DISTINGUISHING A BROWNIAN BRIDGE FROM A BROWNIAN MOTION WITH DRIFT

by

Anirban DasGupta
Purdue University

Technical Report #96-8

Department of Statistics
Purdue University
West Lafayette, IN USA

March 1996

DISTINGUISHING A BROWNIAN BRIDGE FROM A BROWNIAN MOTION WITH DRIFT

by

Anirban DasGupta*

Purdue University

Abstract

Gaussian processes, and in particular the Brownian Bridge and the Brownian motion, are used to model numerous processes in applications. In many of these problems, accurate forecasting of the process at future times is important. We address the problem of distinguishing between the Brownian Bridge and the Brownian motion with possible drift on the basis of observations at n discrete times; both deterministic and random times are considered. The article starts with calculating the mean of the L_1 distance between the two processes; this is done to understand if distinguishing between the two processes is intrinsically difficult. We then derive the likelihood ratio test statistic, its asymptotic null distribution and the fixed sample null and nonnull distributions. These are used to study the critical values and the power and their robustness with respect to the distribution of the times when the times are random. We also derive the Bayes factor and its asymptotics vis-a-vis the likelihood ratio and compare posterior probabilities to the P value of the likelihood ratio test. It is found that the posterior probabilities are not robust with respect to the distribution of times as well, and especially so at large sample sizes. The Bayes rule has the amusing feature that for all small sample sizes, it always rejects Brownian Bridge as the model, for all samples.

1. INTRODUCTION

Many naturally occurring processes in various branches of science, economics, and other disciplines behave like Gaussian processes. In particular, the Brownian motion and the Brownian Bridge are used as models for many processes; for instance, such processes are used to model the ups and downs of the stock market, the movement of cells or subcellular organisms, the diffusion of gas or fluid molecules, etc. See Berg (1993).

In many of these problems, it is important to be able to accurately predict the location

* Research supported by NSF grant DMS 93-077-27

of the process at a (distant) future time. If so, it is clearly important that one can correctly identify the true process generating the observations, from a number of possible processes one can think of. It is also, realistically, often the case that the observations one makes on the process are at discrete times. One therefore has the question: is it possible to accurately distinguish between various Gaussian processes based on observations at a finite sequence of discrete times? In this article, we address the question of testing if observations $X_1, X_2, \ldots, X_n$ corresponding to times $t_1, t_2, \ldots, t_n$ are coming from a standard Brownian Bridge or a Brownian motion with possible drift. The time interval under consideration is the unit interval [0,1]. We consider the case where $t_1, t_2, \ldots, t_n$ are deterministic as well as the case where they are the order statistics of a random sample from a CDF F on [0,1]. In fact, many results we present do not depend on whether the times are deterministic or random.

In section 2, we do some preliminary calculations to anticipate if the distinction is going to be easy or difficult by deriving the mean L_1 distance between the two processes; this calculation has some independent probabilistic interest.

In section 3, we derive the likelihood ratio test, and its asymptotics. In this section, we also consider the fixed sample distribution of the likelihood ratio statistic under both the null and the alternative. These are then used to give critical values and evaluate the power of the test and also to study the robustness of the critical values themselves and the power to the choice of F, when the times of observations are random. We conclude that the critical values lack in robustness with respect to F, making a decision to reject or accept difficult unless the distribution F is known pretty precisely to the practitioner. We also conclude that the power of the test is small, especially if the drift of the Brownian motion is small to moderate, making identification of the true process also difficult.

Section 4 treats the problem from a Bayesian angle; we derive expressions for the Bayes factor and the posterior probability of the null hypothesis and consider the asymptotics of the Bayes factor vis-a-vis the likelihood ratio statistic. We also present tables of posterior probabilities and corresponding P values of the frequentist likelihood ratio test, for comparison. This section also presents a rather interesting phenomenon: if the times are deterministic, then it can happen that for all small n, the Bayes test ALWAYS rejects

H_0, regardless of the sample obtained.

Similar analyses with other Gaussian processes are under investigation. Tests on Wiener processes are considered in Simons et al (1989); some general optimality of likelihood ratio tests is discussed in Brown (1971).

2. SOME ANTICIPATIVE CALCULATIONS

Before we derive the likelihood ratio and Bayes tests and their properties in the subsequent sections, let us do some quick calculations to anticipate what we might see in terms of the ability to distinguish between a Brownian Bridge and a Brownian motion with drift. These calculations are of independent probabilistic interest as well. Let then $B(t)$ and $\xi(t)$ denote respectively a Brownian Bridge and an independent (standard) Brownian motion on [0,1]. It is natural to look at the L_1 norm $\int_0^1 |B(t) - t\mu - \xi(t)|dt$ as a measure of how different are the sample paths of a Brownian Bridge and a Brownian motion with drift. The L_1 norm is a random variable; Johnson and Killeen (1983) consider the L_1 norm of just the Brownian Bridge. In the following, we give a representation for the expected value of the L_1 norm $D_\delta = \int_0^\delta |B(t) - t\mu - \xi(t)|dt$ on the interval $[0, \delta]$; we will then use the expression to compute $E(D_\delta)$ and see how small or large the L_1 norm is on the average.

Theorem 1. $E_{\mu=0}(D_\delta) = \frac{\sqrt{\pi}}{2\sqrt{2}}(\frac{\delta}{2-\delta})^{\frac{3}{2}}$; in general, $E_\mu(D_\delta)$ admits the infinite series representation

$$\begin{aligned} E_\mu(D_\delta) &= (2 - \frac{\delta^2}{2})\mu - 4\mu(2-\delta)\Phi(\frac{\mu\sqrt{\delta}}{\sqrt{2-\delta}}) + \mu(2-\delta)^2\Phi(\frac{\mu\sqrt{\delta}}{\sqrt{2-\delta}}) \\ &+ \frac{4\sqrt{2}}{\sqrt{\pi}}(\frac{\delta}{2-\delta})^{\frac{3}{2}} \sum_{j=0}^{\infty}(-1)^j(\frac{\delta\mu^2}{2(2-\delta)})^j \frac{{}_2F_1(3, j+\frac{3}{2}; j+\frac{5}{2}; -\frac{\delta}{2-\delta})}{j+\frac{3}{2}} \\ &+ \frac{2\sqrt{2}}{\sqrt{\pi}}\mu^2(\frac{\delta}{2-\delta})^{\frac{1}{2}} \sum_{j=0}^{\infty}(-1)^j(\frac{\delta\mu^2}{2(2-\delta)})^j \frac{{}_2F_1(1, j+\frac{1}{2}; j+\frac{3}{2}; -\frac{\delta}{2-\delta})}{j+\frac{1}{2}} \\ &- \frac{\sqrt{2}}{\sqrt{\pi}}\mu^2(\frac{\delta}{2-\delta})^{\frac{1}{2}} \sum_{j=0}^{\infty}(-1)^j(\frac{\delta\mu^2}{2(2-\delta)})^j \frac{{}_2F_1(2, j+\frac{1}{2}; j+\frac{3}{2}; -\frac{\delta}{2-\delta})}{j+\frac{1}{2}}. \end{aligned} \qquad (1)$$

Proof: Although expression (1) looks intimidating, the derivation is easily understood from the following main steps:

Step 1. By Fubini's theorem, $E_\mu(D_\delta) = \int_0^\delta E_\mu|B(t) - t\mu - \xi(t)|dt$.

Step 2. $t\mu + \xi(t) - B(t) \sim N(\theta, \sigma^2)$ where $\theta = \theta(t) = t\mu$ and $\sigma^2 = \sigma^2(t) = 2t - t^2$.

Step 3. By a direct integration by parts, whenever $X \sim N(\theta, \sigma^2)$, $E|X| = 2\sigma\phi(\frac{\theta}{\sigma}) + \theta[2\Phi(\frac{\theta}{\sigma}) - 1]$.

Step 4. The first term in Step 3 corresponds to $\frac{2}{\sqrt{2\pi}} \int_0^\delta \sqrt{2t - t^2} e^{-\frac{t^2\mu^2}{2(2t-t^2)}} dt$. On making the substitution $\frac{t^2}{2t-t^2} = x$, this equals

$$\frac{4\sqrt{2}}{\sqrt{\pi}} \int_0^{\frac{\delta}{2-\delta}} \frac{\sqrt{x} e^{-\frac{\mu^2}{2}x}}{(1+x)^3} dx. \tag{2}$$

Step 5. The second term in Step 3 corresponds to $-\frac{\mu}{2}\delta^2 + 2\mu \int_0^\delta t\Phi(\frac{t\mu}{\sqrt{2t-t^2}})dt$. On making the substitution $\frac{t}{\sqrt{2t-t^2}} = x$, this equals

$$-\frac{\mu}{2}\sigma^2 + 16\mu \int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{x^3\Phi(\mu x)}{(1+x^2)^3} dx. \tag{3}$$

Step 6. If $\int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{x^3\Phi(\mu x)}{(1+x^2)^3} dx$ is integrated by parts twice then (3) becomes

$$\begin{aligned} &-\frac{\mu}{2}\delta^2 + 2\mu - 4\mu(2-\delta)\Phi(\frac{\mu\sqrt{\delta}}{\sqrt{2-\delta}}) + \mu(2-\delta)^2\Phi(\frac{\mu\sqrt{\delta}}{\sqrt{2-\delta}}) \\ &+ \frac{4\sqrt{2}}{\sqrt{\pi}}\mu^2 \int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{e^{-\frac{\mu^2}{2}x^2}}{1+x^2} dx - \frac{2\sqrt{2}}{\sqrt{\pi}}\mu^2 \int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{e^{-\frac{\mu^2}{2}x^2}}{(1+x^2)^2} dx. \end{aligned} \tag{4}$$

Step 7. In (4), $\int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{e^{-\frac{\mu^2}{2}x^2}}{1+x^2} dx$ becomes $\frac{1}{2} \int_0^{\frac{\delta}{2-\delta}} \frac{e^{-\frac{\mu^2}{2}z}}{\sqrt{z}(1+z)} dz$ and $\int_0^{\sqrt{\frac{\delta}{2-\delta}}} \frac{e^{-\frac{\mu^2}{2}x^2}}{(1+x^2)^2} dx$ becomes $\frac{1}{2} \int_0^{\frac{\delta}{2-\delta}} \frac{e^{-\frac{\mu^2}{2}z}}{\sqrt{z}(1+z)^2}$ on making the substitution $z = x^2$.

Step 8. Now one simply combines steps 4 through 7 and uses the following fact:

$$\int_0^u \frac{x^{\mu-1}}{(1+x)^\nu} dx = \frac{u^\mu}{\mu} {}_2F_1(\nu, \mu; 1+\mu; -u). \tag{5}$$

(see, e.g., Gradshteyn and Ryzhik (1980), pp. 284).

Table 1 gives the value of $E_\mu(D_\delta)$ for some values of δ and μ.

Table 1: $E_\mu(D_\delta)$

δ	μ = 0	.5	1
.05	.00257	.00836	.00841
.5	.1206	.25051	.26666
1	.62665	.66174	.76196

Table 1 suggests that even on the whole interval [0,1], the L_1 distance between a Brownian Bridge and a Brownian motion with moderate drift is quite small on the average and distinguishing one from the other may not be very easy.

3. THE LIKELIHOOD RATIO TEST

3.1 Description

In this section, we first derive the likelihood ratio test statistic and its asymptotic null distribution, which is the same whether the times $\{t_i\}$ are deterministic or random. We then give the fixed sample null and nonnull density of the likelihood ratio test statistic and use these to evaluate the critical value and the power function at selected values of μ. These are illustrated for three different choices of F, the distribution of the times. We show that there is indeed a practical difficulty in distinguishing a Brownian bridge from a Brownian motion with drift: the critical value is not robust with respect to F. Thus, unless the practitioner knows according to which F the random times are distributed, he/she will have difficulty in deciding whether to accept or reject the null hypothesis. We also show that to a somewhat less serious extent,the power also lacks in robustness.

3.2. The Test Statistic

Theorem 2. Let $t_1 \leq t_2 \leq \ldots \leq t_n$ be n deterministic or random times in [0,1], and let $x_1, x_2, \ldots, x_n$ be observations from a process $X(t)$ at the corresponding times. For testing $H_0 : X(t) \equiv B(t)$ vs. $H_1 : X(t) \equiv t\mu + \xi(t)$ for some real μ, where $B(t)$ and $\xi(t)$ denote a standard Brownian Bridge and a standard Brownian motion respectively, the likelihood ratio test (LRT) statistic λ is given by $-2\log\lambda = \frac{x_n^2}{t_n(1-t_n)} + \log(1-t_n)$ and $-2\log\lambda - \log(1-t_n) \xrightarrow{\mathcal{L}} \chi^2(1)$ under H_0.

Proof: The second part of the Theorem is an obvious consequence of the first part. To derive the first part, note that

$$\lambda = \frac{\frac{1}{|\Sigma_2|^{1/2}} e^{-\frac{1}{2}\underset{\sim}{x}'\Sigma_2^{-1}\underset{\sim}{x}}}{\sup_{\mu} \frac{1}{|\Sigma_1|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{x}-\mu\underset{\sim}{t})'\Sigma_1^{-1}(\underset{\sim}{x}-\mu\underset{\sim}{t})}}, \tag{6}$$

where $\underset{\sim}{x} = (x_1, \ldots, x_n)$, $\underset{\sim}{t} = (t_1, \ldots, t_n)$, and Σ_1, Σ_2 are the covariance matrices $\Sigma_1 = ((\min(t_i, t_j)))$, $\Sigma_2 = ((\min(t_i, t_j) - t_i t_j))$. Note that the representation (6) for λ is valid regardless of whether the times are deterministic or random, due to the ancillarity of the times.

By calculus, the maximum likelihood estimate of μ equals $\frac{\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{t}}{\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}}$. Substitution of this into (6) yields

$$\log\ \lambda = \frac{1}{2}\log|\Sigma_1| - \frac{1}{2}\log|\Sigma_2| - \frac{1}{2}\underset{\sim}{x}'\Sigma_2^{-1}\underset{\sim}{x} + \frac{1}{2}\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{x} - \frac{1}{2}\frac{(\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{t})^2}{\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}}. \tag{7}$$

Notice now the following facts:

$$\begin{aligned}
&\underline{\text{a}}\ \ \Sigma_2 = \Sigma_1 - \underset{\sim}{t}\underset{\sim}{t}' \\
&\underline{\text{b}}\ \ |\Sigma_2| = |\Sigma_1|(1 - \underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}) \\
&\underline{\text{c}}\ \ \Sigma_2^{-1} = \Sigma_1^{-1} + \frac{\Sigma_1^{-1}\underset{\sim}{t}\underset{\sim}{t}'\Sigma_1^{-1}}{1 - \underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}} \\
&\underline{\text{d}}\ \ \Sigma_1^{-1}\underset{\sim}{t} = \begin{pmatrix} 0 & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{pmatrix}
\end{aligned} \tag{8}$$

(see Rao (1973)). Hence $\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{t} = x_n$ and $\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t} = t_n$ and thus from (7),

$$\begin{aligned}
-2\log\ \lambda &= \log(1 - t_n) + \frac{x_n^2}{1 - t_n} + \frac{x_n^2}{t_n} \\
&= \frac{x_n^2}{t_n(1 - t_n)} + \log(1 - t_n),
\end{aligned} \tag{9}$$

as stated.

3.3. Fixed Sample Distribution of the LRT Statistic

Theorem 3. Let $t_1 \le t_2 \le \ldots \le t_n$ be the order statistics of a random sample from an absolutely continuous CDF F on [0,1] with density f. Then,

$\underline{\text{a}}$ the null distribution of $y = -2\log\ \lambda$ has CDF

$$P_{H_0}(y \le c) = n\int_{(1-e^c)\vee 0}^{1} (2\Phi(\sqrt{c - \log\,(1 - t)}) - 1)F^{n-1}(t)f(t)dt \tag{10}$$

b the nonnull distribution of $y = -2\log\ \lambda$ has CDF

$$P_{H_1}(y \le c) = n \int_{1-e^c)\text{vo}}^{1} \{G_1(c,\mu,t) + G_2(c,\mu,t)\} F^{n-1}(t) f(t) dt$$
$$+ F^n((1-e^c)\text{vo}) - 1, \qquad (11)$$

where

$$G_1(c,\mu,t) = \Phi(\sqrt{(c - \log(1-t))(1-t)} - \mu\sqrt{t})$$

and

$$G_2(c,\mu,t) = \Phi(\sqrt{(c - \log(1-t))(1-t)} + \mu\sqrt{t}) \qquad (12)$$

Proof: Both parts follow from $P(y \le c) = EP(y \le c|t_n)$. Under H_0, given t_n, $y = -2\log\ \lambda$ has the distribution of a $\chi^2(1) + \log(1-t_n)$ random variable and from this, part a follows. Under H_1, $x_n|t_n \sim N(t_n\mu, t_n)$, i.e., $x_n|t_n \stackrel{\mathcal{L}}{=} t_n\mu + z_n\sqrt{t_n}$, where $Z_n \sim N(0,1)$.

$$\therefore P_{H_1}(y \le c|t_n)$$
$$= P(\frac{x_n^2}{t_n(1-t_n)} \le c - \log\ (1-t_n)|t_n)$$
$$= P(N^2(\frac{\mu\sqrt{t_n}}{\sqrt{1-t_n}}, \frac{1}{\sqrt{1-t_n}}) \le c - \log\ (1-t_n))$$
$$= G_1(c,\mu,t_n) + G_2(c,\mu,t_n) - 1 \quad \text{if} \quad c > \log\ (1-t_n) \quad \text{and is zero}$$

otherwise. Part b now follows.

The following table gives the critical values and the power of a 5% test at $\mu = 0, .5, 1, 2$, for $F(t) = t$, $\frac{2}{\pi}\sin^{-1}\sqrt{t}$, and $3t^2 - 2t^3$. Thus F is respectively the $U[0,1]$, the Arc-Sine, and the Beta (2,2) CDF. The critical values and the power were obtained from Theorem 3 by numerical integration on Mathematica.

Table 2

F	n	Critical Value	Power			
			$\mu = 0$	$\mu = .5$	$\mu = 1$	$\mu = 2$
Uniform	10	1.28	.615329	.651321	.740016	.918147
	20	.631	.715502	.745131	.816679	.950607
	30	.236	.764117	.789651	.850801	.962111
	40	−.047	.794125	.816858	.871058	.968254
	50	−.269	.815030	.835704	.884857	.972187
Arc-Sine	10	.1076	.806031	.826464	.875613	.966637
	20	−1.14	.896461	.908233	.936101	.984951
	30	−1.904	.929644	.937783	.956974	.990155
	40	−2.457	.946764	.952963	.967556	.992655
	50	−2.890	.957194	.962195	.973956	.994134
Beta (2,2)	10	2.005	.450021	.493510	.604181	.850991
	20	1.652	.525077	.567427	.673034	.892751
	30	1.445	.565906	.606700	.707427	.910114
	40	1.298	.593301	.632723	.729464	.920163
	50	1.184	.613615	.651862	.745317	.926914

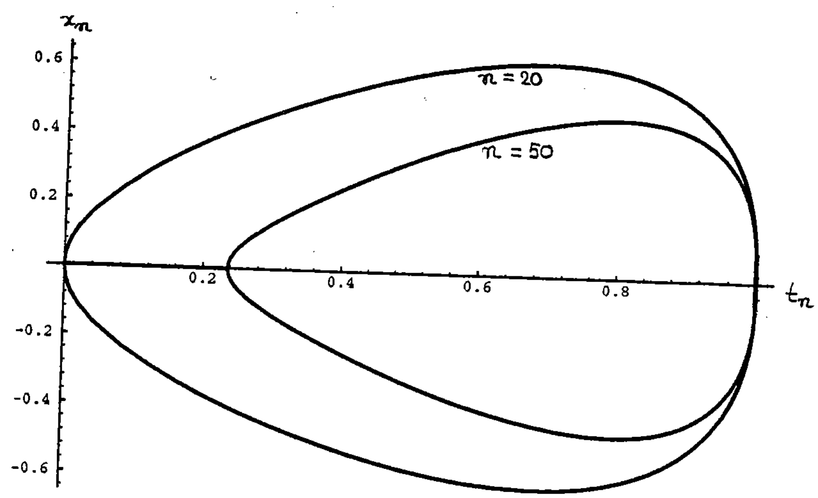
Discussion. The most alarming aspect of Table 2 is the lack of robustness of the critical value to the choice of F. Unless the choice of F was explicitly precontrolled, it would not be easy to tell the difference between the $U[0,1]$ and the Beta [2,2] CDF. Yet, in one case, the critical value is .631 and in another case it is 1.652 for $n = 20$, a moderately large sample size. The consequence is that it will be difficult for the practitioner to decide if H_0 should be rejected. Also, one sees from Table 2 that if F is the $U[0,1]$ or a CDF close to it, i.e., if the times $\{t_i\}$ are quite evenly spaced out, then it would be difficult to distinguish between a Brownian motion with a small drift and a Brownian Bridge; e.g., if F is the Beta [2,2] CDF, then the power of the 5% test is only .65 at $\mu = .5$ with as many as 50 observations.

Corollary 1. Let $A_n = F^{-1}(1-\frac{1}{n})$ and $\alpha_n = \frac{1}{1-A_n}$. Then the CDF of the null distribution of $y = -2\log\lambda$ admits the expansion

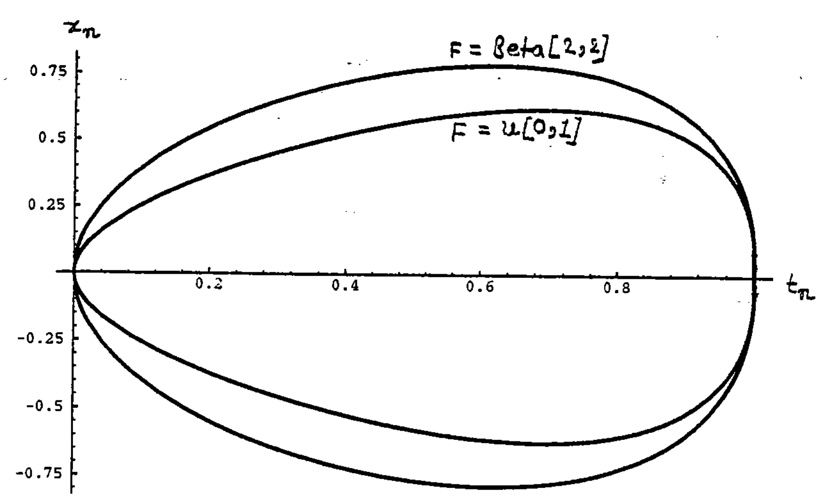
$$P_{H_0}(y \le c) = 1 - \sqrt{\frac{2}{\pi}}\frac{e^{-\frac{c}{2}}}{\sqrt{\alpha_n \log\ \alpha_n}} + o(\frac{1}{\sqrt{\alpha_n \log\ \alpha_n}}).$$

This is useful in the sense numerical evaluation of a critical value can be replaced by use of this asymptotic expansion to approximate the critical value. Corollary 1 follows on using the fact that under H_0, $\frac{x_n^2}{t_n(1-t_n)}$ (given t_n) has a $\chi^2(1)$ distribution, and the fact $1-\Phi(z) = \frac{\phi(z)}{z}(1+o(\frac{1}{z}))$ as $z \to \infty$.

Acceptance region of the Likelihood ratio test
$\alpha = .05 \quad F = U[0,1]$



Acceptance region of the Likelihood ratio test
$\alpha = .05 \quad n = 20$



4. POSTERIOR PROBABILITY AND THE BAYES FACTOR

4.1. Description

The standard Bayesian analysis of testing a point null against a compound alternative assigns a prior probability p to the point null and a conditional prior distribution on the parameters given that the alternative is true and then computes the posterior probability of the null hypothesis; it is also common to compute a Bayes factor, a monotone function of the posterior probability. There are striking differences between the apparent messages of a frequentist and a Bayesian test of a point null hypothesis. Primary among them is a phenomenon commonly called "Lindley's paradox"; the posterior probability of H_0 increases with n to 1 when the classical P-value is very very small. See Robert (1994) for a marvelous discussion. We have the following result.

4.2. The Bayes Test

Theorem 4. Let the prior probability of H_0 be p, $0 < p < 1$, and let μ have a prior distribution G (which may not be absolutely continuous). Then, whether or not the times $\{t_i\}$ are deterministic, the Bayes factor in favor of H_0 equals

$$B = \frac{1}{\sqrt{2\pi}} - e^{-\frac{x_n^2}{2t_n(1-t_n)}} / (\Phi_n * G)(\frac{x_n}{t_n}), \tag{13}$$

where Φ_n denotes the $N(0, \frac{1}{t_n})$ CDF and $\Phi_n * G$ denotes the density of the convolution of Φ_n and G. Furthermore, the posterior probability of H_0 equals $\frac{\frac{p}{q}B}{1+\frac{p}{q}B}$.

Proof: The expression for the posterior probability in terms of B is always true. To see the expression for B, note that by definition,

$$B = \frac{\frac{1}{|\Sigma_2|^{1/2}} e^{-\frac{1}{2}\underset{\sim}{x}'\Sigma_2^{-1}\underset{\sim}{x}}}{\int \frac{1}{|\Sigma_1|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{x}-\mu\underset{\sim}{t})'\Sigma_1^{-1}(\underset{\sim}{x}-\mu\underset{\sim}{t})} dG(\mu)}. \tag{14}$$

Since $-\frac{1}{2}(\underset{\sim}{x}-\mu\underset{\sim}{t})'\Sigma_1^{-1}(\underset{\sim}{x}-\mu\underset{\sim}{t})$ simplifies to $-\frac{\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}}{2}(\mu - \frac{\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{t}}{\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}})^2 + \frac{(\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{t})^2}{2\underset{\sim}{t}'\Sigma_1^{-1}\underset{\sim}{t}}$, (13) follows from (14) because of (8).

Corollary 2. Under the $0-1$ loss, the Bayes test rejects H_0 if and only if

$$\frac{x_n^2}{t_n(1-t_n)} + \log(1-t_n) - \log(t_n) + 2\log\left((\Phi_n * G)(\frac{x_n}{t_n})\right) > 2\log\frac{p}{2\pi q}.$$

The proof of Corollary 1 simply follows from the fact that the Bayes test rejects H_0 when its posterior probability is smaller than $\frac{1}{2}$. This corresponds to the stated rule on rearranging terms in the inequality $B < \frac{q}{p} \Leftrightarrow \log\ B < \log\ \frac{q}{p}$.

Example 1. Corollary 1 leads to the following rather interesting phenomenon. The Bayes rule of Corollary 1 is valid if the times $\{t_i\}$ are deterministic or random. If they are deterministic and t_n is increasing in n, then $2\log\ \frac{p}{2\pi q} + \log\ t_n - \log(1-t_n)$ would be increasing in n and depending on the value of p, there will be a smallest n_0 such that $2\log\ \frac{p}{2\pi q} + \log\ t_n - \log\ (1-t_n) \geq 0$ for $n > n_0$. On the other hand, $\frac{x_n^2}{t_n(1-t_n)} + 2\log\ ((\Phi_n * G)(\frac{x_n}{t_n}))$ will often be *always* positive. Hence, for $n \leq n_0$, the Bayes test will *always* reject H_0 and never accept it. If G is the $N(0,\sigma^2)$ prior distribution, then this phenomenon occurs for all n such that $q > \frac{\sigma}{(1+\sigma)\sqrt{1-t_n}}$; for instance, if $\sigma = .5$, $p = .25$ and $t_n = \frac{n}{n+1}$, then the Bayes test will always reject H_0 for $n \leq 5$.

Corollary 3. The Bayes factor B and the likelihood ratio λ are related as

$$\frac{B}{\lambda} = \frac{\sqrt{t_n}}{\sqrt{2\pi}}/(\Phi_n * G)(\frac{x_n}{t_n}). \tag{15}$$

This is an obvious consequence of (13) and the expression for $-2\log\ \lambda$ given in Theorem 2.

Corollary 4. Under H_0, $\frac{B}{\lambda} \xrightarrow{P} \frac{1}{\sqrt{2\pi}(\Phi * G)(0)} > 1$, provided F is any CDF with sup (support (F)) $= 1$.

Proof: The condition on the support of F ensures $t_n \xrightarrow{\text{a.s.}} 1$. Also $(\Phi_n * G)(\frac{x_n}{t_n}) = \frac{\sqrt{t_n}}{\sqrt{2\pi}} \int e^{-\frac{t_n}{2}(\mu - \frac{x_n}{t_n})^2} dG(\mu)$. Now the function $g(t,y) = \int e^{-\frac{t}{2}(\mu-y)^2} dG(\mu)$ is bounded and jointly continuous in (t,y). Since $(t_n, \frac{x_n}{t_n}) \xrightarrow{P} (1,0)$ under H_0, it follows that $(\Phi_n * G)(\frac{x_n}{t_n}) \xrightarrow{P} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}\mu^2} dG(\mu) = (\Phi * G)(0)$. Corollary 3 now follows from (15).

Remark. Note that the limit in probability of $\frac{B}{\lambda}$ is *always* a number larger than 1. This is indicative of the generally true fact that in assessing if an observed Bayes factor is indeed large, one requires consideration of this matter that Bayes factors tend to be a factor of magnitude larger than the likelihood ratio in favor of the null. In fact, $\sup_G (p\lim \frac{B}{\lambda}) = \infty$ even for the family of normal priors, where $p\lim$ means limit in probability.

4.3. Posterior Probability of H_0

We will now give some values of the posterior probability of H_0, with corresponding values for the P value of the likelihood ratio test. Let us make this more precise. Expression (15) shows that the Bayes factor and hence the posterior probability of H_0 is not a function of the likelihood ratio λ alone. Thus the posterior probability of H_0 is not determined by only the P value; it also depends on the value of the nth time t_n. If the prior distribution G is symmetric about 0 (in the sense μ and $-\mu$ have the same prior distribution), the convolution $\Phi_n * G$ is also symmetric about 0. This gives the following easy representation for B in terms of λ and t_n:

$$B = \lambda \frac{\sqrt{t_n}}{\sqrt{2\pi}} / (\Phi_n * G)|\frac{x_n}{t_n}|$$
$$= \lambda \frac{\sqrt{t_n}}{\sqrt{2\pi}} / (\Phi_n * G)(\sqrt{(-2\log\ \lambda - \log\ (1-t_n))\frac{1-t_n}{t_n}}. \qquad (16)$$

(16) is immediate from (15) and Theorem 2 which says $-2\log\lambda = \frac{x_n^2}{t_n(1-t_n)} + \log\ (1-t_n)$. The P-value determines λ through (10), λ and t_n determine B through (16), and B determines the posterior probability of H_0 as $\frac{\frac{p}{q}B}{1+\frac{p}{q}B}$. In the following table, we use $t_n = \frac{n}{n+1}$, the mean of t_n if F were $U[0,1]$, and p, the prior probability of H_0, is taken as .5. We also need to specify F, as the P value determines λ via (10), which involves F. We take F to be $U[0,1]$, and Beta [2,2] and G is taken as $N(0,1)$.

Table 3

	P value = .01		P value = .05	
F	n	$P(H_0/\text{data})$	n	$P(H_0/\text{data})$
$U[0,1]$	10	.175011	10	.44297
	20	.216438	20	.51957
	30	.247486	30	.562306
	40	.272394	40	.595458
	50	.293271	50	.621113
Beta [2,2]	10	.13005	10	.360201
	20	.142777	20	.39316
	30	.152734	30	.414872
	40	.160753	40	.431066
	50	.167482	50	.443981

Discussion. At higher sample sizes, the posterior probability is less robust to the choice of F. This is especially true when the P value is .05; if $n = 50$, then the posterior probability of H_0 is .62 > .5 when F is $U[0,1]$ and .44 < .5 when F is Beta [2,2], making a difference between acceptance and rejection in the two cases. Notice also the usual "Lindley paradox" phenomenon: $P(H_0/\text{data})$ keeps increasing monotonically even though the P value of the classical test is quite small.

REFERENCES

Berg, H. C. (1993). Random Walks in Biology, Princeton University Press, Princeton.

Brown, L. (1971). Nonlocal asymptotic optimality of likelihood ratio tests, *Ann. Math. Stat.*, **42**, 1206–1240.

Johnson, B. M. and Killeen, T. (1983). An explicit formula for the CDF of the L_1 norm of the Brownian Bridge, *Ann. Prob.*, **11**, 807–808.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications, John Wiley, New York.

Robert, C. P. (1994). The Bayesian Choice, Springer-Verlag,New York.

Simons, G., Yao, Y-C and Wu, X. (1989). Sequential tests for the drift of a Wiener process with a smooth prior, and the heat equation, *Ann. Stat.*, **17**, 783–792.