Statistical Papers
https://doi.org/10.1007/s00362-022-01301-3

REGULAR ARTICLE

A measure of evidence based on the likelihood-ratio statistics

Alexandre Galvão Patriota[1]

Received: 11 May 2021 / Revised: 26 November 2021 / Accepted: 15 February 2022

Abstract
In this paper, we show that the likelihood-ratio measure (a) is invariant with respect to dominating sigma-finite measures, (b) satisfies logical consequences which are not satisfied by standard p values, (c) respects frequentist properties, i.e., the type I error can be properly controlled, and, under mild regularity conditions, (d) can be used as an upper bound for posterior probabilities. We also discuss a generic application to test whether the genotype frequencies of a given population are under the Hardy–Weinberg equilibrium, under inbreeding restrictions or under outbreeding restrictions.

Keywords Classical statistics · Coherency · Evidence measure · Hypothesis testing · Likelihood ratio statistics · Monotonicity · Possibility theory

1 Introduction

The likelihood ratio (LR) approach has a long history in the statistical literature. Neyman and Pearson (1933) and Karlin and Rubin (1956) demonstrated that the test based on the likelihood ratio statistics is the most powerful for a fixed level of significance under simple hypotheses and monotone LR functions, respectively. Birkes (1990) studied the relation between generalized LR tests and uniformly most powerful tests. There is a rich literature about this topic, we refer the reader to Mudholkar and Chaubey (2009), Sprott (2000), Severini (2000), Royall (1997), Royall (2000), Giant and Shenoy (2005), Blume (2008) and Bickel (2012).

The LR statistic is widely known in the classical statistical literature by its optimal properties and its asymptotic distributions, even under nonstandard regular conditions, that can be employed to make approximate inferences (Chernoff 1954; Drton 2009). Typically, p values, test functions and confidence regions can be computed from

---

✉ Alexandre Galvão Patriota
patriota@ime.usp.br

[1] Departamento de Estatística, IME, Universidade de São Paulo, Rua do Matão, 1010, São Paulo, SP 05508-090, Brazil

the asymptotic distribution of this statistic. In this paper, we define the LR measure, discuss some of its random and non-random properties and show that it can be used as an upper bound for the posterior Bayesian probability. An upper bound for the posterior probability can be useful under high dimensional problems where the integration is a cumbersome task. As we shall see, the LR measure is the LR statistic defined as a set function over the subsets of the parameter space.

This LR-measure can be easily applied to verify whether the genotype frequencies of a given population is under the Hardy–Weinberg Equilibrium (HWE). Typically there are three situations for an autosomal biallelic marker: (1) the population is under the HWE; (2) the population is undergoing a regular system of 'inbreeding' (when relatives produce offspring); and (3) the population is undergoing a regular system of 'outbreeding' (when genetically different individuals produce offspring). In order to see these three conditions in terms of the population probabilities, let AA, Aa and aa be the possible genotypes and θ_1, θ_2 and θ_3 their respective population probabilities, where $\theta_1 + \theta_2 + \theta_3 = 1$. From this restriction, we can express the parameter space by using only two parameters, e.g., θ_1 and θ_3 as shown in Fig. 1.

Figure 1 illustrates the HWE in the dashed line, the inbreeding condition in the dotted area below the dashed line and the outbreeding condition in the white area above the dashed line. Let y_1, y_2 and y_3 be the observed frequencies for the respective genotypes, where $y_1+y_2+y_3 = s$, with $s \in \mathbb{N}$. Based on the observed frequencies, one should be able to measure the likelihood the studied population is under the HWE. The farther the point of maximum likelihood $\left(\frac{y_1}{s}, \frac{y_3}{s}\right)$ is from the dashed line, the more implausible would be the hypothesis of Hardy–Weinberg Equilibrium. The degree of implausibility is usually measured by frequentist measures such as p values and confidence regions (see, for instance, Emigh 1980; Graffelman and Weir 2016) and Bayesian procedures (Lindley 1988; Shoemaker et al. 1998; Puig et al. 2017). In this paper, we develop some theoretical results for the likelihood ratio measure and apply them to the HWE problem in Sect. 4.

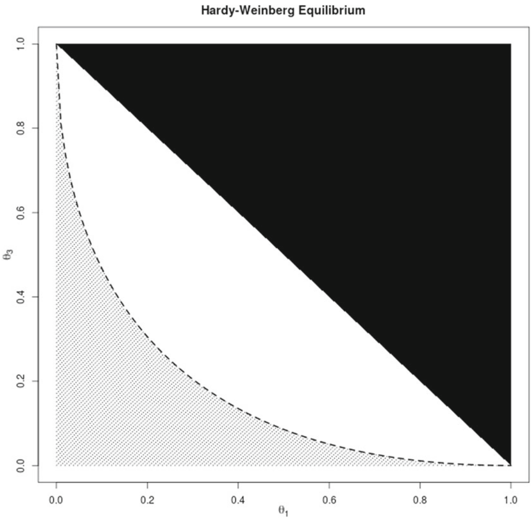

Fig. 1 The dashed curve represents the Hardy–Weinberg equilibrium. The dotted area below the dashed line represents the inbreeding condition and the white area above the dashed line represents the outbreeding condition

The rest of this paper is organized as follows. Section 2 introduces the likelihood-ratio measure, some of its properties and two examples. Section 3 provides a comparison with the Bayesian approach. Section 4 presents an application to the Hardy–Weinberg Equilibrium and Sect. 5 offers Monte Carlo Simulations to illustrate the theoretical results on the Hardy–Weinberg application. Finally, Sect. 6 concludes the paper with some final remarks.

2 The likelihood ratio measure

In order to avoid ambiguity, we adopt the set-measure theory to introduce the likelihood function. Throughout this paper, the statistical model is the triplet $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, where $\mathcal{X} \subseteq \mathbb{R}^n$ is the sample space, $\mathcal{F}$ is a (Borel) sigma-algebra of subsets of $\mathcal{X}$ and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a family of probability measures on $(\mathcal{X}, \mathcal{F})$ that possibly explain the observed data, where Θ is an arbitrary index set. Each measure $P_\theta \in \mathcal{P}$ is dominated by a sigma-finite measure μ, which we represent by $\mathcal{P} \ll \mu$, that is, $\mu(A) = 0 \Rightarrow P_\theta(A) = 0$ for all $P_\theta \in \mathcal{P}$ and $A \in \mathcal{F}$. If $\Theta \subseteq \mathbb{R}^k$, with $k \in \mathbb{N}$, then the model is said to be parametric; otherwise, the model is said to be nonparametric.

Definition 2.1 The likelihood function is defined to be one version of the Radon–Nikodym derivative

$$L_\mu(\theta, \cdot) \in \frac{dP_\theta}{d\mu} \quad \text{such that} \quad P_\theta(A) = \int_A L(\theta, x) d\mu(x),$$

where $A \in \mathcal{F}$, $P_\theta \in \mathcal{P}$ and $\frac{dP_\theta}{d\mu}$ is a set containing all versions that are equal except for a set of μ-measure zero.

The likelihood function is a non-negative function that connects each probability measure $P_\theta \in \mathcal{P}$ with an observable data $x \in \mathcal{X}$ through a version of the Radon–Nikodym derivative with respect to a sigma-finite measure μ. According to the law of likelihood, the larger is the value of $L_\mu(\theta, x)$, the better is the agreement between the probability P_θ and the observed data (Hacking 1965, p. 71). In the discrete case, μ is the counting measure and the likelihood function is $L_\mu(\theta, x) = P_\theta(X = x)$, then $L_\mu(\cdot, x)$ ranks the probabilities in $\mathcal{P}$ from those that make x most probable to those that make x less probable. In the continuous case, μ is the Lebesgue measure and the likelihood function $L_\mu(\cdot, x)$ ranks the probabilities in $\mathcal{P}$ from those with the high instantaneous rate of change at x to those with small instantaneous rate of change at x. We index the likelihood function by the sigma-finite dominating measure μ because different sigma-finite dominating measures may generate different likelihood functions.

Let $\Upsilon = \{\mu : \mathcal{P} \ll \mu\}$ be the set of all sigma-finite dominating measures for the family $\mathcal{P}$. Throughout this paper, the likelihood function $L_\mu(\theta, \cdot)$ must satisfy

$$0 < \sup_{\theta \in \Theta} L_\mu(\theta, x) < \infty, \quad \forall x \in \mathcal{X}, \ \forall \mu \in \Upsilon. \tag{C$_1$}$$

Condition (C_1) guarantees the validity of some properties discussed in this paper. If this conditions is not valid, then the family $\mathcal{P}$ in the statistical model should be replaced by an appropriate family. In what follows, we define the LR function and the LR measure.

Definition 2.2 The likelihood-ratio function with respect to μ is

$$\lambda_\mu(\theta, x) = \frac{L_\mu(\theta, x)}{\sup\limits_{\theta' \in \Theta} L_\mu(\theta', x)}, \tag{2.1}$$

for $\theta \in \Theta$ and $x \in \mathcal{X}$.

The theorem below guarantees that the likelihood ratio function does not depend on a specific dominating sigma-finite measure μ.

Theorem 2.1 *Let $\mu_1, \mu_2 \in \Upsilon$. Under Condition ($C_1$), there exists $A \in \mathcal{F}$ such that $P_\theta(A) = 1$, $\forall \theta \in \Theta$, and*

$$\lambda_{\mu_1}(\theta, x) = \lambda_{\mu_2}(\theta, x), \quad \forall x \in A, \ \forall \theta \in \Theta.$$

Proof Let L_{μ_1} and L_{μ_2} be two likelihood functions based on μ_1 and μ_2, respectively, where $\mu_1, \mu_2 \in \Upsilon$. Therefore, according to the The Likelihood Proportionality Theorem derived by Gonçalves and Franklin (2021), there must exist a measurable function h and a set A such that $P_\theta(A) = 1$, for all $\theta \in \Theta$, and

$$L_{\mu_1}(\theta, x) = h(x) L_{\mu_2}(\theta, x), \quad \forall \theta \in \Theta, \ \forall x \in A.$$

Then, by Condition (C_1),

$$\lambda_{\mu_1}(\theta, x) = \frac{L_{\mu_1}(\theta, x)}{\sup\limits_{\theta' \in \Theta} L_{\mu_1}(\theta', x)} = \frac{L_{\mu_2}(\theta, x)}{\sup\limits_{\theta' \in \Theta} L_{\mu_2}(\theta', x)} = \lambda_{\mu_2}(\theta, x),$$

for all $\theta \in \Theta$ and $x \in A$. □

Corollary 2.1 *Assume valid Condition (C_1), and let A be the set defined in Theorem 2.1, where $x \in A$. If there exists $\mu^* \in \Upsilon$ such that a version $L_{\mu^*}(\cdot, x)$ is continuous on Θ, then $\lambda_\mu(\cdot, x)$ is continuous on Θ for all $\mu \in \Upsilon$.*

Proof It is a direct application of Theorem 2.1. □

As the likelihood-ratio function is well-defined and does not depend on any specific dominating measure, next we define the likelihood measure without indexing by μ.

Definition 2.3 Let A be the set defined in Theorem 2.1. For each fixed $x \in A$, let $\nu_x : 2^\Theta \to [0, 1]$ be a set function such that

$$\nu_x(\Theta_1) = \sup_{\theta \in \Theta_1} \lambda_\mu(\theta, x) \quad \text{and} \quad \nu_x(\varnothing) = 0, \tag{2.2}$$

where $\Theta_1 \subseteq \Theta$ is a non-empty set and $\mu \in \Upsilon$ is any sigma-finite dominating measure. We say that ν_x is the likelihood-ratio measure (LR-measure).

The following properties of ν_x are derived straightforwardly from its definition. Let $\Theta_1, \Theta_2 \subseteq \Theta$, then

$\mathbb{P}_1$. $\nu_x(\Theta) = 1$ and $\nu_x(\varnothing) = 0$;
$\mathbb{P}_2$. $\nu_x(\Theta_1 \cup \Theta_2) = \max\{\nu_x(\Theta_1), \nu_x(\Theta_2)\}$;
$\mathbb{P}_3$. if Θ_1 is non-empty, then $\nu_x(\Theta_1) = \sup_{\theta \in \Theta_1} \nu_x(\{\theta\})$;
$\mathbb{P}_4$. if $\Theta_1 \subseteq \Theta_2$, then $\nu_x(\Theta_1) \leq \nu_x(\Theta_2)$.

Property $\mathbb{P}_4$ is known as the entailment condition or monotonicity. Properties $\mathbb{P}_1 - \mathbb{P}_4$ indicate that ν_x is a possibility measure rather than a probability one, which has non-additive properties; see Denneberg (1994) for more details on non-additive measures and Zadeh (1978) for details on possibility measures (see also Puhalskii 2001).

In order to interpret a ν_x-value, assume that there exists at least one $\theta \in \Theta$ such that $\nu_x(\{\theta\}) = 1$, then:

> *the quantity* $\nu_x(\Theta_0) = \nu_0$ *indicates that the highest likelihood produced by the elements of* Θ_0 *does not exceed* $\nu_0 \cdot 100\%$ *of the largest likelihood.*

A small value of $\nu_x(\Theta_0)$ indicates that all cases in Θ_0 produce small likelihoods compared to the largest likelihood (if it exists). In this context, as the largest likelihood occurs in Θ, the complement Θ_0^c, where $\Theta_0^c = \Theta - \Theta_0$, must have full possibility, i.e., $\nu_x(\Theta_0^c) = 1$. In other words, a small value of $\nu_x(\Theta_0)$ suggests that Θ_0 is "inconsistent with the observed data x" since the observed data are providing evidence against all elements of Θ_0 and no evidence against its complement. It does not mean however that the observed data are providing evidence in favor of its complement Θ_0^c because other models not in $\mathcal{P}$ could be also fully possible. This feature suggests that the LR measure may be used as a degree of inconsistency between models and the observed data as has been advocated by Mudholkar and Chaubey (2009), Sprott (2000), Severini (2000), Royall (1997), Royall (2000), Giant and Shenoy (2005), Blume (2008) and Bickel (2012).

The next theorem provides a threshold to control (asymptotically) the type I error at significance level $\alpha \in (0, 1)$ for testing a general null hypothesis $H_0 : \theta \in \Theta_0$ under some regularity conditions.

Theorem 2.2 *Let* $\Theta_0 \subseteq \Theta$ *be nonempty and* X *be the random sample. Assume that* $-2\log(\nu_X(\Theta_0))$ *converges in distribution to* $Y \sim F_{\Theta_0}$*, as* $n \to \infty$*, for each* $\theta \in \Theta_0$ *where* F_{Θ_0} *is a continuous cumulative distribution. Then,*

$$\lim_{n \to \infty} P_\theta\left(\nu_X(\Theta_0) \leq m(\alpha, \Theta_0)\right) = \alpha$$

for each $\alpha \in (0, 1)$ *and* $\theta \in \Theta_0$*, where* $m(\alpha, \Theta_0) = \exp\left(-0.5 F_{\Theta_0}^{-1}(1 - \alpha)\right)$ *does not depend on* n*.*

Proof Let $\alpha \in (0, 1)$ and $\theta \in \Theta_0$,

$$\begin{aligned}
&\lim_{n\to\infty} P_\theta\Big(\nu_X(\Theta_0) \leq \exp(-0.5F^{-1}_{\Theta_0}(1-\alpha))\Big)\\
&\quad= \lim_{n\to\infty} P_\theta\Big(-2\log(\nu_X(\Theta_0)) \geq F^{-1}_{\Theta_0}(1-\alpha)\Big)\\
&\quad= 1 - F_{\Theta_0}(F^{-1}_{\Theta_0}(1-\alpha))\\
&\quad= \alpha.
\end{aligned}$$

□

In Theorem 2.2, it is assumed that the asymptotic distribution under the null hypothesis F_{Θ_0} does not depend on θ, but in some nonstandard cases, where Θ_0 includes the boundary of the parameter space or singularities, it may depend (see Vu and Zhou 1997; Drton 2009; Patriota and Alves 2021). Under the conditions of Theorem 2.2, the analyst may set the threshold value at $m(\alpha, \Theta_0)$ to control the type I error at significance level α for testing $H_0 : \theta \in \Theta_0$ with the rejection rule: "reject H_0 if $\nu_x(\Theta_0) \leq m(\alpha, \Theta_0)$". This procedure is equivalent to the one based on the asymptotic p value

$$\text{p value}(\Theta_0, x) = 1 - F_{\Theta_0}(-2\log(\nu_x(\Theta_0)))$$

with the rejection rule: "reject H_0 if p value$(\Theta_0, x) \leq \alpha$". As showed by Schervish (1996) and Patriota (2013), the p value does not satisfies property $\mathbb{P}_4$ (the entailment condition or monotonicity) and, therefore, the above procedures may generate conflicting conclusions on different null hypotheses. For instance, if $\tilde{H}_0 \to H_0$, then we would expect that "rejection of H_0" $\implies$ "rejection of $\tilde{H}_0$" with the same data, but (exact and asymptotic) p values do not satisfies this feature (see, for instance, Schervish 1996; Patriota 2013, 2017; Patriota and Alves 2021). As the LR measure satisfies property $\mathbb{P}_4$, we can find another threshold value to avoid this feature and whose rejection rule satisfies the expected logical consequence and controls properly the type I error. The following theorem presents this value by employing the asymptotic distribution of $-2\log(\nu_X(\{\theta\}))$ rather than $-2\log(\nu_X(\Theta_0))$.

Theorem 2.3 *Let $\Theta_0 \subseteq \Theta$ be nonempty and X be the random sample. Assume that $-2\log(\nu_X(\{\theta\}))$ converges in distribution to $Y \sim F_\theta$, as $n \to \infty$, for each $\theta \in \Theta$ where F_θ is a continuous cumulative distribution. Then,*

$$\lim_{n\to\infty} P_\theta\Big(\nu_X(\Theta_0) \leq m(\alpha)\Big) \leq \alpha$$

for each $\alpha \in (0, 1)$ and $\theta \in \Theta_0$, where $m(\alpha) = \exp\Big(-0.5 \sup_{\theta\in\Theta} F^{-1}_\theta(1-\alpha)\Big)$ does not depend on n.

Proof Notice that

$$\left\{ -2\log(\nu_X(\Theta_0)) \geq \sup_{\theta\in\Theta} F_\theta^{-1}(1-\alpha)\right\} \subseteq \left\{ -2\log(\nu_X(\{\theta\})) \geq F_\theta^{-1}(1-\alpha)\right\}$$

for each $\alpha \in (0, 1)$ and $\theta \in \Theta_0$. Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} P_\theta\left(\nu_X(\Theta_0) \leq \exp(-0.5 \sup_{\theta\in\Theta} F_\theta^{-1}(1-\alpha))\right)\\
&= \lim_{n\to\infty} P_\theta\left(-2\log(\nu_X(\Theta_0)) \geq \sup_{\theta\in\Theta} F_\theta^{-1}(1-\alpha)\right)\\
&\leq \lim_{n\to\infty} P_\theta\left(-2\log(\nu_X(\{\theta\})) \geq \sup_{\theta\in\Theta} F_\theta^{-1}(1-\alpha)\right)\\
&\leq \lim_{n\to\infty} P_\theta\left(-2\log(\nu_X(\{\theta\})) \geq F_\theta^{-1}(1-\alpha)\right)\\
&= 1 - F_\theta(F_\theta^{-1}(1-\alpha))\\
&= \alpha
\end{aligned}$$

and the result follows. □

Therefore, under the conditions of Theorem 2.3, the analyst may set the threshold value at $m(\alpha)$ to control the type I error at significance level α for testing $H_0 : \theta \in \Theta_0$ with the rejection rule: "reject H_0 if $\nu_x(\Theta_0) \leq m(\alpha)$". As this threshold value does not depend on the null parameter space Θ_0, the latter rejection rule does not generate conflicting conclusions on different null hypotheses. Because of this feature, the expected logical consequence holds: if $\tilde{H}_0 \to H_0$, then "rejection of H_0" $\Longrightarrow$ "rejection of $\tilde{H}_0$" with the same data. Under standard regularity conditions, the asymptotic distribution F_θ is a chisquared cumulative distribution with $k = \dim(\Theta)$ degrees-of-freedom and therefore it does not depend on θ. However, under a mixed model, Patriota and Alves (2021) showed recently that the asymptotic distribution of $-2\log(\nu_X(\{\theta\}))$ can vary according to whether θ is an interior point or is on the boundary of the paramerter space.

The LR measure can be rewritten as indicated in the next lemma. This characterization will help to illustrate the example presented in Sect. 4.

Lemma 2.1 *Let A be the set defined in Theorem 2.1. Then, for each fixed $x \in A$ and each $\theta \in \Theta$,*

$$\nu_x(\{\theta\}) = \sup\{\alpha \in [0, 1] : \ \Lambda_\alpha(x) \cap \{\theta\} \neq \varnothing\},$$

where $\Lambda_\alpha(x) = \{\theta \in \Theta : \lambda_\mu(\theta, x) \geq \alpha\}$ and $\mu \in \Upsilon$.

Proof Since for each $x \in A$ and $\theta \in \Theta$, we have that

$$\big\{\alpha \in [0,1] : \Lambda_{\alpha}(x) \cap \{\theta\} \neq \varnothing\big\} = \big\{\alpha \in [0,1] : \lambda_{\mu}(\theta, x) \geq \alpha\big\} = \big[0, \lambda_{\mu}(\theta, x)\big],$$

for each $\mu \in \Upsilon$. The proof is completed. $\square$

The set $\Lambda_{\nu_0}(x)$, where $\nu_0 = \nu_x(\Theta_0)$ highlights the distance of Θ_0 from the set $\widehat{\Theta}_x = \{\theta \in \Theta : \ \nu_x(\{\theta\}) = 1\}$ of points that yields maximum likelihoods (if it is nonempty). The more distant is the border of Λ_{ν_0} from its center, the more implausible is Θ_0 according to the likelihood function.

2.1 Two illustrative examples

In the folowing, we consider two simple examples from Binomial and Poisson distributions.

Example 2.1 (Binomial) Let X be a binomial random variable such that

$$P_{\theta}(X = x) = \begin{cases} \binom{8}{x}\theta^x(1-\theta)^{8-x}, & \text{if } x \in A, \\ 0, & \text{otherwise,} \end{cases}$$

where $A = \{0, 1, \ldots, 8\}$ and $\theta \in (0, 1)$. Figure 2 shows the ν_x-values for $\Theta_0 = [0.4, 0.6]$ and all observable values of $x \in A$. The LR-measure is

$$\nu_x(\{\theta\}) = \begin{cases} \dfrac{\theta^x(1-\theta)^{8-x}}{(\frac{x}{8})^x(1-\frac{x}{8})^{8-x}}, & \text{if } x \in \{1, \ldots, 7\}, \\ \theta^x(1-\theta)^{8-x}, & \text{if } x \in \{0, 8\}. \end{cases}$$

As can be seen in Fig. 2e, $\nu_4(\Theta_0) = 1$ and $\nu_4(\Theta_0^c) = 0.85$, that is, when $x = 4$ the set Θ_0 contains the points that generates the highest likelihood. The likelihoods produced by the elements in $\Theta_0^c = (0, 0.4) \cup (0.6, 1)$ attain at most 85% of the maximum likelihood.

In Fig. 2a and i, $\nu_x(\Theta_0) = 0.02$ and $\nu_x(\Theta_0^c) = 1$ for either $x = 0$ and $x = 8$. That is, when $x \in \{0, 8\}$, the set Θ_0^c contains the elements that generates the highest likelihood and the likelihoods produced by the elements of Θ_0 reach at most 2% of the maximum likelihood. Observe that the observed values $x \in \{0, 8\}$ are the ones that provide the most evidence against Θ_0. The interpretations for the other cases are similar.

Example 2.2 (Poisson) Let X be a Poisson random variable, i.e., is probability function is

$$P_{\theta}(X = x) = \begin{cases} \exp(-\theta)\frac{\theta^x}{x!}, & \text{if } x \in A, \\ 0, & \text{otherwise,} \end{cases}$$

where $A = \{0, 1, 2, \ldots\}$ and $\theta \in (0, \infty)$. Figure 3 depicts the ν_x-values for the first eight observable values for $\Theta_0 = (0, 3]$. The LR-measure is

$$\nu_x(\{\theta\}) = \begin{cases} \exp(x-\theta)\big(\frac{\theta}{x}\big)^x, & \text{if } x \in \{1, 2, \ldots\}, \\ \exp(-\theta), & \text{if } x = 0. \end{cases}$$

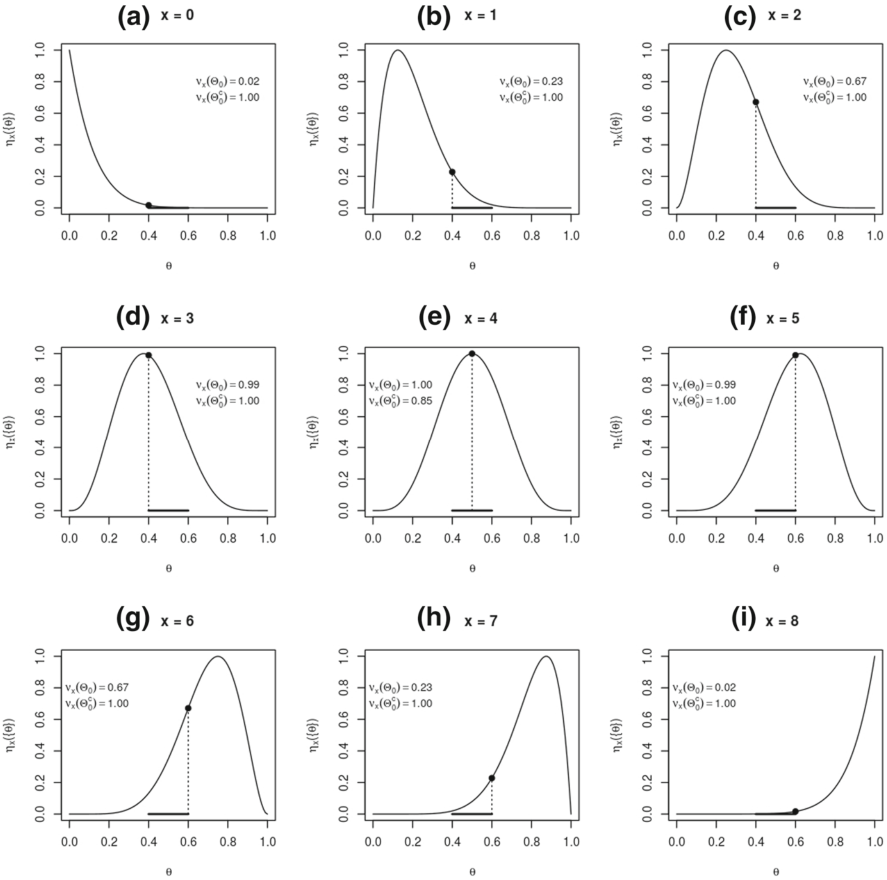

Fig. 2 Normalized likelihood curve for all observable values from a binomial random variable with $n = 8$. The horizontal full segment represents the set $\Theta_0 = [0.4, 0.6]$. The big dot represents the value of $\nu_x(\Theta_0)$ on the curve

Figure 3a shows that $\nu_0(\Theta_0) = 1.00$ and $\nu_0(\Theta_0^c) = 0.05$. Notice that, when $x = 0$, there is no element in Θ that yields the highest likelihood, but the least upper bound of $\mu_0(\{\theta\})$ is $\sup\limits_{\theta>0} \exp(-\theta) = 1$. The elements in $\Theta_0^c = (3, \infty)$ attain at most 5% of the least upper bound for all likelihoods in this model. Figure 3i shows that $\nu_8(\Theta_0) = 0.06$ and $\nu_8(\Theta_0^c) = 1.00$, that is, when $x = 8$, the element that yields the largest likelihood is in Θ_0^c while the elements in Θ_0 attain at most 6% of the largest likelihood. The interpretation for the other cases is similar.

In both examples, as the standard regularity conditions hold, the asymptotic distribution of $-2\log(\nu_X(\{\theta\}))$ is a chisquared distribution with one degree-of-freedom. In this context, the threshold derived in Theorem 2.3 is $m(\alpha) = \exp(-0.5F_{\chi_1^2}^{-1}(1-\alpha))$, where $F_{\chi_1^2}$ is the cumulative chisquared distribution with one degree-of-freedom. In order to reject at 5% significance level and to maintain monotonicity (the entail-

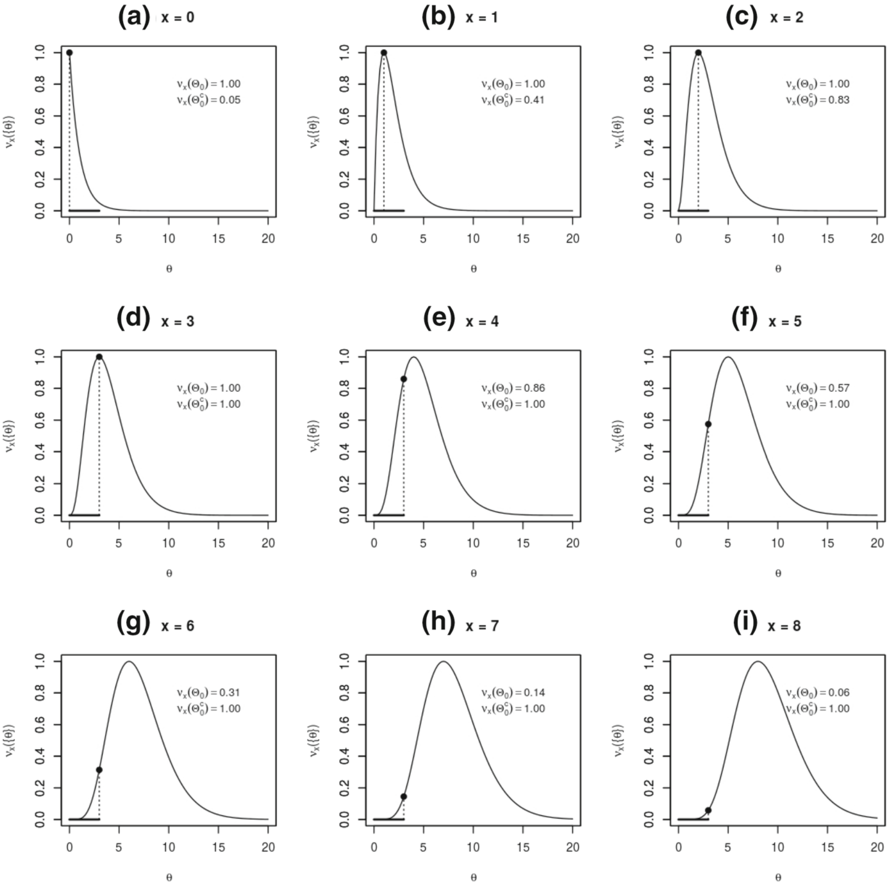

Fig. 3 Normalized likelihood curve for the first eight observable values of a Poisson random variable. The horizontal full segment represents the set $\Theta_0 = (0, 3]$. The big dot represents the value of $\nu_x(\Theta_0)$ on the curve

ment condition or the logical consequences), we need to set the threshold value at $m(0.05) = 0.1465$.

3 Comparison with the Bayesian approach

Hartigan (1998) studied the asymptotic behavior of maximum likelihood and Bayes procedures. In this section, we study some relations of the likelihood measure and the standard Bayesian procedure by considering an observed sample rather than their asymptotic consequences. In the Bayesian framework, a prior probability π, dominated by a sigma-finite measure ξ, is defined over the Borel sigma-field of Θ from which a posterior is attained through

$$h_x(\theta) = \frac{L_\mu(\theta, x)h(\theta)}{m_\mu(x)},$$

where $m_\mu(x) = \int_\Theta L_\mu(\theta, x)d\pi(\theta)$ is the normalizing constant and $h \in \frac{d\pi}{d\xi}$ is one version of the Radon–Nikodym derivative of π with respect to ξ. The posterior probability of Θ_0 given the observed data x is defined by the Lebesgue integral

$$\pi_x(\Theta_0) = \int_{\Theta_0} h_x(\theta)d\xi(\theta).$$

If $\pi_x(\Theta_0)$ is very small, then Θ_0 is improbable according to π_x. In this probabilistic framework, the value $\pi_x(\Theta_0^c)$ is completely determined by $\pi_x(\Theta_0)$, since $\pi_x(\Theta_0^c) = 1 - \pi_x(\Theta_0)$. Then, the following consequence is straightforward

$$0 < \pi_x(\Theta_0) < \pi_x(\Theta_1) < 1 \Longleftrightarrow 0 < \pi_x(\Theta_1^c) < \pi_x(\Theta_0^c) < 1.$$

By properties $\mathbb{P}_1 - \mathbb{P}_4$, this consequence does not hold for the LR-measure ν_x, since $\nu_x(\Theta_0) < 1 \Rightarrow \nu_x(\Theta_0^c) = 1$ which justifies instead the following relation

$$0 < \nu_x(\Theta_0) < \nu_x(\Theta_1) < 1 \Longrightarrow \nu_x(\Theta_0^c) = \nu_x(\Theta_1^c) = 1. \tag{3.1}$$

That is, an increasing in the ν_x-values from Θ_0 to Θ_1 does not necessarily imply a decreasing in the ν_x-values from Θ_0^c to Θ_1^c. They may remain constant. Although ν_x and π_x have different properties, they are connected.

Lemma 3.1 *Let π be a proper prior probability dominated by a sigma-finite measure and defined over a measurable family of subsets of Θ. Assume valid Condition (C_1) and let A be the set defined in Theorem 2.1, then, for any π-measurable Θ_0 such that $\pi(\Theta_0) > 0$,*

$$\nu_x(\Theta_0) \geq \frac{\pi_x(\Theta_0)m_\mu(x)}{\pi(\Theta_0)c_\mu(x)},$$

for any $x \in A$, where $c_\mu(x) = \sup_{\theta\in\Theta} L_\mu(\theta, x)$ and $\mu \in \Upsilon$.

Proof As $0 < c_\mu(x) < \infty$ for each $x \in A$, the posterior probability can be written as

$$\pi_x(\Theta_0) = \frac{c_\mu(x)}{m_\mu(x)} \int_{\Theta_0} \lambda_\mu(\theta, x)d\pi(\theta) \leq \frac{c_\mu(x)}{m_\mu(x)}\pi(\Theta_0) \sup_{\theta\in\Theta_0} \lambda_\mu(\theta, x), \tag{3.2}$$

therefore the result follows

$$\nu_x(\Theta_0) \geq \frac{\pi_x(\Theta_0)m_\mu(x)}{\pi(\Theta_0)c_\mu(x)}.$$

□

It is worth mentioning that the set function ν_x is defined over the power set of Θ, while the posterior distribution is defined only over the family of measurable subsets (in the Lebesgue sense) of Θ. Thus, there may exist subsets of Θ that are not measurable by π_x but are computable by ν_x. Moreover, from Eq. (3.2), $\nu_x(\Theta_0) = 0$ implies that $\pi_x(\Theta_0) = 0$ but the converse is not true, for there exist zero prior probabilities for Θ_0. In this sense, ν_x dominates π_x. This feature is in agreement with the expected reasoning between possibility and probability:

> "a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability imply a low degree of possibility. However, if an event is impossible, it is bound to be improbable." (Zadeh 1978)

Lemma 3.1 provides an upper bound for the posterior probability of Θ_0. The following corollary is straightforwardly obtained from Lemma 3.1.

Corollary 3.1 *Assume valid Lemma 3.1's assumptions. Let Θ_0 be a π-measurable set.*

1. *If $\pi(\Theta_0) \leq \frac{m_\mu(x)}{c_\mu(x)}$, then $\pi_x(\Theta_0) \leq \nu_x(\Theta_0)$,*
2. *If $\nu_x(\Theta_0) \leq \frac{m_\mu(x)}{c_\mu(x)}$, then $\pi_x(\Theta_0) \leq \pi(\Theta_0)$.*

Observe that $\frac{m_\mu(x)}{c_\mu(x)} \leq 1$ by the following

$$m_\mu(x) = \int_\Theta L_\mu(\theta, x) d\pi(\theta) \leq \sup_{\theta \in \Theta} L_\mu(\theta, x)\pi(\Theta) = c_\mu(x).$$

The first part of Corollary 3.1 holds for measurable subsets of Θ with low prior probabilities relative to $m_\mu(x)/c_\mu(x)$. This condition is sufficient but not necessary, since a high prior probability might be assigned for a set Θ_0 with relatively low maximum likelihood such that $\pi_x(\Theta_0) \leq \nu_x(\Theta_0)$ still holds true.

It is worth noting that only optimization procedures are required in the computation of $\nu_x(\Theta_0)$, integration is not required. Computing a posterior probability can be a cumbersome task in some high dimensional problems, therefore, an upper bound of the posterior probability can be attained by Lemma 3.1 and Corollary 3.1 in terms of ν_x. That is, under the specified conditions, if the ν_x-value is small, then the posterior probability must be even smaller, unless in cases where the prior is far away from the region where likelihood is concentrated.

4 Application to the Hardy–Weinberg equilibrium

In this section, we apply the above results to analyze genotype frequencies in a population. We investigate three scenarios, namely: (1) the population is under the Hardy Weinberg equilibrium; (2) the population is undergoing a regular system of 'inbreeding' (when relatives produce offspring); and (3) the population is undergoing a regular system of 'outbreeding' (when very genetically different individuals produce offspring). These scenarios are formalized mathematically in terms of sets in the sequel.

Let AA, Aa and aa be the possible genotypes and θ_1, θ_2 and θ_3 their respective population frequencies, where $\theta_1 + \theta_2 + \theta_3 = 1$. The parameter vector is $\theta = (\theta_1, \theta_2, \theta_3)$.

Denote y_1, y_2 and y_3 the observed frequencies of the genotypes AA, Aa and aa, respectively, where $y_1 + y_2 + y_3 = s$.

The statistical model is then defined in the following. Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be a parametric statistical model, where the set $A \subseteq \mathcal{X}$ such that $P_\theta(A) = 1$ is given by

$$A = \left\{(y_1, y_2, y_3) \in \mathbb{N}^3 : \ y_1 + y_2 + y_3 = s\right\},$$

where $s \in \mathbb{N}$ is a fixed value and each $P_\theta \in \mathcal{P}$ is

$$P_\theta(B) = \sum_{(y_1, y_2, y_3) \in B} \frac{s!}{y_1! y_2! y_3!} \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3},$$

where $B \subseteq A$ is a nonempty set and the parameter space is

$$\Theta = \left\{\theta \in [0, 1]^3 : \ \theta_1 + \theta_2 + \theta_3 = 1\right\}.$$

For the observed sample $x = (y_1, y_2, y_3)$, the likelihood function and the likelihood-ratio function are, respectively, given by

$$L(x, \theta) = \frac{s!}{y_1! y_2! y_3!} \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3} \quad \text{and} \quad \lambda(x, \theta) = c\theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3},$$

where $c = \prod_{i : y_i > 0} \left(\frac{s}{y_i}\right)^{y_i}$.

The population is under the Hardy–Weinberg equilibrium when $\sqrt{\theta_3} = 1 - \sqrt{\theta_1}$, for this situation we define the set

$$\Theta_1 = \left\{\theta \in \Theta : \ \sqrt{\theta_3} = 1 - \sqrt{\theta_1}\right\}.$$

The population is under an inbreeding pressure when $\sqrt{\theta_3} < 1 - \sqrt{\theta_1}$, for this case we define the set

$$\Theta_2 = \left\{\theta \in \Theta : \ \sqrt{\theta_3} < 1 - \sqrt{\theta_1}\right\}.$$

Finally, the population is under an outbreeding pressure when $\sqrt{\theta_3} > 1 - \sqrt{\theta_1}$, for this last case we define the set

$$\Theta_3 = \left\{\theta \in \Theta : \ \sqrt{\theta_3} > 1 - \sqrt{\theta_1}\right\}.$$

For more details and discussion on this topic the reader is referred to Emigh (1980). The LR-measure for each set is

$$\nu_x(\Theta_i) = \sup_{\theta \in \Theta_i} \lambda(\theta, x)$$

for $i = 1, 2, 3$. Notice that

$$\nu_x(\Theta_1) = c \sup_{\theta \in \Theta_1} \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3},$$

where $\theta_3 = (1 - \sqrt{\theta_1})^2$ and $\theta_2 = 1 - \theta_1 - (1 - \sqrt{\theta_1})^2$, that is, $\lambda(x, \theta)$ restricted to Θ_1 depends only on θ_1. Then, we can rewrite the ν_x-value by

$$\nu_x(\Theta_1) = c \sup_{\theta_1 \in [0,1]} f(\theta_1),$$

where $f(z) = z^{y_1}(1 - z - (1 - \sqrt{z})^2)^{y_2}(1 - \sqrt{z})^{2y_3}$. As f is continuous, $f([0, 1])$ is closed and it has a maximum, i.e., $\sup f([0, 1]) = \max f([0, 1])$. It is possible to show that the maximum value of f is attained at

$$z = \frac{(s + y_1 - y_3)^2}{4s^2}.$$

Therefore, the ν_x-value for Θ_1 is

$$\nu_x(\Theta_1) = \lambda(x, \widetilde{\theta}), \tag{4.1}$$

where $\widetilde{\theta} = (\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3)$, $\widetilde{\theta}_1 = \frac{(s+y_1-y_3)^2}{4s^2}$, $\widetilde{\theta}_2 = 1-\widetilde{\theta}_1-(1-\sqrt{\widetilde{\theta}_1})^2$ and $\widetilde{\theta}_3 = (1-\sqrt{\widetilde{\theta}_1})^2$. Notice that,

$$\dim(\Theta_1) = 1, \quad \dim(\Theta_2) = \dim(\Theta_3) = \dim(\Theta) = 2,$$

also note that for each $0 < \theta_1, \theta_2, \theta_3 < 1$, the function $\log(\lambda(x, \cdot))$ is concave. Therefore, the maximum likelihood estimate $\widehat{\theta}$ satisfies

$$\widehat{\theta} \in \overline{\Theta_2} \quad \text{or} \quad \widehat{\theta} \in \overline{\Theta_3},$$

where $\overline{B}$ is the closure of set B. Also, we have the following three situations

1. If $\widehat{\theta} \in \overline{\Theta_2}$, then $\nu_x(\Theta_2) = 1$ and $\nu_x(\Theta_3) = \nu_x(\Theta_1)$.
2. If $\widehat{\theta} \in \overline{\Theta_3}$, then $\nu_x(\Theta_3) = 1$ and $\nu_x(\Theta_2) = \nu_x(\Theta_1)$.
3. If $\widehat{\theta} \in \overline{\Theta_2} \cap \overline{\Theta_3}$, then $\nu_x(\Theta_i) = 1$, for $i = 1, 2, 3$.

Hence, in any cases, we only need to compute numerically $\nu_x(\Theta_1)$. The asymptotic distribution of $-2\log(\lambda(X, \theta))$ is a chisquared distribution with two degrees-of-freedom for each $\theta \in \Theta$. In this context, the asymptotic threshold derived in Theorem 2.3 is

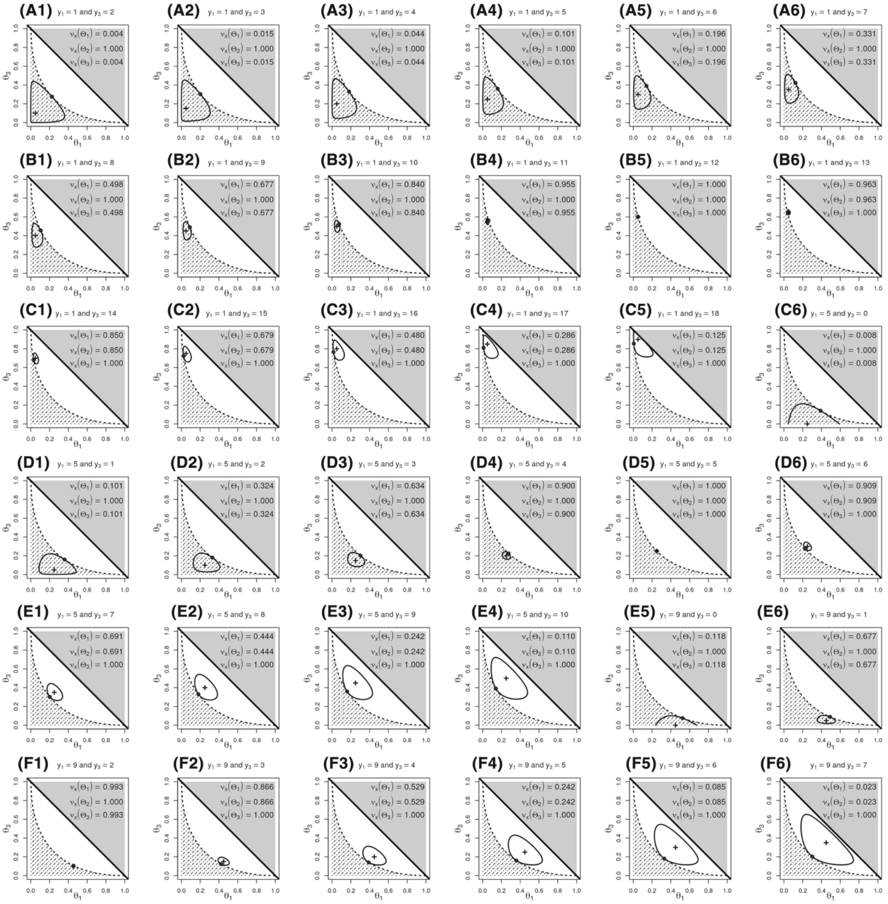

Fig. 4 The dashed lines represent the Hardy–Weinberg equilibrium (Θ_1). The crosshatched area below the dashed line represents the inbreeding condition (Θ_2) and the white area above the dashed line represents the outbreeding condition (Θ_3). The cross mark is the maximum likelihood estimate for each observed sample (y_1, y_3) and its surrounding full line is the smallest contour for which Λ_α has at least one element in common with Θ_1 (dashed line). The associated s-values for Θ_1, Θ_2 and Θ_3 are presented

$m(\alpha) = \exp(-0.5F^{-1}_{\chi^2_2}(1-\alpha)) = \alpha$, where $F_{\chi^2_2}$ is the cumulative chisquared distribution with two degrees-of-freedom. In order to reject at 5% significance level and to maintain monotonicity (the entailment condition or the logical consequences), we choose the threshold value at $m(0.05) = 0.05$.

In Fig. 4, the simplex formed by points (0,0), (0,1) and (1,0) represents the parameter space Θ in a reduced form (only with θ_1 and θ_3, for the restriction $\theta_1 + \theta_2 + \theta_3 = 1$ and also we consider $s = 20$). Each plot presents the three sets Θ_1, Θ_2 and Θ_3 in their respective reduced forms (only with θ_1 and θ_3). The dashed line illustrates the Hardy–Weinberg equilibrium (Θ_1). The crosshatched area (below the dashed curve) stands for the inbreeding restriction (Θ_2) and the white area (above the dashed curve) stands for the outbreeding restriction (Θ_3). Each plot in Fig. 4 refers to a specific observed sample, for instance plot (A1) refers to $y_1 = 1$ and $y_3 = 2$, plot (A2) refers to $y_1 = 1$ and $y_3 = 3$ and so forth. In each plot, the cross mark indicates the maximum likelihood estimate for (θ_1, θ_3) and the full lines surrounding the ML estimates are the smallest contours for which Λ_α has at least one element in common with Θ_1.

On the one hand, in the plots (A1)–(A6), (B1)–(B4), (C6), (D1)–(D4), (E5)–(E6) and (F1) the maximum likelihood lies in Θ_2. That is, if we consider that only the specifed three hypotheses are possible and that the model is correct, these cases are 'favoring' the inbreeding restriction, but only for (A1)–(A3) and (A6) they are relevant at 5% significance level:

- (A1): $y_1 = 1, y_3 = 2 \;\Rightarrow\; \nu_x(\Theta_3) = 0.004$,
- (A2): $y_1 = 1, y_3 = 3 \;\Rightarrow\; \nu_x(\Theta_3) = 0.015$,
- (A3): $y_1 = 1, y_3 = 4 \;\Rightarrow\; \nu_x(\Theta_3) = 0.044$,
- (C6): $y_1 = 5, y_3 = 0 \;\Rightarrow\; \nu_x(\Theta_3) = 0.008$.

On the other hand, in the plots (B6), (C1)–(C5), (D6), (E1)–(E4) and (F2)–(F6) the maximum likelihood lies in Θ_3. That is, if we consider that only the specifed three hypotheses are possible and that the model is correct, these cases are now 'favoring' the outbreeding restriction, but only for (F6) it is relevant at 5% significance level:

- (F6): $y_1 = 9, y_3 = 7 \;\Rightarrow\; \nu_x(\Theta_2) = 0.023$.

It is also noteworthy that in two cases the maximum likelihood is in $\overline{\Theta_2} \cap \overline{\Theta_3} = \Theta_1$, namely: (B5) and (D5). In those cases, we have that $\nu_x(\Theta_1) = \nu_x(\Theta_2) = \nu_x(\Theta_3) = 1$ which means no evidence against any hypotheses.

5 Monte Carlo simulations

As requested by one of the referees, in this section we conduct Monte Carlo simulations to illustrate, on the Hardy–Weinberg application, some of the theoretical results attained in this paper. In our simulations, we consider that the total frequency (which plays the role of the sample size) $y_1 + y_2 + y_3 = s \in \{20, 50, 150, 500, 1000\}$ and we define the null hypothesis to be the Hardy–Weinberg equilibrium, namely, $H_0 : \theta \in \Theta_1$, where Θ_1 is given in the previous section.

We generate 10,000 Monte Carlo simulations, under two scenarios, to compare the LR measure compute at Θ_1, $\nu_x(\Theta_1)$, with the respective asymptotic p value

$$\text{p value}(\Theta_1, x) = 1 - F_{\chi^2_1}(-2\log(\nu_x(\Theta_1))),$$

where $\nu_x(\Theta_1)$ is given in Eq. (4.1).

In the first scenario, we generate samples under the null hypothesis where $(\theta_1, \theta_2, \theta_3) \in \Theta_1$ and $\theta_1 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9\}$. Figure 5 depicts the cumulative empirical distribution of the p values and the ν_x-values. As expected for all settings, under the null and for large sample sizes, the cumulative distributions of p values are close to the 45° line and the cumulative distributions of the ν_x-values are always below the 45° line. These results were predicted by Theorems 2.2 and 2.3 .

In the second scenario, we generate the samples under the alternative hypothesis in order to compare the rejection rates of the null hypothesis defined in the first scenario.

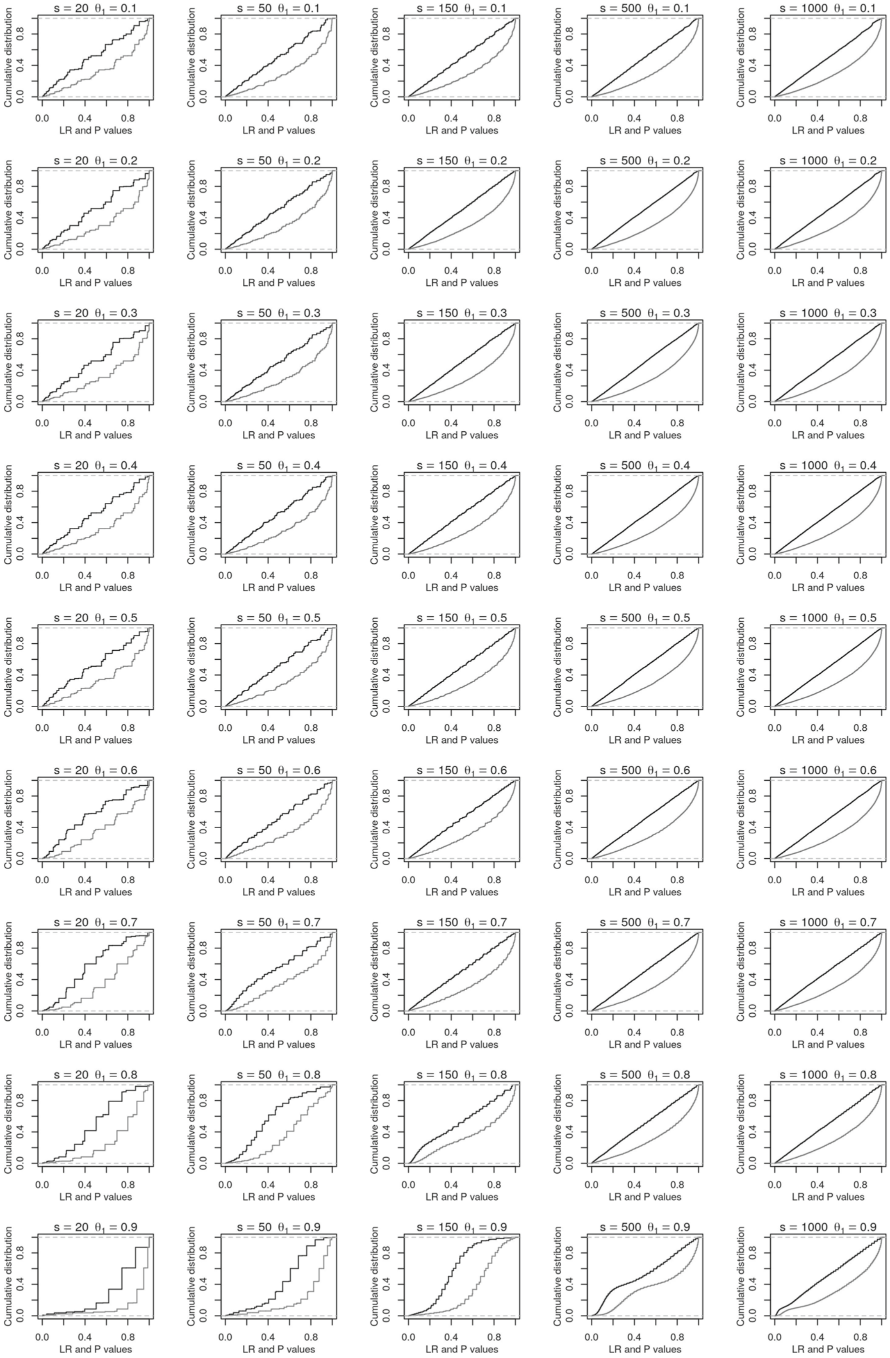

Fig. 5 Cumulative distributions of the p values (black lines) and LR measure (gray lines) computed at Θ_1. Sample sizes $s \in \{20, 50, 150, 500, 1000\}$ and the null values of $\theta_1 \in \{0, 1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

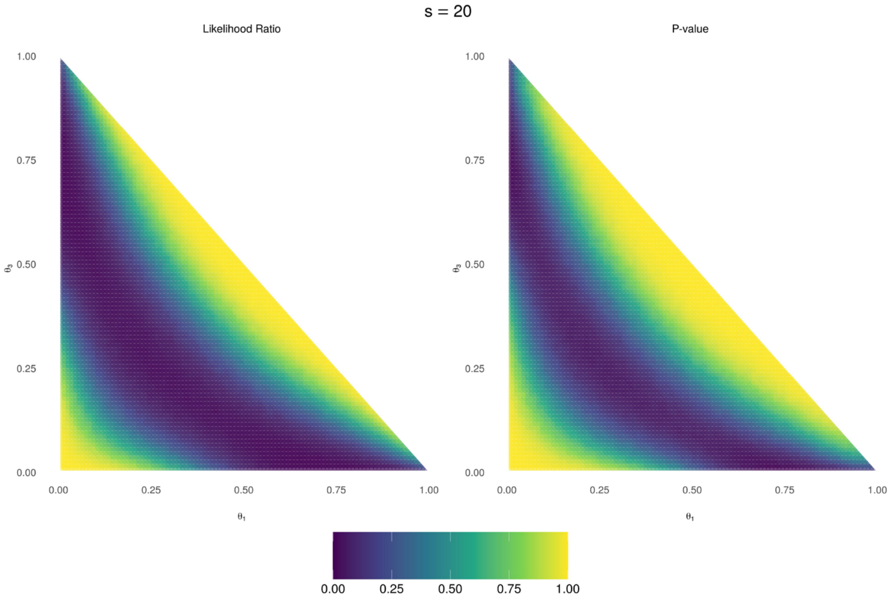

Fig. 6 Rejection rates of the null hypothesis $H_0 : \theta \in \Theta_1$ under local alternative hypotheses by using the Likelihood Ratio and p value rejection rules for the sample size $s = 20$

Two rejection rules are considered, namely, (1) "reject H_0 if p value$(\Theta_1, x) \leq \alpha$" and (2) "reject H_0 if $\nu_x(\Theta_1) \leq m(\alpha)$", where $m(\alpha)$ is given in Theorem 2.3 which is, for this example, equal to α . We generate the data under the inbreeding and outbreeding hypothesis where $(\theta_1, \theta_3) \in \left\{(\frac{i}{100}, \frac{j}{100}), i = 1, \ldots, 99, j = 1, \ldots, 100 - i\right\}$. Figures 6, 7, 8, 9 and 10 present heatmaps of the rejection rates for which the blue/dark color indicates that the rejection rate is close to zero (approximately 0.05) and the yellow/bright color indicates the rejection rate is close to one. As expected, the p value procedure to reject the null hypothesis H_0 is slightly more powerful than the LR one and, also, both procedures increase the power as the sample size increases.

6 Concluding remarks

This paper discussed theoretical properties of the likelihood-ratio measure for random and observed samples. We showed that the LR measure (a) satisfies logical consequences and frequentist principles and (b) it can be used as upper bounds for posterior probabilities with relatively small prior probabilities. One can properly control the type I error in a hypothesis testing procedure without violating logical consequences over the set of null hypotheses. It should be clear that testing different null hypotheses should not be confused with sequential hypothesis testing, since the former does not employ the result of one test to conduct another. An application to the Hardy–Weinberg equilibrium was presented and Monte Carlo simulations were conducted to illustrate the results.

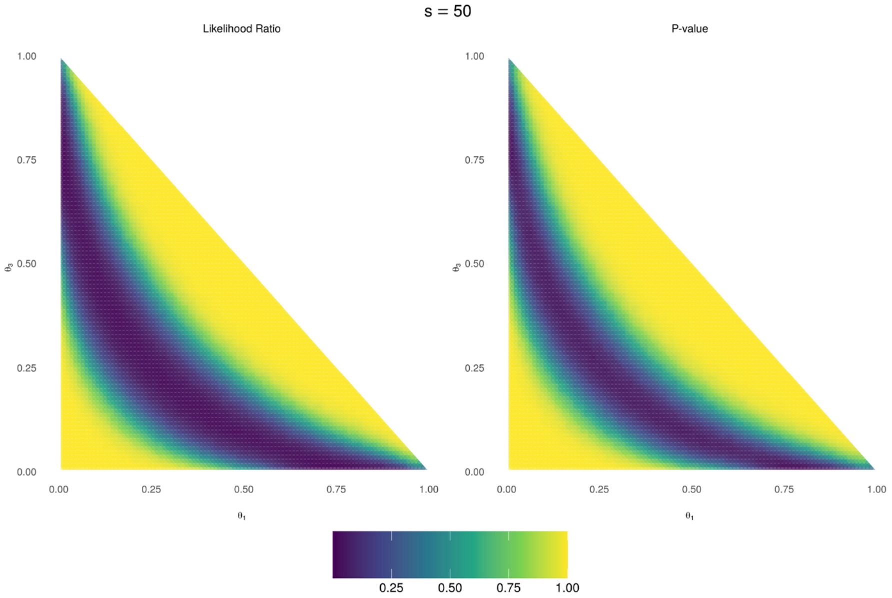

Fig. 7 Rejection rates of the null hypothesis $H_0 : \theta \in \Theta_1$ under local alternative hypotheses by using the Likelihood Ratio and p value rejection rules for the sample size $s = 50$

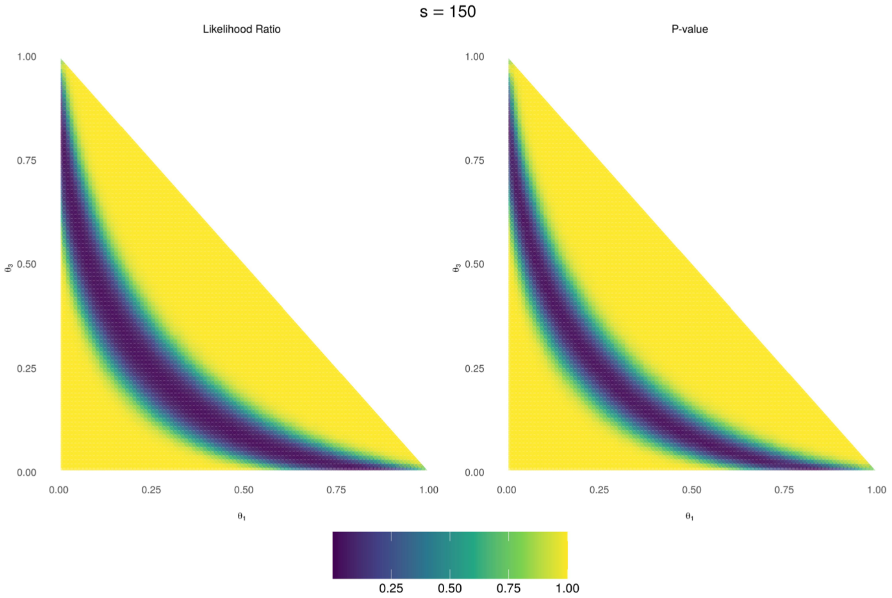

Fig. 8 Rejection rates of the null hypothesis $H_0 : \theta \in \Theta_1$ under local alternative hypotheses by using the Likelihood Ratio and p value rejection rules for the sample size $s = 150$

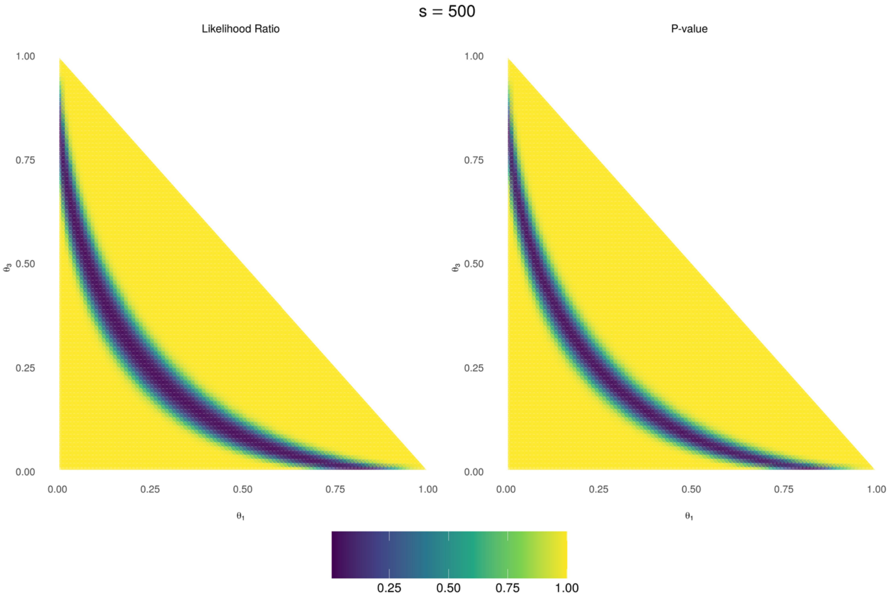

Fig. 9 Rejection rates of the null hypothesis $H_0 : \theta \in \Theta_1$ under local alternative hypotheses by using the Likelihood Ratio and p value rejection rules for the sample size $s = 500$

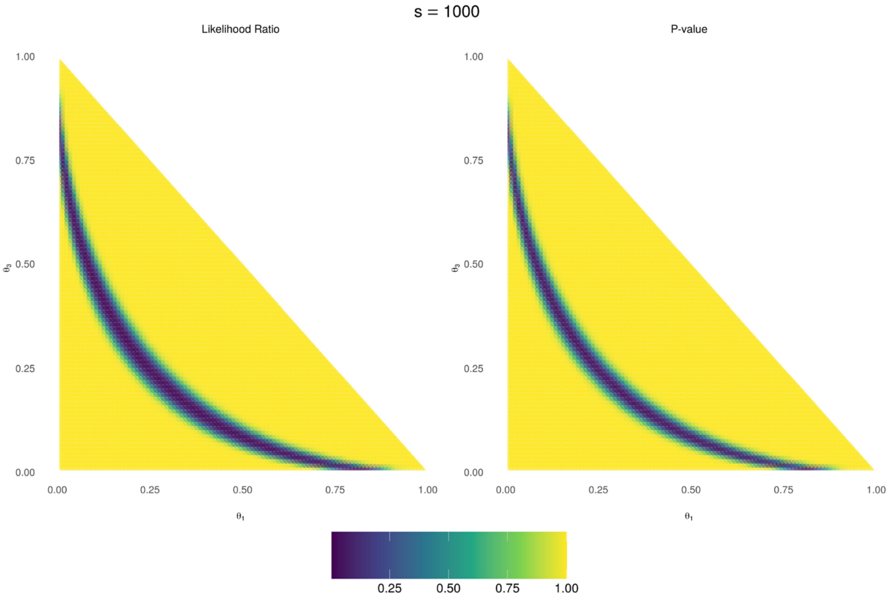

Fig. 10 Rejection rates of the null hypothesis $H_0 : \theta \in \Theta_1$ under local alternative hypotheses by using the Likelihood Ratio and p value rejection rules for the sample size $s = 1000$

Acknowledgements This work received Grants from FAPESP–Brazil (2014/25595-0) and CNPq (200115/2015-4). This paper was partially developed in the Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Canada, and in the Department of Statistics, University of São Paulo, Brazil.

References

Bickel DR (2012) The strength of statistical evidence for composite hypotheses: inference to the best explanation. Stat Sin 22:1147–1198
Birkes D (1990) Generalized likelihood ratio tests and uniformly most powerful tests. Am Stat 44:163–166
Blume JD (2008) Tutorial in biostatistics: likelihood methods for measuring statistical evidence. Stat Med 21:2563–2599
Chernoff H (1954) On the distribution of the likelihood ratio. Ann Math Stat 25:573–578
Denneberg D (1994) Non-additive measure and integral, theory and decision library (Series B mathematical and statistical methods), vol 27. Kluwer Academic Publishers, Dordrecht
Drton M (2009) Likelihood ratio tests and singularities. Ann Stat 37:979–1012
Emigh TH (1980) A comparison of tests for Hardy-Weinberg equilibrium. Biometrics 36:627–642
Giang PH, Shenoy PP (2005) Decision making on the sole basis of statistical likelihood. Artif Intell 165:137–163
Gonçalves FB, Franklin P (2021) On the definition of likelihood function, arXiv:1906.10733
Graffelman J, Weir B (2016) Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. Heredity 116:558–568
Hacking I (1965) Logic of statistical inference. Cambridge University Press, Cambridge
Hartigan JA (1998) The maximum likelihood prior. Ann Stat 26:2083–2103
Karlin S, Rubin H (1956) The theory of decision procedures for distributions with monotone likelihood ratio. Ann Math Stat 27:272–299
Lindley DV (1988) Statistical inference concerning Hardy-Weinberg equilibrium. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) Bayesian statistics, vol 3. Oxford University Press, Oxford, pp 307–326
Mudholkar GS, Chaubey YP (2009) On defining p values. Stat Probab Lett 79:1963–1971
Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. Philos Trans R Soc Lond A 231:289–337
Patriota AG (2013) A classical measure of evidence for general null hypotheses. Fuzzy Sets Syst 233:74–88
Patriota AG (2017) On some assumptions of the null hypothesis statistical testing. Educ Psychol Meas 77:507–528
Patriota AG, Alves JO (2021) A monotone frequentist measure of evidence for testing variance components in linear mixed models. J Stat Plan Inference 219:43–62. https://doi.org/10.1016/j.jspi.2021.11.002
Puhalskii A (2001) Large deviations and idempotent probability, Monographs and surveys in pure and applied mathematics. Chapman & Hall/CRC, Boca Raton
Puig X, Ginebra J, Graffelman J (2017) A Bayesian test for Hardy-Weinberg equilibrium of biallelic X-chromosomal markers. Heredity 119:226–236
Royall R (1997) Statistical evidence: a likelihood paradigm. CRC Press, New York
Royall R (2000) On the probability of observing misleading statistical. J Am Stat Assoc 95:760–768
Severini TA (2000) Likelihood methods in statistics, vol 22. Oxford Statistical Science Series, New York
Schervish MJ (1996) P values: what they are and what they are not. Am Stat 50:203–206
Shoemaker J, Painter I, Weir BS (1998) A bayesian characterization of Hardy-Weinberg disequilibrium. Genetics 149:2079–2088
Sprott DA (2000) Statistical inference in science. Springer, New York
Vu HTV, Zhou S (1997) Generalization of likelihood ratio tests under nonstandard conditions. Ann Stat 25:897–916
Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28