Testing Hardy-Weinberg Equilibrium: an Objective Bayesian Analysis

Guido Consonni
(Università di Pavia)

Elias Moreno
(University of Granada)

Sergio Venturini
(Università Bocconi di Milano)

# 121 (08-10)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Agosto 2010
We analyze the general (multiallelic) Hardy-Weinberg equilibrium problem from an objective Bayesian testing standpoint. We argue that for small or moderate sample sizes the answer is rather sensitive to the prior chosen, and this suggests to carry out a sensitivity analysis with respect to the prior. This goal is achieved through the identification of a class of priors specifically designed for this testing problem. In this paper we consider the class of intrinsic priors under the full model, indexed by a tuning quantity, the training sample size. These priors are objective, satisfy Savage’s continuity condition and have proved to behave extremely well for
many statistical testing problems. We compute the posterior probability of the Hardy-Weinberg equilibrium model for the class of intrinsic priors, assess robustness over the range of plausible answers, as well as stability of the decision in favor of either hypothesis.

KEY WORDS: Bayes factor; Hardy-Weinberg equilibrium; Intrinsic prior; Model posterior probability; Robustness.

1 Introduction

The Hardy-Weinberg law plays a fundamental role in the study of population genetics [?]. For a particular autosomal locus that is always one of \( r \) different alleles \( \{A_i, i = 1, ..., r\} \), the law provides the sampling distribution of genotype counts in a random sample of individuals drawn from a population which is assumed to be in Hardy-Weinberg equilibrium (HWE). We recall that a genotype is an unordered pair of allele combination \( \{A_i, A_j\} \), and consequently a sample of genotype counts is a triangular array of the form \( \{y_{ij}, 1 \leq j \leq i \leq r\} \). Interest centers on testing the null hypothesis that the population is in HWE. Recently, there has been great interest in testing for HWE in genome-wide association studies (GWAS) in which departure from HWE may indicate problems with quality control for the single nucleotide polymorphism (SNP) in question; see [?].

In the frequentist setting, the chi-square test does not provide reliable answers for testing HWE especially when some counts in the sample are zero or the triangular array is scarce; see [?]. This has led to the adoption of exact tests. Algorithms for generating the exact distribution of genotype counts of a sample drawn from a population satisfying the HWE have been developed by [?]. The paper by [?] provided for
the first time general algorithms to perform an exact test using Monte Carlo methods.
Specifically, they presented two methods to estimate the significance level for the exact
test of HWE for multiple alleles: one is a direct Monte Carlo method, while the other
is based on a Markov chain approach. Recently, [?] provided a new and improved
Monte Carlo algorithm for testing HWE.

The HWE testing problem has also been analyzed from a Bayesian standpoint.
Early papers approached the two-allele case as an estimation problem, providing pos-
terior credibility intervals for a specific parametrization; see [?] and [?]. Such intervals
are used as acceptance regions for the HWE null model, even though the Hardy-
Weinberg law does not play any role in their construction. Alternative reparametri-
zations were presented in [?] again for the two-allele case. [?] addressed the issue of
simultaneous estimation of the allelic proportions. More recently, HWE has been an-
alyzed according to an unconventional Bayesian hypothesis testing procedure by [?].
They computed a Bayes factor using a uniform prior on the parameter of the full
model, together with its “projection” prior on the parameter space corresponding to
HWE (the null model). Next they used this Bayes factor to set up an ordering of
the points in the sample space with the objective of computing a “Bayesian p-valu e”.
In [?] the “Full Bayesian Significance Test” is introduced. This is an unconventional
measure of evidence against the null hypothesis that the population is in HWE. The
computation of this quantity does not require a prior distribution for the parameters
of the HWE null model, but only a prior as well as a “reference” distribution (the two
can be equal) on the parameter space of the full model; see also [?]. In a very recent
contribution, [?] discuss notions of compatibility of prior specifications for comparing
nested models, illustrating the methodology with respect to HWE for the two-allele
testing problem, and computing the Bayes factor assuming that the prior under the full
model is a symmetric Dirichlet distribution: the novelty lies in the construction of the prior under the HWE model, which is obtained using a variety of methods, including “Kullback-Leibler projection” and conditioning. Another very recent contribution is [?]. This represents a rich contribution in the area because it clarifies various contexts in which the issue of HWE may arise and deals simultaneously both with testing as well as estimation of specific parameters which may indicate the strength of departure from HWE. There is also a careful consideration of sampling models. For instance, besides the full (saturated) model, alternative intermediate submodels are considered, an interesting one being the inbreeding model; for further details and references see [?, Section 2.2].

In this paper we concentrate on testing the HWE hypothesis. We argue that this problem exhibits a high sensitivity to the choice of the prior whose subjective specification may be problematic in some circumstances; this suggests an objective approach coupled with a robust Bayesian analysis. Here is an outline of our procedure: we start with a default parameter prior both under the full and the null HWE model; next we derive a class of intrinsic priors on the parameter space of the full model, conditional on the HWE null model. This generates objective Bayesian tests by letting the prior distribution vary over the class of intrinsic priors, thus producing an effective robustness analysis. For the notion of intrinsic priors see [?], [?] and [?]; for recent analyzes of discrete data problems using intrinsic prior methodology see [?] and [?]. We implemented our methodology in an R package called HWEintrinsic available from the Comprehensive R Archive Network (www.r-project.org).

The remainder of the paper is organized as follows. In Section 2 we highlight the sensitivity of Bayesian testing to the choice of priors with specific reference to the HWE problem. In Section 3 we obtain the posterior probabilities of the null and the
alternative model using the intrinsic priors. Section 4 presents some applications to simulated and real data sets. Section 5 contains some concluding remarks.

2 Motivating the Class of Intrinsic Priors for the Hardy-Weinberg Testing Problem

The Hardy-Weinberg testing problem may exhibit a high sensitivity to the prior, as illustrated in the following artificial example.

Example 1: Consider the genotype counts $y = \{y_{11}, y_{12}, y_{22}\} = \{6, 8, 6\}$ for a sample of 20 individuals drawn from a population with two alleles $\{A_1, A_2\}$, and unknown genotype probabilities $\{p_{11}, p_{21}, p_{22}\}$, where $p_{11} + p_{21} + p_{22} = 1$. The full sampling model for this data is a trinomial with $n = 20$ and parameters $\{p_{11}, p_{21}, p_{22}\}$, which reduces to a closely related null model (see Section 3.4) indexed by a scalar parameter $p$ under HWE. As prior for the parameters $\{p_{11}, p_{21}, p_{22}\}$, we consider a symmetric Dirichlet distribution $D(p_{11}, p_{21}, p_{22}|\alpha, \alpha, \alpha)$, which contains all default choices (in particular the uniform and Jeffreys prior). Similarly, we take as prior for the parameter $p$ of the HWE null model the symmetric beta distribution $Be(p|\beta, \beta)$.

For $\alpha = \beta = 1/2$, the posterior probability of HWE is $Pr(HWE|y) = 0.64$, while for $\alpha = \beta = 1$ (uniform prior), we have $Pr(HWE|y) = 0.55$; finally for $\alpha = \beta = 3$ the posterior probability turns out to be $Pr(HWE|y) = 0.45$. Therefore, using the standard convention that an hypothesis is accepted whenever it exceeds 0.5, we would accept HWE if $\alpha = \beta = 1/2$ or $\alpha = \beta = 1$, and reject it when $\alpha = \beta = 3$. In fact, it can be shown that $\inf_{\alpha, \beta} Pr(HWE|y, \alpha, \beta) = 0$ while $\sup_{\alpha, \beta} Pr(HWE|y, \alpha, \beta) = 1$. As a curiosity, we note that the Dirichlet priors with $\alpha = 1$ and $\alpha = 3$ essentially produce
the same Bayesian estimates for the data in Example 1. This shows that the Bayesian
tests do not necessarily share the stability exhibited by the Bayesian estimators.

It is a well-documented fact that testing problems in multinomial families are very
sensitive to the choice of priors. In a series of papers, [?], [?], and [?] analyzed in-
dependence in contingency tables and robustified the Bayesian model by considering
mixtures of Dirichlet distributions with respect to the common hyper-parameter $\alpha$.
The recommended mixing distribution was a log-Cauchy distribution. Different mix-
tures of Dirichlet have also been considered by [?] and [?]. In the latter paper, these
mixtures arise as intrinsic priors for analyzing independence in contingency tables.

Intrinsic priors were initially introduced in order to convert objective priors for es-
timation (typically improper) into suitable priors for testing problems [?, ?]. However,
their scope is wider, and this becomes apparent for discrete data problems, wherein
default priors are usually proper. In this context, the intrinsic prior methodology gives
rise to a natural class of priors for testing nested models when prior information on
the parameters is weak. This class represents a suitable environment for evaluating
the robustness of the resulting test.

2.1 Intrinsic Priors

Consider a general model selection problem between two Bayesian models, $M_0$ (null)
and $M_1$ (full),

$$ M_0 : \{f_0(y|\theta_0), \pi_0(\theta_0)\}, M_1 : \{f_1(y|\theta_1), \pi_1(\theta_1)\}, $$

(1)

with $\theta_0 \in \Theta_0, \theta_1 \in \Theta_1, \Theta_0 \subset \Theta_1$. The family of densities $\{f_0(y|\theta_0), \theta_0 \in \Theta_0\}$ is nested
in the family $\{f_1(y|\theta_1), \theta_1 \in \Theta_1\}$. Finally, $\pi_0(\theta_0)$ and $\pi_1(\theta_1)$ are objective (possibly
improper) priors, such as those typically used for estimation purposes, for instance
reference priors [?].
Note that, although the sampling model \( f_0 \) is nested in \( f_1 \), the objective prior \( \pi_i(\theta_i) \) is \textit{not} related to the objective prior \( \pi_0(\theta_0) \) in \( M_0 \) because \( \pi_i(\theta_i) \) only depends on \( f_i(y|\theta_i) \), \( i = 0, 1 \). This is not reasonable, as we expect some connections between the prior distributions of the parameters \( \theta_0 \) and \( \theta_1 \); see also [?] on the general issue of compatibility of prior distributions for Bayesian model choice. In particular, the prior \( \pi_1(\theta_1) \) will typically not concentrate enough probability mass around the null parameter space \( \Theta_0 \). We will elaborate on this point shortly.

At this stage, it is expedient for the subsequent theoretical developments to abstract from the actual data \( y \) and consider \( t \) independent random variables \( x = (x_1, ..., x_t) \) with joint distribution \( f_1(x|\theta_1, t) = \prod_{j=1}^{t} f_1(x_j|\theta_1) \) under model \( M_1 \) (Notice that the notation now involves explicitly the sample size \( t \) as this will play an important role later on).

We assume that the marginal distribution \( m_1(x|t) = \int f_1(x|\theta_1, t) \pi_1(\theta_1)d\theta_1 \) is strictly positive and finite for an integer \( t \), where \( t \) is called the training sample size. Then, the intrinsic prior for \( \theta_1 \), conditional on the null point \( \theta_0 \) and for the given training sample size \( t \), is defined as
\[
\pi^I_1(\theta_1|\theta_0, t) = \pi_1(\theta_1) \mathbb{E}_{\theta_1|\theta_0} \frac{f_0(x|\theta_0, t)}{m_1(x|t)},
\] (2)
where the expectation is taken with respect to the sampling distribution \( f_1(x|\theta_1, t) \).

Further, integrating out the parameter \( \theta_0 \), we obtain the unconditional intrinsic prior for \( \theta_1 \) as
\[
\pi^I_1(\theta_1|t) = \int \pi^I_1(\theta_1|\theta_0, t) \pi_0(\theta_0)d\theta_0.
\] (3)

The pair of distributions \((\pi_0(\theta_0), \pi^I_1(\theta_1|t))\) represents the intrinsic priors for testing model \( M_0 \) versus \( M_1 \) in (1) based on a training sample of size \( t \). When the prior \( \pi_1 \) is improper, \( t \) is usually taken to be equal to the minimal sample size for which \( m_1(x|t) \) is positive and finite, so that \( \pi^I_1(\theta_1|t) \) exists. However, this restriction on \( t \) is not
necessary and other values of $t$ can be of interest, as we will see below. We also note that $x$ is a random vector which is eliminated by integration, so that the intrinsic prior $\pi_1^I(\theta_1|t)$ only depends on the training sample size $t$.

Intrinsic priors enjoy some interesting properties: i) since $\pi_0$ and $\pi_1$ are objective priors they do not require prior elicitation on the side of the user; ii) the intrinsic prior for $\theta_1$ satisfies the “Savage continuity condition” [?], thus providing a fairer comparison between the two hypotheses under investigation, as will be illustrated for the HWE testing problem. This condition is a widely accepted requirement; see [?], [?], [?], [?], [?]. Furthermore, it has also been argued that, if the null model is a reasonable one, it is important to be able to distinguish $f_0$ from close alternatives; on the other hand, putting prior probability on extreme models, far from $f_0$, will discount the more reasonable alternatives [?]; iii) they are invariant to reparameterizations; iv) for the HWE problem they allow us to assess posterior robustness of the test as $t$ varies in the set of integer $\{1, \ldots, n\}$, where $n$ is the size of the observed sample; this is a crucial point when $n$ is small or moderate; v) in the HWE setting, closed-form expressions for the Bayes factor, and posterior model probabilities, can be provided for intrinsic priors; vi) the testing procedure is consistent. Finally, for large sample sizes, the intrinsic testing procedure can be implemented using an efficient Monte Carlo technique.

3 Testing Hardy-Weinberg Equilibrium Using Intrinsic Priors

We focus attention on a particular locus which is always one of $r \geq 2$ different alleles $\{A_i, i = 1, \ldots, r\}$. 

8
3.1 Sampling Models

Suppose we draw a random sample of \( n \) individuals from this population and denote by \( y_{ij} \) the number of genotypes in the sample of the form \( \{A_i, A_j\} \). There are \( R = r(r+1)/2 \) counts \( \{y_{ij}, 1 \leq j \leq i \leq r\} \) satisfying \( \sum y_{ij} = n \). Then, the probability distribution of the observed triangular array \( y \) under the full model is given by

\[
P_1(y|p_1, n) = \frac{n!}{\prod_{ij} y_{ij}!} \prod_{i \geq j} p_{ij}^{y_{ij}},
\]

(4)

where the probability \( p_{ij} \) of occurrence of the genotype \( \{A_i, A_j\} \) ranges over the space

\[
\Theta_1 = \{p_{ij} : p_{ij} \geq 0, \sum_{1 \leq j \leq i \leq r} p_{ij} = 1\}
\]

having \( \text{dim}(\Theta_1) = R - 1 \). A generic point in the space \( \Theta_1 \) is denoted by \( p_1 \).

Under the Hardy-Weinberg equilibrium law the probabilities \( p_{ij} \) are assumed to belong to the subset \( \Theta_0 \subset \Theta_1 \) defined as

\[
\Theta_0 = \{p_{ij} : p_{ii} = p_i^2, p_{ij} = 2p_ip_j\},
\]

(5)

where \( \{p_i, i = 1, \ldots, r\} \) are such that \( p_i \geq 0 \) and \( \sum_{i=1}^r p_i = 1 \), and represent the allele frequencies in the population. We note that \( \text{dim}(\Theta_0) = r - 1 \).

It is straightforward to verify that under law (5) the sampling model (4) reduces to the null model

\[
P_0(y|p_0, n) = \frac{n!}{\prod_{ij} y_{ij}!} \left[ \sum_{i=1}^r y_{ii} \right]^{n-r} \prod_{i=1}^r p_{ii}^{r-1+c_i},
\]

(6)

where \( p_0 = (p_1, \ldots, p_r) \), \( r_i = \sum_{j=1}^i y_{ij} \) and \( c_i = \sum_{k=r}^i y_{ki} \) are the sum of the \( i \)th row and \( i \)th column, respectively, of the triangular array \( y \).

3.2 Intrinsic Priors

To complete the Bayesian specification of the sampling models (4) and (6) we need prior distributions for the parameters \( p_1 \) and \( p_0 \). We start with the most simple
objective prior distribution for the parameter $p_1$, namely a uniform prior which we can write as $D(p_1|1, ..., 1)$. We proceed similarly with the parameter $p_0$ to which we assign a Dirichlet distribution $D(p_0|1, ..., 1)$. Clearly the dimensions of the two distributions are different, although we do not make this explicit in the notation. Thus, for a random array $x = \{x_{ij}, 1 \leq j \leq i \leq r\}$ having total $t$, that is $\sum x_{ij} = t$, the objective Bayesian models involved in the HWE testing problem are

$$M_0 : \{P_0(x|p_0, t), D(p_0|1, ..., 1)\} \tag{7}$$

and

$$M_1 : \{P_1(x|p_1, t), D(p_1|1, ..., 1)\} \tag{8}$$

Applying (2), we obtain the intrinsic prior for $p_1$, conditional on an arbitrary but fixed null point $p_0$,

$$\pi^{(i)}_1(p_1|p_0, t) = D(p_1|1, ..., 1)E_{x|p_1} P_0(x|p_0, t) m_1(x|t), \tag{9}$$

where the sampling model $P_0$ is defined in (6) and the expectation is taken with respect to the sampling distribution $P_1(x|p_1, t)$ with $P_1$ defined in (4). The unconditional intrinsic prior for $p_1$ is given by

$$\pi^{(u)}_1(p_1|t) = \int \pi^{(i)}_1(p_1|p_0, t) D(p_0|1, ..., 1)dp_0. \tag{10}$$

Recall that the quantity $t$ controls the degree of concentration of the conditional intrinsic prior $\pi^{(i)}_1(p_1|p_0, t)$ around the point $p_0$. Consequently, $t$ also controls the degree of concentration of the prior $\pi^{(u)}_1(p_1|t)$ around the null parameter space $\Theta_0$. 
3.3 Bayes Factor

Given the observed triangular array \( y = \{y_{ij}, 1 \leq j \leq i \leq r, \sum y_{ij} = n \} \), we consider the problem of choosing between the two models

\[
M_0 : \{P_0(y|p_0), D(p_0|1, \ldots, 1)\},
\]
\[
M_1 : \{P_1(y|p_1), \pi_1^t(p_1|t)\},
\]

where \( M_1^I \) stands for the intrinsic Bayesian model (omitting for simplicity its dependence on \( t \)). Assuming the common 0-1 loss function and the model prior \( \Pr(M_0) = \Pr(M_1^I) = 1/2 \), the optimal model is the one having the larger posterior probability, with the model posterior probabilities being given by

\[
\Pr(M_0|y, t) = \frac{1}{1 + B_{10}^I(y, t)}, \quad \Pr(M_1^I|y, t) = 1 - \Pr(M_0|y, t),
\]

(12)

with \( B_{10}^I(y, t) = 1/B_{01}^I(y, t) \), where \( B_{01}^I(y, t) \), the Bayes factor to compare \( M_0 \) and \( M_1^I \), is given by

\[
B_{01}^I(y, t) = \frac{m_0(y)}{m_1^t(y, t)} = \frac{\int P_0(y|p_0, n) D(p_0|1, \ldots, 1)dp_0}{\int P_1(y|p_1, n) \pi_1^t(p_1|t)dp_1}.
\]

Thus, for each \( t \), equation (12) will provide the appropriate answer to the HWE problem.

We will let \( t \) vary between 1 and \( n \) (notice that \( t = 0 \) formally returns the standard analysis with a uniform parameter prior under each of the two models). Condition \( t \leq n \), i.e. the training sample size is less than or equal to the actual sample size, ensures that prior precision does not exceed sample precision. Accordingly, we consider the class of priors \( \{\pi_1^t(p_1|t), t = 1, \ldots, n\} \). We compute the posterior probability of \( H_0 \) at \( t \) varies. The smaller the range of such probabilities, the more robust is the analysis. Additionally, if the curve of posterior probabilities does not cross a
conventional decision threshold (e.g. 0.5), then a stable decision in favor of either hypothesis can be reached.

3.4 Two Alleles

For illustrative purposes we consider first the Hardy-Weinberg problem for the simplest case of two alleles. We start from models (7) and (8) with $r = 2$, where

$$P_0(x|p) = \left(\prod_{i \geq j} x_{ij}\right) 2^{x_{21}} p^{x_{21}} (1-p)^{x_{21}+2x_{22}},$$

while $P_1$ is a trinomial model with two free cell-probabilities $\{p_{11}, p_{21}\}$. The default Bayesian models under consideration are therefore

$$M_0 : \{P_0(x|p,t), \pi_0(p) = 1_{(0,1)}(p)\},$$

$$M_1 : \{P_1(x|p_{11}, p_{21}, t), \pi_1(p_{11}, p_{21}) = D(p_{11}, p_{21}, p_{22}|1,1,1)\},$$

where $1_A$ is the indicator function of the set $A$. The intrinsic prior, conditional on an arbitrary but fixed point $p$, is

$$\pi_I(p_{11}, p_{21}, t) = \frac{t!(t+2)!}{(2t+1)!} \sum_{x_{ij} \in C_2(t)} \frac{2^{x_{21}}}{(x_{11}|x_{21}|x_{22})!} 2^{2x_{21}+x_{22}} (1-p)^{x_{21}+2x_{22}} \prod_{i \geq j} p_{ij}^{x_{ij}},$$

(13)

where $C_2(t) = \{x_{ij} : 1 \leq j \leq i \leq 2, \sum x_{ij} = t\}$.

Figure 1 displays the intrinsic prior given $p = 0.5$ and for the values $t = 5$ and $t = 30$, illustrating how the intrinsic prior concentrates more probability mass around the null value $p = 0.5$ as $t$ increases.

Integrating out the parameter $p$ in expression (13) with respect to the uniform prior, we obtain the unconditional intrinsic prior for $\{p_{11}, p_{21}\}$ as

$$\pi'(p_{11}, p_{21}|t) = \frac{t!(t+2)!}{(2t+1)!} \sum_{x_{ij} \in C_2(t)} \left[ \frac{2^{x_{21}}}{(x_{11}|x_{21}|x_{22})!} (2x_{11}+x_{21})! (x_{21}+2x_{22})! \prod_{i \geq j} p_{ij}^{x_{ij}} \right].$$

The null parameter space $\Theta_0$, which is now a curve in the plane, is plotted in Figure 2. The picture of the intrinsic prior $\pi'(p_{11}, p_{21}|t)$ for $t = 30$ is given in Figure 3 and
Figure 1: Conditional intrinsic prior of \((p_{11}, p_{21})\) given \(p = 0.5\), with two distinct degrees of concentration on the null corresponding to \(t = 5\) (left panel) and \(t = 30\) (right panel).

Figure 2: Null space \(\Theta_0\) for the two alleles case; \(\Theta_0 = \{p_{11} = p^2, p_{21} = 2p(1 - p); 0 \leq p \leq 1\}\).
it shows how the prior concentrates mass around $\Theta_0$.

For the observed triangular array $y = \{y_{ij}, 1 \leq j \leq i \leq 2, \sum y_{ij} = n\}$ the marginal distribution of $y$ under model $M_0$ is

$$m_0(y) = \frac{n!}{y_{11}! y_{21}! y_{22}!} \frac{2^{y_{21}} (2y_{11} + y_{21})!(2y_{21} + y_{22})!}{(2n + 1)!}, \quad (14)$$

while under $M_1^\ell$ is

$$m_1^\ell(y|t) = \frac{\ell! (t + 2)!}{(2t + 1)!(t + n + 2)!} \prod_{i \geq j} n! \sum_{x_{ij} \in C_2(t)} \left[ \frac{2^{x_{21}}}{(x_{11}! x_{21}! x_{22}!)^t} (2x_{11} + x_{21})! \right. \left. \times (x_{21} + 2x_{22})! (x_{11} + y_{11})! (x_{21} + y_{21})! (x_{22} + y_{22})! \right] \quad (15)$$

Both the Bayes factor and the posterior probabilities of model $M_0$ and $M_1^\ell$ are now computable.
For the data set used in Example 1, the range of posterior probabilities for the HWE model, as the intrinsic prior varies in the class \( \{ \pi\, p_{11}, p_{21} | t \in \{1, \ldots, 20\} \} \), turns out to be \( \min_{1 \leq t \leq 20} P(M_0 | y, t) = 0.53, \max_{1 \leq t \leq 20} P(M_0 | y, t) = 0.55 \). Therefore, the optimal decision when using the intrinsic priors is to accept HWE, and this decision is robust in the class of intrinsic priors.

### 3.5 The General Case of \( r \) Alleles

Consider models (7) and (8). Using (10), the intrinsic prior for the parameter \( p_1 \) is

\[
\pi^I(p_1 | t) = \frac{(t + R - 1)! (r - 1)! t! \sum_{x_{ij} \in C_r(t)} 2^{t-x_{ij}} \prod_{i=1}^{r} \left( z_i + d_i \right)! \prod_{i \geq j} x_{ij}!}{(2t + R - 1)!} \times \prod_{i \geq j} p_{ij}^{x_{ij}},
\]

where \( C_r(t) = \{ x_{ij} : x_{ij} \geq 0, 1 \leq j \leq i \leq r, \sum x_{ij} = t, \} \) and \( R = r(r + 1)/2 \).

For the observed sample \( y = \{ y_{ij}, 1 \leq j \leq i \leq r, \sum y_{ij} = n \} \), the marginal under model \( M_0 \) becomes

\[
m_0(y) = \frac{(r - 1)! n!}{(2n + r - 1)!} \prod_{i=1}^{r} \left( r_i + c_i \right)! \prod_{i \geq j} y_{ij}!,
\]

where \( r_i = \sum_{j=1}^{i} y_{ij} \) and \( c_i = \sum_{k=1}^{r} y_{ki} \), the sum of the ith row and ith column of the triangular data array. The marginal data distribution under the intrinsic Bayesian model \( M_1^I \) turns out to be

\[
m_1^I(y | t) = \frac{t!(t + R - 1)!(r - 1)!}{(2t + R - 1)!(n + t + R - 1)!} \prod_{i \geq j} y_{ij}! \sum_{x_{ij} \in C_r(t)} 2^{t-x_{ij}} \prod_{i=1}^{r} \left( z_i + d_i \right)! \prod_{i \geq j} x_{ij}! \prod_{i \geq j} \left( x_{ij} + y_{ij} \right)!,
\]

where \( z_i = \sum_{j=1}^{i} x_{ij} \) and \( d_i = \sum_{k=1}^{r} x_{ki} \).

The ratio \( m_0(y) / m_1^I(y | t) \) provides the Bayes factor \( B_{01}^I(y | t) \) to compare \( M_0 \) and \( M_1^I \) based on the intrinsic prior. Using (12) we obtain the posterior probability of model \( M_0 \) and \( M_1^I \) for the given training sample size \( t \).
3.6 Computation of the Bayes Factor

From the structure of $m_1(y|t)$ given in (18) it is clear that the computation of $B_{I0}^I(y|t)$ can become rapidly infeasible if the training sample size $t$ is large (this may easily be the case if $n$ is large because we would like to let $t$ range over the grid $1, \ldots, n$ to evaluate robustness). To overcome this difficulty, we calculate $B_{I0}^I(y, t)$ using a Monte Carlo method, following the approach outlined in [?]. In particular, we use an importance sampling strategy in order to speed up convergence. We choose as candidate distribution a specific $R$-dimensional multinomial $x \sim \text{Multinomial}(t, \hat{p}_1)$ (19) with probabilities $\hat{p}_{ij} = \frac{y_{ij} + 1}{n + R}$, $i = 1, \ldots, r$, $j = 1, \ldots, i$. (20)

We then generate $M$ random deviates $x^{(k)}$, $k = 1, \ldots, M$, from (19) and approximate the Bayes factor using a Monte Carlo average; specifically we evaluate

$$\hat{B}_{I0}^I(y|t) = \frac{1}{M} \sum_{k=1}^{M} \frac{t! (t + R - 1)! (2n + r - 1)!}{(2t - \sum_{i=1}^{r} x_{i}^{(k)})! \prod_{i=1}^{r} (r_i + c_i)!} \sum_{k=1}^{M} \left[ \prod_{i \geq j} \left( x_{ij}^{(k)} + y_{ij} \right) \right] \frac{1}{\prod_{i \geq j} \hat{p}_{ij}^{[x_{ij}^{(k)}]}}.$$ (21)

Apart from $M$ (the number of Monte Carlo iterations), convergence achievement appears to depend critically on the data sparseness, (more sparse observations lead to slower convergence), and the training sample size $t$ (higher values of $t$ typically require a higher number of Monte Carlo iterations $M$ to reach convergence). However, even with a considerable number of iterations (in most of the applications presented in the
next section $M$ is fixed at 300,000), the computational burden of the algorithm is still very reasonable.

4 Examples

In this section we illustrate some features of the testing procedure developed in Section 3 through some examples which for illustration purposes are confined to the case of two-alleles; next we analyze two data sets for the case of multiple alleles also discussed in [?].

4.1 Two Alleles

We consider four different data sets each referring to $n = 20$ artificial observations, and the corresponding counts for the genotypes $(A_1A_1, A_2A_1, A_2A_2)$ are:

- Data set 1: $(3, 9, 8)$,
- Data set 2: $(8, 2, 10)$,
- Data set 3: $(12, 5, 3)$,
- Data set 4: $(2, 13, 5)$.

These data sets have been selected to highlight and better appreciate some features of our methodology. The results are summarized in Figure 4, wherein each row represents a specific data set ordered from top to bottom as above. The first three columns refer to selected intrinsic priors indexed by the fractions $f = 0.1, 0.5, 1$, where $f = t/n$ and $n$ is the actual sample size. Each panel in these three columns report the contour lines of the intrinsic prior (solid black) and those of the (normalized) likelihood (solid gray), together with the null parameter space $\Theta_0$ (represented by the dashed curve). Every individual panel in the last column reports the posterior probability of $M_0$ using
the Monte Carlo approach described in (21) with $M = 300,000$ iterations together with the corresponding exact curve obtained from (14) and (15) (dashed gray): the latter is mostly barely visible because of the excellent approximation provided by our Monte Carlo method. The horizontal dotted line represents the posterior probability of the null model derived from the standard Bayes factor computed using the uniform prior both under the full and the null model. Notice that this value is actually a special case of the Bayes factor based on the intrinsic prior because it can be formally obtained by setting $t = 0$. Focus now on the first data set, i.e. row one of Figure 4. The observations are in very good agreement with the null hypothesis, as one can gather from the fact that the null parameter space intersects the highest likelihood contours. As $f$ increases, panels (a) to (c), the highest contour lines of the intrinsic prior move towards the center of the likelihood surface, thus increasing the marginal data distribution under $M_1, m_1^t(y|t)$, and hence taking away evidence from $M_0$. This explains the monotone decreasing behavior of the null posterior probability in panel (d). Despite the very good agreement of the likelihood with the null hypothesis, the posterior probability of HWE does not exceed 67%: this is due to the fairly moderate sample size. Finally, the small range of variation of the posterior probability of $M_0$, always well above the conventional 0.5-threshold, provides a robust conclusion in favor of the null hypothesis. Consider now the second data set, i.e. row two of Figure 4. It is evident that the null hypothesis is not supported by the data, because the likelihood surface is not crossed by the $\Theta_0$-curve, see panels (e)-(g). This is reflected in the very small values of the posterior probability of $M_0$ which never exceeds 1.2%, see panel (h). The monotone increasing behavior of the curve in the latter panel is easily explained as follows: by increasing $f$ the intrinsic prior is pulled towards $\Theta_0$ and thus away from the likelihood surface as it is apparent from the sequence of panels (e),
(f) and (g). In this way, \( m_1^i(y|t) \) decreases, thus taking away evidence from \( M_1^i \), equivalently adding strength to \( M_0 \). Notice that the support for the HWE model remains negligible throughout the range \( 0 \leq f \leq 1 \), so that a highly robust conclusion against \( M_0 \) can be drawn in this case, too. The third data set, which is represented in the third row of the same figure, presents a non-monotone null posterior probability curve, see panel (l). Here the data are in some accord with the null, because the \( \Theta_0 \)-curve is somewhat tangential to the likelihood surface. Moving from \( f = 0.1 \) - panel (i) - to \( f = 0.5 \) - panel (j), the intrinsic prior starts peaking around \( \Theta_0 \), and in so doing it hits some high likelihood levels, thus producing a reduction in the evidence for \( M_0 \). However by further increasing \( f \) the curve further shrinks on \( \Theta_0 \) but this makes it capture only some peripheral likelihood levels - panel (k) for \( f = 1 \), thus removing evidence from \( M_1 \) and making the posterior probability of \( M_0 \) increase again. The final answer is again robust, because the posterior probability of Hardy-Weinberg equilibrium is always inside the \((0.33, 0.35)\) interval. Finally, let us turn to the fourth row which analyzes the last data set. Similarly to the previous case, the data are in some accord with the null, because there is some overlapping between the null-curve and the likelihood surface, although this occurs only for low-level contours. As we move from panel (m) to (n) some evidence in favor of \( M_1 \) is lost because high level prior contours retract towards \( \Theta_0 \), thus making the posterior probability of \( M_0 \) increase, albeit very slightly. A further increase in \( f \) makes \( M_1 \) more competitive because some high prior contours make their way inside the likelihood surface. The actual changes to the posterior probability of \( M_0 \), panel (p), are however very small, and thus the conclusion is still robust, although evidence for \( M_0 \) is not markedly below the conventional 50%-threshold.
Figure 4: Posterior probability of Hardy-Weinberg equilibrium. First three columns: each panel reports the intrinsic prior (solid black) and the normalized likelihood (solid gray) contours together with the null parameter space $\Theta_0$ (dashed black). Last column: Posterior probability of Hardy-Weinberg equilibrium and comparison between Monte Carlo averages based on $M = 300,000$ iterations (solid black) and exact values (dashed gray) for $0 \leq f = t/n \leq 1$. 

$\bar{t}/n = 0.1$  $\bar{t}/n = 0.5$  $\bar{t}/n = 1$  Null Posterior Prob.
4.2 Multiple Alleles

EXAMPLE 2: This example is concerned with a population of \( r = 8 \) alleles and the
data given in Table 1 represent a sample of size \( n = 30 \) of genotype frequencies
simulated under the Hardy-Weinberg equilibrium when the underlying gene frequencies
are \( .2, .2, .2, .05, .05, .05, .05 \); see \([?, \text{Example 2}]\).

Table 1: Simulated data from \([?, \text{Example 2}]\).

<table>
<thead>
<tr>
<th>(A_1)</th>
<th>(A_2)</th>
<th>(A_3)</th>
<th>(A_4)</th>
<th>(A_5)</th>
<th>(A_6)</th>
<th>(A_7)</th>
<th>(A_8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_2)</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_3)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_4)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_5)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_6)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(A_7)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(A_8)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 5 reports the intrinsic posterior probabilities of Hardy-Weinberg equilibrium
calculated using the Monte Carlo approach \((21)\) with \( M = 300,000 \) iterations together
with the value derived from the standard Bayes factor using uniform priors. Notice
that the curve is (essentially) monotone decreasing. The rationale for this behavior is
analogous to that described for Data set 1 in Subsection 4.1, because the data clearly
support the Hardy-Weinberg equilibrium model \( M_6 \). The intrinsic testing approach
provides strong evidence in favor of the null model. The null posterior probabilities,
Figure 5: Guo and Thompson (1992); \( r = 8 \) alleles; \( n = 30 \); data simulated under Hardy-Weinberg equilibrium. Posterior probability of Hardy-Weinberg equilibrium: Monte Carlo averages based on \( M = 300,000 \) iterations as a function of the ratio \( f = t/n \). The horizontal dashed line is derived from the standard Bayes factor based on uniform priors.

in fact, range from 0.80 to 0.97.

**Example 3.** These data concern the antigen class of 45 French type 1 diabetes patients, with the classes being DR1, DR3, DR4, and Y, a fourth class corresponding to all other antigens. The counts \( (r = 4, n = 45) \) are given in Table 2. These data are discussed in [?, Example 1] and [?]; the latter in particular, reported an exact \( p \)-value of 0.01744, which under conventional levels would indicate some evidence against the Hardy-Weinberg model. Interest here centers in the mode of inheritance of type 1 diabetes, with a hypothesized recessive model being equivalent to the HWE model; see [?, Section 4.1]. This example is interesting because it reveals how the HWE model
can be usefully adapted to a specific scientific context.

Table 2: Genotype frequency data from [?].

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>5</td>
<td>18</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

The null posterior probabilities using the intrinsic prior methodology, as well as the standard approach, are reported in Figure 6. The behavior of the curve, and its explication, is analogous to that described for Data set 2 in Subsection 4.1. The posterior probabilities of the null range from 0.07 to 0.10, thus providing robust substantial evidence against HWE; see [?] for a description of the scale of evidence against the null hypothesis in line with Jeffreys’ recommendations. [?, Section 4.1] also analyzed these data using an informative prior; he obtains a Bayes factor in favor of HWE equal to 0.074 which, with prior odds equal to one, translates to a posterior probability of HWE of 0.07: this coincides with our lower bound. We are thus able to replicate his findings within our objective framework without any prior elicitation.

We also analyzed a more elaborate data set consisting of nine alleles and 8297 individuals (see [?]), thus showing that our method scales up nicely. We do not report the results here because they are wholly comparable with those obtained by [?, Section 4.2], showing that for these data there is overwhelming evidence in favor of the HWE hypothesis. The interested reader can find all the routines and the examples (including the nine alleles one) used to prepare this paper in the R package called HWEintrinsic.
Figure 6: Guo and Thompson (1992, Example 1); \( r = 4; n = 45 \). Posterior probability of Hardy-Weinberg equilibrium: Monte Carlo averages based on \( M = 300,000 \) iterations as a function of the ratio \( f = t/n \). The horizontal dashed line is derived from the standard Bayes factor based on uniform priors available from the Comprehensive R Archive Network (www.r-project.org).

5 Concluding remarks

The Hardy-Weinberg equilibrium law has received considerable attention in recent years, thanks to the increasing availability of genetic data. For instance, [?] claim that detecting departures from Hardy-Weinberg equilibrium of marker-genotype frequencies is a crucial first step in almost all human genetic analyzes; for a related viewpoint see also [?]. It seems therefore appropriate to apply recently developed testing procedures to this problem.

The Bayesian approach to hypothesis testing and model comparison has been ad-
versely affected over the years because of its high sensitivity to prior specification. This feature can be problematic, especially for scientific communication, where objective, or at least benchmark, analyzes are preferred. Unfortunately standard default reference priors do not work for Bayesian hypothesis testing (for instance, the Bayes factor is not even well-defined when the priors are improper). However, it is by now recognized that the methodology based on intrinsic priors represents a sound and viable alternative, especially for nested models. Intrinsic priors are suitably tailored to the hypothesis under investigation and produce sensible Bayes factors. One important reason for this highly satisfactory behavior is that parameter values close to the null receive higher probability mass under the intrinsic prior, a natural desideratum as recognized by several authors.

In this paper we have developed an objective Bayesian testing procedure making systematic use of the notion of intrinsic priors. It turns out that a whole class of intrinsic priors, governed by a single scalar quantity (the training sample size $t$) is a natural class of priors for assessing robustness of the test. The quantity $t$ naturally acts as a concentration parameter (around the null subspace) for the prior under the full model. Making $t$ vary between 1 and the observed sample size $n$, we are able to obtain a whole range of posterior probabilities for the null model which provides a natural sensitivity analysis. The smaller the range of such probabilities, the more robust is the analysis. A separate issue concerns whether the curve does, or does not, cross the 0.5 threshold (or whatever other level is deemed appropriate, depending on the loss function, to make a decision in favor of either hypothesis). If the threshold is crossed, then the experimental evidence does not allow to choose between the two hypotheses, signalling that more data are needed. The \texttt{HWEintrinsic R} package, which implements the methods presented in this paper, is available from the Comprehensive R Archive
Network (www.r-project.org).

Acknowledgements The first author was partially financed by MIUR (Italy), PRIN grant #2007XECZ7L001 and the University of Pavia. The second author was supported by Ministerio de Ciencia y Tecnologa, Grant SEJ-65200 and Junta de Andaluca Grant SEJ-02814. Comments from two referees and an associate editor have helped improve the presentation of the paper.