Sankhyā : The Indian Journal of Statistics
2001, Volume 63, Series B, Pt. 3, p xx-xx

UNIFIED BAYESIAN AND CONDITIONAL FREQUENTIST TESTING FOR DISCRETE DISTRIBUTIONS

By SARAT C. DASS
Michigan State University

SUMMARY. Testing of hypotheses for discrete distributions is considered in this paper. The goal is to develop conditional frequentist tests that allow the reporting of data-dependent error probabilities such that the error probabilities have a strict frequentist interpretation and also reflect the actual amount of evidence in the observed data. The resulting randomized tests are also seen to be Bayesian tests, in the strong sense that the reported error probabilities are also the posterior probabilities of the hypotheses. The new procedure is illustrated for a variety of testing situations, both simple and composite, involving discrete distributions. Testing linkage heterogeneity with the new procedure is given as an illustrative example.

1. **Introduction**

In hypotheses testing situations where the underlying distributions are discrete, a new test is proposed which can be interpreted from both the conditional frequentist and Bayesian viewpoints. We call such tests "unified".

It is desirable for a testing procedure to report error probabilities that reflect the confidence with which a decision (either rejecting or accepting the null hypothesis) is made based on the observed data. The classical frequentist approach to testing constructs acceptance and rejection regions and reports associated error probabilities of Type I and Type II. However, these error probabilities are *unconditional*, in the sense that they depend only on whether the data is in the rejection or acceptance region, and not on the evidentiary strength of the observed data. Thus, if X follows a binomial distribution with $n = 50$ trials and unknown proportion of success, θ, and

---

Paper received August 2001; revised December 2001.
AMS (1991) *subject classifications.*
Key words and phrases. Discrete distributions, randomization, conditional error probabilities, Bayes factors, posterior probabilities, genetic linkage, recombinations.

it is desired to test $H_0 : \theta = 0.5$ versus $H_1 : \theta \neq 0.5$ at level $\alpha = 0.05$, one reports the error probability of 0.05 whether the observed count is $X = 33$ or $X = 50$, even though the latter is a 'stronger' rejection of H_0. As a solution, p-values are introduced as data-dependent measures of the strength of evidence against H_0. However, p-values do not exactly solve the frequentist problem; shortcomings of p-values as data-dependent measures are discussed in detail in Sellke, Bayarri and Berger (2000) and also in earlier references such as Edwards, Lindman and Savage (1963), Berger and Selke (1987), and Berger and Delampady (1987).

The conditional frequentist approach, formalized by Kiefer (1975,1976, 1977) and Brownie and Kiefer (1977), can be used to obtain data-dependent error probabilities which have a proper frequentist interpretation. A statistic measuring 'strength of evidence' in the data, for or against H_0, is found and the Type I and Type II error probabilities are reported conditional on this statistic. The main difficulty here is to find an appropriate conditioning statistic; while constructing conditioning statistics for certain simple testing problems is easy, finding suitable conditioning statistics for general testing problems can be very difficult.

Reported error probabilities that vary with the observed data and reflect evidentiary strength arise naturally in the Bayesian setting. Recently, Berger, Brown and Wolpert (1994) and Wolpert (1996) showed that a conditioning statistic, $S(X)$, that reflects the evidentiary strength in the data can be found in the case of testing simple hypotheses. This leads to a conditional frequentist test that is very easy to implement. Surprisingly, they observed that the ensuing conditional frequentist Type I and Type II error probabilities coincide exactly with the Bayesian posterior probabilities of H_0 and H_1, respectively. Therefore, a frequentist and a Bayesian using this test will not only reach the same decision (rejecting or accepting the null) after observing the data, but will also report the same values for the error probabilities. In this sense, the proposed test represents a unified testing procedure. Berger, Boukai and Wang (1997) generalized this to testing a simple null hypothesis versus a composite alternative. In this testing scenario, the resulting unified test reports conditional Type I error probability that is exactly the same as the posterior probability of the null hypothesis. Berger, Boukai and Wang (1997) also show that the posterior probability of the alternative hypothesis is a weighted average of the conditional Type II error probabilities. A further generalization of the unification for simple versus composite hypotheses using non-informative priors is reported in Dass and Berger (1998). The sequential version of this problem was considered in Berger, Boukai and Wang (1999).

All unified testing procedures that were obtained assumes that the underlying distributions under H_0 and H_1 are continuous and admit densities. However, for discrete distributions, densities do not exist with respect to the Lebesgue measure. We resolve this problem by introducing a randomization independent of the discrete observations. So, essentially, the test proposed is a randomized test. Our proposed (randomized) test has several attractive properties: the error probabilities reported reflect the 'strength of evidence' of the observed data, the reported error probabilities are simulteneously conditional frequentist error probabilities as well as Bayesian posterior probabilities of the hypotheses, and furthermore, for the most part of the decision space, the randomized test decides between accepting or rejecting the null and reports conditional error probabilities *independent* of the randomization.

The rest of the paper is presented as follows. Section 2 discusses simple versus simple hypothesis testing under strict monotone likelihood ratio (MLR) property. Section 3 discusses the general methodology of discrete testing which includes simple versus composite and composite versus composite testing situations. In testing situations involving composite hypotheses, we show that one is able to choose appropriate priors reflecting one's belief while still maintaining a frequentist interpretation of the testing procedure. We illustrate the new methodology for the testing of presence of genetic linkage from n informative offsprings.

2. Simple Versus Simple Hypothesis Testing Under Strict Monotone Likelihood Ratio

We start with an example involving the monotone likelihood ratio (MLR) property to fix ideas and provide motivation to classical statisticians. Later we show that this restriction is really not needed. Let X denote a discrete random variable taking values on the set of integers. Possible models for X can be given by the family of densities (with respect to counting measure) $\{f(\cdot|\theta), \theta \in \Theta\}$ where Θ is a subset of the real line.

We make the following assumptions on the family $\{f(\cdot|\theta), \theta \in \Theta\}$:

(A1) Define $A_\theta = \{x : f(x|\theta) > 0\}$ for $\theta \in \Theta$. Assume that $A_\theta = A$ for all $\theta \in \Theta$.

(A2) If $x, y \in A$ and $x < y$, then $x+1, x+2, \ldots, y-1 \in A$.

(A3) The family $\{f(\cdot|\theta), \theta \in \Theta\}$ has the (strict) Monotone Likelihood Ratio (MLR) property: for any $\theta_0, \theta_1 \in \Theta$ with $\theta_0 < \theta_1$, $f(x|\theta_0)/f(x|\theta_1)$ is a (strictly) decreasing function of x.

Consider the simple versus simple hypothesis testing of

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta = \theta_1, \tag{1}$$

where $\theta_0 < \theta_1$. The Bayesian approach to this problem is to consider $B(x)$, defined by

$$B(x) = \frac{f(x|\theta_0)}{f(x|\theta_1)}, \tag{2}$$

which is the likelihood ratio or *Bayes factor* of H_0 to H_1. This is often regarded by Bayesians as the odds of H_0 to H_1 arising from the data. Thus, values of $B(x)$ greater than 1 lend support to H_0 while values of $B(x)$ less than 1 support H_1. Under the MLR property, $B(x)$ is a decreasing function of x. Thus, small values of x favour H_0 while large values favour H_1.

Let U be a uniform random variable independent of X, and define $\tilde{X} = X + U$. This randomized version of X, $\tilde{X}$, clearly has a continuous distribution function and admits a density with respect to the Lebesgue measure under both H_0 and H_1. The randomized distribution functions are the key for obtaining the Bayesian and conditional frequentist unification. Extend the definition of $B(x)$ to include values of $\tilde{x}$ by

$$B(\tilde{x}) = B(x) \qquad \text{if } x = [\tilde{x}], \tag{3}$$

where $[z]$ denotes the greatest integer less than or equal to z. It is easy to see that $B(\cdot)$ extended as in (3) is the Bayes factor corresponding to $\tilde{X}$.

Let F_0 and F_1 be the distribution functions (dfs) of $\tilde{X}$ under H_0 and H_1, respectively. Let

$$x_L = sup\{\tilde{x} : B(\tilde{x}) > 1\} \tag{4}$$

and

$$x_U = inf\{\tilde{x} : B(\tilde{x}) < 1\}. \tag{5}$$

Under the MLR property, we have $x_L \leq x_U$ with equality if and only if there is no value x such that $B(x) = 1$.

Let ψ be a function given by

$$\psi(\tilde{x}) = F_0^{-1}(1 - F_1(\tilde{x})), \tag{6}$$

and define the quantities x_a and x_r by

$$x_a = \psi(x_U) \quad \text{and} \quad x_r = x_U \qquad \text{if } x_L > \psi(x_U) \text{ or,} \tag{7}$$
$$x_a = x_L \quad \text{and} \quad x_r = \psi^{-1}(x_L) \qquad \text{if } x_L \le \psi(x_U). \tag{8}$$

Consider the following conditional test, T^*, given by

$$T^* = \begin{cases} \text{if } \tilde{x} \ge x_r, & \text{reject } H_0 \text{ and report conditional error} \\ & \text{probability (CEP) } \alpha^*(B(\tilde{x})) = B(\tilde{x})/(1+B(\tilde{x})), \\ \text{if } x_a < \tilde{x} < x_r, & \text{make no decision,} \\ \text{if } \tilde{x} \le x_a, & \text{accept } H_0 \text{ and report conditional error} \\ & \text{probability (CEP) } \beta^*(B(\tilde{x})) = 1/(1+B(\tilde{x})). \end{cases} \tag{9}$$

The interval (x_a, x_r) is called the no-decision region.

We proceed to state two important properties of the test T^* given in (7-9) in the following theorem. Let S be a function of $\tilde{x}$ given by

$$S(\tilde{x}) = \max\{\psi^{-1}(\tilde{x}), \tilde{x}\}, \tag{10}$$

where ψ is as given in (6).

THEOREM 1. *For x_a and x_r given by equations (7) and (8), and S as in (10), we have*

(a) $\alpha^*(B(\tilde{x})) \equiv B(\tilde{x})/(1+B(\tilde{x})) = P_{H_0}(\textit{Reject } H_0 \,|\, S(\tilde{x}))$, *and*

(b) $\beta^*(B(\tilde{x})) \equiv 1/(1+B(\tilde{x})) = P_{H_1}(\textit{Accept } H_0 \,|\, S(\tilde{x}))$
when either $\tilde{x} \ge x_r$ or $\tilde{x} \le x_a$.

We give details of the derivation of equations (7-9) as well as a proof of Theorem 1 in the Appendix. Theorem 1 states that the error probabilities of T^*, $\alpha^*(B(\tilde{x}))$ and $\beta^*(B(\tilde{x}))$, are interpretable, respectively, as frequentist probabilities of Type I and Type II errors conditioned on the statistic, S, in the randomized space of $\tilde{X}$, or as posterior probabilities of H_0 and H_1 for a Bayesian under equal prior probabilities of H_0 and H_1. We will always assume that H_0 and H_1 have equal prior probabilities of 1/2 each for the Bayesian approach for the rest of this paper.

The statistic S defined in (10) is the conditioning statistic that makes unification possible; the conditional frequentist error probabilities are identical to the Bayesian posterior probabilities of hypotheses as in Theorem 1. The statistic S is a special case of general conditioning statistic based on evidential equivalence statistic, E_0 and E_1, of the form

$$S = \max\{E_0, E_1\}. \tag{11}$$

Conditioning statistic based on evidential equivalence statistics serve two purposes: first, if $E_0 > E_1$, H_0 will be accepted (indicating the "evidence" for H_0 is more than the "evidence" for H_1), and rejected otherwise; second, for $S = s$, data in the acceptance region with $E_0 = s$ has the same evidential strength as data in the rejection region with $E_1 = s$. Sellke, Bayarri, and Berger (2001) discuss in detail the choice of the conditioning statistics, S, in general and also the choice of S based on p-values.

We highlight several other attractive features of T^*. One immediate gain of T^* over unconditional tests is that the reported error probabilities of T^* are data-dependent rather than being constant over the rejection and acceptance regions. The error probabilities α^* and β^* are functions of the randomized outcome, $\tilde{x}$, only through the observed data, x, where $x = [\tilde{x}]$. The proofs of these facts are given in the Appendix. The randomization inherent in T^* affects only a small region of the decision space. There is actually little practical difference between the decisions made based on the outcome of the randomization in this small region. Except for this region, the decision made and the reported CEP depend only on the observed value of x.

We will illustrate these key features of the new test in the following section using two testing examples where the underlying discrete distributions possess the MLR property. In the testing situations that follow, we will assume that there are no values of x such that $B(x) = 1$. Subsequently, if we define

$$x^* = \min\{\, x \,:\, B(x) < 1\,,\, x \text{ is an integer}\}, \tag{12}$$

we have that $x^* = x_L = x_U$ where x_L and x_U are as defined in (4) and (5).

2.1. *Testing the unknown proportion for a Binomial distribution.* Here, $f(x|\theta) = Bin(x\,|\,n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ where $\theta \in [0, 1]$ is the parameter of interest. The simple versus simple testing situation is

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta = \theta_1, \tag{13}$$

where $\theta_0 < \theta_1$. The Bayes factor, $B(x)$, is given by

$$B(x) = \frac{Bin(x\,|\,n, \theta_0)}{Bin(x\,|\,n, \theta_1)} \tag{14}$$

$$= \left(\frac{1-\theta_0}{1-\theta_1}\right)^n \cdot \left(\frac{\theta_0/(1-\theta_0)}{\theta_1/(1-\theta_1)}\right)^x, \qquad \text{for } x = 0, 1, 2, \ldots, n. \tag{15}$$

Here, $B(x)$ is a decreasing function of x. In binomial testing, x_a and x_r can explicitly be evaluated using the formulas given below. Define functions

$s_0(\cdot)$ and $s_1(\cdot)$ by

$$s_0(x) = \sum_{k=0}^{x-1} Bin(k|\, n,\, \theta_0) \qquad \text{and} \qquad s_1(x) = \sum_{k=x}^{n} Bin(k|\, n, \theta_1),$$

and x^* is the integer defined in (12).

If $s_0(x^*) \le s_1(x^*)$, denote integer $t = min\{x : s_1(x+1) \le s_0(x^*),\, x$ integer$\}$ and $q = (s_0(x^*) - s_1(t+1))/Bin(t|\, n,\, \theta_1)$, and set

$$x_a = x^* \qquad \text{and} \qquad x_r = t + 1 - q. \tag{16}$$

If $s_0(x^*) > s_1(x^*)$, denote integer $u = max\{x : s_0(x) \le s_1(x^*),\, x \text{ integer}\}$ and $p = (s_1(x^*) - s_0(u))/Bin(u|\, n,\, \theta_0)$, and set

$$x_a = u + p \qquad \text{and} \qquad x_r = x^*. \tag{17}$$

For the binomial testing of $\theta_0 = 0.4$ versus $\theta_1 = 0.5$ ($n = 10$), the decisions and reported CEPs for the conditional test T^* are given in Table 1. The no-decision region (x_a, x_r), from (16) and (17), for this problem is found to be $(4.96, 5.00)$.

Table 1. DECISIONS AND REPORTED CEP BASED ON T^* FOR THE BINOMIAL PROBLEM

Value of x	Decision	CEP	Randomized (if any)	Value of x	Decision	CEP	Randomized (if any)
0	Accept H_0	0.138	-	5	Accept H_1	0.449	-
1	Accept H_0	0.195	-	6	Accept H_1	0.352	-
2	Accept H_0	0.267	-	7	Accept H_1	0.266	-
3	Accept H_0	0.353	-	8	Accept H_1	0.195	-
4	Accept H_0	0.448	0.96	9	Accept H_1	0.140	-
4	No-decision	-	0.04	10	Accept H_1	0.097	-

As mentioned earlier, one immediate gain of T^* over conventional unconditional most powerful testing is that the reported error probabilities of T^* reflect the confidence with which a decision is made based on the observed data. This is clearly seen from the entries in Table 1 where the magnitude of the reported CEP varies according to whether x was observed at the extremes or close to the center of the distribution under H_0.

The randomization inherent in T^* affects only a small region of the decision space. Except for this region, the decision made and the reported CEP depends only on the observed value of x. Thus, in Table 1, the decisions made and the reported CEPs are independent of the randomization except only when $x = 4$. However, even when $x = 4$, there is little practical difference between stating "no-decision" and "accepting H_0 with error probability 0.448". The minimax unconditional test for this example also randomizes at the boundary $x = 4$. However, the decisions made if $x = 4$ obtains are "reject H_0 with probability 0.1" and "accept H_0 with probability 0.9". Thus, the practical effect of randomization in the unified testing situation is minimal, making its use in conditional testing more attractive than in unconditional testing, where the effect can be considerable.

As stated earlier (and proved in the Appendix), the reported error probabilities of T^* also have simultaneous Bayesian and conditional frequentist interpretations. In other words, the quantity α^* in (9) can be interpreted simultaneously as the posterior probability of H_0 and the conditional frequentist Type I error probability (see the Appendix for an explicit form of the conditioning statistic). Similarly, the quantity β^* in (9) can be interpreted as the posterior probability of H_1 as well as the conditional frequentist Type II error probability.

2.2. *Testing for the mean of a Poisson distribution.* Here, $f(x|\theta) = Poi(x|\theta)$ where $Poi(x|\theta)$ represents the Poisson distribution with mean parameter θ. The Poisson distribution is an example of a discrete distribution with infinite range. The simple versus simple testing situation is

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta = \theta_1, \tag{18}$$

where $\theta_0 < \theta_1$. As before, the Bayes factor for the testing problem of (18) is given by

$$\begin{aligned} B(x) &= \frac{Poi(x \mid \theta_0)}{Poi(x \mid \theta_1)} && (19) \\ &= e^{(\theta_1 - \theta_0)} \left(\frac{\theta_0}{\theta_1}\right)^x \quad \text{for } x = 0, 1, 2, \ldots && (20) \end{aligned}$$

For the Poisson testing problem, x_a and x_r can explicitly be evaluated using the formulas given below. The functions $s_0(\cdot)$ and $s_1(\cdot)$ are defined by

$$s_0(x) = \sum_{k=0}^{x-1} Poi(k|\,\theta_0) \qquad \text{and} \qquad s_1(x) = \sum_{k=x}^{\infty} Poi(k|\,\theta_1),$$

and x^* is defined in (12).

If $s_0(x^*) \leq s_1(x^*)$, denote integer $t = min\{x : s_1(x+1) \leq s_0(x^*), x \text{ integer}\}$ and $q = (s_0(x^*) - s_1(t+1))/Poi(t|\,\theta_1)$, and set

$$x_a = x^* \qquad \text{and} \qquad x_r = t + 1 - q. \tag{21}$$

If $s_0(x^*) > s_1(x^*)$, denote integer $u = max\{x : s_0(x) \leq s_1(x^*), x \text{ integer}\}$ and $p = (s_1(x^*) - s_0(u))/Poi(u|\,\theta_0)$, and set

$$x_a = u + p \qquad \text{and} \qquad x_r = x^*. \tag{22}$$

For $\theta_0 = 1.0$ and $\theta_1 = 5.0$, the unified test T^* is given in Table 2. The no-decision region is found to be $(2.76, 3.00)$ using (21) and (22).

Table 2. Decisions and CEP based on T^* for the Poisson problem

Value of x	Decision	CEP	Randomized (if any)	Value of x	Decision	CEP	Randomized (if any)
0	Accept H_0	0.018	-	4	Accept H_0	0.080	-
1	Accept H_0	0.084	-	5	Accept H_1	0.017	-
2	Accept H_0	0.314	0.76	6	Accept H_1	0.003	-
2	No-decision	-	0.24	$x \geq 7$	Accept H_1	$\frac{B(x)}{(1+B(x))}$	-
3	Accept H_0	0.304	-				

The decisions made and the CEPs reported are independent of the randomization for all values of x except for $x = 2$. For $x = 2$, one will randomly choose between the following two decisions: with probability 0.76, "accept H_0 and report CEP 0.314" or,"make no-decision" with probability 0.24. The probability of no-decision in Table 2 is greater than that in Table 1. However, there is still little practical difference between declaring "no-decision" and "accepting H_0 with large CEP of 0.314".

3. General Methodology for Discrete Testing

We discuss the general problem of obtaining unified conditional frequentist and Bayesian tests when both hypotheses are composite. Let X be a discrete random variable with density (with respect to counting measure) $f(\cdot|\theta)$ for $\theta \in \Theta$. Suppose we wish to test

$$H_0 : \theta \in \Theta_0 \qquad \text{vs.} \qquad H_1 : \theta \in \Theta_1, \tag{23}$$

where Θ_0 and Θ_1 are disjoint subsets of Θ. A Bayesian puts proper priors π_0 and π_1 on Θ_0 and Θ_1, respectively, computes the marginals under H_0

and H_1, and reduces the composite testing problem of (23) to the following simple versus simple testing of

$$H_0 \,:\, X \sim m_0 \qquad \text{vs.} \qquad H_1 \,:\, X \sim m_1, \tag{24}$$

where m_0 is the marginal distribution of X under H_0 given by

$$m_0(x) = \int_{\Theta_0} f(x|\theta)\, \pi_0(\theta)\, d\theta \tag{25}$$

and m_1 is the marginal distribution of X under H_1 given by

$$m_1(x) = \int_{\Theta_1} f(x|\theta)\, \pi_1(\theta)\, d\theta. \tag{26}$$

The Bayes factor, $B(X)$, for the testing of (24) is given by

$$B(x) = \frac{m_0(x)}{m_1(x)}. \tag{27}$$

Since, in general, $B(x)$ need not be an decreasing (or increasing) function of x, let $\mathcal{B} = \{\ldots, b_{-2}, b_{-1}, b_0, b_1, \ldots\}$ denote all *distinct* values of $B(x)$ in decreasing order. Define a new variable Y taking values on integers such that $Y = k$ if and only if $b = b_k$ for $b \in \mathcal{B}$. Let f_0^Y and f_1^Y be the densities (with respect to counting measure) of Y induced by m_0 and m_1, respectively. Define $B^Y(y) = f_0^Y(y)/f_1^Y(y)$. B^Y is the likelihood ratio of f_0^Y to f_1^Y. B^Y is also the Bayes factor for the simple versus simple testing of an appropriate null-alternative pair for Y.

As before, let U be a uniform random variable independent of Y and define $\tilde{Y} = Y + U$. Note that $\tilde{Y}$ is an implicit function of the observation X and the uniform random variable U. The distribution functions of $\tilde{Y}$ under H_0 and H_1, say, F_0 and F_1 respectively, are continuous due to the randomization. Replacing X by Y and $\tilde{X}$ by $\tilde{Y}$ in Section 2, we get a unified testing procedure for $\tilde{Y}$ for the testing situation in (24).

Furthermore, for any x, y and $\tilde{y}$ related by $B(x) = B^Y(y)$ and $y = [\tilde{y}]$, we have

$$P\{H_0|\tilde{y}\} = \frac{B^Y(\tilde{y})}{1 + B^Y(\tilde{y})} = \frac{B^Y(y)}{1 + B^Y(y)} = \frac{B(x)}{1 + B(x)} \tag{28}$$

and

$$P\{H_1|\tilde{y}\} = \frac{1}{1 + B^Y(\tilde{y})} = \frac{1}{1 + B^Y(y)} = \frac{1}{1 + B(x)}. \tag{29}$$

The significance of equations (28) and (29) is that the reported conditional error probabilities, α^* and β^*, depend only on the observed value of x, and not on the randomized outcome. We present an example of simple versus composite binomial testing in the following section.

3.1 *Composite Binomial testing.* Let $X \sim Bin(n = 10, \theta)$. We want to test

$$H_0 : \theta = 0.4 \quad \text{vs.} \quad \theta \neq 0.4. \tag{30}$$

Since H_1 is composite, we require a prior on the space of θ in H_1. We choose the conjugate family $Beta(\alpha, \beta)$ for constructing the unified test T^*. The $Beta(\alpha, \beta)$ family can model various types of prior beliefs by appropriate choices of α and β. For the choice of α and β in this example, we select them so that the prior distribution on H_1 is centered around $\theta = 0.4$. Since the quantity $\alpha + \beta$ has the interpretation of 'prior sample size', we choose a small positive value for it, say, $\alpha + \beta = 1$. The choice of total mass, i.e., $\alpha + \beta$ can be a topic for further investigation but we choose it to be 1 here. The values of α and β that satisfy both requirements above are $\alpha = 0.4$ and $\beta = 0.6$.

Using the above prior for H_1 and integrating out, the marginal under H_1 is

$$\begin{aligned} m_1(x) &= \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{10-x} \frac{1}{B(0.4, 0.6)} \theta^{0.4-1} (1-\theta)^{0.6-1} \, d\theta \\ &= \frac{1}{(10-x)!\, x!} \cdot \frac{\Gamma(0.4+x)\Gamma(10-x+0.6)}{\Gamma(0.4)\Gamma(0.6)} \text{ for } x = 0, 1, \ldots, 10, \end{aligned}$$

where $B(\alpha, \beta)$ represents the value of the integral $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$. Also, $m_0(x)$, the distribution of X under the null, is

$$m_0(x) = \binom{n}{x} \theta^x \, (1-\theta)^{10-x} \text{ for } x = 0, 1, \ldots, 10.$$

The Bayes factor, $B(x)$, is given by

$$B(x) = \frac{m_0(x)}{m_1(x)} = \frac{\Gamma(0.4)\Gamma(0.6)\Gamma(11) \cdot (0.4)^x (0.6)^{10-x}}{\Gamma(0.4+x)\Gamma(10-x+0.6)}.$$

Following the general method prescribed in Section 3, we first rearrange the values of $B(x)$ in a decreasing order with b_0 denoting the largest value.

The unified test T^* is given by

$$T^* = \begin{cases} \text{if } \tilde{y} \geq 5.00 & \text{reject } H_0 \text{ and report CEP} \\ & \alpha^*(B(x)) = \frac{B(x)}{1+B(x)}\ , \\ \text{if } 3.09 < \tilde{y} < 5.00 & \text{make no decision,} \\ \text{if } \tilde{y} \leq 3.09 & \text{accept } H_0 \text{ and report CEP} \\ & \beta^*(B(x)) = \frac{1}{1+B(x)}, \end{cases} \tag{31}$$

where x, y and $\tilde{y}$ are related by $B(x) = B^Y(y)$ for $y = [\tilde{y}]$. The no-decision region, (y_a, y_r) is given by $y_a = 3.09$ and $y_r = 5.00$. Table 3 gives the T^* in terms of the original binomial random variable X and the accompanying randomization.

Table 3. Decisions and CEP based on T^* for composite Binomial testing

Value of x	Decision	CEP	Randomized (if any)	Value of x	Decision	CEP	Randomized (if any)
0	Reject H_0	0.022	-	6	Accept H_0	0.336	0.09
1	Reject H_0	0.251	-	6	No-decision	-	0.91
2	No-decision	-	-	7	Reject H_0	0.425	-
3	Accept H_0	0.240	-	8	Reject H_0	0.147	-
4	Accept H_0	0.196	-	9	Reject H_0	0.022	-
5	Accept H_0	0.223	-	10	Reject H_0	0.001	-

The no-decision region for the testing of (30) is larger than the previous two examples. None the less, for $x = 6$, there is still little practical difference between declaring "no-decision" and "accepting H_0 with large CEP of 0.336".

3.2. *An example of testing linkage heterogeneity.* The number of observed recombinations, r, from a total of n informative offsprings follows a Binomial distribution with parameters n and θ, where θ denotes the probability of a recombination. It is commonly assumed that $0 \leq \theta \leq 1/2$ where $\theta = 1/2$ denotes the absence of linkage. A possible null-alternative pair is the testing of no linkage versus presence of linkage, that is,

$$H_0 : \theta = 1/2 \qquad \text{vs.} \qquad H_1 : 0 \leq \theta < 1/2. \tag{32}$$

The distribution of the number of recombinations under the null follows a $Bin(n, \theta = 1/2)$ with the probability mass function, m_0, given by

$$m_0(r) = \binom{n}{r} \frac{1}{2^n} \quad \text{for } r = 0, 1, 2, \ldots, n. \tag{33}$$

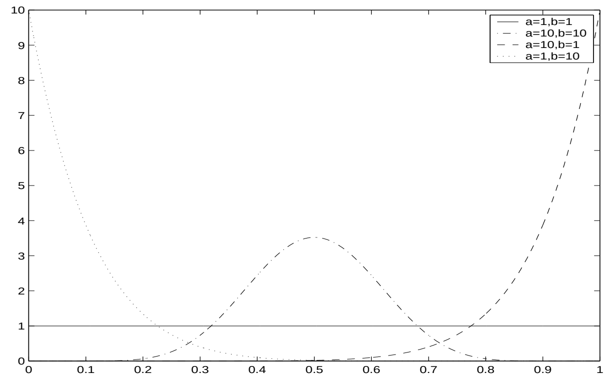

Figure 1: Graphs of $\pi(\cdot)$ for different values of a and b.

Since $0 \leq \theta < 1/2$ for H_1, following Risch (1988), we put a $Beta(a, b)$ prior on 2θ. Beta priors offer great flexibility in modeling prior information on θ by choosing appropriate values for a and b. The prior on θ is

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} 2 \cdot (2\theta)^{a-1}(1-2\theta)^{b-1} \tag{34}$$

for $0 \leq \theta < 1/2$. The graphs of π are plotted for four different choices of the pair (a, b) each representing a certain prior belief on the alternative. See Figure 1. The prior mean and variance is given by

$$u = \frac{1}{2}\frac{a}{a+b} \tag{35}$$

$$v = \frac{1}{4}\frac{ab}{(a+b)^2(a+b+1)} \tag{36}$$

and the marginal distribution of r, m_1, is given by

$$m_1(r) = \binom{n}{r} \sum_{i=0}^{n-r} (-1)^i \binom{n-r}{i} \left(\frac{1}{2}\right)^{r+i} \frac{\Gamma(a+b)\Gamma(a+r+i)}{\Gamma(a)\Gamma(a+r+i+b)} \tag{37}$$

for $r = 0, 1, 2, \ldots, n$. The Bayes factor of H_0 to H_1 for the testing of (32) is

$$B(r) = m_0(r)/m_1(r) \tag{38}$$

$$= \frac{1}{\int_0^{1/2} L^*(\theta)\,\pi(\theta)\,d\theta}, \tag{39}$$

where $L^*(\theta)$ is the antilog of the lod score, see Ott (1991). Therefore, the Bayes factor is the inverse of the weighted average of $L^*(\theta)$, with weights proportional to $\pi(\cdot)$. Large values of $L^*(\theta)$ and hence small values of $B(r)$ provide evidence against H_0. The testing scenario in (32) is a special case of the more general set-up described in the beginning of Section 3 for which we may obtain unified testing procedures. We take the total number of informative offsprings,n, to be 10 in each case. Larger values of n would lead to tables too large to be reported here.

Table 4 gives the unified testing procedure under uniform prior on H_1 $(a = 1, b = 1)$. Table 5 gives the unified testing procedure when a and b are equal and large $(a = 10, b = 10)$. In the limiting case, the distribution under H_1 will be concentrated at $\theta = 1/4$, and hence, the unified testing procedure for large and equal a and b will resemble the one for testing $H_0 : \theta = 1/2$ versus $H_1 : \theta = 1/4$. If b is held fixed with a becoming large, the prior mean moves closer to $1/2$ and the variance goes to 0. In this case, the unified testing procedure will not be able to distinguish between H_0 and H_1. This is reflected by the large conditional error probabilities for all values of r in Table 6. On the other hand, if the prior belief of θ under the alternative is strongly concentrated around a point sufficiently removed from $1/2$, the reported CEPs will be much smaller compared to the entries in Table 6. Table 7 is such a case where a and b are chosen to be 1 and 10, respectively.

We have essentially presented four different situations reflecting various kinds of prior belief on the alternative for the probability of a recombination, θ. The four different priors have mass concentrated around specific mean points of $5/11$ (closest to $1/2$), $1/4$ (intermediate) and $1/22$ (furthest from $1/2$) reflecting various kinds of prior beliefs for θ. This gives the flexibility of tailoring the alternatives to particular theories. In the intermediate case of $1/4$, $(a = 1, b = 1)$ and $(a = 10, b = 10)$, more concentration around the prior mean as measured by the prior variance gives better unified testing procedures in terms of the reported error probabilities. For all the above priors and in fact for any prior π, the derived test T^* is *always* a valid test from the conditional frequentist point of view even though a prior was used in its derivation. The reported error probabilities arising from the use of T^* can be interpreted as the conditional frequentist Type I and Type II error probabilities. In the case when subjective priors are not available, one can use the 'default' choice of uniform prior on the alternative, while retaining the frequentist interpretation of the reported error probabilities.

Table 4. Decisions and CEP based on T^* for the testing of (32) with $a = b = 1$.

Value of r	Decision	CEP	Randomized (if any)	Value of r	Decision	CEP	Randomized (if any)
0	Reject H_0	0.005	-	5	Accept H_0	0.270	-
1	Reject H_0	0.051	-	6	Accept H_0	0.195	-
2	Reject H_0	0.200	-	7	Accept H_0	0.149	-
3	Reject H_0	0.421	-	8	Accept H_0	0.119	-
4	No-decision	-	0.63	9	Accept H_0	0.098	-
4	Accept H_0	0.391	0.37	10	Accept H_0	0.083	-

Table 5. Decisions and CEP based on T^* for the testing of (32) with $a = b = 10$.

Value of r	Decision	CEP	Randomized (if any)	Value of r	Decision	CEP	Randomized (if any)
0	Reject H_0	0.014	-	5	Accept H_0	0.210	-
1	Reject H_0	0.048	-	6	Accept H_0	0.096	-
2	Reject H_0	0.142	-	7	Accept H_0	0.043	-
3	Reject H_0	0.335	-	8	Accept H_0	0.020	-
4	No-decision	-	0.32	9	Accept H_0	0.009	-
4	Accept H_0	0.412	0.68	10	Accept H_0	0.004	-

Table 6. Decisions and CEP based on T^* for the testing of (32) with $a = 10$ and $b = 1$.

Value of r	Decision	CEP	Randomized (if any)	Value of r	Decision	CEP	Randomized (if any)
0	Reject H_0	0.231	-	5	Accept H_0	0.483	0.47
1	Reject H_0	0.299	-	6	Accept H_0	0.444	-
2	Reject H_0	0.364	-	7	Accept H_0	0.411	-
3	Reject H_0	0.422	-	8	Accept H_0	0.381	-
4	Reject H_0	0.473	-	9	Accept H_0	0.356	-
5	No-decision	-	0.53	10	Accept H_0	0.333	-

Table 7. DECISIONS AND CEP BASED ON T^* FOR THE TESTING OF (32) WITH $a = 1$ AND $b = 10$.

Value of r	Decision	CEP	Randomized (if any)	Value of r	Decision	CEP	Randomized (if any)
0	Reject H_0	0.016	-	5	Accept H_0	0.196	-
1	Reject H_0	0.049	-	6	Accept H_0	0.078	-
2	Reject H_0	0.136	-	7	Accept H_0	0.029	-
3	Reject H_0	0.322	-	8	Accept H_0	0.011	-
4	No-decision	-	0.269	9	Accept H_0	0.004	-
4	Accept H_0	0.415	0.731	10	Accept H_0	0.001	-

Acknowledgments. The author wishes to thank Professor J.K. Ghosh for valuable suggestions made by him during a reading of this paper. The author would also like to thank Professors James O. Berger, Rama T. Lingham and Jaeyong Lee for helpful comments and discussions. This work is part of the author's Ph.D. thesis at Purdue University, and was supported by the National Science Foundation, Grants DMS-9303556 and DMS-9802261, and by a Purdue Research Foundation grant.

Appendix

Some details of the derivation of equations (7-9) are presented here. Let the interval $[x_{LB}, x_{UB}]$ denote the common support set of the distributions of $\tilde{X}$. Also, let $I_A = [x_{LB}, x_L)$ and $I_R = [x_U, x_{UB}]$ where x_L and x_U are given by (4) and (5), respectively. The values of $\tilde{x}$ in I_A, with $B(\tilde{x}) > 1$, favor H_0 while $\tilde{x}$ values in I_R, with $B(\tilde{x}) < 1$, favor H_1. The function ψ (defined in (6)) is non-increasing in $\tilde{x}$ with $\psi(x_{UB}) = x_{LB}$. So, we either have $\psi(I_R) \subset I_A$ or $\psi(I_R) \supseteq I_A$. In the first case, $\psi(x_U) < x_L$; thus, we define the boundaries of the acceptance and rejection regions, x_a and x_r, by $x_a = \psi(x_U)$ and $x_r = x_U$. This gives equation (7). In the second case, we have $\psi(x_U) \geq x_L$; thus, we take $x_a = x_L$ and $x_r = x_0$ where $\psi(x_0) = x_L$, or in other words, $x_0 = \psi^{-1}(x_L)$. This gives equation (8).

The conditioning statistic $S(\tilde{x})$ in (10) defines a pairing of points via ψ, with one point from the acceptance region and the other from the rejection region, with equal evidential strength; thus, when $S(\tilde{x}) = s$, the points $\psi(s)$ and s have equal strength of evidence with $\psi(s)$ favoring H_0 and s favoring H_1. Exact pairing is obtained for points $\tilde{x} \geq x_r$ with a point $\tilde{x} \leq x_a$. The remaining points $\tilde{x} \in (x_a, x_r)$ fall into the no-decision region; evidence for or against H_0 is weak for these points. The test T^* given in (9) is defined in terms of the acceptance, rejection and the no-decision regions.

PROOF OF THEOREM 1. Due to the randomization, F_0 and F_1 have densities, say, f_0 and f_1, respectively, with respect to Lebesgue measure, λ, under H_0 and H_1. From the definition of ψ, we have $F_0(\psi(\tilde{x})) = 1 - F_1(\tilde{x})$, or equivalently,

$$f_0(\psi(\tilde{x}))|\psi^{'}(\tilde{x})| = f_1(\tilde{x}), \tag{40}$$

by diffferentiating both sides with respect to $\tilde{x}$ where $\psi^{'}(\cdot)$ denotes the derivative of $\psi(\cdot)$. Note that the equality in (40) is satisfied almost everywhere (with respect to the Lebesgue measure on the real line).

Part (a): We have that

$$\begin{aligned} P_{H_0}(\text{Reject } H_0 \,|\, S(\tilde{x}) = s) &= P_{H_0}(\tilde{X} = s \,|\, \tilde{X} = s \text{ or } \tilde{X} = \psi(s)) \\ &= \frac{f_0(s)}{f_0(s) + f_0(\psi(s))|\psi'(s)|} \\ &= \frac{f_0(s)}{f_0(s) + f_1(s)} \qquad \text{(by (40))} \\ &= \frac{B(s)}{1 + B(s)}, \end{aligned}$$

where the last equality is obtain by dividing by $f_0(s)$ on both the numerator and denominator.

Part (b): In a similar way,

$$\begin{aligned} P_{H_1}(\text{Accept } H_0 \,|\, S(\tilde{x}) = s) &= P_{H_0}(\tilde{X} = \psi(s) \,|\, \tilde{X} = s \text{ or } \tilde{X} = \psi(s)) \\ &= \frac{f_0(\psi(s))|\psi^{'}(s)|}{f_0(s) + f_0(\psi(s))|\psi'(s)|} \\ &= \frac{f_1(s)}{f_0(s) + f_1(s)} \qquad \text{(by (40))} \\ &= \frac{1}{1 + B(s)}, \end{aligned}$$

where the last equality is obtain by dividing by $f_1(s)$ on both the numerator and denominator. The proofs of (a) and (b) are essentially the same as in Berger, Brown and Wolpert (1994). The proofs presented here are based on the randomized observation space, $\tilde{X}$, and not on the space of possible Bayes factors values as was done in Berger, Brown and Wolpert (1994).

To see that the reported error probabilities of T^* do not depend on the randomized outcome, note that

$$\begin{aligned} \alpha^*(B(\tilde{x})) &= \frac{B(\tilde{x})}{1+B(\tilde{x})} = \frac{B(x)}{1+B(x)} \qquad \text{and} \\ \beta^*(B(\tilde{x})) &= \frac{1}{1+B(\tilde{x})} = \frac{1}{1+B(x)} \end{aligned}$$

since $B(x) = B(\tilde{x})$ for $x = [\tilde{x}]$. Thus, α^* and β^* depend only on the observed discrete x and not on the randomized outcome, $\tilde{x}$. The decision made is also independent of the randomized outcome for all values of x except for $x = [x_a]$ if (7) holds, or for $x = [x_r]$ if (8) holds.

For the composite hypothesis testing in Section 3, define the conditioning statistic, $S(\cdot)$, and $\psi(\cdot)$ similarly as before for the random variable $\tilde{Y}$. Following Berger, Boukai and Wang (1997), we have

$$E^{\pi_0(\theta \,|\, S=s)}\,(\,P_\theta\{\,\text{Reject } H_0 \,|\, S(\tilde{y}) = s\,\}) = \frac{B^Y(s)}{1+B^Y(s)} \tag{41}$$

and

$$E^{\pi_1(\theta \,|\, S=s)}\,(\,P_\theta\{\,\text{Accept } H_0 \,|\, S(\tilde{y}) = s\,\}) = \frac{1}{1+B^Y(\psi(s))} \tag{42}$$

where $\pi_0(\cdot \,|\, S = s\,)$ and $\pi_1(\cdot \,|\, S = s\,)$ are the distributions of θ in Θ_0 and Θ_1 respectively, given that $S = s$. Equation (41) states that the Bayesian posterior probability of H_0 is the weighted average of conditional Type I error probabilities with weights equal to $\pi_0(\cdot \,|\, S = s\,)$. Equation (42) relates the Bayesian posterior probability of H_1 and the conditional Type II error probabilities in a similar way. See Berger, Boukai and Wang (1997) for details.

References

Berger, J.O., Boukai, B. and Wang. Y. (1999). Simulteneous Bayesian-frequentist sequential testing of nested hypotheses, *Biometrika*, **86**, 79-92.

Berger, J.O., Brown, L.D. Wolpert, R.L. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing, *Annals of Statistics*, **22**, 1787-1807.

Berger, J.O. and Delampady, M. (1987). Testing precise hypotheses, *Statistical Science*, **2**, 317-335.

Berger, J.O. and Sellke, T. (1987). Testing a point null hypothesis: The irreconciliability of ρ values and evidence, *Journal of the American Statistical Association*, **82**, 112-122.

BERGER, J.O., BOUKAI, B. and WANG, W. (1997). Unified frequentist and Bayesian testing of a precise hypothesis, *Statistical Sciences*, **12**, 133-160.

BROWNIE, C. and KIEFER, J. (1977). The ideas of conditional confidence in the simplest setting, *Communications in Statistics, Series A – Theory and Methods*, **6**, 691-752.

BROWN, L.D. (1978). A contribution of Kiefer's theory of conditional confidence procedures, *Annals of Statistics,* **6**, 59-71.

CASELLA, G. and BERGER, R.L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing proble, *Journal of the American Statistical Association*, **82**, 106-111.

DASS, S.C. and BERGER, J.O. (1998). Unified Bayesian and conditional frequentist testing of composite hypotheses, *Institute of Statistics and Decision Sciences Discussion Paper, Duke University, USA*, **98-43**.

DELAMPADY, M. and BERGER, J.O. (1990). Lower bounds on Bayes factors for multinomial distributions, with application to chi-squared tests of fit, *Annals of Statistics*, **18**, 1295-1316.

EDWARDS, W., LINDMAN, H. and SAVAGE, L.J. (1963). Bayesian statistical inference for psychological research, *Psychological Review*, **70**, 193-242.

KIEFER, J. (1975). Conditional confidence approach in multi decision problems. In *Multivariate Analysis IV*, ed. P.R. Krishnaiah, 143-158. Academic Press, New York.

KIEFER, J. (1976). Admissibility of conditional confidence procedures, *Annals of Statistics*, **4**, 836-865.

KIEFER, J. (1977). Conditional confidence statements and confidence estimators, *Journal of the American Statistical Association*, **72**, 789-807.

RISCH, N. (1988). A new statistical test for linkage heterogeneity, *American Journal of Human Genetics*, **42**, 353-364.

OTT, J. (1991). *Analysis of Human Genetic Linkage (Revised Edition).* The John Hopkins University Press.

SELLKE, T., BAYARRI, M.J. and BERGER, J.O. (2001). Calibration of P-values for testing precise null hypotheses, *The American Statistician*, **55**, no.1, 62-71.

WOLPERT, R.L. (1996). Testing simple hypotheses. In *Data Analysis and Information Systems. Proceedings of the 19th Annual Conference of the Gesellschaft für Klassifikation e.V.*, 289-297. Springer, Berlin.

SARAT C. DASS

DEPARTMENT OF STAT. AND PROBABILITY

MICHIGAN STATE UNIVERSITY

EAST LANSING, MI 48823

USA

E-mail: sdass@msu.edu