Noname manuscript No.
(will be inserted by the editor)

Optimal Hypothesis Testing: From Semi to Fully Bayes Factors

Optimal Hypothesis Testing

Albert Vexler · Chengqing Wu · Kai Fun Yu

Received: date / Accepted: date

Abstract We propose and examine statistical test-strategies that are somewhat between the maximum likelihood ratio and Bayes factor methods that are well addressed in the literature. The paper shows an optimality of the proposed tests of hypothesis. We demonstrate that our approach can be easily applied to practical studies, because execution of the tests does not require deriving of asymptotical analytical solutions regarding the type I error. However, when the proposed method is utilized, the classical significance level of tests can be controlled.

Keywords Likelihood ratio · Maximum likelihood · Bayes factor · Most powerful · Hypotheses testing · Significance level · Type I error

1 Introduction

Testing statistical hypotheses probably has a much longer history than it appears. Our modest attempt in the search notes that Karl Pearson (1900) used significance testing for a simple multinomial hypothesis. Ronald A. Fisher had tremendous contributions in this area of hypothesis testing; see for example Fisher (1925). The statistical hypothesis testing as we know it was mainly due to the important work of Jerzy Neyman and Egon Pearson who laid down the foundation and the formulation. In a series of

Albert Vexler
Department of Biostatistics, The New York State University at Buffalo, Buffalo, NY 14214, USA
E-mail: avexler@buffalo.edu

Chengqing Wu
Department of Epidemiology and Public Health, Yale University, 60 College Street, New Haven, CT 06510, USA
E-mail: chengqing.wu@yale.edu

Kai Fun Yu*
Biostatistics and Bioinformatics Branch
Eunice Kennedy Shriver National Institute of Child Health and Human Development
Department of Health and Human Services
6100 Executive Blvd., Bethesda, MD 20892, USA
E-mail: yukf@mail.nih.gov

papers, starting perhaps in 1928, Neyman and Pearson clearly formulated the classical hypothesis testing problem and obtained very far reaching results which established this area as an important branch of statistical inference. Subsequently many other statistical giants keep making profound contributions, making this paradigm even more important in the statistical discipline. An excellent account can be found in Lehmann and Romano (2005).

It should also be noted that there has been a huge literature on the advocation and on the criticism of the Neyman Pearson formulation of statistical hypothesis testing. The present paper is not to add to the controversies by extolling the virtues or detracting from this classical foundation. We are here to note some difficulty in this classical formulation of statistical hypothesis testing, especially in the case when both the null hypothesis and the alternative hypothesis are composite. We propose an approach which bridges the classical case and the case of the fully Bayes factor. (Marden(2000) has summarized the Bayes approach in the context of hypothesis testing. Section 2 explains the Bayes factor in detail. Among others, the Type I error of tests based on Bayes Factor is difficulat to be controlled, as noted in Berger (1993).) In the following sections we shall give details of an approach we shall call the semi-Bayes approach, and indicate that in some special cases, it reduces to the classical Neyman Pearson statistical hypothesis testing and in other special cases, it becomes a fully Bayes factor approach. This can be viewed as a compromise between the two schools of thoughts. Some examples will be given. A Monte Carlo simulation study will supplement the theoretical results.

2 Stating the Problem

Let the following assumptions hold. A test decision is based on data $\{X_1, \ldots, X_n\}$ from the joint density function $f(x_1, \ldots, x_n)$ which is known to belong to a parametric class $\{f(x_1, \ldots, x_n|\mu, \eta),\ \mu \in \Omega_1 \subseteq \mathcal{R}^q,\ \eta \in \Omega_2 \subseteq \mathcal{R}^d\}$. We want to test the null hypothesis

$$\begin{aligned} H_0 :\ & (\mu, \eta) \in \Theta_0 = \{(\mu_0, \eta) :\ \eta \in \Omega_2\} \\ & \text{versus the alternative} \\ H_A :\ & (\mu, \eta) \in \Theta_A = \{(\mu, \eta) :\ \mu \neq \mu_0, \eta \in \Omega_2\} = \{\Omega_1 \times \Omega_2\} \backslash \Theta_0, \end{aligned}$$

where we shall assume that μ_0 is known and fixed.

The maximum likelihood method (or the likelihood ratio method) proposes the test statistic

$$\Lambda_n^{MLR} = \frac{\sup_{(\mu,\eta)\in\Theta_A} f(X_1, \ldots, X_n|\mu, \eta)}{\sup_{\eta\in\Omega_2} f(X_1, \ldots, X_n|\mu_0, \eta)}. \tag{1}$$

Another approach, which is based on considerations somewhat different from the maximum likelihood methodology, is the method of the *Bayes factor*; that is, the integrated likelihood ratio

$$\Lambda_n^{ILR} = \frac{\int f(X_1, \ldots, X_n|\mu, \eta) d\Psi_A(\mu, \eta)}{\int f(X_1, \ldots, X_n|\mu_0, \eta) d\Psi_0(\eta)} \tag{2}$$

is proposed to be the test statistic, where priors Ψ_0 and Ψ_A of the parameters conditional on it being in the null or alternative, respectively (e.g., Marden, 2000).

The objective of this paper is to propose and examine a class of likelihood ratio test statistics based on both the maximum likelihood estimation and Bayesian approach.

In order to present a likelihood ratio type test, we estimate the denominator of the ratio, whereas the H_A-likelihood, which is known up to parameters, is considered in accordance with the Bayes factor.

Let $\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta})$ be an estimator of the H_0-likelihood $f(X_1,\dots,X_n|\mu_0,\eta)$. Consider, for example, the penalized maximum likelihood estimator of η, say

$$\hat{\eta}(X_1,\dots,X_n) = \arg\max_a \phi(a) f(X_1,\dots,X_n|\mu_0,a), \tag{3}$$

where ϕ is a decreasing function of a penalty (e.g., $\phi(a)=1$, $\phi(a)=\exp(-a^2)$). Note that, from a Bayesian point of view, ϕ can be regarded as a proportion of a prior density of η (i.e. $\phi \propto d\Psi_0$, see Green, 1990). Thus, $\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta})$ can be presented in the forms

$$\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta}) = \phi\left(\hat{\eta}(X_1,\dots,X_n)\right) f\left(X_1,\dots,X_n|\mu_0,\hat{\eta}(X_1,\dots,X_n)\right), \tag{4}$$

$$\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta}) = \frac{1}{D}\phi\left(\hat{\eta}(X_1,\dots,X_n)\right) f\left(X_1,\dots,X_n|\mu_0,\hat{\eta}(X_1,\dots,X_n)\right), \tag{5}$$

where

$$D = \int \phi\left(\hat{\eta}(x_1,\dots,x_n)\right) f\left(x_1,\dots,x_n|\mu_0,\hat{\eta}(x_1,\dots,x_n)\right) \prod_{i=1}^{n} dx_i.$$

Obviously, if, under the null, η is known to be fixed η_0 then we assume that

$$\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta}) = f(X_1,\dots,X_n|\mu_0,\eta_0).$$

Finally, define the semi-Bayes test statistic

$$\Lambda_n^{SBLR} = \frac{\int f(X_1,\dots,X_n|\mu,\eta)d\Psi_A(\mu,\eta)}{\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta})}. \tag{6}$$

Notice that $\hat{f}(X_1,\dots,X_n|\mu_0,\hat{\eta})$ can be an arbitrary funtion of $X_1,\dots,X_n$. Thus, in contrast to the Λ_n^{ILR}-construction, we connect estimation of the H_0-likelihood function, appearing in Λ_n^{SBLR}, with observations and hence we relax dependence on the prior information under H_0. In the next section, we show that this approach leads to procedures with optimal properties and easily controlled levels of significance.

3 Optimality

Let $\delta \in [0,1]$ be a decision rule based on data points $\{X_1,\dots,X_n\}$, which rejects the null hypothesis with probability δ. Suppose that the probability measure and the associated expectation, given parameters (μ,η), are denoted by $Pr_{(\mu,\eta)}$ and $E_{(\mu,\eta)}$, respectively. We propose the following test:

$$\text{reject } H_0 \text{ iff } \quad \Lambda_n^{SBLR} > C, \tag{7}$$

where Λ_n^{SBLR} is denoted by (6) and C is a threshold.

Proposition 1 *For any δ with the fixed estimate of the significance level*

$$\hat{\alpha} = \int \delta(x_1, \ldots, x_n) \hat{f}\left(x_1, \ldots, x_n | \mu_0, \hat{\eta}\right) \prod_{i=1}^{n} dx_i, \tag{8}$$

the statistic $\Lambda_n^{SBLR}(X_1, \ldots, X_n)$ by (6) provides the integrated most powerful test with respect to a prior Ψ_A, i.e.

$$\int Pr_{(\mu,\eta)} \left\{ \Lambda_n^{SBLR} > C_{\hat{\alpha}} \right\} d\Psi_A(\mu, \eta) \geq \int Pr_{(\mu,\eta)} \left\{ \delta \text{ rejects } H_0 \right\} d\Psi_A(\mu, \eta),$$

where the threshold $C_{\hat{\alpha}}$ is chosen by

$$\int I \left\{ \Lambda_n^{SBLR}(x_1, \ldots, x_n) > C_{\hat{\alpha}} \right\} \hat{f}\left(x_1, \ldots, x_n | \mu_0, \hat{\eta}\right) \prod_{i=1}^{n} dx_i = \hat{\alpha}.$$

Proof. By virtue of the inequality: for all A, B and $\delta \in [0, 1]$

$$(A - B)\left(I\left\{A \geq B\right\} - \delta\right) \geq 0, \tag{9}$$

($I\{\cdot\}$ is the indicator function) with $A = \Lambda_n^{SBLR}$ and $B = C$, we have

$$\begin{aligned} &\left(\frac{\int f(X_1, \ldots, X_n | \mu, \eta) d\Psi_A(\mu, \eta)}{\hat{f}(X_1, \ldots, X_n | \mu_0, \hat{\eta})} - C \right) I \left\{ \Lambda_n^{SBLR} > C \right\} \\ \geq &\left(\frac{\int f(X_1, \ldots, X_n | \mu, \eta) d\Psi_A(\mu, \eta)}{\hat{f}(X_1, \ldots, X_n | \mu_0, \hat{\eta})} - C \right) \delta. \end{aligned}$$

And hence

$$\begin{aligned} &\left(\frac{\int f(X_1, \ldots, X_n | \mu, \eta) d\Psi_A(\mu, \eta)}{f(X_1, \ldots, X_n | \mu_0, \eta_0)} - C \frac{\hat{f}(X_1, \ldots, X_n | \mu_0, \hat{\eta})}{f(X_1, \ldots, X_n | \mu_0, \eta_0)} \right) I \left\{ \Lambda_n^{SBLR} > C \right\} \\ \geq &\left(\frac{\int f(X_1, \ldots, X_n | \mu, \eta) d\Psi_A(\mu, \eta)}{f(X_1, \ldots, X_n | \mu_0, \eta_0)} - C \frac{\hat{f}(X_1, \ldots, X_n | \mu_0, \hat{\eta})}{f(X_1, \ldots, X_n | \mu_0, \eta_0)} \right) \delta. \end{aligned} \tag{10}$$

Since

$$\begin{aligned} &E_{(\mu_0,\eta_0)} \frac{\int f(X_1, \ldots, X_n | \mu, \eta) d\Psi_A(\mu, \eta)}{f(X_1, \ldots, X_n | \mu_0, \eta_0)} \delta \\ &= \int \frac{\int f(x_1, \ldots, x_n | \mu, \eta) d\Psi_A(\mu, \eta)}{f(x_1, \ldots, x_n | \mu_0, \eta_0)} \delta(x_1, \ldots, x_n) f(x_1, \ldots, x_n | \mu_0, \eta_0) \prod_{i=1}^{n} dx_i \\ &= \int f(x_1, \ldots, x_n | \mu, \eta) d\Psi_A(\mu, \eta) \delta(x_1, \ldots, x_n) \prod_{i=1}^{n} dx_i \\ &= \int \left(E_{(\mu,\eta)} \delta \right) d\Psi_A(\mu, \eta), \end{aligned}$$

deriving the expectation $E_{(\mu_0,\eta_0)}$ of the inequality (10) with $C = C_{\hat{\alpha}}$, we obtain

$$\int Pr_{(\mu,\eta)} \left\{ \Lambda_n^{SBLR} > C_{\hat{\alpha}} \right\} d\Psi_A(\mu, \eta) - C\hat{\alpha} \geq \int Pr_{(\mu,\eta)} \left\{ \delta reject H_0 \right\} d\Psi_A(\mu, \eta) - C\hat{\alpha}.$$

That completes the proof of Proposition 1.

In the case, where under the simple null hypothesis $\eta = \eta_0$ is known and hence $\hat{f} = f(X_1, \ldots, X_n | \mu_0, \eta_0)$, $\hat{\alpha}$ is the classical definition of the significance level of tests. Note that, in several cases, when η is unknown and considered as a nuisance parameter, the observed data can be transformed to data $T = \{T_1(X_1, \ldots, X_n), \ldots, T_n(X_1, \ldots, X_n)\}$ such that the distribution function of T does not depend on η, under H_0 (e.g., Brown, *et al.*, 1975; Lehmann and Romano, 2005). In these situations, $\hat{\alpha}$ based on the known joint density of T is the type I error.

In accordance with Proposition 1, prior $\Psi_A(\mu, \eta)$ can be chosen with respect to a special area of parameters μ and η, under H_A, where the maximum of the test's power is desired.

When parameters under the null are unknown, the well accepted approach for comparing tests is based on contrasts of the power of decision rules δ that satisfy the following condition: for a given $\alpha \in (0, 1)$

$$\sup_{\eta \in \Omega_2} Pr_{(\mu_0, \eta)} \{\delta \text{ rejects } H_0\} \leq \alpha. \tag{11}$$

Commonly, the monitoring (11) of the type I error is stipulated on a decision rule δ to be a test (e.g., Lehmann and Romano, 2005). To be consistent with requirement (11), we present the next result.

Proposition 2 *Assume that $\hat{f}(X_1, \ldots, X_n | \mu_0, \hat{\eta})$ in (6) is defined by (4) (or by (5)) and limit the set of decision rules to*

$$\left\{ \delta : \quad \textit{for some } \eta_\delta \sup_{\eta \in \Omega_2} Pr_{(\mu_0, \eta)} \{\delta \textit{ rejects } H_0\} = Pr_{(\mu_0, \eta_\delta)} \{\delta \textit{ rejects } H_0\} \right\}.$$

Then

$$\hat{\alpha} \geq \phi(\eta_\delta) \sup_{\eta \in \Omega_2} Pr_{(\mu_0, \eta)} \{\delta \textit{ rejects } H_0\}$$
$$\left(\textit{or} \quad 1 \geq \hat{\alpha} \geq \frac{\phi(\eta_\delta)}{D} \sup_{\eta \in \Omega_2} Pr_{(\mu_0, \eta)} \{\delta \textit{ rejects } H_0\}, \ \textit{if (5) is applied}\right),$$

where $\hat{\alpha}$ by (8) is fixed for tests considered in Proposition 1.

Proof. The proof directly follows from the definition (8) with (3).

In accordance with Proposition 2, a fixed value of $\hat{\alpha}$ can control the type I error of tests.

The proposed test is based on the likelihood ratio technique. It is widely known in the statistical literature that such tests have high power, therefore, evaluation of their significance level is a major issue. Most results dealing with the significance level of the generalized maximum likelihood ratio tests (even in the simplest cases of independent identically distributed observations) are complex and asymptotic ($n \to \infty$) with special conditions on the distribution function of $X_1, \ldots, X_n$. The test (7) has a guaranteed, non asymptotic and distribution free upper bound for the estimated significance level $\hat{\alpha}$. In addition, allowing for Proposition 2, one can request as small as we want, with regard to probabilities that $\hat{\alpha} > 0$. To this end, we present the next proposition.

Proposition 3 *The estimated significance level of the test (7) satisfies the inequality*

$$\hat{\alpha} = \int I\left\{\Lambda_n^{SBLR}(x_1,\ldots,x_n) > C_{\hat{\alpha}}\right\} \hat{f}\left(x_1,\ldots,x_n|\mu_0,\hat{\eta}\right) \prod_{i=1}^{n} dx_i$$
$$\leq \frac{1}{C_{\hat{\alpha}}} \int Pr_{H_A(\mu,\eta)}\left\{\Lambda_n^{SBLR}(X_1,\ldots,X_n) > C_{\hat{\alpha}}\right\} d\Psi_A(\mu,\eta) \ \leq \ \frac{1}{C_{\hat{\alpha}}}.$$

Proof. Since by the definition (6), the event $\left\{\Lambda_n^{SBLR}(x_1,\ldots,x_n) > C_{\hat{\alpha}}\right\}$ is equivalent to

$$\left\{\hat{f}(x_1,\ldots,x_n|\mu_0,\hat{\eta}) \ < \ \frac{1}{C_{\hat{\alpha}}} \int f(x_1,\ldots,x_n|\mu,\eta) d\Psi_A(\mu,\eta)\right\},$$

we have

$$\hat{\alpha} \leq \frac{1}{C_{\hat{\alpha}}} \int I\left\{\Lambda_n^{SBLR}(x_1,\ldots,x_n) > C_{\hat{\alpha}}\right\} \int f(x_1,\ldots,x_n|\mu,\eta) d\Psi_A(\mu,\eta) \prod_{i=1}^{n} dx_i$$
$$= \frac{1}{C_{\hat{\alpha}}} \int\int I\left\{\Lambda_n^{SBLR}(x_1,\ldots,x_n) > C_{\hat{\alpha}}\right\} f(x_1,\ldots,x_n|\mu,\eta) \prod_{i=1}^{n} dx_i \, d\Psi_A(\mu,\eta)$$
$$= \frac{1}{C_{\hat{\alpha}}} \int Pr_{H_A(\mu,\eta)}\left\{\Lambda_n^{SBLR}(X_1,\ldots,X_n) > C_{\hat{\alpha}}\right\} d\Psi_A(\mu,\eta)$$
$$\leq \frac{1}{C_{\hat{\alpha}}} \int\int f(x_1,\ldots,x_n|\mu,\eta) \prod_{i=1}^{n} dx_i \, d\Psi_A(\mu,\eta) \ = \ \frac{1}{C_{\hat{\alpha}}}.$$

The proof of Proposition 3 is complete.

That is, we have the upper bound (that does not involve n and is independent of different conditions on the distribution of $X_1,\ldots,X_n$) for the estimated significance level of test (7): $\hat{\alpha} \leq 1/C$. Thus, selecting $C = 1/\hat{\alpha}$ determines a test with an estimated level of significance that does not exceed $\hat{\alpha}$. Propositions 2 and 3 ensure a p-value of the test. In accordance with the inequality $\hat{\alpha} \leq 1/C$, theoretically, values of $\hat{\alpha}$ can be chosen as small as desired.

One can define the test statistic (6) by

$$\hat{f}(X_1,\ldots,X_n|\mu_0,\hat{\eta}) = \int f(X_1,\ldots,X_n|\mu_0,\eta) d\Psi_0(\eta).$$

In this case, we have $\Lambda_n^{SBLR} = \Lambda_n^{ILR}$, where Λ_n^{ILR} by (2). And hence the proposed method is equivalent to the Bayes factor. Denote the Bayesian significance level of a decision rule δ in the form of

$$\overline{\alpha} = \int \delta(x_1,\ldots,x_n) \int f(X_1,\ldots,X_n|\mu_0,\eta) d\Psi_0(\eta) \prod_{i=1}^{n} dx_i \tag{12}$$
$$= \int Pr_{(\mu_0,\eta)}\{\delta \text{ rejects } H_0\} d\Psi_0(\eta)$$
$$\left(\overline{\alpha} = \hat{\alpha}, \text{ when } \hat{f} = \int f(X_1,\ldots,X_n|\mu_0,\eta) d\Psi_0(\eta)\right).$$

The next result is a simple corollary of Proposition 1

Corollary 1 *For any δ with the fixed Bayesian significance level $\overline{\alpha}$ the statistic Λ_n^{ILR} provides the integrated most powerful test with respect to a prior Ψ_A, i.e.*

$$\int Pr_{(\mu,\eta)}\left\{\Lambda_n^{ILR} > C_{\overline{\alpha}}\right\} d\Psi_A(\mu,\eta) \geq \int Pr_{(\mu,\eta)}\left\{\delta \text{ rejects } H_0\right\} d\Psi_A(\mu,\eta),$$

where the threshold $C_{\overline{\alpha}}$ is chosen by

$$\int Pr_{(\mu_0,\eta)}\{\Lambda_n^{ILR} > C_{\overline{\alpha}}\} d\Psi_0(\eta) = \overline{\alpha}.$$

Remark 1. When η_0 is unknown, comparing decision rules from the set $\{\delta : \sup_{\eta\in\Omega_2} Pr_{(\mu_0,\eta)}(\delta \text{ rejects } H_0) \leq \alpha\}$ (where α is fixed) is a very complex problem. Suppose we have two tests, (T_1) and (T_2), for hypothesis H_0 versus H_A, given by

$$(T_1): \text{ reject } H_0 \text{ iff statistic } S_1 > C_1; \text{ and } (T_2): \text{ reject } H_0 \text{ iff statistic } S_2 > C_2,$$

where statistics S_1, S_2 are based on a sample $\{Z_i,\, i = 1,\ldots,m \geq 1\}$, and C_1, C_2 are both thresholds. We would like to compare (T_1) with (T_2), when the parameters under H_0 and H_A are unknown. The suggestion of fixing the type I errors of (T_1) and (T_2) as α and then contrasting the powers of these tests is problematic. First, in general, we cannot easily choose $C_{1\alpha}, C_{2\alpha}$ such that

$$\sup_\eta Pr_{(\mu_0,\eta)}\{S_1 > C_{1\alpha}\} \leq \alpha \quad \text{and}$$
$$\sup_\eta Pr_{(\mu_0,\eta)}\{S_2 > C_{2\alpha}\} \leq \alpha.$$

Since Monte Carlo evaluations of $\sup_\eta Pr_{(\mu_0,\eta)}$ is complex and biased, analytical presentations of $Pr_{(\mu_0,\eta)}\{L > C\}$ and $Pr_{(\mu_0,\eta)}\{D > C\}$ have to be proposed in order to derive $C_{1\alpha}$ and $C_{2\alpha}$. Second, assuming that $C_{1\alpha}$ and $C_{2\alpha}$ are known or evaluated, then comparing $Pr_{(\mu,\eta)}\{S_1 > C_{1\alpha}\}$ with $Pr_{(\mu,\eta)}\{S_2 > C_{2\alpha}\}$ is an arduous task. Alternatively, we suggest fixing an H_0-likelihood's estimator (say, $\hat{f}_{H_0}(Z_1,\ldots,Z_m)$) as

$$\hat{P}r_{H_0}\left\{H_0 \text{ is rejected}\right\} := \int \hat{f}_{H_0}(z_1,\ldots,z_m) I\{H_0 \text{ is rejected}\} dz_1\cdots dz_m = \hat{\alpha},$$

and then evaluating integrated powers of *(A)* and *(B)*. It is clear that for thresholds $C_{1\hat{\alpha}}$, $C_{2\hat{\alpha}}$:

$$\hat{P}r_{H_0}\{L > C_{1\hat{\alpha}}\} := \int \hat{f}_{H_0}(z_1,\ldots,z_m) I\{L(z_1,\ldots,z_m) > C_{1\hat{\alpha}}\} dz_1\cdots dz_m = \hat{\alpha} \text{ and}$$
$$\hat{P}r_{H_0}\{D > C_{2\hat{\alpha}}\} := \int \hat{f}_{H_0}(z_1,\ldots,z_m) I\{D(z_1,\ldots,z_m) > C_{2\hat{\alpha}}\} dz_1\cdots dz_m = \hat{\alpha},$$

can be easily obtained by Monte Carlo methods. (We consider calculation of $\hat{\alpha}$ in Section 3.1 based on an example.)

Remark 2. Asymptotic approximation. To apply the proposed tests, asymptotic presentation of $Pr_{H_0}\{H_0$ is rejected$\}$ is not of crucial necessity. However, we can demonstrate its behavior under the following scenario. Assume that $X_1, \ldots, X_n$ are independent identically distributed(i.i.d.) random variables. Without loss of generality, let the function ϕ in (4) be equal to 1. O'Hagan(1995) has shown that, under standard regularity conditions (for details see Gelfand and Dey, 1994, O'Hagan, 1995 as well as Kass and Wasserman, 1995), the numerator of (6) can be asymptotically expressed as

$$\psi_A(\hat{\mu}_{MLE}, \hat{\eta}_{MLE}) L_2 n^{-1} 2\pi |V_n|^{\frac{1}{2}}$$

where $L_2 = \max_{\mu,\eta} \prod_{j=1}^{n} f(X_i|\mu,\eta)$ is the maximized likelihood, $-nV_n^{-1}$ is the Hessian matrix of $logf$ at the maximum likelihood estimators $(\hat{\mu}_{MLE}, \hat{\eta}_{MLE})$ $(V_n \to V$ as $n \to \infty$, V is a constant matrix), and $\psi_A(\mu,\eta)d\mu d\eta = d\Psi_A(\mu,\eta)$. Then

$$\Lambda = \frac{\int f(X_1, \ldots, X_n|\mu,\eta) d\Psi_A(\mu,\eta)}{\hat{f}(X_1, \ldots, X_n|\mu_0, \hat{\eta}_{MLE\ H_0})} \approx \frac{L_2}{L_1} \psi_A(\hat{\mu}_{MLE}, \hat{\eta}_{MLE}) n^{-1} 2\pi |V|^{\frac{1}{2}},$$

where $L_1 = f(X_1, \ldots, X_n|\mu_0, \hat{\eta}_{MLE\ H_0})$. Under H_0,

$$2log\Lambda \approx 2log\frac{L_2}{L_1} + 2\log \psi_A(\hat{\mu}_{MLE}, \hat{\eta}_{MLE}) n^{-1} 2\pi |V|^{\frac{1}{2}}$$

which is asymptotically distributed as a χ_1^2 plus $2\log(\psi_A(\mu,\eta) n^{-1} 2\pi |V|^{\frac{1}{2}})$.

3.1 Example: Test for Autoregression.

Consider the autoregressive process AR(1): $X_0 = 0$, $X_i = \mu X_{i-1} + \epsilon_i$, $i = 1, \cdots, n$, where $\{\epsilon_i\}_{i=1}^n$ are independent identically normally distributed random variables with unknown mean η and unit standard deviation. To test the baseline $H_0 : \mu = 0$ versus alternative $H_A : \mu \neq 0$, we apply the proposed method. The test statistic has the form

$$\Lambda_n^{SBLR} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \phi(X_i - \mu X_{i-1} - \eta) d\Psi_A(\mu,\eta)}{\exp(-b\hat{\eta}^2) \prod_{i=1}^{n} \phi(X_i - \hat{\eta})} , \qquad (13)$$

where $\phi()$ and $\Phi()$ are the standard normal density and distribution functions, respectively, and the penalized maximum likelihood estimator of η is

$$\hat{\eta}(X_1, \ldots, X_n) = \arg\max_a e^{-ba^2} \prod_{i=1}^{n} \phi(X_i - a), \quad b > 0 \quad \left(\text{i.e.,} \ \ \hat{\eta} = \frac{\sum_{i=1}^{n} X_i/n}{1 + 2b/n}\right).$$

Note that, if, under H_A, $|\mu| \geq 1$ then H_A corresponds to the non stationary AR(1), and hence we can simply detect H_A. In order to denote a prior Ψ_A, we utilize arguments mentioned in Krieger *et al.* (2003). That is, for some $|\mu_0| < 1, \sigma_\mu > 0, \eta_0$ and $\sigma_\eta > 0$, one can choose the prior

$$\Psi_A(\mu,\eta) = \frac{1}{4} \left\{ \Phi\left(\frac{\mu - \mu_0}{\sigma_\mu^2}\right) + \Phi\left(\frac{\mu + \mu_0}{\sigma_\mu^2}\right) \right\} \left\{ \Phi\left(\frac{\eta - \eta_0}{\sigma_\eta^2}\right) + \Phi\left(\frac{\eta + \eta_0}{\sigma_\eta^2}\right) \right\} ,$$

which simplifies somewhat if, for example, $\mu_0 = 1/2, \sigma_\mu = 1, \eta_0 = 0$ and $\sigma_\eta = 1$. This prior is a commonly used conjugate prior in the context of Bayes Factors (e.g., Aitkin, 1991). In this case, since

$$\int_{-\infty}^{\infty} \exp(tu)\phi(u/v)/\sqrt{v}du = \exp\left(\frac{1}{2}vt^2\right),$$

the numerator of Λ_n^{SBLR} is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{i=1}^{n}\phi(X_i - \mu X_{i-1} - \eta)d\Psi_A(\mu,\eta) \tag{14}$$
$$= \frac{1}{4}\left[A(\mu_0,\eta_0)\exp\left\{\frac{1}{2}B^2(\mu_0,\eta_0)F\right\} + A(\mu_0,-\eta_0)\exp\left\{\frac{1}{2}B^2(\mu_0,-\eta_0)F\right\}\right.$$
$$\left. + A(-\mu_0,\eta_0)\exp\left\{\frac{1}{2}B^2(-\mu_0,\eta_0)F\right\} + A(-\mu_0,-\eta_0)\exp\left\{\frac{1}{2}B^2(-\mu_0,-\eta_0)F\right\}\right],$$

where

$$A(\mu_0,\eta_0) = a\exp\left\{-\frac{\mu_0^2}{2\sigma_\mu^2} - \frac{\sum_{i=1}^n (X_i-\eta_0)^2}{2(1+n\sigma_\eta^2)} - \frac{\sigma_\eta^2(n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)}{2(1+n\sigma_\eta^2)}\right\},$$
$$B(\mu_0,\eta_0) = \frac{\mu_0}{\sigma_\mu^2} + \frac{\sum_{i=1}^n X_{i-1}(X_i-\eta_0)}{1+n\sigma_\eta^2} + \frac{\sigma_\eta^2(n\sum_{i=1}^n X_iX_{i-1} - \sum_{i=1}^n X_i\sum_{i=1}^n X_{i-1})}{(n\sigma_\eta^2+1)},$$
$$F = \left(\frac{1}{\sigma_\mu^2} + \frac{\sum_{i=1}^n X_{i-1}^2}{1+n\sigma_\eta^2} + \frac{\sigma_\eta^2(n\sum_{i=1}^n X_{i-1}^2 - (\sum_{i=1}^n X_{i-1})^2)}{(1+n\sigma_\eta^2)}\right)^{-1},$$
$$a = (2\pi)^{-n/2}F^{1/2}\sigma_\mu^{-1}(1^2+n\sigma_\eta^2)^{-1/2}.$$

Following Proposition 1, the statistic (13) is the most powerful with respect to Ψ_A via tests with fixed $\hat{\alpha}$, where

$$\hat{\alpha} = \frac{\int \exp(-b\hat{\eta}^2(x_1,\cdots,x_n))\prod_{i=1}^n \phi(x_i - \hat{\eta}(x_1,\cdots,x_n))I\left\{\Lambda_n^{SBLR} > C\right\}dx_1\cdots dx_n}{\int \exp(-b\hat{\eta}^2(x_1,\cdots,x_n))\prod_{i=1}^n \phi(x_i - \hat{\eta}(x_1,\cdots,x_n))dx_1\cdots dx_n}.$$

Note that, in the context of Proposition 2, b can be selected in order to minimize the distance between $\hat{\alpha}$ and $Pr_{H_0(\eta_s)}\{\Lambda_n^{SBLR} > C\} = \sup_\eta Pr_{H_0(\eta)}\{\Lambda_n^{SBLR} > C\}$. To this end, b can be chosen to maximize

$$\frac{e^{-b\eta_s^2}}{\int \exp(-b\hat{\eta}^2(x_1,\cdots,x_n))\prod_{i=1}^n \phi(x_i - \hat{\eta}(x_1,\cdots,x_n))dx_1\cdots dx_n}, \tag{15}$$

because by the definitions of $\hat{\alpha}$ and $\hat{\eta}$ we have

$$\hat{\alpha} \geq \frac{e^{-b\eta_s^2} \quad Pr_{H_0(\eta_s)}\{\Lambda_n^{SBLR} > C\}}{\int \exp(-b\hat{\eta}^2(x_1,\cdots,x_n))\prod_{i=1}^n \phi(x_i - \hat{\eta}(x_1,\cdots,x_n))dx_1\cdots dx_n}.$$

One can show that

$$b = \frac{\left(n^2 + 4n/\eta_s^2\right)^{1/2} - n}{4} \sim \frac{1}{2\eta_s^2} \text{ as } n \to \infty$$

maximizes (15).

3.1.1 Monte Carlo Study.

To examine the performance of the proposed test (13), we conduct the following Monte Carlo study. Here we compare the proposed semi-Bayesian test with the maximum likelihood ratio test with $\hat{\alpha}$ in (8) is fixed to be 0.05. Define $\mu_0 = \eta_0 = 0.5$, $\sigma_\mu = \sigma_\eta = 1$ and $b = (n)^{1/2}$ in the numerator (14) and denominator of the test statistic (13), respectively. The operating characteristic $\hat{\alpha}$ of tests corresponding to Section 3.1 can be rewritten as

$$\hat{\alpha} = \left(\frac{2b}{n+2b}\right)^{1/2} \int \exp\left\{\frac{(\sum_{i=1}^n x_i)^2}{2(n+2b)}\right\} \delta(x_1, \cdots, x_n) \prod_{i=1}^n \phi(x_i) dx_1 \cdots dx_n$$
$$= \left(\frac{2b}{n+2b}\right)^{1/2} E \exp\left\{\frac{(\sum_{i=1}^n X_i)^2}{2(n+2b)}\right\} \delta(X_1, \cdots, X_n)$$

where $X_i \sim N(0,1)$ are i.i.d. $(i = 1, \ldots, n)$. This equation and $100,000$ repetitions of samples $\{X_i \sim N(0,1)\}_{i=1}^n$ allowed to calculate the test-thresholds C_{SBLR} and C_{MLR} that are related to $\hat{\alpha} = 0.05$. Table 1 presents these values of C_{SBLR} and C_{MLR}.

Table 1 here

Note that, Proposition 1 shows that the proposed test is the integrated most powerful test with respect to Ψ_A. In this simulation study, we evaluate the power of the tests for fixed η and μ. The Monte Carlo powers of the maximum likelihood ratio test and test (13) were derived via $10,000$ repetitions of samples from each set of parameters (η, μ) and the sample size n. (Note that for the Monte Carlo Power P we can assume for this simulation CI $= P \pm 1.96(P(1-P)/10,000)^{1/2}$.) In accordance with Table 1, while we planned to obtain a test powerful around $(\eta = \pm 0.5, \mu = \pm 0.5)$, the average power (with no respect to Ψ_A) of the proposed test was about 1.35 times better than that of the MLR test, as well as in the considered cases excepting $(\mu, \eta) = (0.1, 0)$ the semi-Bayesian test was also more powerful than the maximum likelihood ratio test was. However, the distance between the powers of the tests asymptotically vanished when $(\mu = 0.1, \eta = 0)$ and the sample size n increases in Table 1.

Since, in the context of the considered example, any reasonable test for $\mu = 0$ vs $|\mu| \geq 1$ is expected to provide high levels of the power and approximated bounds for η can be easily evaluated basing on observed data, the prior Ψ_A can be suggested to be applied. (Our broad Monte Carlo investigation (particularly displayed in Table 1) showed that even when $\eta = 2, 3$ the proposed test could be recommended instead of the MLR test.) Moreover, consideration of μ and η, belonging to

$$\Psi_A(\mu, \eta) = \pi(\mu, \eta) \equiv \frac{I\{\mu \in [-0.9, 0.9]\}}{1.8} \frac{I\{\eta \in [-1, 1]\}}{2},$$

obviously has an independent interest in the terms of the most powerful testing (see Proposition 1), when the alternative parameter assumed to be from $\pi(\mu, \eta)$.

Table 2 here

In comparing with Table 1, Table 2 displays the Monte Carlo powers of the MLR test and test (13) with $\Psi_A(\mu, \eta) = \pi(\mu, \eta)$ based on $10,000$ repetitions of samples from each set of parameters (μ, η) and the sample size n. Although Table 2 corroborates that even when η is not from $\Psi_A(\mu, \eta)$ we can recommend the test (13), we need to indicate when the alternative $\mu = 0.1$ and $n = 25$ the MLR test is slightly superior the proposed test (Table 1 also demonstrated a similar result corresponding to $(\mu, \eta) = (0.1, 0)$).

4 Conclusion

In the general statements of parametric hypothesis testing, the semi-Bayes approach was developed and investigated. We have found that when the test statistic is the likelihood ratio that is supported by both estimation and the Bayes presentation of the likelihood functions under the null and alternative hypotheses, respectively, the test has the following properties.

The proposed test is the integrated most powerful test with respect to a prior distribution on the unknown parameters of the alternative. In particular, this prior can be chosen in accordance with areas where investigators wish to reach the maximum power of the test. The semi-Bayes test can be applied to a real study without analytical presentations of the type I error probability. Generally, the maximum likelihood ratio and the pure Bayes factor tests require evaluating analytical forms of the type I error probability, because the test thresholds have to be fixed. To use the semi-Bayes method, a prior distribution of parameters, under the null, is not necessary.

We proposed tests and investigated their operating characteristics in order that these tests can be easily applied in practice; the proposed tests provide maximum integrated power in the area that can correspond to the tester's interests - it is different from the Bayesian point of view.

Acknowledgements

The authors are grateful to the editor and referee for their helpful comments that clearly improved this paper. This research was supported by the Intramural Research Program of the NIH.

References

Aitkin M (1991) Posterior Bayes Factors. JR Statist Soc B 53:111-142

Berger JO (1993) Statistical Decision Theory and Bayesian Analysis. Second Edition. Springer Series in Statistics, New York

Brown RL, Durbin J, Evans JM (1975) Techniques for testing the constancy of regression relationships over time (with discussion). JR Statist Soc B 37:149-192

Fisher RA (1925) Statistical Methods for Research Workers. First Edition. Edinburgh, Oliver and Boyd

Gelfand AE, Dey DK (1994) Bayesian Model Choice: Asymptotics and Exact Calculations. JR Statist Soc B 56:501-514

Green PJ (1990) On Use of the EM for Penalized Likelihood Estimation. JR Statist Soc B 52:443-452

Kass RE, Wasserman L (1995) A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion. J Am Statist Assoc 90:928-934

Krieger AM, Pollak M, Yakir B (2003) Surveillance of a simple linear regression. J Am Statist Assoc 98:456-469

Lehmann EL, Romano JP (2005) Testing Statistical Hypotheses. Springer, New York

Marden JI (2000) Hypothesis Testing: From p Values to Bayes Factors. J Am Statist Assoc 95:1316-1320

Neyman J, Pearson E (1928) On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika 20A:175-240

Neyman J, Pearson E (1928) On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. Biometrika 20A:263-294

Neyman J, Pearson E (1933) On the testing of statistical hypotheses in relation to probability a priori. Proc Cambridge Phil Soc 29:492-510

Neyman J, Pearson E (1933) On the problem of the most efficient tests of statistical hypotheses. Phil Trans Roy Soc A 231:289-337

Neyman J, Pearson E (1936) Contributions to the theory of testing statistical hypotheses. I. Unbiased critical region of type A and type A. Stat Res Mem 1:1-37

Neyman J, Pearson E (1936) Sufficient statistics and uniformly most powerful tests of statistical hypotheses. Stat Res Mem 1:113-137

Neyman J, Pearson E (1938) Contributions to the theory of testing statistical hypotheses. II. Stat Res Mem 2:25-57

O'Hagan A (1995) Fractional Bayes Factors for Model Comparison. JR Statist Soc B 57:99-138

Pearson K (1900) On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil Mag Ser 5 50:157-172

Table 1 Monte Carlo comparing for the Powers of the tests.

		$n = 25$, $C_{MLR} = 5.99$, $C_{SBLR} = 0.25$		$n = 50$, $C_{MLR} = 6.30$, $C_{SBLR} = 0.12$		$n = 75$, $C_{MLR} = 6.70$, $C_{SBLR} = 0.08$	
μ	η	$Power_{MLR}$	$Power_{SBLR}$	$Power_{MLR}$	$Power_{SBLR}$	$Power_{MLR}$	$Power_{SBLR}$
0	0	0.0456*	0.0336*	0.0492*	0.0408*	0.0481*	0.0428*
0.1	0	0.0519	0.0410	0.0869	0.0765	0.1107	0.1028
0.3	0	0.2365	0.2232	0.4914	0.4880	0.6809	0.6848
0.7	0	0.8312	0.8790	0.9928	0.9961	0.9999	0.9999
0.1	0.25	0.0531	0.0721	0.0837	0.1420	0.109	0.1956
0.3	0.25	0.2351	0.3223	0.4939	0.6337	0.6741	0.8131
0.7	0.25	0.8430	0.9296	0.9941	0.9990	0.9997	1
0.1	0.5	0.0515	0.2647	0.0918	0.5309	0.1125	0.7158
0.3	0.5	0.2365	0.6263	0.4935	0.9202	0.6764	0.9834
0.7	0.5	0.8695	0.9886	0.9954	1	0.9999	1
0.1	1	0.0588	0.2443	0.0848	0.5993	0.1165	0.9875
0.3	1	0.2529	0.6194	0.5047	0.9498	0.6948	1
0.7	1	0.9471	1	0.9980	1	1	1
Average**		0.3890	0.5175	0.5259	0.6946	0.5978	0.7902

* indicates the Monte Carlo Type I Error of corresponding test provided that η is known to be 0 under H_0;
** the case (0,0) is not included.

Table 2 Monte Carlo comparing for the Powers of the tests.

		$n = 25$, $C_{SBLR} = 0.650$		$n = 50$, $C_{SBLR} = 0.295$		$n = 75$, $C_{SBLR} = 0.185$	
μ	η	$Power_{MLR}$	$Power_{SBLR}$	$Power_{MLR}$	$Power_{SBLR}$	$Power_{MLR}$	$Power_{SBLR}$
0	0	0.0523	0.0341	0.0495	0.0412	0.0548	0.0471
0	3	0.0535	0.0419	0.0496	0.0425	0.0548	0.0427
0.1	0	0.0577	0.0484	0.0860	0.0754	0.1194	0.1098
0.1	1.5	0.0622	0.5999	0.0871	0.8409	0.1248	0.9991
0.1	3	0.0711	0.7041	0.0916	0.8971	0.1303	0.9992
0.5	0.5	0.5897	0.9072	0.9041	0.9973	0.9834	1
0.5	2	0.8145	0.8319	0.9634	0.9838	0.9941	0.9995
0.5	3	0.9464	0.9597	0.9828	0.9839	0.9988	0.9989
Average**		0.4236	0.6752	0.5192	0.7964	0.5585	0.8511

** the cases (0,0) and (0,3) are not included.