mathbf{P} \left[\frac{m_*(\mathbf{x})}{m_0(\mathbf{x})} > 2\gamma \right]. \quad (40)$$

Thus, an approximate two-sided UMPBT(γ) can be defined by specifying an alternative hypothesis that equally concentrates its mass on the two UMPBT(2γ) tests.

Additional research is needed to identify classes of models and testing contexts for which UMPBTs can be defined. The UMPBTs described in this article primarily involve tests of point null hypotheses, or tests that can be reduced to a test of a point null hypothesis after marginalizing over nuisance parameters. Whether UMPBTs can be defined in more general settings remains an open question.

8 References

- Berger, J.O., and Selke, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence. *Journal of the American Statistical Association*, **82**, 1121-1122.
- Kass, R.E and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.
- Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd edition. Springer: New York.
- Neyman, J. and Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20A**, 175-240.
- Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A*, **231**, 289-337.
- Johnson, V.E. (2005). Bayes factors based on test statistics. *Journal of the Royal Statistical Society, Series B*, **5**, 689-701.

Johnson, V.E. (2008). Properties of Bayes Factors Based on Test Statistics. *Scandinavian Journal of Statistics*, **35**, 354-368.

Johnson, V.E. and Rossell, D. (2012). Bayesian variable selection in high-dimensional settings. *Journal of the American Statistical Association*, in press.

Prosper, H. (2012), Personal communication to news@bayesian.org, July 11, 2012.



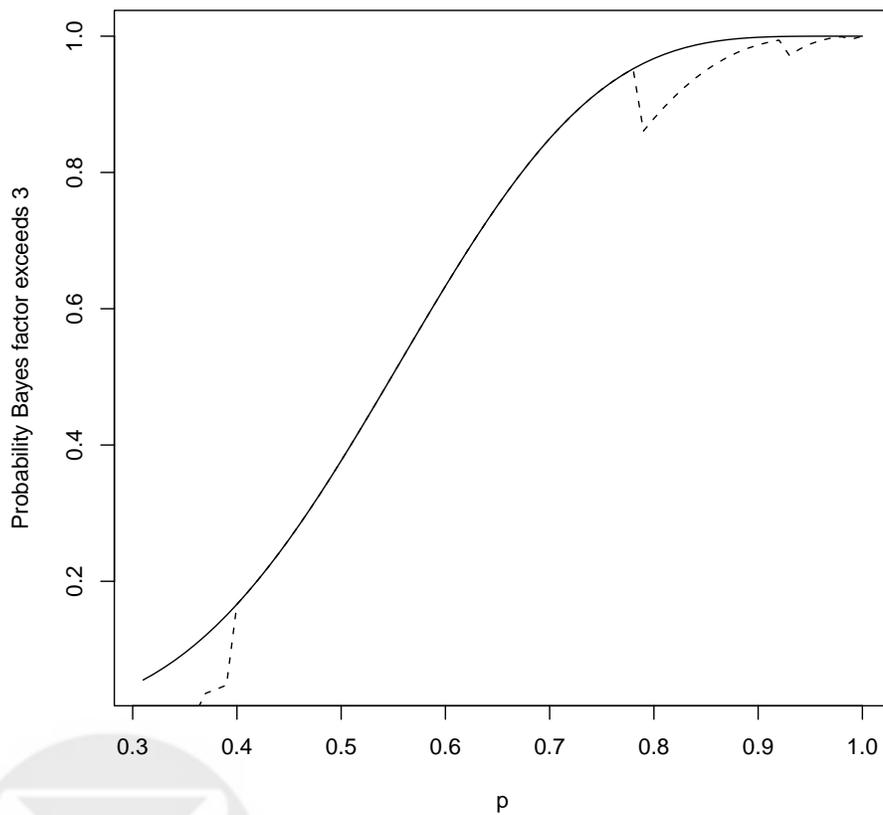


Figure 1: Probability that Bayes factor in favor of UMPBT binary alternative exceeds 3 when $p_0 = 0.3$ and $n = 10$ (solid curve). The dashed curve displays this probability when the Bayes factor is calculated under alternative hypothesis that that p_1 equals the data-generating parameter, which is displayed on the horizontal axis.

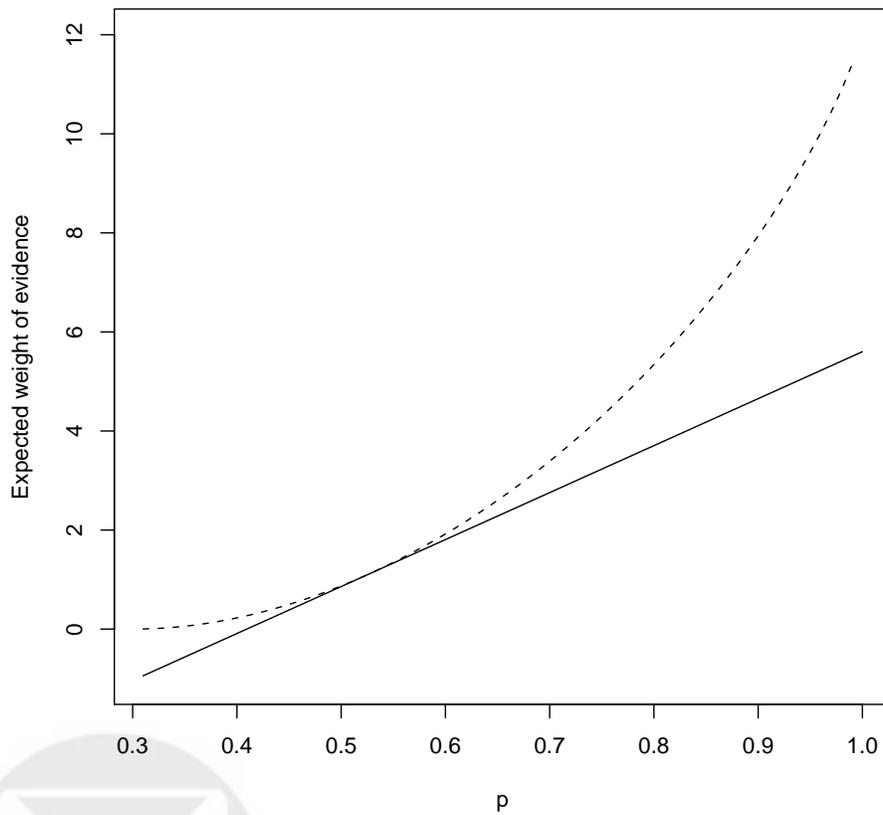


Figure 2: Expected weight of evidence produced by $UMPBT(\gamma)$ against a null hypothesis that $p_0 = 0.3$ when the sample size is $n = 10$, versus the data-generating success probability (solid curve). The dashed curve displays the expected weight of evidence obtained when the alternative hypothesis exactly corresponds to the data-generating parameter value (displayed on the horizontal axis).

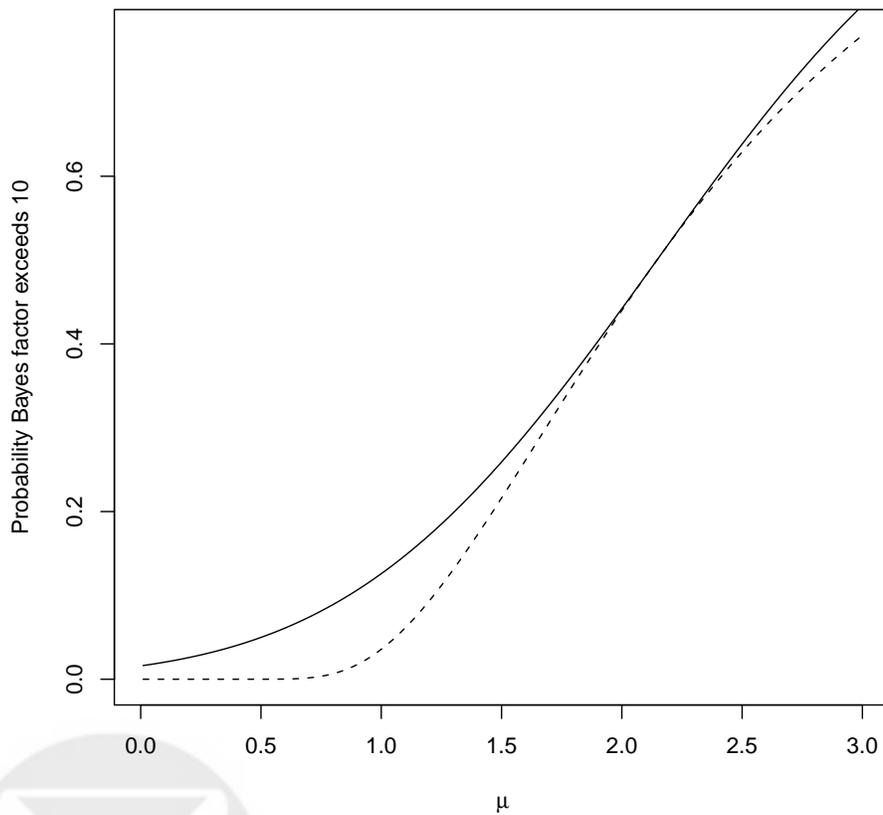


Figure 3: Probability that Bayes factor in favor of UMPBT alternative exceeds 10 when $\mu_0 = 0$ and $n = 1$ (solid curve). The dashed curve displays this probability when the Bayes factor is calculated under the alternative hypothesis that μ_1 equals the data-generating parameter (displayed on the horizontal axis).

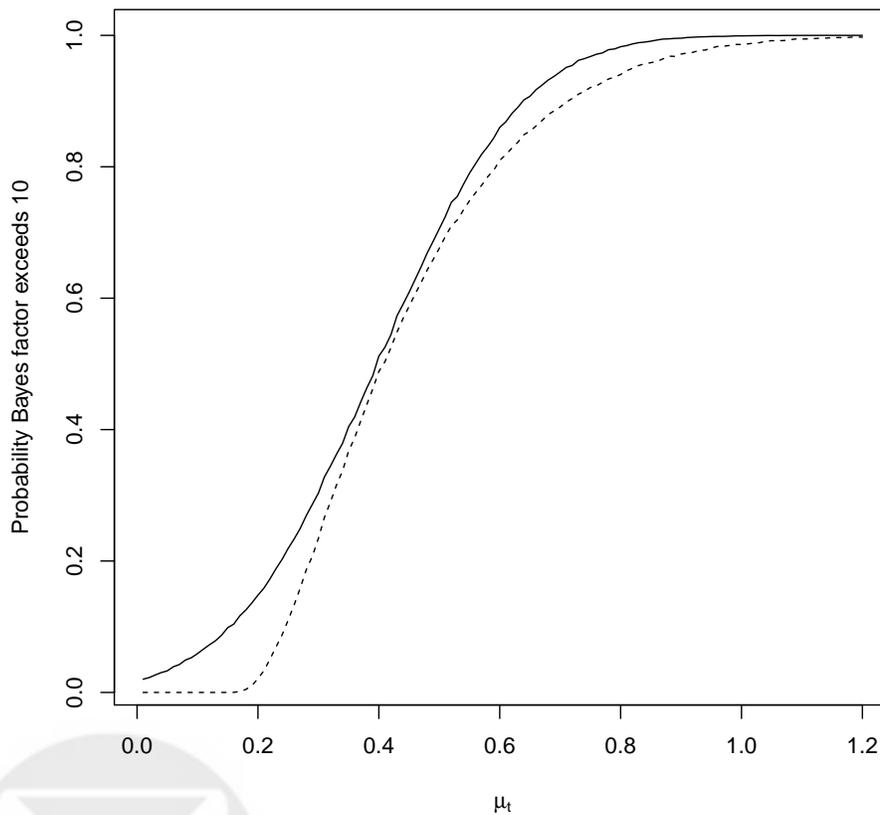


Figure 4: Probability that Bayes factor based on data-dependent, approximate UMPBT alternative exceeds 10 when $\mu_0 = 0$ and $n = 30$ (solid curve). The dashed curve displays this probability when the Bayes factor is calculated under the alternative hypothesis that μ_1 equals data-generating parameter (displayed on the horizontal axis) and $\sigma^2 = 1$.