Non-informative nuisance parameter principle for weighted likelihood test using adaptive significance levels in count data

Andrés Felipe Flórez Rivera

Thesis presented
to
Institute of Mathematics and Statistics
of
University of São Paulo
to
obtain the title
of
Doutor in Science

Program: Statistics

Advisor: Prof. Dr. Luís Gustuavo Esteves

During the development of this work the author received
financial support from CNPq

São Paulo, July of 2020

Princípio de parâmetro nuisance não informativo para teste de verossimilhanças ponderadas usando níveis de significância adaptativos em dados de contagem

Andrés Felipe Flórez Rivera

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
Obtenção do Título
de
Doutor em Ciências

Programa: Estatística

Orientador: Prof. Dr. Luís Gustuavo Esteves

Durante o desenvolvimento deste trabalho o autor recebeu auxílo financeiro do CNPq

São Paulo, Julho de 2020

Non-informative nuisance parameter principle for weighted likelihood test using adaptive significance levels in count data

Esta versão da teses contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 30/07/2020. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Luís Gustavo Esteves (orientador) - IME-USP
- Prof. Dr. Carlos Alberto de Bragança Pereira - IME-USP
- Prof. Dr. Adriano Polpo de Campos- DES-UFSCar
- Prof. Dr. Rafael Bassi Stern - DES-UFSCar
- Prof. Dr. Marcio Alves Diniz - DES-UFSCar

Acknowledgements

I greatly appreciate the support, patience and companionship of my advisor Professor Luís Gustavo Esteves. His human qualities and high academic level made it possible for me to be a Ph.D, undoubtedly, this thesis would not be possible without his guidance. I also thank the cleaners, secretaries and other employees of the USP Institute of Mathematics and Statistics who made possible my stay at the institute. To Ântonia Alessandra Lemos dos Santos for all her help and patience. To my friends from the Department of Statistics of IME-USP. I thank my parents and all family members for their love but also for giving me the strength to chase my dreams and always support me to move on. Ultimately, I thank CAPES and CNPq for the financial support.

Agradecimentos

Agradeço imensamente o apoio, paciência e companheirismo do meu orientador Professor Luís Gustavo Esteves. Suas qualidades humanas e alto nível acadêmico possibilitaram o meu doutorado, sem dúvida essa tese não seria possível sem sua orientação. Agradeço também aos faxineiros, secretários e demais funcionários do Instituto de Matemática e Estatística da USP que possibilitaram minha permanência no instituto. A Ântonia Alessandra Lemos dos Santos por toda a ajuda e paciência. Aos meus amigos do Departamento de Estatística do IME-USP. Agradeço a meus pais e todos os membros da família por seu amor, mas também por me dar força para perseguir meus sonhos e sempre me apoiar para seguir em frente. Por fim, agradeço à CAPES e à CNPq pelo apoio financeiro.

Abstract

The usage of classical *p-value* in significance tests for evaluating statistical hypotheses is a common practice among scientists of different areas of sciences. However, this practice has been widely criticized for its interpretation for many years and from many points of view due to of its misuse. Consequently, alternatives to this procedure are needed. In this work statistical hypothesis testing using weighted likelihood functions and adaptive significance levels are reviewed, with special emphasis on exploring the properties of this procedure. Specifically, it is proved that this procedure follows both the non-informative "nuisance" parameter principle and an invariance property. These properties lead to a reduced model and tractable parametric spaces that allow tackling the problem of testing hypotheses more easily. In addition, the conditional *P-value* is presented as a measure of evidence of the hypotheses. The proposed test is applied to test independence and diagonal symmetry on contingency tables, compare two Poisson means and to test the Hardy-Weinberg Equilibrium hypothesis. The advantages of this methodology are discussed and possible future works are suggested.

Keywords: P-values, Contingency Tables, Bayes factor, Homogeneity, Independence, Diagonal symmetry, Hardy–Weinberg equilibrium, Poisson means comparison.

Resumo

O uso do *valor-p* clássico em testes de significância para avaliar hipóteses estatísticas é um procedimento comum entre cientistas de diferentes áreas das ciências. No entanto, esse procedimento tem sido amplamente criticado por sua interpretação há muitos anos e de muitos pontos de vista devido ao seu mau uso. Consequentemente, são necessárias alternativas para esse procedimento. Neste trabalho, o teste de hipóteses estatísticas usando funções de verossimilhança ponderadas e níveis de significância adaptativos é revisado, com ênfase especial na exploração de propriedades desse procedimento. Especificamente, prova-se que este procedimento segue o princípio do parâmetro "nuisance" não informativo e uma propriedade de invariância. Essas propriedades levam a um modelo reduzido e a espaços paramétricos mais tratáveis que permitem enfrentar o problema de testar hipóteses com mais facilidade. Além disso, é apresentado o *valor-P* condicional como uma medida de evidência das hipóteses. Os resultados são aplicados para testar a independência e simetria diagonal em tabelas de contingência, também para comparar duas médias de Poisson e o teste de Equilíbrio de Hardy-Weinberg. A discussão geral apresenta as vantagens dessa metodologia e sugere possíveis trabalhos futuros.

Palavras-chave: Valor-P, Tabelas de contingencia, fator de Bayes, Homogeneidade, Independência, Simetria diagonal, Equilíbrio de Hardy–Weinberg, Comparação de médias de Poisson.

Contents

Abbreviation and Symbols

Abbreviation

Bf:	Bayes Factor.
H:	Null hypothesis.
A:	Alternative Hypothesis.
P-P test:	Methodology proposed by Pereira *et al.* (2017) to hypothesis testing.

Symbols

Θ :	Parametric space.
Λ :	Alternative parametric space.
f_H :	Predictive function under null hypothesis.
f_A :	Alternative function under null hypothesis.
$\mathbb{P}_H(\cdot)$:	Measure over the parameter $(\cdot)$ given the null hypothesis H.
$\mathbb{P}_A(\cdot)$:	Measure over the parameter $(\cdot)$ given the alternative hypothesis A.
$\xi(\cdot)$:	Posterior distribution of the parameter $(\cdot)$ given x.
$\mathcal{E}$:	Experiment in which the data x are generated.

List of Figures

List of Tables

Chapter 1

Introduction

In experimental sciences, tests of significance are the most commonly used method for inference. However, the significance test has been generally criticized for their interpretation for many years and from many points of view because of its misuse, see for example Berger and Delampady (1987); Demidenko (2016); Schervish (1996a); Trafimow and Marks (2015) and references therein. In view of this, we present some ideas from the point of view of Fisher and Neyman involving p-value and some of its limitations, which motivated to implement new measures of evidence.

According to Schervish (1996a), the p-value is interpreted as an evidence measure of the null hypothesis. This interpretation distinguishes Fisher's theory from the theory of Jerzy Neyman who refused the idea that p-values can be seen as a measure evidence . According to Neyman, the result in a hypothesis test is a decision between alternate actions (Johnstone *et al.*, 1986), i.e, to accept or reject the null hypothesis (Neyman, 1960). Neyman's interpretation of tests of significance was not well received in its epoch, although the mathematical concepts he developed with Egon Pearson, especially the concept of "power", are the concepts used explicitly by every orthodox (not-Bayesian) statistician to characterize and compare tests of hypotheses (Johnstone *et al.*, 1986).
As said by Berger and Delampady (1987), for many years, it has been alleged that Fisher's logic, specifically his much-vaunted disjunction, is fundamentally misconceived, since if the probability of an event E given the null hypothesis H, $p(E|H)$, is small, it is not enough to discredit the null hypothesis H unless there is an alternative hypothesis A such that the probability $p(E|H)$ is relatively large, i.e. such that the likelihood ratio $p(E|H)/p(E|A)$ is small, certainly less than 1 (Berger and Delampady, 1987; Johnstone *et al.*, 1986).

The idea that the null hypothesis is not discredited unless the likelihood ratio $p(E|H)/p(E|A)$ is less than 1 is one version of the so-called "law of likelihood" presented by Hacking (1965). If the law of likelihood is logical, which most theorists accept, at least intuitively, then Fisher's disjunction, which takes into account $p(E|H)$, but not $p(E|A)$, makes no sense. Thus, Fisher's disjunction conflicts with Bayes' theorem. This inconsistency, which is often described as a "paradox", was elaborated by Lindley (1957) who demonstrated that for any prior probability $p(H) > 0$, and for any level of significance p, no matter how small, there is always a sample size n such that the posterior probability of the null hypothesis H equals $1-p$. As explained, first by Jeffreys (1998) and later by Cornfield (1966), the interpretation of the level of significance p-values based depends on the sample space. The sample space is defined in part by the sampling rule, i.e. the rule (routine) with which the sample is drawn. Thus, the p-value depends not only on the discrepancy of the sample with the null hypothesis but also on the stopping rule, the scheme of experimental randomization and the type of null hypothesis being tested. Note that the inferences based on the likelihood does not depend on the stopping rules and if one accepts the likelihood principle one must reject *p-value* as evidence against the null hypothesis (Berger and Delampady, 1987; Johnstone *et al.*, 1986).

Another inconsistency in Fisher's philosophy is its lack of coherence. Suppose that one hypothesis implies another, e.g., $H_1 : \theta \in (-\infty, 0]$ implies $H_2 : \theta \in (-\infty, 10]$ because $(-\infty, 0]$ is a subset of$(-\infty, 10]$. Then, H_1 and H_2 are coherent if rejection of H_2 always entails rejection of H_1 (Gabriel, 1969), and we say that a measure of support for hypotheses is coherent if, whenever H_1 implies H_2, the measure of support for H_2 is at least as large as that for H_1 (Schervish, 1996a). For example, a coherent measure's support for $H_2 : \theta \in (-\infty, 10]$ is at least as large as its support for $H_1 : \theta \in (-\infty, 0]$. Thus, *p-values* are incoherent for point hypotheses, for one-sided hypotheses and for hypotheses of the bounded interval (Barber and Ogle, 2014; Schervish, 1996a). In addition, notice that, the *p-value* is computed not only using the observed result, but also considering all unobserved results more extreme than the observed result. It seems sensible to include the observed result in a measure of evidence. However, to include more extreme results that are somehow hypothetically observable, but are actually unobserved, seems unreasonable. The p-value often provides only a measure of how well the observed data fit the null hypothesis. This seems useful for comparing data sets relative to a single hypothesis, however, we typically have just one data set and possibly multiple hypotheses (Barber and Ogle, 2014).

As mentioned by Kass and Raftery (1995) and Etz and Wagenmakers (2017), the Bayesian approach to hypothesis testing was developed by Wrinch and Jeffreys (1919, 1923a,b) and Jeffreys (1935, 1936) as a major part of his program for scientific inference. Ever since, there is growing evidence that this approach is very useful in practice, and not just a cudgel for bashing frequentists. See for example, Dickey and Lientz (1970); Good (1950, 1965); Irony and Pereira (1995); Lindley (1968); Montoya *et al.* (2001); Pereira and Wechsler (1993); Pereira *et al.* (2017); Savage (1961); Wald (1947). Among them, there is a consensus on the advantages of Bayesian procedure through the Bayes factor. For instance, they emphasize that the Bayes factor not only is an exact test, that is, it does not depend on asymptotic results but also that it offers a way of evaluating evidence in favor of a null hypothesis and its results are simpler to interpret.

As most of the Bayesian procedures, the Bayes factor follows the likelihood principle. Despite that, the Bayes factor is often criticized for its sensitivity to the choice of priors. However, today there exists considerable literature in eliciting probabilistic information from experts and sensitivity analysis that can help us with this limitation (see for example Flórez (2015), Kass and Raftery (1995) and references therein). Another criticism is that, in some cases, no alternative hypothesis can be specified. Consequently, without an explicit alternative hypothesis, no Bayes factor or posterior probability can be calculated. Thus, the justification is, there is no other option except to use the *p-values*. But, if we are conscious that *p-values* overstate the evidence against the null hypothesis H when the alternative hypothesis is known, how can we argue that no problem exists when the alternatives are unknown? To the contrary, what we have learned about testing hypothesis when we have alternatives should serve as overwhelming evidence that a small *p-value* against a null hypothesis simply may not indicate strong evidence to doubt the hypothesis (Berger and Delampady, 1987).

Bayes factor has the form of a likelihood ratio, where the densities under each hypothesis involved are obtained by integrating (not maximizing) over either whole or partial parametric space. Up to now, many integration techniques have been adapted to problems of Bayesian inference, including, of course, the computation of Bayes factors. For instance, when the dimension of the subspace defined by any of the hypotheses is smaller that of the original parameter space, it is possible to determine the Bayes factor using either line or surface integrals. Pereira *et al.* (2017) proposed such an approach for hypothesis testing that, roughly speaking, reduces to test simple versus simple hypothesis.

Nowadays there is a large number of statistical methodologies that over the years have been resistant to attempts to correct them. The use of fixed significance levels is one of them, even when different authors agree in that there are no scientific justification to use fixed or nominal significance level, see for example (DeGroot, 1986; Pericchi and Pereira , 2016; Pérez and Pericchi, 2014; Rosnow and Rosenthal, 1992) and references therein. As explained by Gill (1999) these arbitrary rejection thresholds were presented by Fisher at a time when computers were rare and expensive, thereby, the table of significance levels cannot give every possible value in the range of a test statistic, so particular discrete values had to be provided by convention. However, this logic no longer applies due to the ubiquity of computers in research and teaching environments. On what basis do we decide that p = 0.051 is unacceptable but p = 0.049 is ? Such a distinction relies on the assumption that there is virtually no measurement error, and also that in most situations there is no precise limit to the type I probability error that can be tolerated (Lehmann and Romano, 2006). Furthermore, the use of nominal levels in discontinuous data can be seriously misleading. This warning is made by Yates (1984), where he showed how in the toss of a fair coin 10 times, the probability of getting 8 or more heads is 0.055 and for 9 or more is 0.011, ie., only 1.1 percent of all samples on average will be declared significant at a nominal level of 5 percent.

Previous criticisms generated dissatisfaction among the scientific community, see for example Trafimow and Marks (2015); Wasserstein and Lazar (2016). Part of that dissatisfaction is due to a phenomenon known as “p-hacking”. The “p-hacking” occurs when researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results (Head *et al.*, 2015). We shall address the methodologies presented by Irony and Pereira (1995); Pereira and Pereira (2005); Pereira (1985); Pereira *et al.* (2017); Pericchi and Pereira (2016) as a solution to overcome those criticisms. We will also explore new properties that were not yet explored in the literature and show its applicability in count data arranged in contingency tables.

Analysis of contingency tables is a statistical area where the adequacy of a hypothesis is generally evaluated by the outcome of a classical p-value against a nominal significance level through the standard tests likelihood ratio from Neyman and Pearson (1957) and Pearson's chi-squared from Wilks (1935), although other methods have also been developed (see, for example, Agresti (2013); Upton (1982) and references therein). Nevertheless, in situations where the sample size is small or the count cells are close to zero, these standard tests may have poor performance because they are based on asymptotic results, that is, for a large sample size. Therefore, when the sample size is small, the performances of this test are not optimal. Indeed, when there are count cells close to zero it is not possible to use the asymptotic likelihood ratio test. In these cases, the exact tests are a natural alternative (Graziadei, 2015; Klein and Linton, 2013). Even if the sample size is large and there are no probabilities close to zero, the "standard" methods inherit the problems previously discussed about tests of significance and, also, the Lindley's paradox Lindley (1957).

This work is organized as follows: in chapter 2, we explore the methodology presented by Pereira *et al.* (2017) and the NNP and Invariance principles. These principle are used to develop new solutions for count data. In chapter 3, we apply these developments for count data, to test homogeneity, independence, and symmetry on contingency tables, comparison of means between populations that follow Poisson distribution and in Hardy Weinberg equilibrium problem. Examples and simulations of practical problems are presented in chapter 4 and in the final chapter, conclusions and suggestions for future work are presented.

Chapter 2

Preliminaries

In this chapter we start with the traditional definitions of statistical model and hypothesis testing. Later we introduce the Bayes factor and the adaptive significance levels. With these definitions, we prove that the P-P test obeys both the invariance principle and the non-informative nuisance parameter principle, which are the main results of our work. Other definitions are introduced and we finish the chapter with definitions of conditional version of the adaptive level of significance and the Bayesian P-value.

2.1 Statistical model

As usually described, the mathematical formulation of the statistical model requires some notions of measure theory. We assume that there is a probability space

$$(\mathfrak{X}, \mathcal{F}, \mathcal{P}), \tag{2-1}$$

where $\mathfrak{X}$ denotes the sample space, $\mathcal{F}$ a σ-field of subsets of $\mathfrak{X}$ and $\mathcal{P}$ a parametric family of distributions. Sets in $\mathcal{F}$ are often known as random events and generic elements $x \in \mathfrak{X}$ as the data coming from an experiment which are assumed to follow a joint probability distribution, $P(\cdot)$, belonging to $\mathcal{P}$. Frequently, the distributions are indexed by a parameter, say $\theta : \mathfrak{X} \to \Theta$, where $\Theta \subseteq \mathbb{R}^k$ is a parameter space with Borel $\sigma-$field given by $\sigma(\Theta)$, so that

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}. \tag{2-2}$$

Usually, Θ will be a subset of some finite-dimensional Euclidean space. For more details see Pace and Salvan (1997); Schervish (1996b). Under this setting, we define for each $x \in \mathfrak{X}$, $L_x : \Theta \to \mathbb{R}_+$ as

$$L_x = L(\theta) = L(\theta|x) = P(x|\theta), \tag{2-3}$$

the likelihood function where $c(x) > 0$ is a constant of proportionality. The function $L(\theta|x)$

gives information on θ for each observed x. It is assumed that the family of likelihood functions indexed by x must be measurable in the prior σ-field.

In the Bayesian paradigm, in addition to the conditional distribution of $P(x \mid \theta)$, θ has prior probability measure $\pi(\cdot)$ over $\sigma(\Theta)$. From the prior distribution and the likelihood function we obtain the posterior probability function $\xi(\theta \mid x)$.

2.2 Hypotheses test

Statistical hypotheses testing consists of a decision problem in which the objective is to choose a statistical hypothesis among two complementary hypotheses, say, H and A. Each one of these hypotheses defines a specific subset of the parameter space Θ, that is, $H: \theta \in \Theta_H$; $A: \theta \in \Theta_A$, where $\Theta_H \cup \Theta_A = \Theta$ and $\Theta_H \bigcap \Theta_A = \emptyset$ with $\Theta_A \in \sigma(\Theta)$ and $\Theta_H \in \sigma(\Theta)$ (for further explanation see Schervish (1996b)). Thus, following the above definition, the hypotheses to be tested in this work are given by:

$$\begin{aligned} H:&\quad \theta \in \Theta_0 \\ A:&\quad \theta \in \Theta_0^c. \end{aligned} \tag{2-4}$$

Where, Θ_0 is some subset of the parametric space Θ and Θ_0^c is its complement. Let us define a test function $\delta : X \to \{0, 1\}$ by

$$\delta(x) = \begin{cases} 1 & \text{if we reject H when } x \text{ is observed.} \\ 0 & \text{if we do not reject H when } x \text{ is observed.} \end{cases} \tag{2-5}$$

In frequentist approach, it is common to use the Neyman-Pearson procedure for testing simple hypotheses, say $H: \theta \in \Theta_0$ versus $A: \theta \in \Theta_0^c$ hypotheses. In that case, the type I and type II error probabilities for the test δ are expressed as

$$\alpha(\delta) = \mathbb{P}(\delta(x) = 1 \mid \theta \in \Theta_0) \tag{2-6}$$

and

$$\beta(\delta) = \mathbb{P}(\delta(x) = 0 \mid \theta \in \Theta_0^c). \tag{2-7}$$

Under their theory, a rejection region is shaped by fixing beforehand an upper bound value for $\alpha(\delta) \in [0, 1]$, and, based on this rejection region, the decision between H and A is made, i.e, the choice of a test procedure is restricted to those tests δ' where:

$$\alpha(\delta') < \alpha_0,\ \theta \in \Theta_0 \text{ and } Min\left(\beta_\theta(\delta)\right),\ \theta \in \Theta_0^c, \tag{2-8}$$

with $\alpha_0 \in [0, 1]$ fixed beforehand (this idea also has been generalized for composite hypothesis). In the next section we examine the problem of testing (2-5) from a Bayesian perspective. Actually, there are different Bayesian methodologies to help researchers decide about the null hypothesis H. For instance, the Bayes factor presented by Jeffreys (1935) which is based on likelihood ratio, methods based on posterior probabilities such as the posterior predictive *p-value* developed by Rubin (1984) or methods based on the highest posterior density region like the Full Bayesian Significance Test presented by Pereira and Stern (1999). Each one of these methods has its own properties and depending on the situation they could perform better than other statistical tests. Despite this, we limit our work to the tests based on the Bayes factor proposed by Pereira *et al.* (2017).

2.2.1 Bayes Factor

The Bayes factor has its origins in Jeffreys (1935, 1936) and until today it is one of the main tools used by Bayesian statisticians for hypotheses testing. According to Hacking (1965)'s *Likelihood Law*

> *If the hypothesis H implies that the probability that a random variable X takes the value x is $f_H(x|\theta)$, while hypothesis A implies that the probability is $f_A(x|\theta)$, then the observation $X = x$ is evidence supporting H over A if and only if $f_H(x|\theta) > f_A(x|\theta)$, and the likelihood ratio $LR = \frac{f_H(\theta|x)}{f_A(\theta|x)}$, measure the strength of that evidence.*

In this work, we use the Bayes factor as the index of evidence in favor of the null hypothesis H in (2-5). Let us define the predictive function of the data under each hypothesis as

$$f_H(x) = E(L(\theta \,|\, x) \,|\, \theta \in \Theta_0) = \int_\Theta L(\theta \,|\, x) d\mathbb{P}_H(\theta) \tag{2-9}$$

and

$$f_A(x) = E(L(\theta \,|\, x) \,|\, \theta \in \Theta_0^c) = \int_\Theta L(\theta \,|\, x) d\mathbb{P}_A(\theta). \tag{2-10}$$

where $\mathbb{P}_H(\cdot)$ and $\mathbb{P}_A(\cdot)$ are, respectively, the conditional measures of θ given the hypothesis H and A. For each $x \in \mathfrak{X}$, (2-9) and (2-10) can be seen as evidence measures for the data x under each hypothesis. Hence, the Bayes factor can then be defined as

$$Bf(x) = \frac{f_H(x)}{f_A(x)}, \tag{2-11}$$

and it measures the evidence of $x \in \mathfrak{X}$ supporting H over A. Note from (2-9) and (2-10) that $f_H(x)$ ($f_A(x)$) has the form of a likelihood weighted by the prior distribution under each hypothesis. In what follows we will refer to the predictive function under each hypothesis as weighted likelihoods, where, the prior measure may be seen as a preference system in the parameter space (Lindley *et al.*, 1979; Winkler, 1967). This preference system could suggest that some subset of values in Θ are, a priori more likely than others. This can be described by a probability density function $\pi(\theta)$. And for any event this probability can be evaluated by

$$P(\theta \in \Theta_0) = \int_{\Theta_0} \pi(\theta) d\theta = \pi_H \qquad \text{and} \qquad P(\theta \in \Theta_0^c) = \int_{\Theta_0^c} \pi(\theta) d\theta = \pi_A. \tag{2-12}$$

When we are testing "simple versus simple" hypothesis, the Bayes factor is the likelihood ratio. In other cases, when either one or both of the hypotheses are composite, the Bayes factor is still given by (2-11), where, for the case of sharp hypotheses, the integral symbol in (2-9) or (2-10) will possibly represent either line or surface integrals.

In this work, we limit our comments on the Bayes factor to those already presented in the introductory chapter. For an in-depth discussion about the Bayes factor we recommend Etz and Wagenmakers (2017); Kass and Raftery (1995); Lavine and Schervish (1999).

2.2.2 Adaptive Significance Levels

In both Neyman-Pearson's and Fisher's approaches for hypothesis testing, the choice of α_0 is totally arbitrary, since in most situations there is no precise limit to the type I probability error that can be tolerated (Lehmann and Romano, 2006). As an alternative to these approaches, Cornfield (1966) and DeGroot (1986) recommend that, instead of fixing type I error probability and minimizing type II error probability to test hypothesis, it is better to use a test that minimizes the linear combination of the two error probabilities, in the case of simple versus simple hypotheses testing, that is

$$\underset{\delta}{Min}[a \times \alpha_\theta(\delta) + b \times \beta_\theta(\delta)], \quad \text{where,} \quad a, b > 0. \tag{2-13}$$

Pericchi and Pereira (2016) showed that using weighted likelihoods it is possible to construct a test $\delta^* : \mathfrak{X} \to \{0, 1\}$ that minimizes the linear combination of weighted error probabilities. This optimal test is given by

$$\delta^*(x) = \begin{cases} 1 & \text{if} \quad af_H(x) > bf_A(x) \\ 0 & \text{if} \quad af_H(x) < bf_A(x) \end{cases} \tag{2-14}$$

and accept any of them if $af_H(x) = bf_A(x)$, with $a > 0$ and $b > 0$. Pereira *et al.* (2017) uses this test $\delta^*(x)$ to calculate an optimal level of significance that depends on the sample size, thus avoiding the "Lindley paradox" in Bayesian hypothesis testing. Under their proposal the adaptive significance level α_{δ^*} is defined as

$$\begin{aligned} \alpha_{\delta^*} &= \mathbb{P}(Bf(X) \leq b/a \,|\, X \sim f_H) \\ &= \sum_{\mathfrak{D}} \int_\Theta L(\theta \,|\, x) d\mathbb{P}_H(\theta) = \sum_{\mathfrak{D}} f_H(x), \end{aligned} \tag{2-15}$$

where $\mathfrak{D} = \{x \in \mathfrak{X} : Bf(x) \leq b/a\}$. Similarly, the adaptive type II error probability β_{δ^*} is defined as

$$\begin{aligned} \beta_{\delta^*} &= \mathbb{P}(Bf(X) > b/a \,|\, X \sim f_A) \\ &= \sum_{\mathfrak{D}^c} \int_\Theta L(\theta \,|\, x) d\mathbb{P}_A(\theta) = \sum_{\mathfrak{D}^c} f_A(x). \end{aligned} \tag{2-16}$$

The above idea can also be useful in determining the sample size. For instance, if one wants to calculate a sample size for which the adaptive type I probability error does not exceed, say, 0.05. On the other hand, we want to point out that the condition $X \sim f_H$ $(X \sim f_A)$ tells us that the decision problem has been reduced to choose one of these densities $f_H(f_A)$ as being the true generator of the observed data. In a sense, the problem of testing $H : \theta \in \Theta_0$ against $A : \theta \in \Theta_0^c$ is replaced by testing $H^{'} : X \sim f_H(x)$ versus $A^{'} : X \sim f_A(x)$. In other words, we are always transforming a general hypotheses (simple v.s composite and composite v.s composite) to a "simple versus simple" hypothesis testing. In addition, Pereira *et al.* (2017) define a new *P-value* for the observed value x_0 as

$$P - value(x_0) = \sum_{\mathfrak{D}_0} f_H(x), \tag{2-17}$$

where $\mathfrak{D}_\circ = \{x \in \mathfrak{X} : Bf(x) \leq Bf(x_0)\}$. In order to get a better understanding of this new method consider the next example.

Example 1. Let X be a random variable that follows a Binomial distribution with parameters n and θ. Then, the likelihood function generated by $x \in \{0, 1, \ldots, n\}$ is given by

$$L(\theta|x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{2-18}$$

Assume that $\theta \in [0, 1]$ has prior distribution $\pi(\theta)$ such that

$$\pi(\theta) = \frac{\Gamma(c_1 + c_2)}{\Gamma(c_1)\Gamma(c_2)} \theta^{c_1 - 1}(1-\theta)^{c_2 - 1} \underset{[0,1]}{\mathbb{1}}(\theta), \tag{2-19}$$

where, $c_1, c_2 > 0$. Then, the hypotheses which we are interested can then be represented by

$$\begin{aligned} H: &\quad \theta \leq \theta_0 \\ A: &\quad \theta > \theta_0, \end{aligned} \tag{2-20}$$

with $\theta_0 \in (0, 1)$. Hence, to perform the P-P test, the first step is to compute the weighted likelihood under each hypothesis, i.e.,

$$\begin{aligned} f_H(x) &= \int_\Theta L(\theta|x) d\theta \\ &= \binom{n}{x} \frac{\displaystyle\int_0^{\theta_0} \theta^{x+c_1-1}(1-\theta)^{n-x+c_2-1} d\theta}{\displaystyle\int_0^{\theta_0} \theta^{c_1-1}(1-\theta)^{c_2-1} d\theta}, \end{aligned} \tag{2-21}$$

and

$$\begin{aligned} f_A(x) &= \int_\Theta L(\theta|x) d\mathbb{P}_A(\theta) \\ &= \binom{n}{x} \frac{\displaystyle\int_{\theta_0}^{1} \theta^{x+c_1-1}(1-\theta)^{n-x+c_2-1} d\theta}{\displaystyle\int_{\theta_0}^{1} \theta^{c_1-1}(1-\theta)^{c_2-1} d\theta}. \end{aligned} \tag{2-22}$$

The second step is to compute the Bayes factor

$$Bf(x) = \frac{\displaystyle\int_0^{\theta_0} \theta^{x+c_1-1}(1-\theta)^{n-x+c_2-1}d\theta}{\displaystyle\int_{\theta_0}^{1} \theta^{x+c_1-1}(1-\theta)^{n-x+c_2-1}d\theta} \frac{\displaystyle\int_{\theta_0}^{1} \theta^{c_1-1}(1-\theta)^{c_2-1}d\theta}{\displaystyle\int_0^{\theta_0} \theta^{c_1-1}(1-\theta)^{c_2-1}d\theta},$$

for simplicity, we will assume a uniform distribution for θ, i.e, $c_1 = c_2 = 1$. Now, if our experiment yields the result $x = 6$ for a sample size $n = 10$ and we want to know if $\theta_0 \leq 0.5$, then, the Bayes factor is expressed as

$$Bf(x) = \frac{\displaystyle\int_0^{0.5} \theta^x(1-\theta)^{10-x}d\theta}{\displaystyle\int_{0.5}^{1} \theta^x(1-\theta)^{10-x}d\theta}.$$

Table **2-1** presents the weighted likelihood functions and the Bayes factor for all possible results of x.

Table 2-1 Weighted likelihoods functions and the Bayes factor for all possible results of Binomial population with sample size $n = 10$.

x	$f_H(x)$	$f_A(x)$	$Bf(x)$
0	0.18	0.00	2047
1	0.18	0.00	170
2	0.18	0.01	30
3	0.16	0.02	7.8
4	0.13	0.05	2.6
5	0.09	0.09	1
6	0.05	0.13	0.4
7	0.02	0.16	0.1
8	0.01	0.18	0.03
9	0.00	0.18	0.01
10	0.00	0.18	0

Thus, the adaptive significance level considering $b/a = 1$ is given by

$$\alpha_{\delta^*} = \sum_{\mathfrak{D}} f_H(x) = 0.1685,$$

and the Bayesian *P-value* by

$$P - value(6) = \sum_{\mathfrak{D}_0} f_H(x) = 0.0071.$$

As in other standard tests, we can decide between to reject or not reject the hypothesis H (in 2-20) if the Bayesian *P-value*$(x_0) < \alpha_{\delta^*}$ or *P-value*$(x_0) > \alpha_{\delta^*}$ respectively. In this case, the decision is to reject H. Note that, to compare *P-value*(x_0) with α_{δ^*} is equivalent to testing by using $\delta^*(x)$ in (2-14). This equivalence is proved in Pereira *et al.* (2017).

In this work, we adopt adaptive significance levels and the Bayesian *P-value* as tools to make a decision (reject/not reject) about null hypothesis of interest. In the next section, properties of this new procedure are presented.

2.3 The resulting hypothesis tests and some properties

The P-P test procedure has been explored by Gannon *et al.* (2019); Irony and Pereira (1995); Montoya *et al.* (2001); Olivera (2014); Pereira and Pereira (2005); Pereira *et al.* (2017); Pericchi and Pereira (2016). The idea is basically to compute the Bayes factor for all possible values of x, then, compute the adaptive type I error probability (2-15) by summing up the weighted likelihood under the null hypothesis $f_H(x)$ of those values x for which the Bayes factor is less than b/a and compute the Bayesian *P-value* (2-17) for those values x for which the Bayes factor is smaller than the Bayes factor evaluated in the observed value x_0. Finally, one chooses between one of the hypotheses in (2-4) if the Bayesian *P-value*$<(>)\alpha_{\delta^*}$.

This new hypothesis test is an exact test that does not require any constraints on the dimensionalities of the sample space and parameter space, and that takes advantage of the ordering of sample space made by the Bayes factor to compute the Bayesian *P-value*. This Bayesian *P-value* may be used as a test statistic (as well as the frequentist *p-value*), but the difference here is that instead of comparing it to nominal values such as 0.05 or 0.01, the Bayesian *P-value* is compared to an adaptive significance level that is a function of the sample size, obtained from (2-13), avoiding the Lindley paradox Lindley (1957). In addition to this, the P-P test has several interesting properties.

2.3.1 Likelihood Principle

The *Likelihood Principle* concerns about either justification or evaluation of statistical inference procedures. It is frequently invoked in arguments about correct statistical reasoning. In simpler terms, the likelihood principle tells us that all information (resulting from an experiment) about the parameter θ should be contained in the Likelihood function. In addition, if two likelihood functions have the same information about θ, they must be proportional each to other. A formal definition of the likelihood principle is given by Berger and Wolper (1988):

> *The Formal Likelihood Principle* : Consider two experiments $\mathcal{E}_1 = (\mathfrak{X}_1, \Theta, \mathcal{P}^1)$ and $\mathcal{E}_2 = (\mathfrak{X}_2, \Theta, \mathcal{P}^2)$, where, $\mathfrak{X}_i$, for $i = 1, 2.$ is the sample space, Θ is the parameter space (θ is the same quantity in each experiment) and $\mathcal{P}^i$ a family of probability density functions indexed by the conditioning parameter $\theta \in \Theta$. Suppose that for the particular realizations $x_1^* \in \mathfrak{X}_1$ and $x_2^* \in \mathfrak{X}_2$ from $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively,
>
> $$L(\theta|x_1^*) = k \times L(\theta|x_2^*), \tag{2-23}$$

where $k = k(x_1^*, x_1^*)$ does not depend on θ. That is, the likelihood generated by data x_1^* in experiment $\mathcal{E}_1$ differs by a constant k from the likelihood generated by data x_2^* in experiment $\mathcal{E}_2$. Then

$$Ev(\mathcal{E}_1, x_1^*) = Ev(\mathcal{E}_2, x_2^*), \tag{2-24}$$

where $Ev(\cdot)$ represents the inferences made from the experiments and the corresponding observed sample point. In the context of hypotheses testing it means the decision about the null hypothesis (2-4) always must be the same under both $Ev(\mathcal{E}_1, x_1^*)$ and $Ev(\mathcal{E}_2, x_2^*)$, regardless of the details of the experiment that produced the observations, if (2-23) holds. Pereira *et al.* (2017) showed that the P-P test does not violate the Likelihood Principle, i.e,

$$Ev\left(Bf^{(1)}(x_1^*), x_1^*\right) = Ev\left(Bf^{(2)}(x_2^*), x_2^*\right), \tag{2-25}$$

the decision about the hypotheses (2-4) is the same when the $Bf^i(x_i^*)$ is computed under proportional likelihoods $L^1(\theta|x_1^*) = k \times L^2(\theta|x_2^*)$. In the following sections we shall present other two useful properties of this new procedure.

2.3.2 Invariance Principle

The invariance principle is not a new topic of discussion in Statistics. Initially formalized by Hotelling (1936) and Pitman (1939), its use and applications have been extensively explored. For instance, as is indicated by Berger (1985), in fiducial inference of Fisher (1935), the structural inference of Fraser (1968, 1979) and non informative priors. According to Lehmann and Romano (2008) and Lehmann (2012), in an unpublished work by Hunt and Stein (1946), this principle began to be used in hypothesis testing. Different definitions of invariant tests are presented in Eaton (1989) and Lehmann (2012). Since our aim is to show that the P-P test is invariant under transformations, we do not delve into everything related with invariance topic. We recommend the interested reader in this topic to see the mentioned references.

As is explained by Pace and Salvan (1997), through of his *Parameterization Invariance Principle*,

> *Parameterization invariance principle*: If θ and λ are two alternative parameterizations for the parametric model $\mathcal{P}$, $g(x)$ is an inferential procedure, C^Θ and C^Λ are the conclusions that $g(\cdot)$ leads to, expressed, respectively in the parameterizations θ and λ, the same conclusion C^Λ should be reached both by the application of $g(\cdot)$ in the parameterization λ and by the translation into the parameterization λ of the conclusions C^Θ.

When the inferential procedures can be considered as "intrinsic functions" that do not depend on the coordinate system expressed by the parameterization defined on $\mathcal{P}$, the *Parameterization invariance principle* holds, that is, no matter which one parameterization is chosen, the conclusion will be the same. In the next theorem, we shall formalize the idea that the P-P test has the property of invariance when we are going from a fixed measure to another fixed measure in another parametric space, consequently, the decision about the null hypothesis is the same no matter which one parameterization is chosen.

Theorem 1. Consider the probability spaces where $\Theta \subseteq \mathbb{R}^k$ and $\Lambda \subseteq \mathbb{R}^k$ are the full parametric spaces, $\sigma(\cdot)$ is the Borel $\sigma-$field from each parametric space and $\mathbb{P}_{\tilde{H}(\tilde{A})}$ is obtained from $\mathbb{P}_{H(A)}$ by direct random transformation of the form $\lambda = h(\theta)$, where H(A) represents the evaluated hypothesis. Then, define $\varphi^\Theta : X \to \{0,1\}$ and $\varphi^\Lambda : X \to \{0,1\}$ as the P-P test under $(\Theta, \sigma(\Theta), \mathbb{P}_{H(A)})$ and $(\Lambda, \sigma(\Lambda), \mathbb{P}_{\tilde{H}(\tilde{A})})$, respectively, where

$$\varphi^\Theta = \begin{cases} 1 & \text{if} \quad Bf^\Theta(x) \leq c \\ 0 & \text{if} \quad Bf^\Theta(x) > c \end{cases} \qquad \text{and} \qquad \varphi^\Lambda = \begin{cases} 1 & \text{if} \quad Bf^\Lambda(x) \leq c \\ 0 & \text{if} \quad Bf^\Lambda(x) > c. \end{cases} \tag{2-26}$$

Let $h : \Theta \to \Lambda$ be a measurable functions, i.e., $h^{-1}(\Lambda_0) \in \sigma(\Theta)$ for all $\Lambda_0 \in \sigma(\Lambda)$ and in addition, let h be a measure-preserving transformation, such that,

$$\mathbb{P}_{\tilde{H}(\tilde{A})}(h(\Theta_0)) = \mathbb{P}_{H(A)}(h^{-1}(h(\Theta_0))) = \mathbb{P}_{H(A)}(\Theta_0), \tag{2-27}$$

where $\tilde{H}(\tilde{A})$ represent the hypotheses under the alternative parameterization Λ, for $\Theta_0 \in \sigma(\Theta)$ and $\Lambda_0 \in \sigma(\Lambda)$. Then, equation (2-27) implies that $Bf^\Theta(x) = Bf^\Lambda(x)$, consequently, the P-P test is invariant under reparameterizations.

Thus, Theorem 1 imply that the diagram (**2-1**) holds for the P-P test,

$$\begin{tikzcd}
\sigma(\Theta) \arrow[r, "h"'] \arrow[d, "\mathbb{P}_{H(A)}"'] & \sigma(\Lambda) \arrow[d, "\mathbb{P}_{\tilde{H}(\tilde{A})}"] \\
\varphi^\Theta \arrow[r, "\mathbb{P}_{H(A)} \circ h^{-1}"'] & \varphi^\Lambda
\end{tikzcd}$$

Figure 2-1: The statistical procedure $\mathbb{P}_{(\cdot)}$ is invariant under reparameterization if the conclusions φ^Θ and φ^Λ are the same, no matter what parameterization Θ or Λ was chosen.

It means that no matter which parameterization is chosen, the conclusion about the null hypothesis will be the same by considering the measure $\mathbb{P}_H$ in the original parametrization (θ) or by considering $\mathbb{P}_{\tilde{H}}$. This is a useful result, because in cases where the hypotheses of interest are “complex” under the original parameterization, a reparameterization could turn it into a "straightforward" hypotheses test in another coordinate system.

Example 2. To illustrate the principle we consider three cases, each one representing different sets of full parametric space $\Lambda \subseteq \mathbb{R}^k$. We show that for a measure-preserving transformation h the P-P test obeys the *Parameterization invariance principle* in all of them. For this, let us define the hypotheses in the alternative parameterization as

$$\begin{aligned} \tilde{H}: &\quad \lambda \in \Lambda_0 \\ \tilde{A}: &\quad \lambda \in \Lambda_0^c. \end{aligned} \tag{2-28}$$

In addition, let $\tilde{f}_{(\cdot)}$ and $\tilde{\pi}(\cdot)$ be the predictive function and prior distribution respectively for the alternative parameterization.

Case 1: Let Λ_0 in (2-28) be $\Lambda_0 = B$, with $B \subseteq \Lambda$, that is, B is a subset of Λ. Hence, the transformation h has jacobian J given by

$$J = \begin{bmatrix} \frac{\partial \theta_1}{\partial \lambda_1} & \cdots & \frac{\partial \theta_1}{\partial \lambda_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_k}{\partial \lambda_1} & \cdots & \frac{\partial \theta_k}{\partial \lambda_k} \end{bmatrix}.$$

Therefore, the prior distributions for λ under each hypothesis in (2-28) are given by:

$$\tilde{\pi}(\lambda \,|\, \lambda \in \Lambda_0) = \frac{\tilde{\pi}\left(h(\theta)\right)}{\displaystyle\int_\Lambda \tilde{\pi}\left(h(\theta)\right) |J_{\tilde{H}}| d\mathbb{P}_{\tilde{H}}\left(\lambda\right)},$$

and

$$\tilde{\pi}(\lambda | \lambda \in \Lambda_0^c) = \frac{\tilde{\pi}\left(h(\theta)\right)}{\displaystyle\int_\Lambda \tilde{\pi}\left(h(\theta)\right) |J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}\left(\lambda\right)}.$$

The predictive functions under each hypothesis in (2-28) can then be represented as

$$\tilde{f}_{\tilde{H}}(x) = \int_\Lambda L(h(\theta) \,|\, x) \pi(h(\theta) \,|\, h(\theta) \in \Lambda_0) |J_{\tilde{H}}| d\mathbb{P}_{\tilde{H}}(\lambda),$$

and

$$\tilde{f}_{\tilde{A}}(x) = \int_\Lambda L(h(\theta) \,|\, x) \pi(h(\theta) \,|\, h(\theta) \in \Lambda_0^c) |J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda),$$

thus, the $Bf^\Lambda(x)$ is expressed by

$$\begin{aligned} Bf^\Lambda(x) =& \frac{\displaystyle\int_\Lambda L(h(\theta) \,|\, x) \pi(h(\theta) \,|\, h(\theta) \in \Lambda_0) |J_{\tilde{H}}| d\mathbb{P}_{\tilde{H}}(\lambda)}{\displaystyle\int_\Lambda L(h(\theta) \,|\, x) \pi(h(\theta) \,|\, h(\theta) \in \Lambda_0^c) |J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda)} \\ =& \frac{\displaystyle\int_\Theta L(h^{-1}(\lambda) \,|\, x) \pi(h^{-1}(\lambda) \,|\, h^{-1}(\lambda) \in \Theta_0) |J_H|^{-1} d\mathbb{P}_H(\theta)}{\displaystyle\int_\Theta L(h^{-1}(\lambda) \,|\, x) \pi(h^{-1}(\lambda) \,|\, h^{-1}(\lambda) \in \Theta_0^c) |J_A|^{-1} d\mathbb{P}_A(\theta)} \\ =& \frac{\displaystyle\int_\Theta L(\theta | x) \pi(\theta | \theta \in \Theta_0) d\mathbb{P}_H(\theta)}{\displaystyle\int_\Theta L(\theta | x) \pi(\theta | \theta \in \Theta_0^c) d\mathbb{P}_A(\theta)} = Bf^\Theta(x). \end{aligned} \tag{2-29}$$

Case 2: Let Λ_0 in (2-28) be $\Lambda_0 = \{b\}$, where $b \in \Lambda$, that is, b is a singleton in Λ. Hence, λ has degenerate distribution in b, consequently, this leads to the following equation for prior distributions:

$$\tilde{\pi}(\lambda \,|\, \lambda \in \Lambda_0) = \frac{\tilde{\pi}\,(h(b))}{\sum_b \tilde{\pi}\,(h(b))}$$
$$=1$$

and

$$\tilde{\pi}(\lambda \,|\, \lambda \in \Lambda_0^c) = \frac{\tilde{\pi}(h(\theta))}{\displaystyle\int_\Lambda \tilde{\pi}(h(\theta))|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda)}$$
$$=\tilde{\pi}(h(\theta)).$$

The predictive functions under each hypothesis are given by:

$$\tilde{f}_{\tilde{H}}(x) = L(h(b) \,|\, x)$$

and

$$\tilde{f}_{\tilde{A}}(x) = \int_\Lambda L(h(\theta) \,|\, x)\pi(h(\theta))|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda),$$

hence, the Bayes factor $Bf^\Lambda(x)$ can be expressed as

$$\begin{aligned} Bf^\Lambda(x) =& \frac{L(h(b)|x)}{\displaystyle\int_\Lambda L(h(\theta)\,|x)\pi(h(\theta))|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda)} \\ =& \frac{L(h^{-1}(b)|x)}{\displaystyle\int_\Theta L(h^{-1}(\lambda)\,|x)\pi(h^{-1}(\lambda))|J_A|^{-1} d\mathbb{P}_A(\theta)} \\ =& \frac{L(\theta = b|x)}{\displaystyle\int_\Theta L(\theta|x)\pi(\theta) d\mathbb{P}_A(\theta)} = Bf^\Theta(x). \end{aligned} \tag{2-30}$$

Case 3: Let Λ_0 in (2-28) be $\Lambda_0 = \{|b|\}$, where $b \in \Lambda$, that is, b is a singleton in Λ. Thus, the prior distributions under each hypothesis can be represented as

$$\tilde{\pi}(\lambda \,|\, \lambda \in \Lambda_0) = \frac{\tilde{\pi}\,(h(\theta))}{\sum_{|b|} \tilde{\pi}\,(h(\theta))}$$
$$=\pi(\lambda \,|\, \lambda \in |b|),$$

and

$$\tilde{\pi}(\lambda \,|\, \lambda \in \Lambda_0^c) = \frac{\tilde{\pi}(h(\theta))}{\displaystyle\int_\Lambda \tilde{\pi}(h(\theta))|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda)}$$
$$=\pi(h(\theta)),$$

accordingly, the prior predictive functions under each hypothesis can be written as

$$\tilde{f}_{\tilde{H}}(x) = \sum_{\{|b|\}} L\left(h(\theta) \,|\, x\right) \pi\left(h(\theta) \,|\, h(\theta) \in \{|b|\}\right),$$

and

$$\tilde{f}_{\tilde{A}}(x) = \int_{\Lambda} L(h(\theta) \mid x)\pi\left(h(\theta)\right)|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda),$$

consequently, the Bayes factor $Bf^{\Lambda}(x)$ is given by:

$$\begin{aligned} Bf^{\Lambda}(x) =& \frac{\sum_{\{|b|\}} L(h(\theta) \mid x)\pi\left(h(\theta) \mid h(\theta) \in \{|b|\}\right)}{\displaystyle\int_{\Lambda} L(h(\theta) \mid x)\pi\left(h(\theta)\right)|J_{\tilde{A}}| d\mathbb{P}_{\tilde{A}}(\lambda)} \\ =& \frac{\sum_{\{|b|\}} L(h^{-1}(\lambda) \mid x)\pi\left(h^{-1}(\lambda) \mid h^{-1}(\lambda) \in \{|b|\}\right)}{\displaystyle\int_{\Theta} L(h^{-1}(\lambda) \mid x)\pi\left(h^{-1}(\lambda)\right)|J_{A}|^{-1} d\mathbb{P}_{A}(\theta)} \\ =& \frac{\sum_{\{|b|\}} L(\theta \mid x)\pi\left(\theta \mid \theta \in \{|b|\}\right)}{\displaystyle\int_{\Theta} L(\theta \mid x)\pi\left(\theta\right) d\mathbb{P}_{A}(\theta)} = Bf^{\Theta}(x). \end{aligned}$$

We just show for the continuous case since for discrete cases the procedure is similar.

2.3.3 NNP Principle

When a multidimensional parameter, say (θ_1, θ_2) takes values in the cartesian product of the corresponding ranges, say $(\Theta_1, \Theta 2)$, and the Likelihood function (2-3) factorizes as

$$L\left(\theta|x\right) = L^1\left(\theta_1|x\right) \cdot L^2\left(\theta_2|x\right), \tag{2-31}$$

it seems reasonable that inferences about θ_1 and θ_2 can be performed independently. On the other hand, if θ is partitioned as $\theta = (\theta_1, \theta_2)$, and just θ_1 is of interest, then it seems sensible under (2-31) that we achieve the same conclusion either by using just $L^1\left(\theta_1|x\right)$ or the whole likelihood function $L\left(\theta|x\right)$. Thus makes sense to disregard the information contained in $L^2\left(\theta_2|x\right)$ and focus the interest on θ_1. As mentioned by Pace and Salvan (1997), examples of Likelihood functions with separable parameters like (2-31) are rare, but if (2-31) holds, it would be a useful property for Bayesian and non-Bayesian schools, especially in the presence of nuisance parameters. We use the equation (2-31) to motivate and revisit the principle of inference named the *Noninformative Nuisance Parameter Principle* (*NNPP*) by Berger and Wolper (1988),

> *Noninformative Nuisance Parameter Principle*: Suppose $\mathcal{E}$ is an experiment such that (2-31) is satisfied. Let $\mathcal{E}^{\theta_2}$ be the "thought" experiment in which, in addition to x, θ_2 is observed. Then θ_2 is a **noninformative nuisance parameter** if $Ev_{\theta_1}(\mathcal{E}^{\theta_2}, (x, \theta_2))$ is independent of θ_2. The NNPP states that if $\mathcal{E}$ and x are as in (2-31) and θ_2 is as **noninformative nuisance parameter**, then
>
> $$Ev_{\theta_1}(\mathcal{E}, x) = Ev(\mathcal{E}^{\theta_2}, (x, \theta_2)).$$
>
> That is, if one were to reach identical conclusion for every θ_2, were θ_2 is known, then the same conclusion should be reached even if θ_2 is unknown.

Despite its relevance in eliminating nuisance parameters in the analysis of data with nuisance parameter, Berger and Wolper (1988) presented the NNP principle in their remarkable book, but they have not explored this principle in-depth as far as we have examined in the literature. In this work, we examine this principle and put it in the context of the hypothesis testing problem.

Elimination of nuisance parameters and different notions of non-information have been studied in more detail in Barndorff-Nielsen (1976, 1978); Basu (1977) and Jørgensen (1994), where, based on suitable statistics T, the concepts of B, S and G non-information are presented. The generalized Sufficiency and Conditionality principles are also discussed. To sketch those ideas we will quote from Basu (1977) the following paragraph:

> Corresponding to any statistic T we can conceive of a decomposition of the experiment $\mathcal{E}$ into a two-stage experimental setup in which the marginal experiment $\mathcal{E}_T$ is followed by the conditional experiment $\mathcal{E}_t^T$ that corresponds to the observed value $t = T(x)$ of T. The original data $(\mathcal{E}, x)$ may then be viewed as $\{(\mathcal{E}_T, t), (\mathcal{E}_t^T, x)\}$.

Let us to introduce some definitions:

Definition 2.3.1. Variation independent : A multidimensional parameter Θ is said to be variation independent if its range is the Cartesian product of the separate ranges of its coordinates Θ_1 and Θ_2, i.e, $\Theta = \Theta_1 \times \Theta_2$.

Definition 2.3.2. p-Sufficiency (Basu): The statistic T is partially sufficient (denoted by p-sufficient) for θ_1 if for each fixed $\theta_2 \in \Theta_2$, the statistic T is sufficient with respect to the model $(\mathfrak{X}, \mathcal{F}, \mathcal{P}_{\theta_2})$, where, $\mathcal{P}_{\theta_2} = \{P_\theta : \theta_1 \in \Theta_1, \theta_2 \,\text{fixed}\}$ and if the marginal distribution of T only depends on θ_1.

Definition 2.3.3. s-Ancillarity (Basu): The statistic T is a partial ancillary (s-ancillary) for θ_1 if for each fixed $\theta_1 \in \Theta_1$, the statistic T is sufficient with respect to the model $(\mathfrak{X}, \mathcal{F}, \mathcal{P}_{\theta_1})$, where, $\mathcal{P}_{\theta_1} = \{P_\theta : \theta_1 \,\text{fixed}, \theta_2 \in \Theta_2\}$ and if the marginal distribution of T only depends on θ_2.

Then, Basu (1977) argues that if the statistics T is p-sufficient for the parameter which we are interested for inference, say θ_1, where the general vector is given by $\theta = (\theta_1,\, \theta_2)$, then it makes good statistical sense just to use the marginalized part of the data $(\mathcal{E}_T, t)$. Conversely, if T is s-ancillary for θ_1, then it appear logical to use the conditional part of the data $(\mathcal{E}_t^T, x)$. Note that Basu's approach is an application of the NNP principle: which the same conclusion should be achieved either by using the marginal experiment (or conditional experiment depending on the type of statisitcs T) or the whole experiment. Consequently, for a non-Bayesian analyst, Basu's reasoning implies that when the likelihood comes factored as in (2-31) and $\theta = (\theta_1, \theta_2)$ is variation independent (see definition below) there exist a statistic $T : \mathfrak{X} \to \mathcal{T}$ such that it is either p-sufficient or s-ancillary for θ. As a consequence of this, we can state the following theorem.

Theorem 2. Let X be a random vector with likelihood function $L : \Theta \to \mathbb{R}_+$. Assume the parameter vector $\theta = (\theta_1, \theta_2)$ is variation independent. Then, if $\exists T : \mathfrak{X} \to \mathcal{T}$ such that T is either p-sufficient or s-ancillary for θ_1, then for all $x \in \mathfrak{X}$ the likelihood function $L(\theta|x)$ can be factored as (2-31).

On the other hand, Bayesian methods for eliminating nuisance parameters based on a suitable statistics T involve different definitions of sufficiency, for instance, K-Sufficiency, Q-Sufficiency, L-Sufficiency (see for example Basu (1977) and references therein). However, we are concerned with the particular fact that for a Bayesian statistician who faces an analysis that involves nuisance parameters, it will be ideal that the posterior distribution of θ given x can be factored as $\xi(\theta_1, \theta_2|x) = \xi(\theta_1|x)\xi(\theta_2|x)$. And thus (by using Basu's words), it makes good statistical sense to work only with the posterior distribution that involves the parameter of interest θ_1. Note that if the likelihood comes factored in the manner (2-31) above and the prior distribution can be as written $\pi(\theta) = \pi(\theta_1)\pi(\theta_2)$, then the posterior distribution will also have a factored form. We formalize this in the following theorem.

Theorem 3. Let X be random variable with likelihood function $L : \Theta \to \mathbb{R}_+$. Assume the parameter vector $\theta = (\theta_1, \theta_2)$ is variation independent and let π be the prior distribution for θ with θ_1 independent of θ_2, then for all $x \in \mathfrak{X}$,

$$\theta_1 \perp\!\!\!\perp \theta_2 \,|\, x \Longleftrightarrow \exists L^i_x : \Theta \to \mathbb{R}_+, \tag{2-32}$$

such that $L(\theta|x) = L^1(\theta_1|x)L^2(\theta_2|x)$ for $i = 1.2$.

Based on the theorems and definitions given above, we present the NNP principle applied to the problem of statistical hypothesis testing. In what follows, we adapt the concepts introduced by Berger and Wolper (1988) (presented in the beginning of this section) for hypothesis testing. A formal study of the principle proposed by Berger and Wolper (1988) for more general inferential procedures is the subject of investigation for future works.

Suppose $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ is an experiment such that (2-31) is satisfied. Let $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \bar{\mathcal{P}})$ be the "thought" experiment in which, in addition to x, θ_2 is observed where the parametric space $\Theta = \Theta_1 \times \Theta_2$ is variation independent. Then, consider the following definition.

Definition 2.3.4. Non-Informative Nuisance Parameter: Let $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be test for the hypotheses

$$\begin{aligned} \bar{H}: &\quad \theta \in B \\ \bar{A}: &\quad \theta \notin B^c \end{aligned} \tag{2-33}$$

with $B \subseteq \Theta_1$. Then, for every $\theta = (\theta_1, \theta_2)$ and $x \in \mathfrak{X}$ such that (2-31) holds, we will say that θ_2 is Non-Informative Nuisance Parameter (NNP) to testing $\bar{H}$ versus $\bar{A}$ if $\bar{\varphi}\,(x, \theta_2)$ is independent of θ_2, that is, it only depends on x.

In Definition (2.3.4) we have introduced the concept of Non-Informative Nuisance Parameter in the context of hypotheses testing as a preamble to a formal definition of Non-Informative Nuisance Parameter principle. In a nutshell, this definition tells us about something that appears intuitive, that is, if θ_2 is a Non-informative parameter for θ_1, then $\bar{\varphi}$ should not depend on θ_2. In the following example, we illustrate this idea.

Example 3. Consider the experiment $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \mathcal{P})$ and let $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be the test for the hypotheses

$$\begin{aligned} \bar{H}: &\quad \theta_1 \in B \\ \bar{A}: &\quad \theta_1 \notin B, \end{aligned} \tag{2-34}$$

such that the null hypothesis will be rejected when the conditional probability of B given x and θ_2 is small, that is,

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow P(\theta_1 \in B \,|\, x, \theta_2) < \delta, \tag{2-35}$$

Let us verify that when θ_1 and θ_2 are independent, θ_2 is NNP for testing these hypotheses by means of $\bar{\varphi}$, where the right side of (2-35) leads to the following equation

$$\begin{aligned} P(\theta_1 \in B \,|\, x, \theta_2) &= \int_B \pi(\theta_1 \,|\, x, \theta_2) d\theta_1 \\ &= \frac{\displaystyle\int_B L^1(\theta_1 \,|\, x) \pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x) \pi(\theta_1) d\theta_1}, \end{aligned} \tag{2-36}$$

then, we have that

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow \frac{\displaystyle\int_B L^1(\theta_1 \,|\, x) \pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x) \pi(\theta_1) d\theta_1} < \delta. \tag{2-37}$$

Note that the equation (2-37) is independent of θ_2, so, θ_2 is NNP for θ_1.

Example 4. As in the example 3, consider the experiment $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \mathcal{P})$ where in addition to x, θ_2 is also observed and let $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be the FBST test for the hypotheses

$$\begin{aligned} \bar{H}: &\quad \theta_1 \in B \\ \bar{A}: &\quad \theta_1 \notin B. \end{aligned} \tag{2-38}$$

Then, the null hypothesis will be rejected when

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow P(\theta_1 \in T(x,\, \theta_2) \,|\, x, \theta_2) > \delta, \tag{2-39}$$

where,

$$T(x,\, \theta_2) = \{\theta_1 \in \Theta_1 : \pi(\theta_1 | \, x, \theta_2) > \underset{\theta_1 \in B}{Sup}\, \pi(\theta_1 \,|\, x, \theta_2)\}. \tag{2-40}$$

Thus, from Theorem 3 we have that for all $x \in \mathfrak{X}$ such that (2-31) holds and $\theta_1 \perp\!\!\!\perp \theta_2$, then, $\theta_1 \perp\!\!\!\perp \theta_2 |\, x$. Hence

$$\begin{aligned} T(x,\,\theta_2) =& \{\theta_1 \in \Theta_1 : \pi(\theta_1|\, x, \theta_2) > \underset{\theta_1 \in B}{Sup}\, \pi(\theta_1 \,|\, x, \theta_2)\} \\ =& \{\theta_1 \in \Theta_1 : \pi(\theta_1|\, x) > \underset{\theta_1 \in B}{Sup}\, \pi(\theta_1 \,|\, x)\} = T^*(x), \end{aligned} \tag{2-41}$$

consequently,

$$\begin{aligned} \bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow & P(\theta_1 \in T(x,\,\theta_2) \,|\, x, \theta_2) > \delta \\ \Leftrightarrow & P(\theta_1 \in T^*(x) \,|\, x) > \delta. \end{aligned} \tag{2-42}$$

Again, note that the equation (2-42) is independent of θ_2, so, θ_2 is NNP for $\bar{\varphi}(x, \theta_2)$. Now, if θ_1 and θ_2 are not independents, then θ_2 is not NNP for $\bar{\varphi}(x, \theta_2)$, even if (2-31) holds.

Hence, based on the Definition (2.3.4) it is essential to present a formal definition of the NNP principle for hypothesis testing.

Definition 2.3.5. Non-Informative Nuisance Parameter Principle: Assume $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ and $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \bar{\mathcal{P}})$ experiments as in definition (2.3.4). Let the parametric space $\Theta = \Theta_1 \times \Theta_2$ be variation independent. In addition, let $\varphi : \mathfrak{X} \to \{0, 1\}$ and $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be the tests for the hypotheses

$$\begin{aligned} H:&\quad \theta \in B \times \Theta_2 \\ A:&\quad \theta \notin B \times \Theta_2, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \bar{H}:&\quad \theta_1 \in B \\ \bar{A}:&\quad \theta_1 \notin B, \end{aligned} \tag{2-43}$$

respectively, where $B \subseteq \Theta_1$ and for $x \in \mathfrak{X}$ the equation (2-31) holds. Then, if θ_2 is NNP for test $\bar{\varphi}\,(x, \theta_2)$, so $\varphi(x) = 1 \Leftrightarrow \bar{\varphi}\,(x, \theta_2) = 1$.

Basically, the NNP principle for statistical hypothesis testing says that in both experiments, in the first one just x is observed while in the second one in addition to x, θ_2 is also observed, the decision about $\theta_1 \in B$ should be the same. Note that the previous definition is limited to hypothesis testing, but it can be extended to other inferential procedures. The following two examples illustrate these ideas.

Example 5. (continuation of example 3) Consider $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ and let $\varphi : \mathfrak{X} \to \{0, 1\}$ be the test for the hypotheses

$$\begin{aligned} H:&\quad \theta \in B \times \Theta_2 \\ A:&\quad \theta \notin B \times \Theta_2, \end{aligned} \tag{2-44}$$

with the parametric space $\Theta = \Theta_1 \times \Theta_2$ variation independent, $\theta_1 \perp\!\!\!\perp \theta_2$ and $x \in \mathfrak{X}$ such that (2-31) holds. Then, the null hypothesis will be rejected when

$$\varphi(x) = 1 \Leftrightarrow P(\theta \in B \times \Theta_2 \,|\, x) < \delta. \tag{2-45}$$

We can write the right side of equation (2-45) as

$$
\begin{aligned}
P(\theta \in B \times \Theta_2 \,|\, x) = \int_{B\times\Theta_2} \pi(\theta \,|\, x) d\theta &= \frac{\displaystyle\int_{B\times\Theta_2} L(\theta \,|\, x)\pi(\theta) d\theta}{\displaystyle\int_{\Theta} L(\theta \,|\, x)\pi(\theta) d\theta} \\
&= \frac{\displaystyle\int_{B} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1} \frac{\displaystyle\int_{\Theta_2} L^2(\theta_2 \,|\, x)\pi(\theta_2) d\theta_2}{\displaystyle\int_{\Theta_2} L^2(\theta_2 \,|\, x)\pi(\theta_2) d\theta_2} \\
&= \frac{\displaystyle\int_{B} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1}.
\end{aligned}
\tag{2-46}
$$

Hence, we will reject the hypothesis if

$$
\varphi(x) = 1 \Leftrightarrow \frac{\displaystyle\int_{B} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1} < \delta. \tag{2-47}
$$

Now, from Example 3 we have that

$$
\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow \frac{\displaystyle\int_{B} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \,|\, x)\pi(\theta_1) d\theta_1} < \delta. \tag{2-48}
$$

As a result from equations (2-47) and (2-48) we have that $\varphi(x) = 1 \Leftrightarrow \bar{\varphi}(x, \theta_2) = 1$, thus, the NNP principle is obeyed.

Example 6. (continuation of example 4) Assume $X = (X_1, X_2)$ given $\theta = (\theta_1, \theta_2)$ be a conditionally independent random variables such that each $X_i \,|\, \theta \sim Bernoulli(\theta_i)$, para $i = 1, 2$. In this case $\Theta = [0, 1] \times [0, 1]$. Let $\varphi : \mathfrak{X} \to \{0, 1\}$ be the FBST procedure for the hypotheses

$$
\begin{aligned}
H: &\quad \theta \in \{1/2\} \times [0,1] \\
A: &\quad \theta \notin \{1/2\} \times [0,1].
\end{aligned}
\tag{2-49}
$$

Here, the null hypothesis will be rejected when

$$
\varphi(x) = 1 \Leftrightarrow P(\theta \in T_x \,|\, X = x) > \delta, \tag{2-50}
$$

where

$$
T_x = \{\theta \in [0,1]^2 : \pi(\theta \,|\, x) > \underset{\theta \in \{1/2\} \times \Theta_2}{Sup} \pi(\theta \,|\, x)\}, \tag{2-51}
$$

is the tangent region used in the FBST test from Pereira and Stern (1999). Assuming that $\theta_1 \perp\!\!\!\perp \theta_2$ with $\theta_i \sim U(0, 1)$ for $i = 1, 2$. Then, for $x_1 = x_2 = 1$, we have that the likelihood function is given by

$$
\begin{aligned}
L(\theta \,|\, x_1 = 1, x_2 = 1) &= \theta_1\theta_2 \\
&= L^1(\theta_1 \,|\, x_1 = 1, x_2 = 1) L^2(\theta_2 \,|\, x_1 = 1, x_2 = 1)
\end{aligned}
\tag{2-52}
$$

Note that the equation (2-52) satisfies (2-31), in addition, by Theorem 3, $\theta_1 \perp\!\!\!\perp \theta_2 | \, x_1 = x_2 = 1$ and $\theta_i | \, x_1 = x_2 = 1 \sim Beta(2,1)$ for $i = 1, 2$. Then

$$\underset{\theta \in \{1/2\} \times \Theta_2}{Sup} \pi(\theta \,|\, x_1 = x_2 = 1) = \pi(1/2, 1) = 2. \tag{2-53}$$

The tangent region T_x (displayed in Figure **2-2**) is given by

$$\begin{aligned} T_x =& \{(\theta_1, \theta_2) \in [0,1]^2 : 4\theta_1\theta_2 \geq 2\} \\ =& \{(\theta_1, \theta_2) \in [0,1]^2 : \theta_2 \geq \frac{1}{2\theta_1}\}. \end{aligned} \tag{2-54}$$

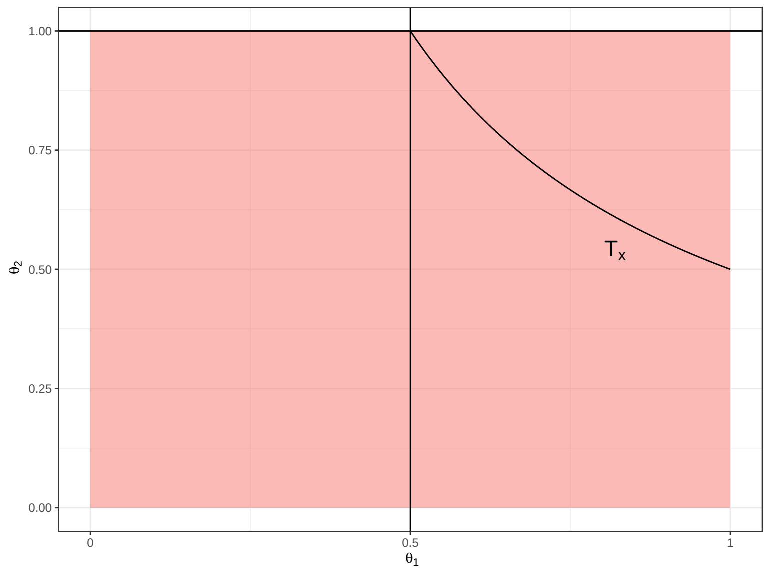

Figure 2-2: Tangent Region for FBST example of non-informative nuisance parameter.

Thus,

$$\begin{aligned} P(\theta \in T_x \,|\, x_1 = x_2 = 1) &= \int_{1/2}^{1} \int_{1/2\theta_1}^{1} 4\theta_1\theta_2 d\theta_2 d\theta_1 \\ &= \int_{1/2}^{1} 2\theta_1 \times \theta_2^2 \Big|_{1/2\theta_1}^{1} d\theta_1 \\ &= \int_{1/2}^{1} 2\theta_1 - \frac{1}{2\theta 1} d\theta_1 = \theta_1^2 - \frac{log(\theta 1)}{2} \Bigg|_{1/2}^{1} = 0,4035. \end{aligned} \tag{2-55}$$

Then, for $\delta = 1/2$ we have that $\varphi(1,1) = 0$, hence the null hypothesis is not rejected. Now, let $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \rightarrow \{0,1\}$ be the test for the hypotheses

$$\begin{aligned} \bar{H}_0 : &\quad \theta_1 = 1/2 \\ \bar{H}_1 : &\quad \theta_1 \neq 1/2. \end{aligned} \tag{2-56}$$

Analogously, as in the Example 4, the null hypothesis will be rejected (recall the tangent set constructed on the posterior for θ_1 given (x, θ_2) does not depend on θ_2 and, θ_1 and θ_2 are independent given x) when

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow P(\theta_1 \in T_x^* \,|\, X = x) > \delta. \tag{2-57}$$

Once more, for $X = (1, 1)$, we have that

$$P\left(\theta \in T_x \,|\, X = (1,1)\right) = P\left(\theta \in \left\{\theta' \in [0,1] : 2\theta' > 2\frac{1}{2}\right\} |X = (1,1)\right). \tag{2-58}$$

Because

$$\underset{\theta \in \{1/2\}}{Sup}\ \pi(\theta \,|\, x_1 = x_2 = 1) = \pi(1/2 \,|\, x_1 = x_2 = 1) = 2\frac{1}{2}, \tag{2-59}$$

Then, $T_x^* = (1/2, 1]$. As a result,

$$P\left(\theta_1 \in T_{(1,1)}^* \,|\, X = (1,1)\right) = P\left(\theta_1 \in (1/2, 1] \,|\, X = (1,1)\right) = 3/4, \tag{2-60}$$

and consequently $\bar{\varphi}(x, \theta_2) = 1$, that is, the null hypothesis is rejected. Now, since $\bar{\varphi}(x, \theta_2) = 1 \not\Leftrightarrow \varphi(x) = 1$, we can say that the FBST test does not obey the NNP principle.

Theorem 4. Let $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ and $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \bar{\mathcal{P}})$ be two experiments as defined in (2.3.4) and (2.3.5) respectively. In addition, let $\varphi : \mathfrak{X} \to \{0, 1\}$ and $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be the P-P test for the hypotheses

$$\begin{aligned} H:&\quad \theta \in \Theta_0 \times \Theta_2 \\ A:&\quad \theta \notin \Theta_0 \times \Theta_2 \end{aligned} \qquad \text{and} \qquad \begin{aligned} \bar{H}:&\quad \theta \in \Theta_0 \\ \bar{A}:&\quad \theta \notin \Theta_0, \end{aligned} \tag{2-61}$$

respectively, where $\Theta_0 = \{\theta_0\}$ and $\theta_0 \in \Theta_1$. Assume $\theta = (\theta_1, \theta_2)$ absolutely continuous with density function f and marginal density f_i for $i = 1, 2$, with θ_1 independent of θ_2. Then, the P-P test satisfies the NNP principle to test φ. That is to say, θ_2 is NNP and for all $x \in \mathfrak{X}$, $\varphi^*(x) = 1 \Leftrightarrow \bar{\varphi}^*(x, \theta_2) = 1$.

Corollary 1. Assume the same conditions of Theorem 4 and suppose that $\exists T : \mathfrak{X} \to \mathcal{T}$ such that T is p-sufficient for θ_2 and s-ancillary for θ_1. Then, for all $x \in \mathfrak{X}$, $\varphi^*(x) = 1 \Leftrightarrow \bar{\varphi}^*(x, \theta_2) = 1$.

Theorem 4 tells us that when the likelihood function may be factored as (2-31) the P-P test obeys the NNP principle. This is a good property since it reduces the statistical model and consequently the test of hypotheses that involves intensive calculations may become simpler. However, the predictive functions under the the original null hypothesis $\Theta_0 \times \Theta_2$ and under the reduced null hypothesis Θ_0 are different for both the reduced model and the whole model, hence, it becomes necessary to define the *P-value* and adaptive significance levels α^* for the reduced problem in such a way that these new quantities preserve the original decision-making between the hypotheses. Next, we present the definitions of conditional *P-value* and conditional α^* and we show that those new definitions are conditional versions of (2-17) and (2-15).

2.3.4 Conditional P-value and conditional adaptive significance level

Following Basu's approach, for the case where $\mathfrak{X}$ is a countable set, we use a suitable statistic T to define the conditional P_T-value and the conditional α^*_T. Let us define $\mathbb{F}_H$ as the probability measure associated with $f_H(\cdot)$. That is, $\mathbb{F}_H : \mathcal{P}(\mathfrak{X}) \to [0,1]$ is such that for each $B \in \mathcal{P}(\mathfrak{X})$, $\mathbb{F}_H(B) = \sum_{x\in B} f_H(x)$. Analogously, for $t \in T(\mathfrak{X})$ define $\mathbb{F}_{H,t} : \mathcal{P}(\mathfrak{X}) \to [0,1]$ as $\mathbb{F}_{H,t}(B) = \frac{\mathbb{F}_H(B\cap\{T(x)=t\})}{\mathbb{F}_H(\{T(x)=t\})}$. Based on these probability measures, consider the following definitions.

Definition 2.3.6. Conditional P-value: Assume the same conditions of Corollary 1 and Theorem 4. Then, we define the *P-value* conditional on T, P_T-value : $\mathfrak{X} \to [0,1]$ for testing the hypotheses H versus A as

$$\begin{aligned} P_T\text{-value}\,(x_0) =& \mathbb{F}_{H,T(x_0)}\left(\{x \in \mathfrak{X} : Bf(x) \leqslant Bf\,(x_0)\}\right) \\ =& \frac{\mathbb{F}_H\left(\{x \in \mathfrak{X} : Bf\,(x) \leqslant Bf\,(x_0)\,, T\,(x) = T\,(x_0)\}\right)}{\mathbb{F}_H\left(\{x \in \mathfrak{X} : T(x) = T\,(x_0)\}\right)} \\ =& \frac{\sum\limits_{\mathfrak{D}_0^*} f_H(x)}{\sum\limits_{\mathfrak{D}_T} f_H(x)}. \end{aligned} \tag{2-62}$$

Where $\mathfrak{D}_0^* = \{x \in \mathfrak{X} : Bf(x) \le Bf(x_0), T(x) = T(x_0)\}$ and $\mathfrak{D}_T = \{x \in \mathfrak{X} : T(x) = T(x_0)\}$ with x_0 as the observed value. It is worth mentioning that in case T is p-sufficient for θ_2 and s-ancillary for θ_1 with $H : \theta_1 = \theta_0$ against $A : \theta_1 \neq \theta_0$, the P_T-value can be written as

$$\begin{aligned} \frac{\sum\limits_{\mathfrak{D}_0^*} f_H(x)}{\sum\limits_{\mathfrak{D}_T} f_H(x)} =& \frac{\sum\limits_{\mathfrak{D}_0^*}\left[L^1\left(\theta_0|\,x\right)\int_{\Theta_2} L^2\left(\theta_2|\,x\right) f_2\left(\theta_2\right) d\theta_2\right]}{\sum\limits_{\mathfrak{D}_T}\left[L^1\left(\theta_0|\,x\right)\int_{\Theta_2} L^2\left(\theta_2|\,x\right) f_2\left(\theta_2\right) d\theta_2\right]} \\ =& \frac{\sum\limits_{\mathfrak{D}_0^*}\left[L^1\left(\theta_0|\,x\right)\int_{\Theta_2} P\left(T(X) = T(x)|\,\theta_2\right) f_2\left(\theta_2\right) d\theta_2\right]}{\sum\limits_{\mathfrak{D}_T}\left[L^1\left(\theta_0|\,x\right)\int_{\Theta_2} P\left(T(X) = T(x)|\,\theta_2\right) f_2\left(\theta_2\right) d\theta_2\right]}, \end{aligned} \tag{2-63}$$

and consequently is reduced to

$$\begin{aligned} \frac{\sum\limits_{\mathfrak{D}_0^*} L^1\left(\theta_0|\,x\right) g\left(T(x_0)\right)}{\sum\limits_{\mathfrak{D}_T} L^1\left(\theta_0|\,x\right) g\left(T(x_0)\right)} =& \frac{\sum\limits_{\mathfrak{D}_0^*} P(X = x|\,T(X) = T(x_0), \theta_0)}{\sum\limits_{\mathfrak{D}_T} P(X = x|\,T(X) = T(x_0), \theta_0)} \\ =& \sum_{\mathfrak{D}_0^*} P(X = x|\,T(X) = T(x_0), \theta_0). \end{aligned} \tag{2-64}$$

Definition 2.3.7. Conditional Adaptive significance level: Assume the same conditions of Corollary 1 and Theorem 4. Then, we define the conditional adaptive significance level given T, α_T^*, as

$$\begin{aligned} \alpha_T^* =& \mathbb{F}_{H,T(x_0)}\left(\{x \in \mathfrak{X} : Bf(x) \leqslant b/a\}\right) \\ =& \frac{\mathbb{F}_H\left(\{x \in \mathfrak{X} : Bf\left(x\right) \leqslant b/a, T\left(x\right) = T\left(x_0\right)\}\right)}{\mathbb{F}_H\left(\{x \in \mathfrak{X} : T(x) = T\left(x_0\right)\}\right)} \\ =& \frac{\sum\limits_{\mathfrak{D}^*} f_H(x)}{\sum\limits_{\mathfrak{D}_T} f_H(x)}. \end{aligned} \tag{2-65}$$

Where $\mathfrak{D}^* = \{x \in \mathfrak{X} : Bf(x) \leq b/a, T(x) = T(x_0)\}$. As it was commented before, Definitions (2.3.6) and (2.3.7) are conditional versions of the Definitions (2-17) and (2-15) for the reduced model when θ_2 is NNP. Note that P_T- value and α_T^* can be seen, respectively, as an alternative evidence measure of the reduced null hypothesis and an alternative threshold value for testing hypotheses in the reduced setting since they are statistics such as (2-17) and (2-15). In addition, those conditional measures are easier to calculate than the original ones (as we will see in some examples in the following chapter). From these new definitions we can obtain the following theorem.

Theorem 5. Assuming the same conditions of Corollary 1 and Theorem 4, then, for all $x_0 \in \mathfrak{X}$ we have that $\varphi^*\left(x_0\right) = 1 \Leftrightarrow P_T\text{-value}\left(x_0\right) \leqslant \alpha_T^*$.

Finally, we close this section reinforcing that although Theorem 4 has been shown for the case where Θ_0 is a singleton, this theorem holds for other types of hypotheses, as mentioned earlier in this chapter. In the next examples, we show how the P-P test follows Theorem 4 in two different scenarios.

Example 7. Consider the experiment $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ where only x was observed and $\varphi : \mathfrak{X} \to \{0, 1\}$ is the P-P test for the hypotheses

$$\begin{aligned} H: &\quad \theta \in \Theta_0 \\ A: &\quad \theta \in \Theta_0^c, \end{aligned} \tag{2-66}$$

where $\Theta_0 = B \times \Theta_2$ with $B \subseteq \Theta_1$ for the whole parametric space $\Theta = \Theta_1 \times \Theta_2$ variation independent with $\theta_1 \perp\!\!\!\perp \theta_2$ and $x \in \mathfrak{X}$ such that (2-31) holds. Then, the null hypothesis will be rejected when

$$\varphi(x) = 1 \Leftrightarrow Bf(x) < b/a, \tag{2-67}$$

where, the $Bf(x) = \frac{f_H(x)}{f_A(x)}$. Thus, the prior distribution for θ under the null hypothesis can be expressed as

$$\pi\left(\theta \,|\, \theta \in \Theta_0\right) = \begin{cases} 0 & \text{if } \theta \in \Theta_0^c \\ \dfrac{\pi\left(\theta\right)}{\displaystyle\int_{\Theta} \pi\left(\theta\right) d\mathbb{P}_H\left(\theta\right)} & \text{if } \theta \in \Theta_0, \end{cases}$$

and note that

$$\begin{aligned}\pi\left(\theta\,|\,\theta\in\Theta_0\right) &= \frac{\pi\left(\theta\right)}{\displaystyle\int_{\Theta}\pi\left(\theta\right)d\mathbb{P}_H\left(\theta\right)} = \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_{\Theta_0}\pi\left(\theta_1\right)\pi\left(\theta_2\right)d\theta_1 d\theta_2}\\ &= \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_B\int_{\Theta_2}\pi\left(\theta_1\right)\pi\left(\theta_2\right)d\theta_1 d\theta_2}\\ &= \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_B\pi\left(\theta_1\right)d\theta_1\int_{\Theta_2}\pi\left(\theta_2\right)d\theta_2}\\ &= \pi\left(\theta_1\,|\,\theta_1\in B\right)\pi\left(\theta_2\,|\,\theta 2\in\Theta_2\right),\end{aligned}$$

which means that, under the null hypothesis, the prior distribution for θ can be written as a marginal product (Theorem 3). Analogously, for the alternative hypothesis we have that

$$\pi\left(\theta\,|\,\theta\in\Theta_0^c\right) = \begin{cases} 0 & \text{if } \theta\in\Theta_0 \\ \dfrac{\pi\left(\theta\right)}{\displaystyle\int_{\Theta}\pi\left(\theta\right)d\mathbb{P}_A\left(\theta\right)} & \text{if } \theta\in\Theta_0^c, \end{cases}$$

then,

$$\pi\left(\theta\,|\,\theta\in\Theta_0^c\right) = \frac{\pi\left(\theta\right)}{\displaystyle\int_{\Theta}\pi\left(\theta\right)d\mathbb{P}_A\left(\theta\right)} = \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_{\Theta_0^c}\pi\left(\theta_1\right)\pi\left(\theta_2\right)d\theta_1 d\theta_2},$$

since $\Theta_0^c = (B\times\Theta_2)^c$ we have that

$$\begin{aligned}(B\times\Theta_2)^c &= (B^c\times\Theta_2)\cup(B\times\Theta_2^c)\cup(B^c\times\Theta_2^c)\\ &= (B^c\times\Theta_2),\end{aligned}$$

thus,

$$\begin{aligned}&= \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_{(B^c\times\Theta_2)}\pi\left(\theta_1\right)\pi\left(\theta_2\right)d\theta_1 d\theta_2}\\ &= \frac{\pi\left(\theta_1\right)\pi\left(\theta_2\right)}{\displaystyle\int_{B^c}\pi\left(\theta_1\right)d\theta_1\int_{\Theta_2}\pi\left(\theta_2\right)d\theta_2}\\ &= \pi\left(\theta_1|\theta_1\in B^c\right)\pi\left(\theta_2|\theta_2\in\Theta_2\right).\end{aligned}$$

Hence, the predictive functions under each hypothesis can be written as

$$\begin{aligned}f_H(x) &= \int_{\Theta_1}\int_{\Theta_2}L^1(\theta_1\,|\,x)L^2(\theta_2\,|\,x)\pi(\theta_1\,|\,\theta_1\in B)\pi(\theta_2\,|\,\theta_2\in\Theta_2)d\mathbb{P}_H\left(\theta_2\right)d\mathbb{P}_H\left(\theta_1\right)\\ &= \int_B L^1(\theta_1\,|\,x)\pi(\theta_1\,|\,\theta_1\in B)\left(\int_{\Theta_2}L^2(\theta_2\,|\,x)\pi(\theta_2\,|\,\theta_2\in\Theta_2)d\theta_2\right)d\theta_1\\ &= \int_B L^1(\theta_1\,|\,x)\pi(\theta_1\,|\,\theta_1\in B)d\theta_1\int_{\Theta_2}L^2(\theta_2\,|\,x)\pi(\theta_2\,|\,\theta_2\in\Theta_2)d\theta_2,\end{aligned}$$

and

$$\begin{aligned}
f_A(x) &= \int_{\Theta_1}\int_{\Theta_2} L^1(\theta_1 \mid x)L^2(\theta_2 \mid x)\pi(\theta_1 \mid \theta_1 \in B^c)\pi(\theta_2 \mid \theta_2 \in \Theta_2)d\mathbb{P}_A(\theta_2)\, d\mathbb{P}_A(\theta_1) \\
&= \int_{B^c} L^1(\theta_1 \mid x)\pi(\theta_1 \mid \theta_1 \in B^c)\left(\int_{\Theta 2} L^2(\theta_2 \mid x)\pi(\theta_2 \mid \theta_2 \in \Theta_2)d\theta_2\right) d\theta_1 \\
&= \int_{B^c} L^1(\theta_1 \mid x)\pi(\theta_1 \mid \theta_1 \in B^c)d\theta_1 \int_{\Theta_2} L^2(\theta_2 \mid x)\pi(\theta_2 \mid \theta_2 \in \Theta_2)d\theta_2,
\end{aligned}$$

as a result, the Bayes factor $Bf(x)$ can then be represented as

$$\begin{aligned}
Bf(x) &= \frac{\displaystyle\int_B L^1(\theta_1|x)\pi(\theta_1 \mid \theta_1 \in B)d\theta_1 \int_{\Theta_2} L^2(\theta_2 \mid x)\pi(\theta_2 \mid \theta_2 \in \Theta_2)d\theta_2}{\displaystyle\int_{B^c} L^1(\theta_1 \mid x)\pi(\theta_1 \mid \theta_1 \in B^c)d\theta_1 \int_{\Theta_2} L^2(\theta_2 \mid x)\pi(\theta_2 \mid \theta_2 \in \Theta_2)d\theta_2} \\
&= \frac{\displaystyle\int_B L^1(\theta_1 \mid x)\pi(\theta_1 \mid \theta_1 \in B)d\theta_1}{\displaystyle\int_{B^c} L^1(\theta_1 \mid x)\pi(\theta_1 \mid \theta_1 \in B^c)d\theta_1} \\
&= \frac{\displaystyle\int_B f(x \mid \theta_1)\pi(\theta_1 \mid \theta_1 \in B)d\theta_1}{\displaystyle\int_{B^c} f(x \mid \theta_1)\pi(\theta_1 \mid \theta_1 \in B^c)d\theta_1}.
\end{aligned} \tag{2-68}$$

Now, consider the experiment $\bar{\mathcal{E}} = (\mathfrak{X} \times \Theta_2, \Theta_1, \mathcal{P})$ where in addition to x, θ_2 also is observed. Let $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \to \{0, 1\}$ be the P-P test for the hypotheses

$$\begin{aligned}
\bar{H} :&\quad \theta_1 \in \Theta_0 \\
\bar{A} :&\quad \theta_1 \notin \Theta_0^c,
\end{aligned} \tag{2-69}$$

for $\Theta_0 = B$ with $B \subseteq \Theta_1$ and $x \in \mathfrak{X}$ such that (2-31) holds. Then, the null hypothesis will be rejected when

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow \bar{B}f(x) < b/a, \tag{2-70}$$

where, the $\bar{B}f(x) = \frac{f_{\bar{H}}(x)}{f_{\bar{A}}(x)}$. Let the likelihood for θ_1 generated by (x, θ_2) be written as

$$\begin{aligned}
\bar{L}(\theta_1 \mid x, \theta_2) &= f(x, \theta_2 \mid \theta_1) \\
&= f(x \mid \theta_2, \theta_1)f(\theta_2 \mid \theta_1),
\end{aligned}$$

as $\theta_1 \perp\!\!\!\perp \theta_2$ we have that

$$\begin{aligned}
\bar{L}(\theta_1 \mid x, \theta_2) &= \bar{L}^1(\theta_1 \mid x)\bar{L}^2(\theta_2 \mid x)f(\theta_2) \\
&= f(x \mid \theta_1)f(x \mid \theta_2)f(\theta_2).
\end{aligned} \tag{2-71}$$

Let $\pi(\theta_1)$ be the prior distribution for θ_1, thus, the predictive functions under each hypothesis are given by

$$\begin{aligned}
f_{\bar{H}}(x) &= \int_{\Theta_1} \bar{L}(\theta_1 \mid x, \theta_2)\pi(\theta_1 \mid \theta_1 \in B)\, d\mathbb{P}_{\bar{H}}(\theta_1) \\
&= \int_{\Theta_1} f(x \mid \theta_1)f(x \mid \theta_2)f(\theta_2)\pi(\theta_1 \mid \theta_1 \in B)\, d\mathbb{P}_{\bar{H}}(\theta_1) \\
&= f(x \mid \theta_2)f(\theta_2)\int_B f(x \mid \theta_1)\pi(\theta_1 \mid \theta_1 \in B)\, d\theta_1,
\end{aligned}$$

and

$$
\begin{aligned}
f_{\bar{A}}(x) &= \int_{\Theta_1} \bar{L}(\theta_1 \,|\, x, \theta_2)\pi\left(\theta_1 \,|\, \theta_1 \in B^c\right) d\mathbb{P}_{\bar{A}}\left(\theta_1\right) \\
&= \int_{\Theta_1} f(x \,|\, \theta_1) f(x \,|\, \theta_2) f(\theta_2)\pi\left(\theta_1 \,|\, \theta_1 \in B^c\right) d\mathbb{P}_{\bar{A}}\left(\theta_1\right) \\
&= f(x \,|\, \theta_2) f(\theta_2) \int_{B^c} f(x \,|\, \theta_1)\pi\left(\theta_1 \,|\, \theta_1 \in B^c\right) d\theta_1.
\end{aligned}
$$

The Bayes factor $\bar{B}f(x)$ can then be represented as

$$
\begin{aligned}
\bar{B}f(x) &= \frac{f(x \,|\, \theta_2) f(\theta_2) \displaystyle\int_{B} f(x \,|\, \theta_1)\pi\left(\theta_1 \,|\, \theta_1 \in B\right) d\theta_1}{f(x \,|\, \theta_2) f(\theta_2) \displaystyle\int_{B^c} f(x \,|\, \theta_1)\pi\left(\theta_1 \,|\, \theta_1 \in B^c\right) d\theta_1} \\
&= \frac{\displaystyle\int_{B} f(x \,|\, \theta_1)\pi\left(\theta_1 \,|\, \theta_1 \in B\right) d\theta_1}{\int_{B^c} f(x \,|\, \theta_1)\pi\left(\theta_1 \,|\, \theta_1 \in B^c\right) d\theta_1}.
\end{aligned} \tag{2-72}
$$

Consequently, we have that

$$
\begin{aligned}
\varphi(x) = 1 &\Leftrightarrow Bf(x) < b/a \\
&\Leftrightarrow \frac{\displaystyle\int_{B} f(x \,|\, \theta_1)\pi(\theta_1 \,|\, \theta_1 \in B) d\theta_1}{\displaystyle\int_{B^c} f(x \,|\, \theta_1)\pi(\theta_1 \,|\, \theta_1 \in B^c) d\theta_1} < b/a \\
&\Leftrightarrow \bar{B}f(x) < b/a.
\end{aligned} \tag{2-73}
$$

Hence,

$$
\varphi(x) = 1 \Leftrightarrow \bar{\varphi}(x, \theta_2) = 1. \tag{2-74}
$$

Example 8. Consider $\mathcal{E} = (\mathfrak{X}, \Theta, \mathcal{P})$ the experiment in which only x was observed and $\varphi : \mathfrak{X} \to \{0, 1\}$ is the P-P test for the hypotheses

$$
\begin{aligned}
H:&\quad \theta \in \Theta_0 \\
A:&\quad \theta \in \Theta_0^c,
\end{aligned} \tag{2-75}
$$

where $\Theta_0 = \{|b|\} \times \Theta_2$ with $b \in \Theta_1$ for the whole parametric space $\Theta = \Theta_1 \times \Theta_2$ variation independent with $\theta_1 \perp\!\!\!\perp \theta_2$ and $x \in \mathfrak{X}$ such that (2-31) holds. Then, the null hypothesis will be rejected if

$$
\varphi(x) = 1 \Leftrightarrow Bf(x) < b/a, \tag{2-76}
$$

where, the $Bf(x) = \frac{f_H(x)}{f_A(x)}$. Then, the prior distribution for θ under the null hypothesis is given by

$$\begin{aligned}\pi(\theta\mid\theta\in\Theta_0)=&\frac{\pi(\theta)}{\displaystyle\int_{\Theta}\pi(\theta)\,d\mathbb{P}_H(\theta)}\\=&\frac{\pi(\theta_1)}{\displaystyle\sum_{\{|b|\}}\pi(\theta_1)}\times\frac{\pi(\theta_2)}{\displaystyle\int_{\Theta_2}\pi(\theta_2)\,d\theta_2}\\=&\pi(\theta_1\mid\theta_1\in\{|b|\})\,\pi(\theta_2\mid\theta_2\in\Theta_2)\,,\end{aligned}$$

analogously, for the alternatively hypothesis we have that

$$\begin{aligned}\pi(\theta\mid\theta\in\Theta_0^c)=&\frac{\pi(\theta)}{\displaystyle\int_{\Theta}\pi(\theta)\,d\mathbb{P}_A(\theta)}\\=&\frac{\pi(\theta_1)}{\displaystyle\int_{\{|b|\}^c}\pi(\theta_1)\,d\theta_1}\frac{\pi(\theta_2)}{\displaystyle\int_{\Theta_2}\pi(\theta_2)\,d\theta_2}\\=&\pi(\theta_1\mid\theta\in\{|b|\}^c)\,\pi(\theta_2\mid\theta_2\in\Theta_2)\,.\end{aligned}$$

Hence, the predictive functions under each hypothesis can be written as

$$\begin{aligned}f_H(x)&=\int_{\Theta}L^1(\theta_1\mid x)L^2(\theta_2\mid x)\pi(\theta|\theta\in\Theta_0)d\mathbb{P}_H(\theta)\\&=\int_{\Theta}L^1(\theta_1\mid x)L^2(\theta_2\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\})\,\pi(\theta_2\mid\theta_2\in\Theta_2)d\mathbb{P}_H(\theta)\\&=\sum_{\{|b|\}}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\})\times\int_{\Theta_2}L^2(\theta_2\mid x)\pi(\theta_2\mid\theta_2\in\Theta_2)d\theta_2,\end{aligned}$$

and

$$\begin{aligned}f_A(x)&=\int_{\Theta_1}\int_{\Theta_2}L^1(\theta_1\mid x)L^2(\theta_2\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\}^c)\,\pi(\theta_2\mid\theta_2\in\Theta_2)\,d\mathbb{P}_A(\theta_2)\,d\mathbb{P}_A(\theta_1)\\&=\int_{\{|b|\}^c}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\}^c)\left(\int_{\Theta 2}L^2(\theta_2\mid x)\pi(\theta_2\mid\theta_2\in\Theta_2)\,d\theta_2\right)d\theta_1\\&=\int_{\{|b|\}^c}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\}^c)\,d\theta_1\int_{\Theta_2}L^2(\theta_2\mid x)\pi(\theta_2\mid\theta_2\in\Theta_2)\,d\theta_2.\end{aligned}$$

This leads to the following equation for the $Bf(x)$,

$$\begin{aligned}Bf(x)=&\frac{\sum_{\{|b|\}}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\})\times\displaystyle\int_{\Theta_2}L^2(\theta_2\mid x)\pi(\theta_2\mid\theta_2\in\Theta_2)\,d\theta_2}{\displaystyle\int_{\{|b|\}^c}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in b^c)\,d\theta_1\int_{\Theta_2}L^2(\theta_2|x)\pi(\theta_2|\theta_2\in\Theta_2)\,d\theta_2}\\=&\frac{\sum_{\{|b|\}}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\})}{\displaystyle\int_{\{|b|\}^c}L^1(\theta_1\mid x)\pi(\theta_1\mid\theta_1\in\{|b|\}^c)\,d\theta_1}.\end{aligned}\tag{2-77}$$

Now, consider the experiment $\bar{\mathcal{E}}=(\mathfrak{X}\times\Theta_2,\Theta_1,\mathcal{P})$ where in addition to x, θ_2 also is observed. Let $\bar{\varphi}:\mathfrak{X}\times\Theta_2\to\{0,1\}$ be the P-P test for the hypotheses

$$\begin{aligned}\bar{H}:&\quad\theta_1\in\Theta_0\\\bar{A}:&\quad\theta_1\notin\Theta_0^c,\end{aligned}\tag{2-78}$$

for $\Theta_0 = \{b\}$ with $B \in \Theta_1$ and $x \in \mathfrak{X}$ such that (2-31) holds. Then, the null hypothesis will be rejected when

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow \bar{B}f(x) < b/a, \tag{2-79}$$

where, the $\bar{B}f(x) = \frac{f_{\bar{H}}(x)}{f_{\bar{A}}(x)}$. Let the likelihood for θ_1 generated by (x, θ_2) be written as

$$\begin{aligned} \bar{L}(\theta_1 \,|\, x, \theta_2) &= f(x, \theta_2 \,|\, \theta_1) \\ &= f(x \,|\, \theta_2, \theta_1) f(\theta_2 \,|\, \theta_1), \end{aligned}$$

as $\theta_1 \perp\!\!\!\perp \theta_2$ we have that

$$\begin{aligned} \bar{L}(\theta_1 \,|\, x, \theta_2) &= \bar{L}^1(\theta_1 \,|\, x) \bar{L}^2(\theta_2 \,|\, x) f(\theta_2) \\ &= f(x \,|\, \theta_1) f(x \,|\, \theta_2) f(\theta_2). \end{aligned} \tag{2-80}$$

Hence, the prior predictive functions under each hypothesis are given by

$$\begin{aligned} f_{\bar{H}}(x) =& \int_{\Theta_1} f(x \,|\, \theta_1) f(x \,|\, \theta_2) f(\theta_2) \pi\left(\theta_1 | \theta_1 \in \{|b|\}\right) d\mathbb{P}_{\bar{H}}(\theta_1) \\ =& f(x \,|\, \theta_2) f(\theta_2) \sum_{\{|b|\}} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in |b|\right), \end{aligned}$$

and

$$\begin{aligned} f_{\bar{A}}(x) &= \int_{\Theta_1} f(x \,|\, \theta_1) f(x \,|\, \theta_2) f(\theta_2) \pi\left(\theta_1 | \theta_1 \in \{|b|\}^c\right) d\mathbb{P}_{\bar{A}}(\theta_1) \\ &= f(x \,|\, \theta_2) f(\theta_2) \int_{\{|b|\}^c} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}^c\right) d\theta_1. \end{aligned}$$

As a result, the Bayes Factor $\bar{B}f(x)$ is given by

$$\begin{aligned} \bar{B}f(x) &= \frac{f(x \,|\, \theta_2) f(\theta_2) \sum_{\{|b|\}} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}\right)}{f(x \,|\, \theta_2) f(\theta_2) \displaystyle\int_{\{|b|\}^c} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}^c\right) d\theta_1} \\ &= \frac{\sum_{\{|b|\}} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}\right)}{\displaystyle\int_{\{|b|\}^c} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}^c\right) d\theta_1}. \end{aligned} \tag{2-81}$$

Consequently, we have that

$$\begin{aligned} \varphi(x) = 1 &\Leftrightarrow Bf(x) < b/a \\ &\Leftrightarrow \frac{\sum_{\{|b|\}} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}\right)}{\displaystyle\int_{\{|b|\}^c} f(x \,|\, \theta_1) \pi\left(\theta_1 \,|\, \theta_1 \in \{|b|\}^c\right) d\theta_1} < b/a \\ &\Leftrightarrow \bar{B}f(x) < b/a. \end{aligned} \tag{2-82}$$

Hence,

$$\varphi(x) = 1 \Leftrightarrow \bar{\varphi}(x, \theta_2) = 1. \tag{2-83}$$

In this chapter, we have introduce the P-P test and we have shown that it obeys both principles, the noninformative nuisance parameters (NNP) and invariance. In addition, we saw that when the NNP principle is applied, the predictive functions are reduced: this is the reason why it is necessary to redefine the Bayesian P-value and the adaptive significance levels to the conditional version. All of these properties are useful to compute the Bayes factor when it involves complex integrals. In the next chapter, we will use these properties and we will show how the Bayes factor in hypotheses that are initially difficult to test can be estimated in a simpler way.

Chapter 3

Application to count data

In this chapter we study alternative solutions to traditional problems of hypotheses testing that involve count data such as homogeneity, independence and symmetry on contingency tables, comparison of Poisson means and Hardy–Weinberg equilibrium test. In all of them, we use the P-P test considering alternative parameterizations of the corresponding models. The alternative parameterizations result in likelihood functions that satisfy equation (2-31) and allow the use of the NNP Principle in those settings which at last may reduce the calculations involved in the determination of P-values and adaptive significance levels.

3.1 Contingency Tables

Contingency tables are formed with frequency counts of categorical data. They are common in fields such as social and health sciences, although, not restricted to those areas. More precisely, the cells of the contingency table represent the number of units observed under a cross-classification of two or more categorical variables. The analysis of contingency tables consist of identifying the dependence structure between the variables using statistical models. The statistical models depend on the sampling scheme, i.e., the way the data are collected (Agresti, 2007; Andersen, 2012). Data arranged on a contingency table are usually modeled by the sample scheme Poisson, Binomial, Multinomial or Hypergeometric. Here, we use the results from Chapter 2 to analyze the dependence structure on two-way contingency tables assuming both Multinomial and product of Multinomial schemes.

3.1.1 Homogeneity

As commented by Upton (1982), the comparison of proportions in contingency tables is a frequent practice in diverse areas of knowledge. There exists an extensive literature on this topic which reveals the importance of the problem, but also reflects the lack of universally accepted methods for testing the most common hypotheses of interest. In frequentist statistics, the common way for testing the compatibility of the data with the hypothesis of homogeneity of two or more proportions is through the standard likelihood ratio tests from Neyman and Pearson (1957) and Pearson's chi-squared from Wilks (1935), although other methods have also been developed (see, for example, Upton (1982) and references therein). Nevertheless, in situations where the sample size is small or the count cells are close to zero, these standard tests may have poor performance because they are based on asymptotic results, that is, for a large sample size. Therefore, when the sample size is small, its performance is not optimal. Indeed, when there are count cells close to zero it is not possible to use the asymptotic likelihood ratio test. In these cases, the exact tests are a natural alternative (Agresti, 2013; Graziadei, 2015; Klein and Linton, 2013). Even if the sample size is large and there are no counts close to zero, the "standard" methods inherit the problems previously discussed about tests of significance and, also, the Lindley's paradox Lindley (1957).

Next we use both Bayes factors and adaptive significance levels for the Binomial case as presented by Pereira *et al.* (2017) to decide about equality of proportions between m populations. Although it is a special case of the Multinomial model: it is more instructive to begin with it and then to extend it to Multinomial model and in this way to introduce this methodology applied to contingency tables.

Binomial Proportion

Consider m populations with a specific characteristic. Consider that samples from these populations are observed with respect to this characteristic. Let X_i be a random variable denoting if the characteristic is observed and θ_i be the probability of success in each population, for $i = 1, 2, \ldots, m$. Assume $X_1, X_2, \ldots X_m$ follow independent Binomial distribution with parameters (n_1, θ_1), (n_2, θ_2) $\ldots (n_m, \theta_m)$ respectively. We want to test whether or not the success probability θ_i are the same among of this m populations. Let us formulate the following hypotheses

$$\begin{aligned} &H : \theta_1 = \theta_2 = \ldots = \theta_m \\ &A : \theta_i \neq \theta_j \quad \text{ for } \quad i \neq j \in 1, 2, \ldots m. \end{aligned} \tag{3-1}$$

In this case, the likelihood function is given by

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = \prod_{i=1}^{m} \binom{n_i}{x_i} \theta_i^{x_i} (1-\theta_i)^{n_i - x_i} \mathbb{I}(\theta_i \in (0,1)). \tag{3-2}$$

Also, assume $(\theta_1, \theta_2, \ldots, \theta_m)$ are independent and that the prior distribution for each θ_i is given by

$$\pi(\theta_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta_i^{a_i - 1} (1-\theta_i)^{b_i - 1} \mathbb{I}(\theta_i \in (0,1)), \qquad i = 1, 2 \ldots, m, \tag{3-3}$$

that is, $\pi(\theta_i)$ is a Beta distribution with parameters $a_i > 0$ and $b_i > 0$. Thus, the predictive function under the null hypothesis H can be expressed as (for the details see the subsection A.3.1)

$$\begin{aligned} f_H(\boldsymbol{x}) =& \int_{\Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x}) d\mathbb{P}_H(\theta) \\ =& \frac{\prod_{i=1}^m \binom{n_i}{x_i} \Gamma(B)\Gamma(C)\Gamma\left(\sum_{i=1}^m (a_i+b_i) - 2(m-1)\right)}{\Gamma(B+C)\Gamma\left(\sum_{i=1}^m a_i - (m-1)\right)\Gamma\left(\sum_{i=1}^m b_i - (m-1)\right)}, \end{aligned} \tag{3-4}$$

where $B = \sum_{i=1}^m (a_i + x_i) - (m-1)$, $C = \sum_{i=1}^m (n_i + b_i - x_i) - (m-1)$. Analogously, the predictive function under alternative hypothesis A can be expressed as

$$\begin{aligned} f_A(\boldsymbol{x}) =& \int_{\Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x}) d\mathbb{P}_A(\theta) \\ =& \prod_{i=1}^m \left[\binom{n_i}{x_i} \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \frac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)} \right], \end{aligned} \tag{3-5}$$

Thus, the Bayes factor $Bf(\boldsymbol{x})$ in favor of H is given by:

$$Bf(\mathbf{x}) = \frac{\Gamma(C)\Gamma(B)\Gamma\left(\sum_{i=1}^m (a_i+b_i) - 2(m-1)\right)}{\Gamma(C+B)\Gamma\left(\sum_{i=1}^m a_i - (m-1)\right)\Gamma\left(\sum_{i=1}^m b_i - (m-1)\right)\prod_{i=1}^m D_i}, \tag{3-6}$$

where, $D_i = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \frac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)}$ for $i = 1, 2, \ldots, m$. Example 9 illustrates how homogeneity can be tested using this methodology.

Example 9. Consider three populations ($m = 3$), and suppose that we are interested in testing if those populations have the same success proportions. It is of interest to test if $H : \theta_1 = \theta_2 = \theta_3$ vs. $A : \theta_i \neq \theta_j$ for $i \neq j$. Let X_i for $i = 1, 2, 3.$, represents the number of positive outcomes (or success outcomes) in the i-th population following a Binomial distribution with likelihood function (3-2) and sample sizes $n_1 = n_2 = n_3 = 3$. For simplicity, we consider in (3-3) each $a_i = b_i = 1$. Hence, the predictive function under each hypothesis is given by

$$\begin{aligned} f_H(\mathbf{x}) =& \prod_{i=1}^3 \binom{n_i}{x_i} \frac{\left(\sum_{i=1}^3 x_i\right)! \left(\sum_{i=1}^3 n_i - x_i\right)!}{\left(\sum_{i=1}^3 n_i + 1\right)!} \\ =& \frac{\prod_{i=1}^3 \binom{n_i}{x_i}}{\binom{\sum_{i=1}^3 n_i}{\sum_{i=1}^3 x_i} \left(\sum_{i=1}^3 n_i + 1\right)}, \end{aligned} \tag{3-7}$$

and

$$\begin{aligned} f_A(\mathbf{x}) =& \prod_{i=1}^3 \binom{n_i}{x_i} \prod_{i=1}^3 \frac{x_i!(n_i-x_i)!}{(n_i+1)!} \\ =& \frac{1}{\prod_{i=1}^3 (n_i+1)}. \end{aligned} \tag{3-8}$$

Table **3-1** presents the $Bf(x)$ for all possible results of x_1, x_2, x_3.

Table 3-1 $Bf(x)$ for homogeneity hypotheses in 3 Binomial populations with sample size of $n = 3$.

		X_3			
X_1	X_2	0	1	2	3
0	0	6.4	2.13	0.533	0.0762
	1	2.13	1.6	0.686	0.152
	2	0.533	0.686	0.457	0.152
	3	0.0762	0.152	0.152	0.0762
1	0	2.13	1.6	0.686	0.152
	1	1.6	2.06	1.37	0.457
	2	0.686	1.37	1.37	0.686
	3	0.152	0.457	0.686	0.533
2	0	0.533	0.686	0.457	0.152
	1	0.686	1.37	1.37	0.686
	2	0.457	1.37	2.06	1.6
	3	0.152	0.686	1.6	2.13
3	0	0.0762	0.152	0.152	0.0762
	1	0.152	0.457	0.686	0.533
	2	0.152	0.686	1.6	2.13
	3	0.0762	0.533	2.13	6.4

Suppose $x_0 = (3, 3, 1)$ was observed. Then, in this case, $Bf(x_0) = 0.533$ for $b/a = 1$. The adaptive significance level and the Bayesian P-value are given by

$$\alpha_{\delta^*} = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}:\\ BF(\boldsymbol{x})\leq 1}} f_H(\boldsymbol{x}) = 0.257 \qquad \text{and} \qquad P\text{-}value(x_0) = \sum_{\substack{x\in\mathcal{X}:\\ BF(\boldsymbol{x})\leq 0.533}} f_H(\boldsymbol{x}) = 0.129.$$

Since $P\text{-}value(x_0) < \alpha_{\delta^*}$, the decision in this case is to reject the null hypothesis in (3-1).

Multinomial Proportions

Now, we will extend the methodology from Binomial to Multinomial proportions. We are going to evaluate whether frequency counts are distributed identically across M different populations. Consider $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M)$ as the observations of M independent Multinomial random quantities each one with k categories and parameters $(n_1, \boldsymbol{\theta_1})$, $(n_2, \boldsymbol{\theta_2}), \ldots, (n_M, \boldsymbol{\theta_M})$ respectively, with $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{i(k-1)})$, $\theta_{ij} > 0$ and $\theta_{i1} + \theta_{i2} + \ldots + \theta_{i(k-1)} \leq 1$ (considering $\theta_{ik} = 1 - (\theta_{i1} + \theta_{i2} + \ldots + \theta_{i(k-1)})$ where θ_{ij} is the probability of each category and $\boldsymbol{n} = (n_1, \ldots, n_M)$ the sample sizes for the M populations. We want to test the following hypotheses:

$$\begin{aligned} &H: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \ldots = \boldsymbol{\theta}_M \\ &A: \boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j \quad \text{for} \quad i \neq j \in 1, 2, \ldots M. \end{aligned} \tag{3-9}$$

Then, in this case, the likelihood function is given by

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i=1}^{M} \prod_{j=1}^{k} \left[\frac{n_j!}{x_{1j}! x_{2j}! \ldots x_{kj}!} \theta_{ij}^{x_{ij}} \right]. \tag{3-10}$$

Assume that the prior distribution for $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}) = \prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\prod_{i=1}^{k}\theta_{ij}^{\alpha_{ij}-1}\right], \tag{3-11}$$

where, $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_M$ are independent and follows a Dirichlet distributions with parameters vector $\alpha = (\alpha_{i1} > 0,\ldots,\alpha_{iM} > 0)$. Hence, the predictive function (2-9) under the null hypothesis H is given by (for details see subsection A.3.2)

$$\begin{aligned} f_{H}(\boldsymbol{x}) &= \int_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_{H}(\boldsymbol{\theta}) \\ &= C_1 \times \left[\frac{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)}{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\right)}\frac{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)\right)}{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)}\right], \end{aligned} \tag{3-12}$$

with $C_1 = \prod_{1}^{M}\frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}$. Analogously, the predictive function under the alternative hypothesis A is

$$\begin{aligned} f_{A}(\boldsymbol{x}) &= \int_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_{A}(\boldsymbol{\theta}) \\ &= \prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\frac{\Gamma(\sum_{i=1}^{k}x_{ij}+1)}{\prod_{i=1}^{k}\Gamma(x_{ij}+1)}\right]\prod_{j=1}^{M}\left[\frac{\prod_{i=1}^{k}\Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^{k}\alpha_{ij}+x_{ij})}\right]. \end{aligned} \tag{3-13}$$

As a result, the Bayes factor $Bf(\boldsymbol{x})$ in favor of H is given by

$$\begin{aligned} Bf(\boldsymbol{x}) &= \frac{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}x_{ij}+1)}{\prod_{i=1}^{k}\Gamma(x_{ij}+1)}\right]\frac{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)}{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\right)}\frac{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)\right)}{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)}}{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\frac{\Gamma(\sum_{i=1}^{k}x_{ij}+1)}{\prod_{i=1}^{k}\Gamma(x_{ij}+1)}\right]\prod_{j=1}^{M}\left[\frac{\prod_{i=1}^{k}\Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^{k}\alpha_{ij}+x_{ij})}\right]} \\ &= \frac{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)\right)}{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\frac{\prod_{i=1}^{k}\Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^{k}\alpha_{ij}+x_{ij})}\right]\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\right) \times \prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)}. \end{aligned} \tag{3-14}$$

Let us illustrate the result with the following example.

Example 10. Now, we are interested in testing if two populations, each one with three categories $k = 3$, according to a characteristic of interest, have the same parameter vector of proportions. We want to test $H : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ vs. $A : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, as in Example 9. We consider sample sizes $n_1 = n_2 = 3$. Let $\boldsymbol{X}_i$ for $i = 1, 2$, represent the number of outcomes in the populations. For simplicity, we consider in (3-11) that $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2} \sim Dirichlet(1,1,1)$. The predictive function under each hypothesis can be expressed as

$$\begin{aligned} f_{H}(\boldsymbol{x}) &= \prod_{j=1}^{2}\left[\frac{(\sum_{i=1}^{3}x_{ij})!}{\prod_{i=1}^{3}x_{ij}!}\right]\frac{\prod_{i=1}^{3}(\alpha_{i1}+\alpha_{i2}+x_{i1}+x_{i2}-2)!}{\left((\sum_{i=1}^{3}\alpha_{i1}+\alpha_{i2}+x_{i1}+x_{i2}-1)-1\right)!}\frac{\left((\sum_{i=1}^{3}\alpha_{i1}+\alpha_{i2}-1)-1\right)!}{\prod_{i=1}^{3}(\alpha_{i1}+\alpha_{2j}-2)!} \\ &= \prod_{j=1}^{2}\left[\frac{(\sum_{i=1}^{3}x_{ij})!}{\prod_{i=1}^{3}x_{ij}!}\right]\frac{\Gamma(3)\prod_{i=1}^{3}(x_{i1}+x_{i2})!}{\left(\sum_{i=1}^{3}x_{i1}+x_{i2}+2\right)!}, \end{aligned}$$

and

$$f_A(\boldsymbol{x}) = \prod_{j=1}^{2} \left[\frac{\Gamma\left(\sum_{i=1}^{3}\alpha_{ij}\right)}{\prod_{i=1}^{3}\Gamma\left(\alpha_{ij}\right)} \frac{\Gamma\left(\sum_{i=1}^{3}x_{ij}+1\right)}{\prod_{i=1}^{3}\Gamma\left(x_{ij}+1\right)} \right] \frac{\prod_{i=1}^{3}\Gamma(\alpha_{i1}+x_{i1})}{\Gamma\left(\sum_{i=1}^{3}\alpha_{i1}+x_{i1}\right)} \frac{\prod_{i=1}^{3}\Gamma(\alpha_{i2}+x_{i2})}{\Gamma\left(\sum_{i=1}^{3}\alpha_{i2}+x_{i2}\right)}$$
$$= \frac{4}{(n_1+2)(n_1+1)(n_2+2)(n_2+1)}.$$

Table **3-2** presents the $Bf(\boldsymbol{x})$ for all possible results of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and $n_1 = n_2 = 3$.

Table 3-2 $Bf(x)$ for homogeneity hypotheses in 2 Multinomial populations with $n_1 = n_2 = 3$ and 3 levels.

	X_2									
X_1	$(0,0,3)$	$(0,1,2)$	$(0,2,1)$	$(0,3,0)$	$(1,0,2)$	$(1,1,1)$	$(1,2,0)$	$(2,0,1)$	$(2,1,0)$	$(3,0,0)$
$(0,0,3)$	7.143	3.571	1.429	0.357	3.571	1.429	0.357	1.429	0.357	0.357
$(0,1,2)$	3.571	4.286	3.214	1.429	2.143	2.143	1.071	1.071	0.714	0.357
$(0,2,1)$	1.429	3.214	4.286	3.571	1.071	2.143	2.143	0.714	1.071	0.357
$(0,3,0)$	0.357	1.429	3.571	7.143	0.357	1.429	3.571	0.357	1.429	0.357
$(1,0,2)$	3.571	2.143	1.071	0.357	4.286	2.143	0.714	3.214	1.071	1.429
$(1,1,1)$	1.429	2.143	2.143	1.429	2.143	2.857	2.143	2.143	2.143	1.429
$(1,2,0)$	0.357	1.071	2.143	3.571	0.714	2.143	4.286	1.071	3.214	1.429
$(2,0,1)$	1.429	1.071	0.714	0.357	3.214	2.143	1.071	4.286	2.143	3.571
$(2,1,0)$	0.357	0.714	1.071	1.429	1.071	2.143	3.214	2.143	4.286	3.571
$(3,0,0)$	0.357	0.357	0.357	0.357	1.429	1.429	1.429	3.571	3.571	7.143

Suppose the following outcome was observed for each x_i

Table 3-3 Observed outcomes for each level in the populations.

	k		
	0	1	2
x_1	3	0	0
x_2	2	0	1

Then, in this case $Bf(x_1, x_2) = 3.572$ and the adaptive type I error probability and the Bayesian *P-value* for $b/a = 1$ are given

$$\alpha_{\delta(x)^*} = \sum_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ BF(\boldsymbol{x})\leq 1}} f_H(\boldsymbol{x}) = 0.013 \qquad \text{and} \qquad P\text{-}value(x_1,x_2) = \sum_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ BF(\boldsymbol{x})\leq 3.57}} f_H(\boldsymbol{x}) = 0.195.$$

Since $P\text{-}value(x_1, x_2) > \alpha_{\delta^*}$, the decision is to not reject the hypothesis H in (3-9).

3.1.2 Independence

In different applied fields, statistical data can be presented in a two-way contingency table, and it is common that a scientist wants to answer the question of whether there exists or not an association between the variables arranged in a two-way contingency table. To test association, scientists typically use the chi-squared test, log-likelihood ratio test, Neyman-modified chi-squared test, Kullback-Leibler test, Freeman-Tukey test and Cressie-Read test (Agresti, 2013; Kullback, 1971; Menéndez *et al.*, 2005). All of them are performed with asymptotic χ^2

distribution under the null hypothesis, and due to, the performances of these tests are poor when the contingency tables are sparse or with small samples. The classical alternative for small samples is the Fisher exact test. However, it is developed only for 2×2 contingency tables. Here, we provide a new solution to the problem of testing independence between two categorical variables by using the P-P test taking into account a parameterization of the model for which the likelihood function has the form (2-31). At the end, the problem of testing independence becomes an elementary test of homogeneity.

Independence hypothesis for 2×2 contingency table

Under a Multinomial sampling scheme, each θ_{ij} in Table **3-4** represents the probability of observing the i-th category of X_1 and j-th category of X_2, with X_1 and X_2 random variables. We want to test if there exists independence between this variables,

Table 3-4 Proportion of outcomes for independence hypothesis in the 2×2 case

	X_2		
X_1	θ_{11}	θ_{12}	$\theta_{1.}$
	θ_{21}	θ_{22}	$\theta_{2.}$
	$\theta_{.1}$	$\theta_{.2}$	1

where $0 \leq \theta_{ij} \leq 1$ and $\sum \theta ij = 1$. In this case, the hypothesis of independence is represented by the following equation.

$$\begin{aligned} H &: \theta_{11} = \theta_{1.}\theta_{.1} \\ A &: \theta_{11} \neq \theta_{1.}\theta_{.1}, \end{aligned} \tag{3-15}$$

Let $X = (X_{11}, X_{12}, X_{21}, X_{22})$, where X_{ij} is the number of outcomes for the row i and column j of Table **3-4**. We assume that X has a Multinomial distribution with parameter n and $\boldsymbol{\theta} = (\theta_{11}, \theta_{12}, \theta_{21})$, where n represents the sample size (under Multinomial sampling n is known) and each θ_{ij} the probability presented in the Table **3-4**. The likelihood function generated by $x = (x_{11}, x_{12}, x_{21}, x_{22})$ can be defined as

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = \frac{n!}{x_{11}! \ldots x_{22}!} \theta_{11}^{x_{11}} \ldots \theta_{22}^{x_{22}}. \tag{3-16}$$

Also, consider $\pi(\boldsymbol{\theta})$ as the prior distribution for $\boldsymbol{\theta}$ following a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$, that is

$$\pi(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_{11} + \ldots + \alpha_{22})}{\Gamma(\alpha_{11}) \ldots \Gamma(\alpha_{22})} \theta_{11}^{\alpha_{11}-1} \ldots \theta_{22}^{\alpha_{22}-1}. \tag{3-17}$$

Hence, the predictive function under the null hypothesis of independence can be expressed by

$$
\begin{aligned}
f_H(\boldsymbol{x}) =& \int_{\Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x}) d\mathbb{P}_H(\boldsymbol{\theta}) \\
=& \frac{n!}{x_{11}! \ldots x_{22}!} \left[\frac{\oint_{(0,1)^3} \theta_{11}^{x_{11}+\alpha_{11}-1} \theta_{12}^{x_{12}+\alpha_{12}-1} \theta_{21}^{x_{21}+\alpha_{21}-1} \theta_{22}^{x_{22}+\alpha_{22}-1} d\mathbb{P}_H(\boldsymbol{\theta})}{\oint_{(0,1)^3} \theta_{11}^{\alpha_{11}-1} \theta_{12}^{\alpha_{12}-1} \theta_{21}^{\alpha_{21}-1} \theta_{22}^{\alpha_{22}-1} d\mathbb{P}_H(\boldsymbol{\theta})} \right],
\end{aligned} \tag{3-18}
$$

where the symbol $\oint$ represents the line integral. Note that given the occurrence of $\theta \in \Theta_0$, the possible results of the random variable X preserve the same relative odds as they had before the experiment was carried out Barry (1996). Then, when the predictive functions have different dimensions we use the line integral to get these relatives odds. For further information about line integral, the reader can see Courant and Fritz (1974). Thus, the predictive function under alternative hypothesis A, is given by

$$
\begin{aligned}
f_A(\boldsymbol{x}) =& \int_{\Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\
=& \frac{n!}{x_{11}! \ldots x_{22}!} \frac{\Gamma(\alpha_{11} + \ldots + \alpha_{22})}{\Gamma(\alpha_{11}) \ldots \Gamma(\alpha_{22})} \left[\int_{(0,1)^3} \theta_{11}^{x_{11}+\alpha_{11}-1} \theta_{12}^{x_{12}+\alpha_{12}-1} \theta_{21}^{x_{21}+\alpha_{21}-1} \theta_{22}^{x_{22}+\alpha_{22}-1} d\boldsymbol{\theta} \right],
\end{aligned} \tag{3-19}
$$

where $\theta_{11} + \theta_{12} + \theta_{21} \leq 1$. Note that the integral in (3-18) is not simple to compute directly. But, considering a suitable reparameterization we obtain a likelihood function of the form (2-31) and then we can make use of the NNP principle for the P-P test. The test of independence can then be tackled in an easier way. Consider the new parameter $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ defined by

$$
\lambda_1 = \frac{\theta_{11}}{\theta_{12} + \theta_{11}}, \qquad \lambda_2 = \frac{\theta_{21}}{\theta_{22} + \theta_{21}} \qquad \text{and} \qquad \lambda_3 = \theta_{11} + \theta_{12}, \tag{3-20}
$$

with the new parametric space $\Lambda = [0,1]^3$. Hence, the hypotheses (3-15) can be rewritten as:

$$
\begin{aligned}
\tilde{H} :& \boldsymbol{\lambda} \in \Lambda_0 \\
\tilde{A} :& \boldsymbol{\lambda} \in \Lambda_0^c,
\end{aligned} \tag{3-21}
$$

with $\Lambda_0 = B \times \Lambda^*$, where $B = \{(\lambda_1, \lambda_2) \in [0,1]^2 : \lambda_1 = \lambda_2\}$ and $\Lambda^* = [0,1]$. Then, with abuse of notation we use $L(\cdot)$ to denote the likelihood function for $\boldsymbol{\lambda}$ generated by $\boldsymbol{x} = (x_{11}, x_{12}, x_{21}, x_{22})$ as

$$
\begin{aligned}
L(\boldsymbol{\lambda}|x) =& \tfrac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!} \left(\tfrac{\theta_{11}}{\theta_{11}+\theta_{12}}\right)^{x_{11}} \left(\tfrac{\theta_{12}}{\theta_{11}+\theta_{12}}\right)^{x_{12}} \left(\tfrac{\theta_{21}}{\theta_{21}+\theta_{22}}\right)^{x_{21}} \left(\tfrac{\theta_{22}}{\theta_{21}+\theta_{22}}\right)^{x_{22}} \\
& \qquad \times (\theta_{11}+\theta_{12})^{x_{11}+x_{12}} (\theta_{21}+\theta_{22})^{x_{21}+x_{22}} \\
=& \frac{n!}{x_{11}! \ldots x_{22}!} \lambda_1^{x_{11}} (1-\lambda_1)^{x_{12}} \lambda_2^{x_{21}} (1-\lambda_2)^{x_{22}} \lambda_3^{x_{11}+x_{12}} (1-\lambda_3)^{x_{21}+x_{22}} \\
=& \tbinom{x_{11}+x_{12}}{x_{11}} \lambda_1^{x_{11}} (1-\lambda_1)^{x_{12}} \tbinom{x_{21}+x_{22}}{x_{21}} \lambda_2^{x_{21}} (1-\lambda_2)^{x_{22}} \\
& \qquad \times \tbinom{n}{x_{11}+x_{12}} \lambda_3^{x_{11}+x_{12}} (1-\lambda_3)^{x_{21}+x_{22}}.
\end{aligned} \tag{3-22}
$$

Observe that the new hypotheses (3-21) are independent of λ_3 and that the likelihood function (3-22) can be factorized into two parts, one containing λ_1 and λ_2 and the other containing λ_3. Hence, given that the prior distribution of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ can be written as a product of Beta distributions, see (Equation A-44), the conditions of Theorem 4 are satisfied. Consequently, now we are in a straightforward problem as we have solved the problem of testing homogeneity by means of P-P test in the previous section. Note that under this new parameterization, the Bayes factor $Bf(\boldsymbol{x})$ is "simpler" to compute. Thus, the predictive function under the null hypothesis (3-18) is given by

$$
\begin{aligned}
f_{H}(\boldsymbol{x} \mid T = x_{i.}) &= \int_{\Lambda} L(\boldsymbol{\lambda} \mid \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
&= C_2 \times \left[\frac{\Gamma(n_{.1} + \alpha_{11} + \alpha_{21} - 1)\Gamma(n_{.2} + \alpha_{12} + \alpha_{22} - 1)\Gamma(\sum_{i,j}^{2} \alpha_{ij} - 2)}{\Gamma(n + \sum_{i,j=1}^{2} \alpha_{ij} - 2)\Gamma(\alpha_{11} + \alpha_{21} - 1)\Gamma(\alpha_{12} + \alpha_{22} - 1)} \right]
\end{aligned}
\tag{3-23}
$$

where, $C_2 = \binom{x_{11}+x_{12}}{x_{11}}\binom{x_{21}+x_{22}}{x_{21}}$, $n_{.j} = \sum_{i=1}^{2} x_{ij}$ and $n_{i.} = \sum_{j=1}^{2} x_{ij}$, with $i, j = 1, 2$. Now, the predictive function under the alternative hypothesis (3-19) is

$$
\begin{aligned}
f_{A}(\boldsymbol{x} \mid T = x_{i.}) &= \int_{\Lambda} L(\boldsymbol{\lambda} \mid \boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
&= C_2 \times \left[\frac{\prod_{i=1}^{2} \Gamma(\alpha_{i1} + \alpha_{i2})\Gamma(x_{11} + \alpha_{11})\Gamma(x_{12} + \alpha_{12})\Gamma(x_{21} + \alpha_{21})\Gamma(x_{22} + \alpha_{22})}{\prod_{i,j=1}^{2} \Gamma(\alpha_{ij})\Gamma(x_{11} + x_{12} + \alpha_{11} + \alpha_{12})\Gamma(x_{21} + x_{22} + \alpha_{21} + \alpha_{22})} \right],
\end{aligned}
\tag{3-24}
$$

where $x_{i.} = \sum_{j=1}^{l} x_{ij}$ and $C_2 = \binom{x_{11}+x_{12}}{x_{11}}\binom{x_{21}+x_{22}}{x_{21}}$. Hence, the $Bf(\boldsymbol{x})$ can be expressed as

$$
Bf(\boldsymbol{x}) = \left[\frac{\begin{array}{r} \Gamma(n_{.1} + \alpha_{11} + \alpha_{21} - 1)\Gamma(n_{.2} + \alpha_{12} + \alpha_{22} - 1)\Gamma(\sum_{i,j=1}^{2} \alpha_{ij} - 2) \\ \times \Gamma(n_{1.} + \alpha_{11} + \alpha_{12})\Gamma(n_{2.} + \alpha_{21} + \alpha_{22}) \prod_{i,j=1}^{2} \Gamma(\alpha_{ij}) \end{array}}{\begin{array}{r} \Gamma(n + \sum_{i,j=1}^{2} \alpha_{ij} - 2)\Gamma(\alpha_{11} + \alpha_{21} - 1)\Gamma(\alpha_{12} + \alpha_{22} - 1)\Gamma(x_{11} + \alpha_{11}) \\ \times \Gamma(x_{12} + \alpha_{12})\Gamma(x_{21} + \alpha_{21})\Gamma(x_{22} + \alpha_{22}) \prod_{i=1}^{2} \Gamma(\alpha_{i1} + \alpha_{i2}) \end{array}} \right]. \tag{3-25}
$$

Independence hypothesis for $r \times l$ contingency tables

We should note that the problem in the new parameterization is not exactly the same problem specified in the original parameter space. However, we have chosen the alternative model in order to meet the conditions of Theorem 4 and apply the NNP principle to the P-P test which reduces, at first sight, the calculations involved in the obtainment of the relevant quantities such as Bayes factors, P-values and adaptive significance levels.

As in the 2×2 case, under the Multinomial sampling scheme, each θ_{ij} in (Table **3-5**) represents the probability of the i-th category of X_1 and j-th category of X_2, where $0 \leq \theta_{ij} \leq 1$ and $\sum \theta ij = 1$. Thus, we want to test if the variables X_1 and X_2 are independent,

Table 3-5 Proportion of outcomes for independence hypothesis in the $r \times l$ case.

	X_2				
X_1	θ_{11}	θ_{12}	$\dots$	θ_{1l}	$\theta_{1.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	θ_{r1}	θ_{r2}	$\dots$	θ_{rl}	$\theta_{r.}$
	$\theta_{.1}$	$\theta_{.2}$	$\dots$	$\theta_{.l}$	1

The hypotheses for testing independence in the $r \times l$ case are given

$$\begin{aligned} H &: \theta_{ij} = \theta_{i.}\theta_{.j} \\ A &: \theta_{ij} \neq \theta_{i.}\theta_{.j}, \end{aligned} \tag{3-26}$$

for ($i = 1, \dots, r; \quad j = 1, \dots, l$). Let X_{ij} be the number of outcomes in the i-th row and j-th column of Table **3-5**. We assume that $X = (X_{11}, \dots, X_{rl})$ has a Multinomial distribution with parameter n and $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{rl})$, where n represents the sample size and each θ_{ij} the probability presented in Table **3-5**. The likelihood function generated by $\boldsymbol{x} = (x_{11}, \dots, x_{rl})$ can be expressed by

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = \frac{n!}{x_{11}! \dots x_{rl}!} \theta_{11}^{x_{11}} \dots \theta_{rl}^{x_{rl}}. \tag{3-27}$$

As in the 2×2 case, we assume the prior distribution for $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_{11} + \dots + \alpha_{rl})}{\Gamma(\alpha_{11}) \dots \Gamma(\alpha_{rl})} \theta_{11}^{\alpha_{11}-1} \dots \theta_{rl}^{\alpha_{rl}-1}, \tag{3-28}$$

where, $\pi(\boldsymbol{\theta})$ is a Dirichlet distribution with parameter vector $\alpha = (\alpha_{11}, \dots, \alpha_{rl})$ with $\alpha_{ij} > 0$. The predictive function under null the hypothesis H in (3-26) is defined by

$$\begin{aligned} f_H(\boldsymbol{x}) &= \int_{\Theta} L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_H(\boldsymbol{\theta}) \\ &= \frac{n!}{\prod_{i=1}^{r}\prod_{j=1}^{l} x_{ij}!} \left[\frac{\oint_{\Theta} \prod_{i=1}^{r}\prod_{j=1}^{l} \theta_{ij}^{x_{ij}+\alpha_{ij}-1} d\mathbb{P}_H(\boldsymbol{\theta})}{\oint_{\Theta} \prod_{i=1}^{r}\prod_{j=1}^{l} \theta_{ij}^{\alpha_{ij}-1} d\mathbb{P}_H(\boldsymbol{\theta})} \right], \end{aligned} \tag{3-29}$$

and under the alternative hypothesis A in (3-26) by

$$\begin{aligned} f_A(x) &= \int_{\Theta} L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\ &= \frac{n!}{\prod_{i=1}^{r}\prod_{j=1}^{l} x_{ij}!} \frac{\Gamma\left(\sum_{i=1}^{r}\sum_{j=1}^{l} \alpha_{ij}\right)}{\prod_{i=1}^{r}\prod_{j=1}^{l} \Gamma(\alpha_{ij})} \left[\int_{\Theta} \prod_{i=1}^{r}\prod_{j=1}^{l} \theta_{ij}^{x_{ij}+\alpha_{ij}-1} d\mathbb{P}_A(\boldsymbol{\theta}) \right]. \end{aligned} \tag{3-30}$$

Again, notice that under H the integral (3-29) is not simple to compute and for this reason we adopt as alternative for the model. Consider a reparameterization as in (2-27), where the new parameters are defined as

$$\begin{array}{llll}
\lambda_{11} = \frac{\theta_{11}}{\theta_{1.}}; & \lambda_{12} = \frac{\theta_{12}}{\theta_{1.}}; & \ldots & ; \lambda_{1(l-1)} = \frac{\theta_{1(l-1)}}{\theta_{1.}} \\
\lambda_{21} = \frac{\theta_{21}}{\theta_{2.}}; & \lambda_{22} = \frac{\theta_{22}}{\theta_{2.}}; & \ldots & ; \lambda_{2(l-1)} = \frac{\theta_{2(l-1)}}{\theta_{2.}} \\
\vdots & \vdots & \ldots & \vdots \\
\lambda_{r1} = \frac{\theta_{r1}}{\theta_{r.}}; & \lambda_{r2} = \frac{\theta_{r2}}{\theta_{r.}}; & \ldots & ; \lambda_{r(l-1)} = \frac{\theta_{r(l-1)}}{\theta_{r.}} \\
\eta_1 = \theta_{1.}; & \eta_2 = \theta_{2.}; & \ldots & ; \eta_{r-1} = \theta_{(r-1).}.
\end{array} \tag{3-31}$$

Note that the new parametric space is given by $\Lambda = \{(\lambda_{ij}, \eta_i) : \sum_{i=1}^{r-1} \eta_i \in [0,1], \sum_{j=1}^{l-1} \lambda_{ij} \in [0,1] \text{ for } \lambda_{ij} > 0, \eta_i > 0 \text{ and } i = 1, \ldots, r.\}$. With this reparameterization, the hypotheses (3-26) can be reformulated as

$$\begin{aligned}
\tilde{H} &: \boldsymbol{\lambda} \in \Lambda_0 \\
\tilde{A} &: \boldsymbol{\lambda} \in \Lambda_0^c,
\end{aligned} \tag{3-32}$$

where $\Lambda_0 = B \times \Lambda^*$, with $B = \{\lambda_{il} : \lambda_{1l} = \lambda_{2l} = \ldots = \lambda_{rl};\ l = 1, 2, \ldots, r-1.\}$ and $\Lambda^* = \{(\eta_1, \ldots, \eta_{(r-1)}) \in [0,1]^{r-1} : \sum_{i=1}^{r-1} \eta_i \leq 1\}$. Thus, with abuse of notation we use $L(\cdot)$ to denote the likelihood for $\boldsymbol{\lambda}$ generated by $\boldsymbol{x} = (x_{11}, \ldots, x_{rl})$, that is

$$\begin{aligned}
L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) = & \tfrac{n!}{x_{11}! \ldots x_{rl}!} (\lambda_{11}\eta_1)^{x_{11}} (\lambda_{12}\eta_1)^{x_{12}} \ldots (1 - \eta_1(\lambda_{11} - \lambda_{12} - \ldots - \lambda_{1(l-1)}))^{x_{1(l-1)}} \ldots (\lambda_{(r-1)1}\eta_{(r-1)})^{x_{(r-1)1}} \\
& \times (\lambda_{(r-1)2}\eta_{(r-1)})^{x_{(r-1)2}} \ldots (1 - \eta_{(r-1)}(\lambda_{(r-1)1} - \lambda_{(r-1)2} - \ldots - \lambda_{(r-1)(l-1)}))^{x_{(r-1)}} \\
= & \prod_{i=1}^{r} \left[\frac{(\sum_{j=1}^{l} x_{ij})!}{\prod_{j=1}^{l} x_{ij}!} \prod_{j=1}^{l} \lambda_{ij}^{x_{ij}} \right] \left[\frac{n!}{\prod_{i=1}^{r} (\sum_{j=1}^{l} x_{ij})!} \right] \prod_{i=1}^{r} \eta_i^{\sum_{j=1}^{l} x_{ij}}.
\end{aligned} \tag{3-33}$$

As in the 2×2 case, the new hypotheses (3-32) are independent of η_i and the likelihood function (3-33) can be factorized in two parts: one depending on λ_{ij} and the other depending on η_i only (not depending on λ_{ij}). In addition, it follows from Equation A-55 that λ_{ij} are independent of η_i. Hence, we are under the conditions of Theorem 4, consequently, the problem is now simpler because under reparameterization (3-31) and from Theorem 4 one needs now to perform a test of homogeneity which we discussed above and the $Bf(\boldsymbol{x})$ is simpler to compute. The predictive function under the null hypothesis (3-29) can be rewritten as

$$\begin{aligned}
f_{\tilde{H}}(\boldsymbol{x} | T = x_{i.}) = & \int_{\Lambda} L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
= & \prod_{i=1}^{r} \left[\frac{(\sum_{j=1}^{l} x_{ij})!}{\prod_{j=1}^{l} x_{ij}!} \right] \left[\frac{\prod_{j=1}^{l} \Gamma(\alpha_j^* - r + 1) \Gamma\left(\sum_{j=1}^{l} (\alpha_j^{**} - r + 1)\right)}{\Gamma\left(\sum_{j=1}^{l} (\alpha_j^* - r + 1)\right) \prod_{j=1}^{l} \Gamma(\alpha_j^{**} - r + 1)} \right],
\end{aligned} \tag{3-34}$$

and under the alternative hypothesis (3-30) as

$$f_{\tilde{A}}(\boldsymbol{x}|T = x_{i.}) = \int_{\Lambda} L(\boldsymbol{\lambda}\,|\,\boldsymbol{x})d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda})$$
$$= \prod_{i=1}^{r}\left[\frac{(\sum_{j=1}^{l} x_{ij})!}{\prod_{j=1}^{l} x_{ij}!}\right]\left[\frac{\prod_{i=1}^{r}\left[\prod_{j=1}^{l}\Gamma\left(x_{ij}+\alpha_{ij}\right)\Gamma\left(\alpha_i^{**}\right)\right]}{\prod_{i=1}^{r}\left[\Gamma\left(\alpha_i^{*}\right)\prod_{j=1}^{l}\Gamma\left(\alpha_{ij}\right)\right]}\right], \tag{3-35}$$

where $x_{i.} = \sum_{j=1}^{l} x_{ij}$, $\alpha_i^* = \sum_{j=1}^{l} x_{ij} + \alpha_{ij}$, $\alpha_j^* = \sum_{i=1}^{r} x_{ij} + \alpha_{ij}$, $\alpha_i^{**} = \sum_{j=1}^{l}\alpha_{ij}$ and $\alpha_j^{**} = \sum_{i=1}^{r}\alpha_{ij}$. Thus, the Bayes factor is reduced to

$$Bf(\boldsymbol{x}) = \left[\frac{\prod_{j=1}^{l}\Gamma\left(\alpha_j^* - r + 1\right)\Gamma\left(\sum_{j=1}^{l}(\alpha_j^{**} - r + 1)\right)\prod_{i=1}^{r}\left[\Gamma\left(\alpha_i^*\right)\prod_{j=1}^{l}\Gamma\left(\alpha_{ij}\right)\right]}{\Gamma\left(\sum_{j=1}^{l}(\alpha_j^* - r + 1)\right)\prod_{j=1}^{l}\Gamma\left(\alpha_j^{**} - r + 1\right)\prod_{i=1}^{r}\left[\prod_{j=1}^{l}\Gamma\left(x_{ij}+\alpha_{ij}\right)\Gamma\left(\alpha_i^{**}\right)\right]}\right]. \tag{3-36}$$

In Example 11 we illustrate the test of independence on a 3×3 contingency table.

Example 11. Let Table **3-6** represent the observed frequencies of the variables X_1 and X_2. We want to test the hypotheses (3-32), i.e., to test if X_1 and X_2 are independent.

Table 3-6 Observed frequencies of X_1 and X_2 for independence hypothesis for the Example 11.

	X_2			
X_1	2	0	1	3
	1	1	0	2
	2	2	1	5
	5	3	2	10

For simplicity, we assume that $\boldsymbol{\theta}$ in (3-28) follows a Dirichlet distribution with parameter $\alpha_{ij} = 1, \forall\, i, j$. Hence, from (3-34) and (3-35), the predictive functions under each hypothesis are defined as

$$f_{\tilde{H}}(\boldsymbol{x}) = \Gamma(3) \times \prod_{i=1}^{r}\left[\frac{(\sum_{j=1}^{l} x_{ij})!}{\prod_{j=1}^{l} x_{ij}!}\right]\left[\frac{\prod_{j=1}^{3} n_{.j}!}{(n+2)!}\right] \tag{3-37}$$

and

$$f_{\tilde{A}}(\boldsymbol{x}) = \Gamma(3)^3 \times \prod_{i=1}^{r}\left[\frac{(\sum_{j=1}^{l} x_{ij})!}{\prod_{j=1}^{l} x_{ij}!}\right]\frac{\prod_{i=1}^{3}\prod_{j=1}^{3} x_{ij}!}{\prod_{i=1}^{3}(n_{i.}+2)!}, \tag{3-38}$$

For the observed data, the Bayes factor is $Bf(\boldsymbol{x}) = 1.364$. Considering $b/a = 1$ we do not reject the null hypothesis since $Bf(\boldsymbol{x}) > 1$. In this case, the conditional adaptive significance level and the conditional Bayesian P_T-*value* (x_0) are given

$$\alpha_T^* = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}:\\ BF(\boldsymbol{x})\leq 1}} f_H(\boldsymbol{x}) = 0.284 \qquad \text{and} \qquad P_T\text{-}value\,(x_0) = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}:\\ BF(\boldsymbol{x})\leq 1.364}} f_H(\boldsymbol{x}) = 0.395.$$

Again, since P-*value*$(x_0) > \alpha_{\delta^*}$, the decision is to not reject the null hypothesis H (3-32).

3.1.3 Diagonal Symmetry

The analysis of the cell frequencies about the main diagonal in a $r \times r$ two-way contingency table is also an important problem. For instance, scientists who work with data that represent paired values or repeated values in time are frequently focused on whether or not the row classification is symmetric with the column classification, and how the row classification is symmetric or asymmetric with the column classification (as it is said by Tahata and Tomizawa (2014), in square contingency tables there is a strong association between two classifications and there is not statistical independence between them). For testing diagonal symmetry several methods have been proposed, see for example, Bowker (1948), Ireland *et al.* (1969), Kullback (1971), Bernardo *et al.* (2012) and references therein. A complete revision about symmetry and asymmetry models applied to contingency tables are found in Agresti (2013) and Tahata and Tomizawa (2014). Next, we propose a solution to this problem by means of the P-P test and its properties studied in the previous chapter.

We show, first for the 3×3 case and then for $r \times r$ case, that by applying a suitable transformation on the parameter vector $\boldsymbol{\theta}$, the likelihood function associated to the new parameter factorizes as (2-31). Then, the predictive functions (2-9) and (2-10) can be easily integrated and consequently, an exact test for testing diagonal symmetry through Bayes factor (2-11) is achieved. Also, when the symmetry hypothesis does not hold, the *P-value* (2-17) provides measure of the degree of departure from symmetry.

Diagonal Symmetry hypothesis for 3×3 contingency tables

Let Table **3-7** represent the observed frequencies of cross classified cases, according to variables X_1 and X_2,

Table 3-7 Observed frequencies of X_1 and X_2 for diagonal symmetry hypothesis in the 3×3 case.

	X_2			
X_1	x_{11}	x_{12}	x_{13}	$n_{1.}$
	x_{21}	x_{22}	x_{23}	$n_{2.}$
	x_{31}	x_{32}	x_{33}	$n_{3.}$
	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

where, as in the former examples, we assume that the vector $\boldsymbol{X} = (X_{11}, \ldots, X_{33})$ follows a Multinomial distribution with parameters n and $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{32})$ with $\theta_{ij} \geq 0$ and $\sum \theta_{ij} = 1$. The hypotheses for testing diagonal symmetry are formulated as

$$\begin{aligned} H :& \theta_{ij} = \theta_{ji} \quad \forall i \neq j. \\ A :& \theta_{ij} \neq \theta_{ji} \quad \forall i \neq j. \end{aligned} \tag{3-39}$$

where θ_{ij} denotes the probability that an observation will fall in the i-th row and j-th column of Table **3-7**. The likelihood function generated by $\boldsymbol{x} = (x_{11}, \ldots, x_{33})$ is given by

$$L(\boldsymbol{\theta}\,|\,x) = \frac{n!}{x_{11}!\ldots x_{33}!}\theta_{11}^{x_{11}}\ldots\theta_{33}^{x_{33}}. \tag{3-40}$$

Assume that the probability vector $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{32})$ has prior distribution $\pi(\boldsymbol{\theta})$ Dirichlet with parameter vector $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{33})$ for $\alpha_{ij} > 0$ such that

$$\pi(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_{11}+\ldots+\alpha_{33})}{\Gamma(\alpha_{11})\ldots\Gamma(\alpha_{33})}\theta_{11}^{\alpha_{11}-1}\ldots\theta_{32}^{\alpha_{32}-1}\left(1-\theta_{11}-\ldots-\theta_{32}\right)^{\alpha_{33}-1}, \tag{3-41}$$

In the symmetry test, the predictive function under the null hypothesis $f_H(\boldsymbol{x})$ involves calculating a surface integral that may be somewhat cumbersome. To avoid that, we use a suitable reparameterization as in (2-27), such that, the conditions of the Theorem 4 are satisfied and in this way calculations are simplified. Let us define the new variables as

$$\begin{array}{llll} \lambda_{12} = \frac{\theta_{12}}{\theta_{12}+\theta_{21}}; & \lambda_{13} = \frac{\theta_{13}}{\theta_{13}+\theta_{31}}; & \lambda_{23} = \frac{\theta_{23}}{\theta_{23}+\theta_{32}}; & \eta_{12} = \theta_{12}+\theta_{21}; \\ \eta_{13} = \theta_{13}+\theta_{31}; & \eta_{23} = \theta_{23}+\theta_{32}; & \eta_{11} = \theta_{11}; & \eta_{22} = \theta_{22}. \end{array} \tag{3-42}$$

Thus, the new parametric space is given by $\Lambda = \{(\lambda_{ij}, \eta_i) : \sum \eta_i \in [0,1]\,, \sum \lambda_{ij} \in [0,1] \text{ for } i = 1,2,3. \text{ and } i < j\}$. Then, we can reformulate the hypotheses (3-39) as

$$\begin{array}{ll} \tilde{H}: & \boldsymbol{\lambda} \in \Lambda_0 \\ \tilde{A}: & \boldsymbol{\lambda} \in \Lambda_0^c. \end{array} \tag{3-43}$$

where, defining $\Psi = \{(i,j) : i < j \text{ for } i,j = 1,2,3.\}$. $\Lambda_0 = B \times \Lambda^*$, where $B = \{\lambda_{ij} : \lambda_{ij} = 1/2,\ \forall\,(i,j) \in \Psi.\}$ and $\Lambda^* = \{(\eta_{12}, \ldots, \eta_{22}) \in (0,1)^5 : \sum \eta_{ij} \leq 1\}$. Thus, with abuse of notation we use $L(\cdot)$ to denote the likelihood function for $\boldsymbol{\lambda}$ generated by $\boldsymbol{x} = (x_{11}, \ldots, x_{33})$ as

$$L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) = \frac{n!}{\prod_{i,j=1}^{3} x_{ij}!}\left[\prod_{\Psi}\lambda_{ij}^{x_{ij}}(1-\lambda_{ij})^{x_{ji}}\right]\left[\prod_{\Psi}\eta_{ij}^{x_{ij}+x_{ji}}\prod_{i=1}^{3}\eta_{ii}^{x_{ii}}\right]. \tag{3-44}$$

As in the case of independence, the new hypotheses (3-43) are independent of η_{ij} and the new likelihood function (3-44) can be factorized in two parts, one part depending on λ_{ij} and other depending on η_{ij}. Hence, given that the prior distribution for $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{13}, \lambda_{23})$ can be written as a product of Beta distributions, see (Equation A-67), once again the conditions of Theorem 4 are satisfied. Consequently, we are in a straightforward problem, since the reparameterization (3-42) transforms (3-39) into a simple hypothesis and, as we discussed above, conditional on that hypotheses, it is easier to compute the $Bf(\boldsymbol{x})$. The predictive function under the null hypothesis can be rewritten as

$$\begin{aligned} f_{\tilde{H}}(\boldsymbol{x}|T = x_{\Psi}) &= \int_{\Lambda} L(\lambda\,|\,\boldsymbol{x})d\mathbb{P}_{\tilde{H}}(\lambda) \\ &= \left[(1/2)^{\sum_{\Psi} x_{ij}+x_{ji}}\prod_{\Psi}\binom{x_{ij}+x_{ji}}{x_{ij}}\right], \end{aligned} \tag{3-45}$$

and

$$\begin{aligned} f_{\tilde{A}}(\boldsymbol{x}|T = x_{\Psi}) &= \int_{\Lambda} L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\ &= \left[\prod_{\Psi} \binom{x_{ij}+x_{ji}}{x_{ij}} \frac{\prod_{\Psi}\Gamma(\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi}\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})} \frac{\prod_{\Psi}\Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_{\Psi}\Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}\right], \end{aligned} \tag{3-46}$$

where $x_{\Psi} = \sum_{\Psi} x_{ij} + x_{ji}$. Finally, the $Bf(\boldsymbol{x})$ is given by

$$\begin{aligned} BF(\boldsymbol{x}) =& \frac{\int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda})}{\int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda})} \\ =& (1/2)^{\sum_{\Psi} x_{ij}+x_{ji}} \left[\frac{\prod_{\Psi}\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})}{\prod_{\Psi}\Gamma(\alpha_{ij}+\alpha_{ji})} \frac{\prod_{\Psi}\Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi}\Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}\right]. \end{aligned} \tag{3-47}$$

The above methodology is easy to extend for the general case, which will be done in the sequel.

Diagonal Symmetry hypothesis for $r \times r$ contingency tables

Notice that for the general case the new reparameterization (3-42) can be defined as

$$\begin{aligned} \lambda_{ij} &= \tfrac{\theta_{ij}}{\theta_{ij}+\theta_{ji}} \quad \text{for} \quad i<j \quad \text{and} \quad i,j = 1,2,\ldots,r. \\ \eta_{ij} &= \theta_{ij}+\theta_{ji} \quad \text{for} \quad i<j \quad \text{and} \quad i,j = 1,2,\ldots,r. \\ \eta_{ij} &= \theta_{ij} \quad \text{for} \quad i=j \quad \text{and} \quad i,j = 1,2,\ldots,r-1. \end{aligned} \tag{3-48}$$

Then, the hypotheses have the same form as in (3-43). That is

$$\begin{aligned} \tilde{H}:&\quad \boldsymbol{\lambda} \in \Lambda_0 \\ \tilde{A}:&\quad \boldsymbol{\lambda} \in \Lambda_0^c. \end{aligned} \tag{3-49}$$

where, defining $\Psi^* = \{(i,j) : i<j \text{ for } i,j = 1,\ldots,r.\}$. $\Lambda_0 = B \times \Lambda^*$, where $B = \{\lambda_{ij} : \lambda_{ij} = 1/2,\ \forall\,(i,j) \in \Psi^*.\}$ and $\Lambda^* = \{(\eta_{12},\ldots,\eta_{r(r-1)}) \in (0,1)^{\frac{r^2+r}{2}} : \sum \eta_{ij} \leq 1\}$. Thus, the new likelihood function for the general case is given by

$$L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) = \frac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \left[\prod_{\Psi^*} \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}}\right] \left[\prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}+\alpha_{ii}-1}\right]. \tag{3-50}$$

Note that for the general case the likelihood may be factorized. Hence, once again we are under conditions of Theorem 4. Then, the predictive function under the null hypothesis H can be written as

$$\begin{aligned} f_{\tilde{H}}(\boldsymbol{x}|T = x_{\Psi^*}) &= \int_{\Lambda} L(\lambda\,|\,\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\lambda) \\ &= (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}} \prod_{\Psi^*} \binom{x_{ij}+x_{ji}}{x_{ij}}, \end{aligned} \tag{3-51}$$

and, the predictive function under the alternative hypothesis A as

$$
\begin{aligned}
f_{\tilde{A}}(\boldsymbol{x}|\, T = x_{\Psi^*}) &= \int_\Lambda L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
&= \left[\prod_{\Psi^*} \binom{x_{ij}+x_{ji}}{x_{ij}} \frac{\prod_{\Psi^*} \Gamma(\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi^*} \Gamma(\alpha_{ij})\Gamma(\alpha_{ji})} \frac{\prod_{\Psi^*} \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_{\Psi^*} \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}\right],
\end{aligned} \tag{3-52}
$$

where $x_{\Psi^*} = \sum_{\Psi^*} x_{ij} + x_{ji}$. Finally, the $Bf(\boldsymbol{x})$ for the general case is given by

$$
\begin{aligned}
Bf(\boldsymbol{x}) =& \frac{\int_\Lambda L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda})}{\int_\Lambda L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda})} \\
=& (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}} \left[\frac{\prod_{\Psi^*} \Gamma(\alpha_{ij})\Gamma(\alpha_{ji})}{\prod_{\Psi^*} \Gamma(\alpha_{ij}+\alpha_{ji})} \frac{\prod_{\Psi^*} \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi^*} \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}\right].
\end{aligned} \tag{3-53}
$$

In the next example we illustrate how to apply this methodology to test symmetry on a 3×3 contingency table.

Example 12. Assume that Table **3-8** represents the observed frequencies of cross classified cases, according to two variables X_1 and X_2.

Table 3-8 Observed frequencies of X_1 and X_2 for diagonal symmetry hypothesis for the Example 12.

	X_2			
X_1	0	1	2	3
	1	1	2	4
	2	2	3	7
	3	4	7	14

For simplicity, suppose that $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{32})$ a prior follows a Dirichlet distribution with parameters $\alpha_{ij} = 1\,\forall\, i, j$. Hence, the predictive functions under each hypothesis are given by

$$
f_{\tilde{H}}(\boldsymbol{x}) = (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}} \prod_{\Psi^*} \binom{x_{ij}+x_{ji}}{x_{ij}}, \tag{3-54}
$$

and

$$
f_{\tilde{A}}(\boldsymbol{x}) = \left[\frac{\prod_{\Psi^*} x_{ij}!x_{ij}!}{\prod_{\Psi^*} (x_{ij}+x_{ji}+1)!}\right] \prod_{\Psi^*} \binom{x_{ij}+x_{ji}}{x_{ij}}. \tag{3-55}
$$

Again, for the observed data, the Bayes factor is $Bf(\boldsymbol{x}) = 5.273$. Considering $a = b(b/a = 1)$ we do not reject the null hypothesis since $Bf(\boldsymbol{x}) > 1$. In this case, conditional the adaptive significance level and the conditional Bayesian P_T-*value* (x_0) are

$$
\alpha^*_T = \sum_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ BF(\boldsymbol{x})\le 1}} f_H(\boldsymbol{x}) = 0.048 \qquad \text{and} \qquad P_T\text{-}value\,(x_0) = \sum_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ BF(\boldsymbol{x})\le BF(x_0)}} f_H(\boldsymbol{x}) = 0.252.
$$

Since P_T-*value*$(x_0) > \alpha^*_T$, the decision is not to reject the null hypothesis in (3-32).

3.2 Other Applications

Our purpose so far has been to use the P-P test (with weighted likelihoods and adaptive significance levels) on count data in contingency tables. We showed how the *NNP* principle can be used to achieve a simpler form of the original problem. Next, we show the use of the procedure developed by (Pericchi and Pereira, 2016) in other cases. The first one is the comparison of means between two populations that follow Poisson distributions, the second one is the test of Hardy-Weinberg equilibrium in relative genotype frequencies and, finally, we apply the results of independence and symmetry to a real data set from Brazilian aeronautics.

3.2.1 Poisson means comparison

Suppose we are interested in testing the equality between two Poisson means. For this purpose, let $\mathbf{X} = (X_1, X_2)$ be independent observations that follow a Poisson distribution with parameters $m\theta_1$ and $n\theta_2$ respectively, with $m, n \in \mathbb{N} = 1, 2, 3, \ldots$ as the sample size for each random variable. The hypotheses to be tested can be formulated as

$$\begin{aligned} H : &\boldsymbol{\theta} \in \Theta_0 \\ A : &\boldsymbol{\theta} \in \Theta_0^c, \end{aligned} \tag{3-56}$$

where $\Theta_0 = \{(\theta_1, \theta_2) \in \mathbb{R}_+^2 : \theta_1 = \theta_2\}$. The likelihood function for $\boldsymbol{\theta} = (\theta_1, \theta_2)$ generated by $x = (x_1, x_2) \in \mathbb{N}^2$ is defined by

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = \frac{(m\theta_1)^{x_1}}{x_1!} \frac{(n\theta_2)^{x_2}}{x_2!} e^{-\theta_1 m} e^{-\theta_2 n}. \tag{3-57}$$

Assuming that $\boldsymbol{\theta}$ has prior distribution as

$$\pi(\boldsymbol{\theta}) = \frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \theta_1^{a-1} \theta_2^{c-1} e^{-\theta_1 b} e^{-\theta_2 d}, \tag{3-58}$$

that is, θ_1 and θ_2 are distributed as a Gamma distributions with parameters $(a > 0, b > 0)$ and $(c > 0, d > 0)$ respectively and $\theta_1 \perp\!\!\!\perp \theta_2$. Then, the predictive function under the null hypothesis H can be expressed by

$$\begin{aligned} f_H(\boldsymbol{x}) = &\int_\Theta L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_H(\boldsymbol{\theta}) \\ = &\frac{m^{x_1} n^{x_2}}{x_1! x_2!} \left[\frac{\Gamma(x_1 + x_2 + a + c - 1)(b + d)^{(a+c-1)}}{(b + d + m + n)^{(x_1 + x_2 + a + c - 1)} \Gamma(a + c - 1)} \right]. \end{aligned} \tag{3-59}$$

Analogously, the predictive function under the alternative hypothesis A is given by

$$\begin{aligned} f_A(\boldsymbol{x}) = &\int_\Theta L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\ = &\frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \frac{m^{x_1} n^{x_2}}{x_1! x_2!} \left[\frac{\Gamma(x_1 + a)}{(m + b)^{(x_1 + a)}} \frac{\Gamma(x_2 + c)}{(n + d)^{(x_2 + c)}} \right]. \end{aligned} \tag{3-60}$$

Then, the Bayes factor $Bf(\boldsymbol{x})$ is expressed by

$$Bf(\boldsymbol{x}) = \frac{\Gamma(a)\Gamma(c)}{b^a d^c}\left[\frac{\Gamma(x_1+x_2+a+c-1)(b+d)^{(a+c-1)}}{(b+d+m+n)^{(x_1+x_2+a+c-1)}\Gamma(a+c-1)}\frac{(m+b)^{(x_1+a)}(n+d)^{(x_2+c)}}{\Gamma(x_1+a)\Gamma(x_2+c)}\right]. \tag{3-61}$$

The problem of comparing two Poisson means by using (3-61) was first presented by Irony and Pereira (1995). They explain how a production process can be monitored through hypotheses testing using weighted likelihoods. Note that in order to calculate P-values and adaptive significance levels for the P-P test for the equality of means, one should conceptually determine the Bayes factor in (3-61) for infinitely many points in the sample space $\mathbb{N}^2$, which is not viable from any point of view. To overcome this difficult, we reparametrize the model, and make use of Theorem 4 to reduce the amount of calculations and to obtain conditional quantities instead. Next, we show that assuming equal sample sizes and equal rate parameters, i.e.,

$$n = m \qquad \text{and} \qquad b = d, \tag{3-62}$$

and by using a suitable reparameterization in order to obtain a factored likelihood, the P-P test procedure meets the condition of Theorem 4, and consequently, this problem will be reduced to a sharp hypothesis testing problem. Thus, the new parameters can defined as

$$\lambda_1 = \frac{\theta_1}{\theta_1+\theta_2} \quad \text{and} \quad \lambda_2 = \theta_1 + \theta_2. \tag{3-63}$$

The new parametric space is $\Lambda = [0,1]\times\mathbb{R}_+$. Now, the hypotheses (3-56) can be rewritten as

$$\begin{aligned} \tilde{H} &: \lambda_1 \in \Lambda_0 \\ \tilde{A} &: \lambda_1 \in \Lambda_0^c, \end{aligned} \tag{3-64}$$

with $\Lambda_0 = B\times\Lambda^*$, where $B = \{\lambda_1 : \lambda_1 = \{1/2\}\}$ and $\Lambda^* = (0,\infty)$. Note that, since X_1 and X_2 are independent, the likelihood function (3-57) can be expressed as

$$\begin{aligned} P(X_1=x_1, X_2=x_2\,|\,\boldsymbol{\theta}) =& P(X_1=x_1\,|\,X_1+X_2=x_1+x_2,\boldsymbol{\theta})P(X_1+X_2=x_1+x_2\,|\,\boldsymbol{\theta}) \\ =& n^{x_1+x_2}\binom{x_1+x_2}{x_1}\left(\frac{\theta_1}{\theta_1+\theta_2}\right)^{x_1}\left(\frac{\theta_2}{\theta_1+\theta_2}\right)^{x_2}\frac{e^{-n(\theta_1+\theta_2)}(\theta_1+\theta_2)^{x_1+x_2}}{(x_1+x_2)!}. \end{aligned} \tag{3-65}$$

Hence, the new likelihood function for $\boldsymbol{\lambda} = (\lambda_1,\lambda_2)$ generated by $\boldsymbol{x} = (x_1,x_2)$ can be written as

$$L(\boldsymbol{\lambda}\,|\,\boldsymbol{x}) = \left[\binom{x_1+x_2}{x_1}\lambda_1^{x_1}(1-\lambda_1)^{x_2}\right]\left[\frac{n^{x_1+x_2}}{(x_1+x_2)!}\lambda_2^{x_1+x_2}e^{-\lambda_2 n}\right]. \tag{3-66}$$

and the prior distribution for $\boldsymbol{\lambda}$ is given by

$$\pi(\boldsymbol{\lambda}) = \left[\frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)}\lambda_1^{a-1}(1-\lambda_1)^{c-1}\right]\left[\frac{b^a d^c}{\Gamma(a+c)}\lambda_2^{a+c-2}e^{-\lambda_2\lambda_1(b-d)-\lambda_2 d}\right]. \tag{3-67}$$

Thus, the predictive function under the null hypothesis $\tilde{H}$ can be expressed by

$$
\begin{aligned}
f_{\tilde{H}}(\boldsymbol{x}) &= \int_{\Lambda} L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
&= \frac{n^{x_1+x_2}}{x_1!x_2!} \left[\frac{(1/2)^{x_1+x_2} \displaystyle\int_0^{\infty} \lambda_2^{x_1+x_2+a+c-1} e^{-\lambda_2 \frac{(m+n+b+d)}{2}} d\lambda_2}{\displaystyle\int_0^{\infty} \lambda_2^{a+c-1} e^{-\lambda_2 \frac{(b+d)}{2}} d\lambda_2} \right],
\end{aligned} \tag{3-68}
$$

and the predictive function under the alternative hypothesis $\tilde{A}$ as

$$
\begin{aligned}
f_{\tilde{A}}(\boldsymbol{x}) &= \int_{\Lambda} L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
&= C \times \left[\int_0^1 \int_0^{\infty} \lambda_1^{x_1+a-1} (1-\lambda_1)^{x_2+c-1} \lambda_2^{x_1+x_2+a+c-1} e^{-\lambda_2(\lambda_1(m+b-n-d)+n+d)} d\lambda_2 \lambda_1 \right],
\end{aligned} \tag{3-69}
$$

where $C = \frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \frac{m^{x_1} n^{x_2}}{x_1! x_2!}$. Now, note that assuming $n = m$ and $b = d$ the predictive functions are reduced to

$$
\begin{aligned}
f_{\tilde{H}}(\boldsymbol{x}) &= \frac{n^{x_1+x_2}}{x_1!x_2!} \left[\frac{(1/2)^{x_1+x_2} \displaystyle\int_0^{\infty} \lambda_2^{x_1+x_2+a+c-1} e^{-\lambda_2 (n+b)} d\lambda_2}{\displaystyle\int_0^{\infty} \lambda_2^{a+c-1} e^{-\lambda_2 b} d\lambda_2} \right] \\
&= (1/2)^{x_1+x_2} \left[\frac{n^{x_1+x_2}}{x_1!x_2!} \right] \left[\frac{\Gamma(x_1+x_2+a+c)\, b^{a+c}}{\Gamma(a+c)(n+b)^{x_1+x_2+a+c}} \right],
\end{aligned} \tag{3-70}
$$

and

$$
\begin{aligned}
f_{\tilde{A}}(\boldsymbol{x}) &= C_5 \times \left[\int_0^1 \int_0^{\infty} \lambda_1^{x_1+a-1} (1-\lambda_1)^{x_2+c-1} \lambda_2^{x_1+x_2+a+c-1} e^{-\lambda_2 (n+b)} d\lambda_2 \lambda_1 \right] \\
&= C_5 \times \left[\int_0^1 \lambda_1^{x_1+a-1} (1-\lambda_1)^{x_2+c-1} d\lambda_1 \int_0^{\infty} \lambda_2^{x_1+x_2+a+c-1} e^{-\lambda_2 (n+b)} d\lambda_2 \right], \\
&= \frac{n^{x_1+x_2}}{x_1!x_2!} \left[\frac{\Gamma(x_1+a)\Gamma(x_2+c)}{(n+b)^{(x_1+x_2+a+c)}} \right] \left[\frac{b^{a+c}}{\Gamma(a)\Gamma(c)} \right].
\end{aligned} \tag{3-71}
$$

As in the previous cases, the likelihood function (3-66) comes factorized as (2-31) and the prior distribution of $\boldsymbol{\lambda}$ can be written as a product of distributions, in this case, of a Beta and a Gamma distributions. Then, by Theorem 4, the predictive functions can be calculated disregarding λ_2. It follows that

$$
f_{\tilde{H}}(\boldsymbol{x} | T = x_1 + x_2) = (1/2)^{x_1+x_2} \binom{x_1+x_2}{x_1}, \tag{3-72}
$$

and

$$
f_{\tilde{A}}(\boldsymbol{x} | T = x_1 + x_2) = \binom{x_1+x_2}{x_1} \left[\frac{\Gamma(x_1+a)\Gamma(x_2+c)}{\Gamma(x_1+x_2+a+c)} \right] \left[\frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} \right]. \tag{3-73}
$$

Consequently, the Bayes factor $Bf(\boldsymbol{x})$ is given by

$$
Bf(\boldsymbol{x}) = (1/2)^{x_1+x_2} \frac{\Gamma(a)\Gamma(c)}{\Gamma(a+c)} \frac{\Gamma(x_1+x_2+a+c)}{\Gamma(x_1+a)\Gamma(x_2+c)}. \tag{3-74}
$$

In Example 13 we present an application of the previous results.

Example 13. Irony and Pereira (1995) says that a methodology to detect a shift in a production process is to compare the quality index of the current rating period, λ_T, with the quality index of the previous rating period, λ_{T-1}, using (3-61). We take this idea to elaborate a numerical example. Suppose that we want to know if a process is under control. For this, two audit samples of size $n = 10$ are collected at rating periods $T-1$ and T respectively. Let X_1 represent the number of defects found in the first sample and X_2 represent the number of defects found in the second sample. Also let, X_1 and X_2 follow a Poisson distribution with parameters $n\lambda_1$ and $n\lambda_2$, respectively. For simplicity, we will assume the hyperparameters in (3-58) as $a = b = c = d = 1$. Hence, the predictive functions under each hypothesis are given by:

$$f_H(\boldsymbol{x}) = (1/2)^{x_1+x_2}\left[(x_1+x_2+1)\binom{x_1+x_2}{x_1}\right]\left[\left(\frac{n}{n+1}\right)^{x_1+x_2}\left(\frac{1}{n+1}\right)^2\right], \tag{3-75}$$

and

$$f_A(\boldsymbol{x}) = \left(\frac{n}{n+1}\right)^{x_1+x_2}\left(\frac{1}{n+1}\right)^2. \tag{3-76}$$

Consequently, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed by

$$Bf(\boldsymbol{x}) = (1/2)^{x_1+x_2}(x_1+x_2+1)\binom{x_1+x_2}{x_1}. \tag{3-77}$$

Now, suppose that $x_0 = (2, 9)$ was found. In this case $Bf(x_0) = 0.322$ and the conditional adaptive significance level and the conditional Bayesian P_T*-value* are given, respectively by

$$\alpha_T^* = \frac{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leqslant b/a,\\ T(x)=T(\boldsymbol{x}_0)}} f_H(\boldsymbol{x})}{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ T(x)=T(x_0)}} f_H(\boldsymbol{x})} = 0.227 \qquad \text{and} \qquad P_T\textit{-value}\,(x_0) = \frac{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq Bf(\boldsymbol{x_0}),\\ T(x)=T(x_0)}} f_H(\boldsymbol{x})}{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:\\ T(x)=T(x_0)}} f_H(\boldsymbol{x})} = 0.065.$$

Since P_T*-value*$(x_0) < \alpha_T^*$, the decision is to reject the null hypothesis in (3-64). Note that $(X_1, X_2) \in \mathbb{N}^2$. Therefore, computing the *P-value* and α_{δ^*} from (2-17) and (2-15) respectively, is an exhaustive work since it is necessary to find the Bayes factor for the whole cartesian product $\mathbb{N}^2$. In this case, it make sense to us to use the conditional P_T*-value* and the conditional α_T^* since the comparison between them is equivalent to the comparison between the original *P-value* and α_{δ^*} with the advantage of performing finitely many calculations.

3.2.2 Hardy–Weinberg Equilibrium

The "Hardy-Weinberg law" was initially formulated by Hardy (1908) and Weinberg (1908) and later, the term "Hardy-Weinberg equilibrium" (HWE) appeared in Stern (1943). As mentioned by Graffelman (2019), this law plays an important role in the context of genetic association studies since disequilibrium may be the result of genotyping error, most typically the confusion of heterozygotes and homozygotes. Test for HWE is a useful tool in the analysis of DNA evidence, used in human identification and paternity studies (Council *et al.*, 1996). The conclusions reached by analyzing such evidence depend on the probabilistic evaluation of this evidence, where, the evaluation is simplified if HWE holds (Montoya *et al.*, 2001). On the other hand, disequilibrium among cases in a case-control study may be indicative of disease association. Thus, tests for HWE may also provide clues in marker-disease association studies (Graffelman, 2019; Montoya *et al.*, 2001; Shoemaker *et al.*, 1998).

In short, the HWE may be described as follows: Consider a single autosomal biallelic locus, comprising alleles a and b with respective frequencies p and $1-p$. This locus is said to be in Hardy-Weinberg equilibrium if the relative genotype frequencies of aa, ab and bb, (f_{aa}, f_{ab}, f_{bb}), are given by p^2, $2p(1-p)$ and $(1-p)^2$, respectively, for some $0 \leq p \leq 1$. Suppose that the system is codominant where all alleles can be recognized by genotypes, so that direct estimation of the frequencies of alleles is possible (Thomson *et al.*, 2009). Then, in a sample of size n, the frequencies of members in each class x_1, x_2, and x_3, satisfy the condition $\sum_{i=1}^{3} x_i = n$ (Graffelman, 2019; Montoya *et al.*, 2001). Thus, the test for the HWE verifies if a population follows the genotypes proportions given by the HWE proportions. Then, considering $\boldsymbol{\theta} = \{(\theta_1, \theta_2) \in [0,1]^2 : \theta_1 + \theta_2 \leq 1, \theta_3 = 1 - \theta_1 - \theta_2\}$ the equilibrium hypotheses can be written as

$$\begin{aligned} H &: \boldsymbol{\theta} \in \Theta_0 \\ A &: \boldsymbol{\theta} \in \Theta_0^c, \end{aligned} \tag{3-78}$$

where $\Theta_0 = \{(p^2, (1-p)^2) : 0 \leq p \leq 1\}$, see figure (**3-1**).

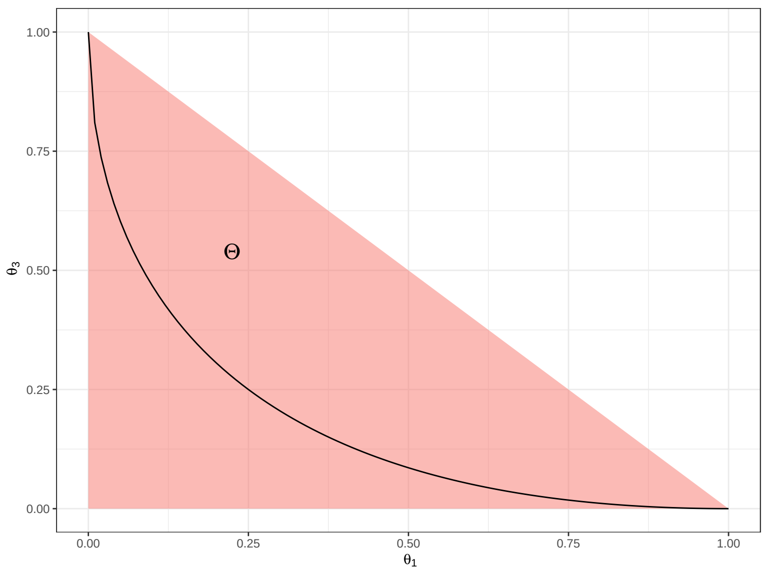

Figure 3-1: The HWE curve. The complete parametric space is shown by the shaded area and the Hardy-Weinberg equilibrium is represented by the curve.

For testing (3-78), we consider that the sample size n is known and sample elements of genotypes frequencies are obtained independently in such a way that the vector $\boldsymbol{X} = (X_1, X_2, X_3)$ follows a Multinomial distribution with parameters n and $\boldsymbol{\theta}$. Thus, the Likelihood function for $\boldsymbol{\theta}$ generated by $\boldsymbol{x} = (x_1, x_2, x_3)$ is defined by

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = \frac{n!}{\prod_{i=1}^{3} x_i!} \prod_{i=1}^{3} \theta_i^{x_i} \tag{3-79}$$

In addition, we assume that the parameter vector $\boldsymbol{\theta}$ follows a Dirichlet distribution, with parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ for $\alpha_i > 0$, that is

$$\pi(\boldsymbol{\theta}) = \frac{\Gamma(\sum_{i=1}^{3} \alpha_i)}{\prod_{i=1}^{3} \Gamma(\alpha_i)} \prod_{i=1}^{3} \theta_i^{\alpha_i - 1} \tag{3-80}$$

Then, the predictive function under the null hypothesis H can be expressed by

$$\begin{aligned} f_H(\boldsymbol{x}) &= \int_{\Theta} L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_H(\boldsymbol{\theta}) \\ &= 2^{x_2} \frac{n!}{\prod_{i=1}^{3} x_i!} \frac{\int_0^1 \left(\sqrt{1 - 3p(1-p)}\right) p^{2A_1 + A_2 - 2} (1-p)^{2A_3 + A_2 - 2} dp}{\int_0^1 \left(\sqrt{1 - 3p(1-p)}\right) p^{2\alpha_1 + \alpha_2 - 2} (1-p)^{2\alpha_3 + \alpha_2 - 2} dp}, \end{aligned} \tag{3-81}$$

and under the alternative hypothesis A by

$$\begin{aligned} f_A(\boldsymbol{x}) &= \int_{\Theta} L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\ &= \left[\frac{\Gamma(\sum_{i=1}^{3} \alpha_i)}{\prod_{i=1}^{3} \Gamma(\alpha_i)} \frac{n!}{\prod_{i=1}^{3} x_i!}\right] \int_{\Theta_0^c} \theta_1^{A_1 - 1} \theta_2^{A_2 - 1} \theta_3^{A_3 - 1} d\boldsymbol{\theta} \\ &= \left[\frac{\Gamma(\sum_{i=1}^{3} \alpha_i)}{\prod_{i=1}^{3} \Gamma(\alpha_i)} \frac{n!}{\prod_{i=1}^{3} x_i!}\right] \frac{\prod_{i=1}^{3} \Gamma(A_i)}{\Gamma(\sum_{i=1}^{3} A_i)}, \end{aligned} \tag{3-82}$$

where $A_i = x_i + \alpha_i$ for $i = 1, 2, 3$. Note that the exact calculation of (3-81) is not feasible. Thereby, as in the previous cases, we apply a suitable reparameterization that makes the task of obtaining (3-81) easier. For this, we consider the parameter (θ_1, θ_3) in place of (θ_1, θ_2) (and, in this case the equilibrium hypothesis is $\{(p^2, (1-p)^2) : p \in [0, 1]\}$ sketched in Figure **3-2**) and define the new parameters

$$\lambda_1 = \frac{\sqrt{\theta_1}}{\sqrt{\theta_1} + \sqrt{\theta_3}}; \qquad \lambda_2 = \sqrt{\theta_1} + \sqrt{\theta_3}. \tag{3-83}$$

The new parametric space Λ is defined by $\Lambda = \{(\lambda_1, \lambda_2) \in \mathbb{R}_+^2 : \lambda_1 \in [0, 1] \,\text{and}\, 0 \leq \lambda_2 \leq \{\lambda_1^2 + (1 - \lambda_1)^2\}^{-1/2}\}$ and the HWE hypothesis corresponds to the line segment $\lambda_2 = 1$ (that is, $[0, 1] \times \{1\}$), on the right side of Figure **3-2**. This choice will make the computation of predictive function easier.

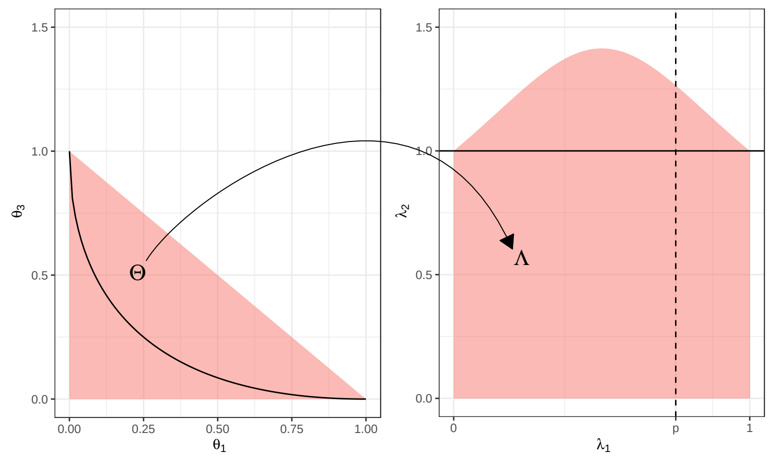

Figure 3-2: HWE - variable re-parametrization .The complete parametric space is shown by the shaded area and the HWE is represented by the black line.

Thus, the hypotheses (3-78) can be rewritten as:

$$\begin{aligned}\tilde{H} &: \boldsymbol{\lambda} \in \Lambda_0 \\ \tilde{A} &: \boldsymbol{\lambda} \in \Lambda_0^c,\end{aligned} \tag{3-84}$$

with $\Lambda_0 = B \times \{1\}$ and $B = [0, 1]$. Then, the predictive function under the null hypothesis $\tilde{H}$ is given by

$$\begin{aligned}f_{\tilde{H}}(\boldsymbol{x}) &= \int_{\Lambda} L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ &= 2^{x_2} \frac{n!}{\prod_{i=1}^{3} x_i!} \left[\frac{\Gamma(2A_1 + A_2 - 1)\Gamma(2A_3 + A_2 - 1)\Gamma(2\sum_{i=1}^{3} \alpha_i - 2)}{\Gamma(2\alpha_1 + \alpha_2 - 1)\Gamma(2\alpha_3 + \alpha_2 - 1)\Gamma(2\sum_{i=1}^{3} A_i - 2)} \right].\end{aligned} \tag{3-85}$$

From Theorem 1, it follows that

$$\begin{aligned}f_{\tilde{A}}(\boldsymbol{x}) &= \int_{\Lambda} \bar{L}(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{A}} \\ &= \int_{\Theta} L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) d\mathbb{P}_{A} \\ &= f_A(\boldsymbol{x}),\end{aligned} \tag{3-86}$$

that is the final result of (3-82). Therefore, the Bayes factor $Bf(\boldsymbol{x})$ can expressed as

$$Bf(\boldsymbol{x}) = 2^{x_2} \left[\frac{\Gamma(2A_1 + A_2 - 1)\Gamma(2A_3 + A_2 - 1)\Gamma(2\sum_{i=1}^{3} \alpha_i - 2)}{\Gamma(2\alpha_1 + \alpha_2 - 1)\Gamma(2\alpha_3 + \alpha_2 - 1)\Gamma(2\sum_{i=1}^{3} A_i - 2)} \frac{\prod_{i=1}^{3} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{3} \alpha_i)} \frac{\Gamma(\sum_{i=1}^{3} A_i)}{\prod_{i=1}^{3} \Gamma(A_i)} \right], \tag{3-87}$$

where $A_i = x_i + \alpha_i$. If we assume that the prior distribution function for $\boldsymbol{\theta}$ is uniform, that is, considering $\alpha_i = 1$, for $i = 1, 2, 3.$, $Bf(\boldsymbol{x})$ is given by:

$$Bf(\boldsymbol{x}) = 2^{x_2 - 1} \frac{(n+2)!}{\prod_{i=1}^{3} x_i!} \left[\frac{3!(2x_1 + x_2 + 1)!(2x_3 + x_2 + 1)!}{\Gamma(3)(2n+3)!} \right]. \tag{3-88}$$

In order to analyze the performance of the P-P test, we compare its results with the frequently used statistical tests for HWE (Chi-Square, Likelihood Ratio, Exact test and Approximated Bayes Factor). Table 3-9 shows the Rejection/Not Rejection proportions for all allelic combinations of 10 genotype samples. Under this scenario, the Bayes factor $Bf(\boldsymbol{x})$ and the Aprox Bayes factor presented the minor proportion of not rejection.

Table 3-9 Rejection/Not Rejection proportions of different tests against Bayes factor $Bf(\boldsymbol{x})$ for HWE hypothesis.

		Bayes factor	
		Not Reject	Reject
Exact test	Not Reject	0.45	0.30
	Reject	0	0.25
LR test	Not Reject	0.45	0.18
	Reject	0	0.37
Chi-Square test	Not Reject	0.42	0.36
	Reject	0.03	0.19
Aprox Bayes Factor	Not Reject	0.27	0.03
	Reject	0.18	0.52

In Table 3-9 *Approximated Bayes Factor* refers to the methodology presented by Montoya *et al.* (2001) and Exact test refers to the methodology implemented by Wigginton *et al.* (2005), the latter methodology is implemented in the *HardyWeinberg* package of R Core Team (2017) and can be invoked with the function *HWExact*(). The other two are the well-known tests, *Likelihood Ratio* and *Chi-Square* from Neyman and Pearson (1957) and Wilks (1935) respectively.

Next, we apply all tests in simulated data by the *HWData()* function from Graffelman (2015) R package. Since this function returns data being in HWE, the tests should not reject the hypothesis of HWE.

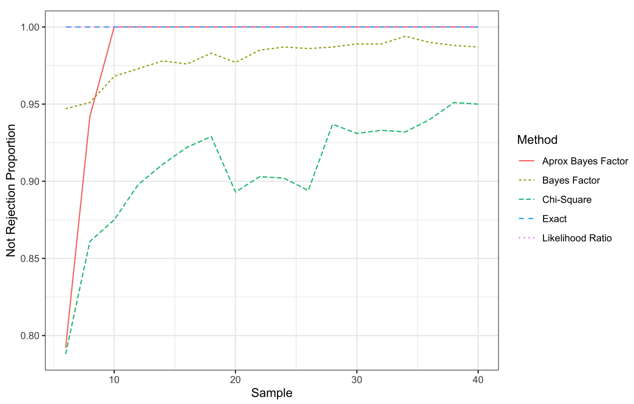

Figure 3-3: Rejection proportion for simulated data (under HWE) from the HardyWeinberg R package.

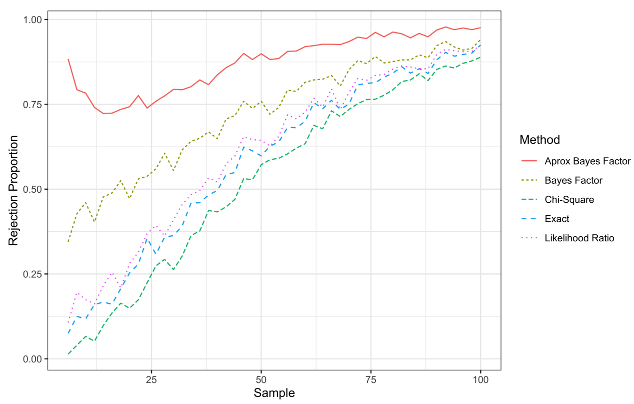

Figure 3-4: Rejection proportion for simulated data (not under HWE) by using rmultinom R function.

Figure (3-3) shows the performance of all tests for different samples sizes. Note that, the Chi-Square test presented the worst performance for all sample sizes, and in small samples, the approximated Bayes Factor test presented slow performance too.

3.2.3 CENIPA data application

To illustrate the results derived in Subsections (3.1.2) and (3.1.3) we shall use a data set from aeronautical occurrences. This data set is managed by the Brazilian Aeronautical Accident Investigation and Prevention Center (CENIPA, in Portuguese). The data set is of public access and contains aeronautical events notified to CENIPA in the last 10 years which occurred into Brazilian territory. The available information from each event includes data on aircraft involved, fatalities, location, date and time of events, and typical taxonomy information from accident investigations (AIG) (Cenipa, 2019).

Independence

Table **3-10** presents the number of air accidents, incidents and severe incidents by kind of engines in February 2019. We want to know if there exists an association between the kind of event and the kind of engines of the airplanes.

Table 3-10 Number of accidents and Incidents by kind of engines.

Aircraft Engines	Accident	Incident	Serious Incident
Jet	0	9	0
Piston	8	7	6
TurboJet	1	0	1
Turboprop	1	3	0

The adaptive type I error probability and the Bayesian *P-value* are given

$$\alpha_{\delta^*} = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}:\\ Bf(\boldsymbol{x})\leq 1}} f_H(\boldsymbol{x}) = 0.0263 \qquad \text{and} \qquad P\text{-}value(x_0) = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}:\\ Bf(\boldsymbol{x})\leq BF(x_0)}} f_H(\boldsymbol{x}) = 0.0003.$$

Since P-value$(x_0) < \alpha_{\delta^*}$, the decision, in this case, is to reject the null hypothesis, i.e, there exists some kind of association between the kind of accident and the engines. Table **3-11** presents the results of other test procedures for the same data set. Notice that all tests rejected the null hypothesis, except for the LRT test that appears not to work when the table has cell counts equal to zero. Even though the chi-square test rejects the null hypothesis at the 5% level, it does not reject it at 1%. In addition, note that the approximations should be poor because some cells have small counts.

Table 3-11 Significance indices for test independence between accidents and kind of engines.

Test	Measure	Value	Significance Level
Bayes factor	P-value	0.0003	0.0263
FBST	e-value	0.000	0.05
LRT	p-value	0.9969	0.05
Chi-square	P-value	0.0219	0.05

Symmetry

Table **3-12** presents the number of planes that suffered accidents twice. We are interested in knowing if the kind of accident that occurred for the first time is the same that occurred in the second.

Table 3-12 Number of accidents involving aircraft failed twice.

First Time	Second Time		
	Accident	Incident	Serious Incident
Accident	0	1	0
Incident	1	22	1
Serious Incident	1	7	0

The adaptive significance level and the Bayesian *P-value* are given

$$\alpha_{\delta^*} = \sum_{\substack{\boldsymbol{x}\in\mathcal{X}: \\ Bf(\boldsymbol{x})\leq 1}} f_H(\boldsymbol{x}) = 0.1386 \qquad \text{and} \qquad P\text{-}value(x_0) = \sum_{\substack{x\in\mathcal{X}: \\ BF(\boldsymbol{x})\leq Bf(x_0)}} f_H(\boldsymbol{x}) = 0.0521.$$

Since $P\text{-}value(x_0) < \alpha_{\delta^*}$, the decision in this case is to reject the null hypothesis, i.e, there are differences between orders of two types of accidents. Table **3-13** presents the results of other test procedures for the same data set. Notice different decisions are made if the Bayes factor or McNemar test is used.

Table 3-13 Significance indices for test diagonal symmetry between two-time accidents.

Test	Measure	Value	Significance Level
Bayes factor	P-value	0.0521	0.1386
FBST	e-value	0.000	0.05
McNemar's chi-squared	p-value	0.1386	0.05

Chapter 4

Discussion

Statistical hypothesis testing is present in the day-to-day of scientists in different areas of knowledge. However, with the computational advances and data revolution, different misuses of this important tool have come to light. As discussed in Chapter 1, over the years, problems with tests of significance and fixed significance levels generated a growing dissatisfaction among statisticians and the modern scientific community. On the other hand, credibility in Bayesian framework for hypotheses testing increases in this community for its advantages in application and interpretation. Under the Bayesian approach, we want to emphasize that the use of the Bayes factor along with adaptive significance levels is a practical and powerful tool for hypothesis testing since this methodology can be used in sharp, one-sided and two-sided hypotheses while it allows us to avoid all the problems inherent of significance tests and fixed significance levels. Nonetheless, there exist hypotheses for which the computation of Bayes factor is complex. the computation of In several cases the integrals involved in the Bayes factor are not easy to calculate: for example, the independence hypothesis in a two-way contingency table and the HWE hypothesis. We showed that besides being in a sense invariant under reparameterization, the Bayes factor also obeys the non-informative nuisance parameter principle (NNPP). Based on this, we use reparameterizations of standard models for count data (analysis of contingency tables and comparison of Poisson means) to perform simpler (reduced) versions of the P-P test: it usually demands the determination of complex integrals and after reparameterization, some parameters are considered non-informative and consequently removed from the calculation of the Bayes factor, thus, its calculation is simplified.

Generally, in hypothesis tests of homogeneity, independence and symmetry on two-way contingency tables traditional methods such as Chi-squared, LRT and McNemar statistical tests are used. When these procedures deal with small samples sizes or sparsity (cells counts close to zero), their performances are very poor, as we can see in the LRT test (Table **3-11**). Although some modifications have been made (see, for example, Sulewski (2019) and their references), there is not a complete solution yet. From a Bayesian perspective, there is a wide literature on methods for analyzing data with contingency tables

(Agresti and David (2005); Lindley and Smith (1972)). However, we focused on the works presented by Irony and Pereira (1995); Pereira and Pereira (2005); Pereira *et al.* (2017); Pericchi and Pereira (2016) and we used them as a baseline for applying our main results concerning mainly the NNPP for hypothesis testing and a kind of invariance. Under this approach, we have proposed a solution to the problem of testing independence in contingency tables that is easier to implement by means of the P-P test: it reduces the original problem to a simpler test of homogeneity , although the problem we deal with is not exactly the same as the original one because of the reparameterization taken into account. In a similar way, the diagonal symmetry hypothesis was transformed into a simple hypothesis. In addition, we derived an exact test for the HWE. Furthermore, we used our proposal to carry out a Poisson means comparison, where we found that, when both the sample sizes and rate parameters (in the prior distribution) are equal, the hypothesis test for means comparison turns into a test of a simple hypothesis.

The use of the NNP principle along with the invariance property on Bayes factor is an unexplored methodology that could be used in many other problems that involve complex hypotheses. However, note that both adaptive significance levels and *P-value* need the evaluation of the Bayes factor for all sample points and this implies high computational costs in large sample sizes. For example, tables with a sample size greater than 100 could demand much time to calculate the adaptive significance level and the *P-value*. In addition, when the P-P test has been applied some parameters are considered as non-informative and therefore removed from the Bayes factor calculation, simplifying the computation of relevant quantities. The resulting expression in the predictive function under each hypothesis H (A) is not equivalent to original predictive function with all parameters, consequently, is necessary to use the conditional α^*_T and conditional P_T*-value* for the analysis.

Although we believe that the properties of the P-P test examined in this work may be useful and contribute to hypothesis testing involving count data, we are aware that there are still many things to do. For example, to develop this methodology for multi-way contingency tables and other statistical models, to explore other statistical areas where the NNP principle could be applied to transform the original hypotheses into simpler ones, to explore another logical properties of the NNPP and its extensions for other inferential procedures and its behavior in large samples.

Appendix A

Proofs

A.1 Invariance

Let $(\Theta, \sigma(\Theta))$ and $(\Lambda, \sigma(\Lambda))$ be measurable spaces and $h : \Theta \to \Lambda$ a measurable and invertible function. Then, consider the hypotheses

$$\begin{array}{lllll} H: & \theta \in \Theta_0 & \qquad & \tilde{H}: & \lambda \in \Lambda_0 \\ A: & \theta \notin \Theta_0 & \qquad & \tilde{A}: & \lambda \notin \Lambda_0, \end{array} \tag{A-1}$$

where, $\Theta_0 \in \sigma(\Theta)$ and $\Lambda_0 = \{h(\theta) : \theta \in \Theta\}$. Let $\mathbb{P}_H$ and $\mathbb{P}_A$ be probability measures conditioned on the hypotheses H and A for θ respectively. In addition, define $\mathbb{P}_{\tilde{H}} : \sigma(\Lambda) \to \mathbb{R}_+$ as $\mathbb{P}_{\tilde{H}}(B) = \mathbb{P}_H\left(h^{-1}(B)\right)$ and $\mathbb{P}_{\tilde{A}} : \sigma(\Lambda) \to \mathbb{R}_+$ as $\mathbb{P}_{\tilde{A}}(B) = \mathbb{P}_A\left(h^{-1}(B)\right)$. Let $L : \Theta \to \mathbb{R}_+$ be the likelihood function generated by x for θ and $\tilde{L} : \Lambda \to \mathbb{R}_+$ the likelihood function over the alternative parameterization given by $\tilde{L}(\lambda|\, x) = L(h^{-1}(\lambda)|\, x)$. Then, we define $\varphi^\Theta : X \to \{0, 1\}$ and $\varphi^\Lambda : X \to \{0, 1\}$ as the P-P tests for the hypotheses (A-1) under the parameterization Θ and Λ respectively by

$$\varphi^\Theta = \begin{cases} 1 & \text{if} \quad Bf^\Theta(\boldsymbol{x}) \leq c \\ 0 & \text{if} \quad Bf^\Theta(\boldsymbol{x}) > c \end{cases} \qquad \text{and} \qquad \varphi^\Lambda = \begin{cases} 1 & \text{if} \quad Bf^\Lambda(\boldsymbol{x}) \leq c \\ 0 & \text{if} \quad Bf^\Lambda(\boldsymbol{x}) > c \end{cases} \tag{A-2}$$

where,

$$Bf^\Theta(\boldsymbol{x}) = \frac{\displaystyle\int_\Theta L(\theta|\, x) d\mathbb{P}_H(\theta)}{\displaystyle\int_\Theta L(\theta|\, x) d\mathbb{P}_A(\theta)} \qquad \text{and} \qquad Bf^\Lambda(\boldsymbol{x}) = \frac{\displaystyle\int_\Lambda \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{H}}(\theta)}{\displaystyle\int_\Lambda \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{A}}(\theta)}. \tag{A-3}$$

We shall prove that for all $x \in \mathfrak{X}$, $\varphi^\Theta(x) = 1 \Leftrightarrow \varphi^\Lambda(x) = 1$. Then, we have that

$$\varphi^\Theta(x) = 1 \Leftrightarrow \frac{\displaystyle\int_\Theta L(\theta|\, x) d\mathbb{P}_H(\theta)}{\displaystyle\int_\Theta L(\theta|\, x) d\mathbb{P}_A(\theta)} \leq c. \tag{A-4}$$

However, note that the predictive function under the null hypothesis in the alternative parameterization can be written as

$$\begin{aligned}
\int_{\Lambda} \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{H}}(\lambda) &= \int_{\Theta} \left(\tilde{L} \circ h\right)(\theta|\, x) d\mathbb{P}_{H}(\theta) \\
&= \int_{\Theta} \left(\left(L \circ h^{-1}\right) \circ h\right)(\theta|\, x) d\mathbb{P}_{H}(\theta) \\
&= \int_{\Theta} \left(L \circ \left(h^{-1} \circ h\right)\right)(\theta|\, x) d\mathbb{P}_{H}(\theta) \\
&= \int_{\Theta} L(\theta|\, x) d\mathbb{P}_{H}(\theta),
\end{aligned} \tag{A-5}$$

and applying the analogous procedure (note that this result follows directly from the properties of the integral of a transformation), we have that $\int_{\Lambda} \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{A}}(\lambda) = \int_{\Theta} L(\theta|\, x) d\mathbb{P}_{A}(\theta)$, hence

$$\begin{aligned}
\varphi^{\Theta}(x) = 1 &\Leftrightarrow \frac{\int_{\Theta} L(\theta|\, x) d\mathbb{P}_{H}(\theta)}{\int_{\Theta} L(\theta|\, x) d\mathbb{P}_{A}(\theta)} \leq c \\
&\Leftrightarrow \frac{\int_{\Lambda} \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{H}}(\lambda)}{\int_{\Lambda} \tilde{L}(\lambda|\, x) d\mathbb{P}_{\tilde{A}}(\lambda)} \leq c \\
&\Leftrightarrow \varphi^{\Lambda}(x) = 1 \quad \square.
\end{aligned} \tag{A-6}$$

A.2 NNP principle.

Proof Theorem 2: Suppose there exists a statistic $T : \mathfrak{X} \to \mathcal{T}$ such that it is both p-Sufficient for θ_1 and s-Ancillary for θ_2, then

$$\begin{aligned}
P(X = x|\, \theta) =& P(X = x, T(X) = T(x) \,|\, \theta) \\
=& P(X = x, T(X) = T(x) \,|\, \theta_1, \theta_2) \\
=& g(T(x)|\, \theta_1,\, \theta 2) f(x|\, T(X) = T(x),\, \theta_1,\, \theta 2) \\
=& g(T(x)|\, \theta_1) f(x|\, T(X) = T(x),\, \theta 2) \\
=& L^1(\theta_1|\, x) L^2(\theta_2|\, x).
\end{aligned} \tag{A-7}$$

The proof when T is both p-Sufficient for θ_2 and s-Ancillary for θ_1 is analogous. □

Proof Theorem 3: Suppose that $x \in \mathfrak{X}$ is such that the posterior distribution of θ can be factorized as $\pi(\theta|\, x) = \pi_1(\theta_1|\, x)\pi_2(\theta_2|\, x)$. Then,

$$\begin{aligned}
\frac{L(\theta|x)\pi(\theta)}{\int_{\Theta} L(\theta|x)\pi(\theta) d\theta} &= \pi(\theta|x) \\
&= \pi_1(\theta_1|x)\pi_2(\theta_2|x) \\
&= \frac{\int_{\Theta_2} L(\theta|x)\pi(\theta) d\theta_2}{\int_{\Theta} L(\theta|x)\pi(\theta) d\theta} \frac{\int_{\Theta_1} L(\theta|x)\pi(\theta) d\theta_1}{\int_{\Theta} L(\theta|x)\pi(\theta) d\theta}.
\end{aligned} \tag{A-8}$$

Due to the fact that $\theta_1 \perp\!\!\!\perp \theta_2$, equation (A-8) can be written as

$$\begin{aligned} L(\theta|x) =& \int_{\Theta_2} L(\theta|x)\pi_2(\theta_2)d\theta_2 \frac{\displaystyle\int_{\Theta_1} L(\theta|x)\pi_1(\theta_1)d\theta_1}{\displaystyle\int_{\Theta} L(\theta|x)\pi(\theta)d\theta} \\ =& L^1(\theta_1|x)L^2(\theta_2|x), \end{aligned} \tag{A-9}$$

where,

$$L^1(\theta_1|x) = \int_{\Theta_2} L(\theta|x)\pi_2(\theta_2)d\theta_2 \quad \text{and} \quad L^2(\theta_2|x) = \frac{\displaystyle\int_{\Theta_1} L(\theta|x)\pi(\theta_1)d\theta_1}{\displaystyle\int_{\Theta} L(\theta|x)\pi(\theta)d\theta} \quad \Box. \tag{A-10}$$

This implies that if the prior distribution for θ_1 and θ_2 are independent and the posterior distribution for θ can be factorized, then the likelihood can also be factorized in two parts, one that depends just on θ_1 and other that depend only on θ_2. Now, suppose that the likelihood can be factorized as $L(\theta|\, x) = L^1(\theta_1|\, x)L^2(\theta_2|\, x)$. Then

$$L(\theta|\, x) = L^1(\theta_1|\, x)L^2(\theta_2|\, x) \Longrightarrow \tag{A-11}$$

$$L(\theta|\, x)\pi(\theta) = L^1(\theta_1|\, x)L^2(\theta_2|\, x)\pi(\theta) \Longrightarrow \tag{A-12}$$

$$\begin{aligned} \xi(\theta|x) = \frac{L(\theta|x)\pi(\theta)}{\displaystyle\int_{\Theta} L(\theta|x)\pi(\theta)d\theta} =& \frac{L^1(\theta_1|x)L^2(\theta_2|x)\pi(\theta)}{\displaystyle\int_{\Theta_1\times\Theta_2} L^1(\theta_1|x)L^2(\theta_2|x)\pi(\theta)d\theta} \\ =& \frac{L^1(\theta_1|x)L^2(\theta_2|x)\pi^1(\theta_1)\pi^2(\theta_2)}{\displaystyle\int_{\Theta_1\times\Theta_2} L^1(\theta_1|x)L^2(\theta_2|x)\pi^1(\theta_1)\pi^2(\theta_2)d\theta}. \end{aligned} \tag{A-13}$$

Under these conditions, we can use Tonelli's theorem in (A-9). Then

$$\begin{aligned} \xi(\theta|x) =& \frac{L^1(\theta_1|x)\pi^1(\theta_1)}{\displaystyle\int_{\Theta_1} L^1(\theta_1|x)\pi^1(\theta_1)d\theta_1} \frac{L^2(\theta_2|x)\pi^2(\theta_2)}{\displaystyle\int_{\Theta_2} L^2(\theta_2|x)\pi^2(\theta)d\theta_2} \\ =& \xi(\theta_1|x)\xi(\theta_2|x), \end{aligned} \tag{A-14}$$

where the last equality follows from direct calculation of the posterior marginal distributions from (by integration) the posterior joint density. The equation (A-14) implies that if θ_1 and θ_2 are independent and the likelihood function can be factorized as $L(\theta|\, x) = L^1(\theta_1|\, x)L^2(\theta_2|\, x)$, then the posterior distribution for θ will be factorized too $\Box$.

Proof Theorem 4: We want to show that for $x \in \mathfrak{X}$ such that (2-31) holds, $\varphi^*(x) = 1 \Leftrightarrow \bar{\varphi}^*(x, \theta_2) = 1$. We first verify that if θ_2 is NNP. Then, we know that

$$\bar{\varphi}^*(x,\theta_2)=1\Leftrightarrow\frac{\bar{f}_{\bar{H}}(x,\theta_2)}{\bar{f}_{\bar{A}}(x,\theta_2)}<\frac{b}{a},\tag{A-15}$$

where $\bar{f}_{H\,(A)}(x)$ is the predictive function under each hypothesis. Thus, consider the likelihood function generated by (x_0,θ_2) with $x_0\in\mathfrak{X}$ and $\theta_2\in\Theta_2$ as

$$\begin{aligned}\bar{L}(\theta_1\,|\,x_0,\theta_2)=&P\left(X=x_0\,|\,\theta_1,\theta_2\right)f\left(\theta_2\,|\,\theta_1\right)=L\left(\theta_1,\theta_2\,|\,x\right)f_2\left(\theta_2\right)\\=&L^1\left(\theta_1\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right).\end{aligned}\tag{A-16}$$

Then, the predictive function under null hypothesis $\theta_1=\theta_0$ can be calculated as

$$\begin{aligned}\bar{f}_{\bar{H}}\left(x_0,\theta_2\right)=&\int_{\Theta_1}\bar{L}\left(\theta_1\,|\,x_0,\theta_2\right)d\mathbb{P}_{\bar{H}_0}\left(\theta_1\right)\\=&\int_{\Theta_1}L^1\left(\theta_1\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)d\mathbb{P}_{\bar{H}_0}\left(\theta_1\right),\end{aligned}\tag{A-17}$$

where $P_{\bar{H}}$ is the conditional distribution of $\theta_1|\theta_1=\theta_0$. Due to the fact that the distribution of $\theta_1\,|\,\theta_1=\theta_0$ is degenerate at θ_0, we have that

$$\bar{f}_{\bar{H}}\left(x_0,\theta_2\right)=L^1\left(\theta_0\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right).\tag{A-18}$$

In addition, the predictive function under the alternative hypothesis is given by

$$\begin{aligned}\bar{f}_{\bar{A}}\left(x_0,\theta_2\right)=&\int_{\Theta_1}L^1\left(\theta_0\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)dP_{\bar{H}_1}\left(\theta_1\right)\\=&\int_{\Theta_1}L^1\left(\theta_0\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)f_1\left(\theta_1\right)d\theta_1\\=&L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)\int_{\Theta_1}L^1\left(\theta_1\,|\,x_0\right)f_1\left(\theta_1\right)d\theta_1.\end{aligned}\tag{A-19}$$

Thus, the Bayes factor can be expressed by

$$\begin{aligned}\frac{\bar{f}_{\bar{H}}\left(x_0,\theta_2\right)}{\bar{f}_{\bar{A}}\left(x_0,\theta_2\right)}=&\frac{L^1\left(\theta_0\,|\,x_0\right)L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)}{L^2\left(\theta_2\,|\,x_0\right)f_2\left(\theta_2\right)\int_{\Theta_1}L^1\left(\theta_1\,|\,x_0\right)f_1\left(\theta_1\right)d\theta_1}\\=&\frac{L^1\left(\theta_0\,|\,x_0\right)}{\int_{\Theta_1}L^1\left(\theta_1\,|\,x_0\right)f_1\left(\theta_1\right)d\theta_1}.\end{aligned}\tag{A-20}$$

Note that Equation (A-20) is independent of θ_2. As a result, the test in (A-15) does not depend on θ_2, consequently, θ_2 is NPP for the test $\bar{\varphi}^*$ and the hypotheses $\bar{H}$ versus $\bar{A}$. Now, we shall determine the test φ^* for H versus A. Thus, the likelihood function under null hypothesis is defined

$$f_H\left(x_0\right)=\int_{\Theta_1\times\Theta_2}L(\theta_1,\theta_2\,|\,x_0)d\mathbb{P}_H\left(\theta_1,\theta_2\right).\tag{A-21}$$

Where it is possible to verify that for fixed $\theta_0 \in \Theta_1$ the conditional distribution of θ given $\theta_1 = \theta_0$ satisfies that

- $\theta_1 \mid \theta_1 = \theta_0$ is degenerated at θ_0
- $\theta_2 \mid \theta_1 = \theta_0$ is absolutely continuous, with density function f_2 (when θ_1 and θ_2 are not independent the density function for $\theta_2 \mid \theta_1 = \theta_0$ is $\frac{f(\theta_0,\theta_2)}{f_1(\theta_0)}$)

Then, the predictive function for the null hypothesis is given by

$$\begin{aligned} f_H(x_0) &= \int_{\Theta_1, \times \Theta_2} L(\theta_1, \theta_2 \mid x_0)\, d\mathbb{P}_{H_0}(\theta_1, \theta_2) \\ &= \int_{\Theta_2} L^1(\theta_0 \mid x_0)\, L^2(\theta_2 \mid x_0)\, f_2(\theta_2)\, d\theta_2 \\ &= L^1(\theta_0 \mid x_0) \int_{\Theta_2} L^2(\theta_2 \mid x_0)\, f_2(\theta_2)\, d\theta_2. \end{aligned} \tag{A-22}$$

And, for the alternative hypothesis we have that

$$\begin{aligned} f_A(x_0) &= \int_{\Theta_1 \times \Theta_2} L^1(\theta_1 \mid x_0)\, L^2(\theta_2 \mid x_0)\, d\mathbb{P}_A(\theta_1, \theta_2) \\ &= \int_{\Theta_1 \times \Theta_2} L^1(\theta_1 \mid x_0)\, L^2(\theta_2 \mid x_0)\, d\mathbb{P}(\theta_1, \theta_2) \\ &= \int_{\Theta_1} \int_{\Theta_2} L^1(\theta_1 \mid x_0)\, L^2(\theta_2 \mid x_0)\, f_1(\theta_1)\, f_2(\theta_2)\, d\theta_1 d\theta_2 \\ &= \int_{\Theta_1} L^1(\theta_1 \mid x_0)\, f_1(\theta_1)\, d\theta_1 \int_{\Theta_2} L^2(\theta_2 \mid x_0)\, f_2(\theta_2)\, d\theta_2. \end{aligned} \tag{A-23}$$

Finally,

$$\begin{aligned} \frac{f_H(x_0)}{f_A(x_0)} &= \frac{L^1(\theta_0 \mid x_0) \displaystyle\int_{\Theta_2} L^2(\theta_2 \mid x_0)\, f_2(\theta_2)\, d\theta_2}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \mid x)\, f_1(\theta_1)\, d\theta_1 \int_{\Theta_2} L^2(\theta_2 \mid x_0)\, f_2(\theta_2)\, d\theta_2} \\ &= \frac{L^1(\theta_0 \mid x_0)}{\displaystyle\int_{\Theta_1} L^1(\theta_1 \mid x_0)\, f_1(\theta_1)\, d_{\theta_1}} = \frac{\bar{f}_{\bar{H}}(x_0, \theta_2)}{\bar{f}_{\bar{A}}(x_0, \theta_2)}, \end{aligned} \tag{A-24}$$

hence,

$$\frac{f_{H_0}(x_0)}{f_{H_1}(x_0)} < \frac{a}{b} \Leftrightarrow \frac{\bar{f}_{\bar{H}_0}(x_0, \theta_2)}{\bar{f}_{\bar{H}_1}(x_0, \theta_2)} < \frac{a}{b}, \tag{A-25}$$

and consequently

$$\varphi^*(x_0) = 1 \Leftrightarrow \bar{\varphi}^*(x_0, \theta_2) = 1 \quad \square. \tag{A-26}$$

Proof Corollary 1: The proof of the corollary follows directly from Theorem 2 and Theorem 4.

Proof Theorem 5: In fact, from Theorem 4 and Corollary 1 we have that for all $x_0 \in \mathfrak{X}$

$$\varphi^*(x_0) = 1 \Leftrightarrow Bf(x_0) = \frac{f_H(x_0)}{f_A(x_0)} \leqslant \frac{b}{a}, \tag{A-27}$$

which is equivalent to

$$\sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq Bf(\boldsymbol{x_0})} \mathbb{F}_{H,T(x_0)}(\{\boldsymbol{x}\}) \leqslant \sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq b/a} \mathbb{F}_{H,T(x_0)}(\{\boldsymbol{x}\}), \tag{A-28}$$

where the last equivalence comes from a similar argument presented by Pereira *et al.* (2017). Then

$$\varphi^*(x_0) = 1 \Leftrightarrow \sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq Bf(\boldsymbol{x_0})} \frac{\mathbb{F}_H(\{x\}\cap\{T(x)=T(x_0)\})}{\mathbb{F}_H(\{T(x)=T(x_0)\})} \leqslant \sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq b/a} \frac{\mathbb{F}_H(\{x\}\cap\{T(x)=T(x_0)\})}{\mathbb{F}_H(\{T(x)=T(x_0)\})}. \tag{A-29}$$

But,

$$\mathbb{F}_H(\{x\}\cap\{T(x)=T(x_0)\}) = \begin{cases} f_H(x) & \text{if} \quad T(x)=T(x_0) \\ 0 & \text{if} \quad T(x)\neq T(x_0). \end{cases} \tag{A-30}$$

Then

$$\begin{aligned} \varphi^*(x_0) = 1 &\Leftrightarrow \sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq Bf(\boldsymbol{x_0})} \frac{f_H(x)\,\mathbb{1}(T(x)=T(x_0))}{\mathbb{F}_H(\{T(x)=T(x_0)\})} \leqslant \sum_{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq b/a} \frac{f_H(x)\,\mathbb{1}(T(x)=T(x_0))}{\mathbb{F}_H(\{T(x)=T(x_0)\})} \\ &\Leftrightarrow \frac{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq Bf(\boldsymbol{x_0}) \\ T(x)=T(x_0)}} f_H(x)}{\sum\limits_{\boldsymbol{x}\in\mathfrak{X}:T(x)=T(x_0)} f_H(x)} \leqslant \frac{\sum\limits_{\substack{\boldsymbol{x}\in\mathfrak{X}:Bf(\boldsymbol{x})\leq b/a \\ T(x)=T(x_0)}} f_H(x)}{\sum\limits_{\boldsymbol{x}\in\mathfrak{X}:T(x)=T(x_0)} f_H(x)} \\ &\Leftrightarrow P_T\text{-value}(x_0) \leqslant \alpha_T^* \qquad \square. \end{aligned} \tag{A-31}$$

A.3 Homogeneity

A.3.1 Binomial

To test the hypotheses (3-1) we compute the predictive functions under each hypothesis. Hence, the predictive function under the null hypothesis is given by

$$\begin{aligned} f_H(\boldsymbol{x}) &= \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_H(\theta) \\ &= \frac{\oint_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})\pi(\boldsymbol{\theta})d\mathbb{P}_H(\theta)}{\oint_\Theta \pi(\boldsymbol{\theta})d\mathbb{P}_H(\theta)} \\ &= \frac{\int_0^1 t^{\sum_{i=1}^m x_i+a_i-1}(1-t)^{\sum_{i=1}^m n_i+b_i-x_i-1}\sqrt{\Delta}\,dt}{\int_0^1 t^{\sum_{i=1}^3 a_i-1}(1-t)^{\sum_{i=1}^3 b_i-1}\sqrt{\Delta}\,dt}, \end{aligned} \tag{A-32}$$

where, $\Delta = \left(\frac{\partial\theta_1}{\partial t}\right)^2 + \left(\frac{\partial\theta_2}{\partial t}\right)^2 + \ldots + \left(\frac{\partial\theta_m}{\partial t}\right)^2$. Hence, the arch $\sqrt{\Delta}$ that represents the hypothesis H is given by the equations

$$\theta_1 = \theta_2 = \ldots = \theta_m = t \quad \text{then,} \quad \frac{\partial\theta_1}{\partial t} = \frac{\partial\theta_2}{\partial t} = \ldots = \frac{\partial\theta_m}{\partial t} = 1, \tag{A-33}$$

thus, the predictive function under H can be written as

$$
\begin{aligned}
f_H(\boldsymbol{x}) =& \frac{\displaystyle\int_0^1 t^{\sum_{i=1}^m x_i+a_i-1}(1-t)^{\sum_{i=1}^m n_i+b_i-x_i-1}\sqrt{m}\,dt}{\displaystyle\int_0^1 t^{\sum_{i=1}^3 a_i-1}(1-t)^{\sum_{i=1}^3 b_i-1}\sqrt{m}\,dt} \\
=& \frac{\prod_{i=1}^m \binom{n_i}{x_i}\Gamma(B)\Gamma(C)\Gamma\left(\sum_{i=1}^m (a_i+b_i)-2(m-1)\right)}{\Gamma(B+C)\Gamma\left(\sum_{i=1}^m a_i-(m-1)\right)\Gamma\left(\sum_{i=1}^m b_i-(m-1)\right)},
\end{aligned} \tag{A-34}
$$

where, $B=\sum_{i=1}^m (a_i+x_i)-(m-1)$ and $C=\sum_{i=1}^m (n_i+b_i-x_i)-(m-1)$. Analogously, the predictive function under the alternative hypothesis A can be written as

$$
\begin{aligned}
f_A(\boldsymbol{x}) &= \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_A(\theta) \\
&= \prod_{i=1}^m \binom{n_i}{x_i}\tfrac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\int_\Theta \prod_{i=1}^m \theta_i^{x_i}(1-\theta_i)^{n_i-x_i}d\mathbb{P}_A(\theta) \\
&= \prod_{i=1}^m \binom{n_i}{x_i}\tfrac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\prod_{i=1}^m \int_0^1 \theta_i^{a_i+x_i-1}(1-\theta_i)^{n_i+b_i-x_i-1}d\theta_i \\
&= \prod_{i=1}^m \binom{n_i}{x_i}\tfrac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\prod_{i=1}^m \tfrac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)}.
\end{aligned} \tag{A-35}
$$

Note that in this case the distribution of θ under A coincides with the distribution of θ on the whole parametric space since $A=\Theta\setminus H$, where H is a set of probability zero. Finally, Bayes factor $Bf(\boldsymbol{x})$ is given by

$$
Bf(\boldsymbol{x}) = \frac{\Gamma(C)\Gamma(B)\Gamma\left(\sum_{i=1}^m (a_i+b_i)-2(m-1)\right)}{\Gamma(A+B)\Gamma\left(\sum_{i=1}^m a_i-(m-1)\right)\Gamma\left(\sum_{i=1}^m b_i-(m-1)\right)\prod_{i=1}^m D_i}, \tag{A-36}
$$

where, $D_i=\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\frac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)}$.

A.3.2 Multinomial

As a generalization of Binomial test, we compute the predictive functions under each hypothesis for testing (3-9). Hence, the predictive function under null hypothesis is given by

$$
\begin{aligned}
f_H(\boldsymbol{x}) =& \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_H(\boldsymbol{\theta}). \\
=& \frac{\displaystyle\oint_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})\pi(\boldsymbol{\theta})d\mathbb{P}_H(\boldsymbol{\theta})}{\displaystyle\oint_\Theta \pi(\boldsymbol{\theta})d\mathbb{P}_H(\boldsymbol{\theta})}. \\
=& \frac{\prod_{j=1}^T\left[\frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}\right]\displaystyle\int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^T \alpha_{ij}+x_{ij}-1\right)}\sqrt{\Delta}\,dt}{\displaystyle\int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^T \alpha_{ij}-1\right)}\sqrt{\Delta}\,dt},
\end{aligned} \tag{A-37}
$$

in this case, $\Delta=\sum_{i=1}^m D_i$, where D_i represents the determinant of the i-th squared submatrix of order $(k-1)$ of the partial matrix of derivatives J_θ. Hence, the arch $\sqrt{\Delta}$ that represents the hypothesis H for the Multinomial case is given by the equations

$$\begin{aligned}
t_1 =& \theta_{11} = \theta_{12} = \ldots = \theta_{1T} \\
t_2 =& \theta_{21} = \theta_{22} = \ldots = \theta_{2T} \\
\vdots \quad & \vdots \ldots\ldots \vdots \ldots\ldots \vdots \\
t_{(k-1)} =& \theta_{(k-1)1} = \theta_{(k-1)2} = \ldots = \theta_{(k-1)T},
\end{aligned} \tag{A-38}$$

where, the partial derivatives matrix J_θ is given by

$$J_\theta = \begin{bmatrix}
\frac{\partial \theta_{11}}{\partial t_1} & \cdots & \frac{\partial \theta_{11}}{\partial t_{(k-1)}} \\
\vdots & \ddots & \vdots \\
\frac{\partial \theta_{(k-1)1}}{\partial t_1} & \cdots & \frac{\partial \theta_{(k-1)1}}{\partial t_{(k-1)}} \\
\frac{\partial \theta_{12}}{\partial t_1} & \cdots & \frac{\partial \theta_{12}}{\partial t_{(k-1)}} \\
\vdots & \ddots & \vdots \\
\frac{\partial \theta_{(k-1)2}}{\partial t_1} & \cdots & \frac{\partial \theta_{(k-1)2}}{\partial t_{(k-1)}} \\
\vdots & \ddots & \vdots \\
\vdots & \cdots & \vdots \\
\frac{\partial \theta_{1T}}{\partial t_1} & \cdots & \frac{\partial \theta_{1T}}{\partial t_{(k-1)}} \\
\vdots & \cdots & \vdots \\
\frac{\partial \theta_{(k-1)T}}{\partial t_1} & \cdots & \frac{\partial \theta_{(k-1)T}}{\partial t_{(k-1)}}
\end{bmatrix} = \begin{bmatrix}
1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1 \\
1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1 \\
\vdots & \ddots & \vdots \\
\vdots & \ldots & \vdots \\
1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1
\end{bmatrix}. \tag{A-39}$$

Therefore, the predictive function under the null hypothesis H is given by

$$\begin{aligned}
f_H(\boldsymbol{x}) =& \frac{\prod_{j=1}^M \left[\frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}\right] \int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^M \alpha_{ij}+x_{ij}-1\right)} \sqrt{\sum_{i=1}^m D_i}\, dt}{\int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^M \alpha_{ij}-1\right)} \sqrt{\sum_{i=1}^m D_i}\, dt} \\
=& \frac{\prod_{j=1}^M \left[\frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}\right] \int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^M \alpha_{ij}+x_{ij}-1\right)}\, dt}{\int_0^1 \prod_{i=1}^k t_i^{\left(\sum_{j=1}^M \alpha_{ij}-1\right)}\, dt} \\
=& C_1 \times \left[\frac{\prod_{i=1}^k \Gamma\left(\sum_{j=1}^M (\alpha_{ij}+x_{ij}) - (M-1)\right)}{\Gamma\left(\sum_{i=1}^k \left(\sum_{j=1}^M (\alpha_{ij}+x_{ij}) - (M-1)\right)\right)} \frac{\Gamma\left(\sum_{i=1}^k \left(\sum_{j=1}^M \alpha_{ij} - (M-1)\right)\right)}{\prod_{i=1}^k \Gamma\left(\sum_{j=1}^M \alpha_{ij} - (M-1)\right)}\right],
\end{aligned} \tag{A-40}$$

with $C_1 = \prod_{j=1}^M \frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}$. Analogously, the predictive function under the alternative hypothesis A can be written as

$$\begin{aligned}
f_A(\boldsymbol{x}) =& \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\
=& \prod_{j=1}^M \left[\frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}\right] \int_\Theta \prod_{j=1}^M \prod_{i=1}^k \theta_{ij}^{x_{ij}}\, d\mathbb{P}_A(\boldsymbol{\theta}) \\
=& \prod_{j=1}^M \left[\frac{\Gamma(\sum_{i=1}^k \alpha_{ij})}{\prod_{i=1}^k \Gamma(\alpha_{ij})} \frac{n_j!}{x_{1j}!x_{2j}!\ldots x_{kj}!}\right] \prod_{j=1}^M \int_0^1 \prod_{i=1}^k \theta_{ij}^{\alpha_{ij}+x_{ij}-1} d\theta_j \\
=& \prod_{j=1}^M \left[\frac{\Gamma(\sum_{i=1}^k \alpha_{ij})}{\prod_{i=1}^k \Gamma(\alpha_{ij})} \frac{\Gamma(\sum_{i=1}^k x_{ij}+1)}{\prod_{i=1}^k \Gamma(x_{ij}+1)}\right] \prod_{j=1}^M \left[\frac{\prod_{i=1}^k \Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^k \alpha_{ij}+x_{ij})}\right].
\end{aligned} \tag{A-41}$$

As a result, the Bayes factor $Bf(\boldsymbol{x})$ in favor of the null hypothesis H is

$$Bf(\boldsymbol{x}) = \frac{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k} x_{ij}+1)}{\prod_{i=1}^{k}\Gamma(x_{ij}+1)}\right] \frac{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)}{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\right)} \frac{\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)\right)}{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)}}{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\frac{\Gamma(\sum_{i=1}^{k}x_{ij}+1)}{\prod_{i=1}^{k}\Gamma(x_{ij}+1)}\right]\prod_{j=1}^{M}\left[\frac{\prod_{i=1}^{k}\Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^{k}\alpha_{ij}+x_{ij})}\right]}$$

$$= \frac{\prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)\right)}{\prod_{j=1}^{M}\left[\frac{\Gamma(\sum_{i=1}^{k}\alpha_{ij})}{\prod_{i=1}^{k}\Gamma(\alpha_{ij})}\frac{\prod_{i=1}^{k}\Gamma(\alpha_{ij}+x_{ij})}{\Gamma(\sum_{i=1}^{k}\alpha_{ij}+x_{ij})}\right]\Gamma\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{M}(\alpha_{ij}+x_{ij})-(M-1)\right)\right) \times \prod_{i=1}^{k}\Gamma\left(\sum_{j=1}^{M}\alpha_{ij}-(M-1)\right)}. \tag{A-42}$$

A.4 Independence

Independence hypothesis for 2×2 contingency table

Under the hypotheses (3-15) the predictive function (3-18) is not easy to compute. Then, we use the reparameterization (3-20) to factorize the likelihood function as (2-31). Thus it satisfies the conditions of Theorem 4 allowing the computation of $Bf(\boldsymbol{x})$. Then, with the reparameterization (3-20) the likelihood function for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ generated by $\boldsymbol{x} = (x_{11}, x_{12}, x_{21}, x_{22})$ is defined as

$$\begin{aligned} L(\lambda|x) &= \frac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!}\left(\frac{\theta_{11}}{\theta_{11}+\theta_{12}}\right)^{x_{11}}\left(\frac{\theta_{12}}{\theta_{11}+\theta_{12}}\right)^{x_{12}}\left(\frac{\theta_{21}}{\theta_{21}+\theta_{22}}\right)^{x_{21}}\left(\frac{\theta_{22}}{\theta_{21}+\theta_{22}}\right)^{x_{22}} \\ &\quad \times (\theta_{11}+\theta_{12})^{x_{11}+x_{12}}(\theta_{21}+\theta_{22})^{x_{21}+x_{22}} \\ &= \frac{n!}{x_{11}!\ldots x_{22}!}\lambda_1^{x_{11}}(1-\lambda_1)^{x_{12}}\lambda_2^{x_{21}}(1-\lambda_2)^{x_{22}}\lambda_3^{x_{11}+x_{12}}(1-\lambda_3)^{x_{21}+x_{22}} \\ &= \binom{x_{11}+x_{12}}{x_{11}}\lambda_1^{x_{11}}(1-\lambda_1)^{x_{12}}\binom{x_{21}+x_{22}}{x_{21}}\lambda_2^{x_{21}}(1-\lambda_2)^{x_{22}} \\ &\quad \times \binom{n}{x_{11}+x_{12}}\lambda_3^{x_{11}+x_{12}}(1-\lambda_3)^{x_{21}+x_{22}}. \end{aligned} \tag{A-43}$$

And the prior distribution of θ under the reparameterization (3-20) is given by

$$\begin{aligned} \pi(\boldsymbol{\lambda}) = &\left[\frac{\Gamma(\alpha_{11}+\alpha_{12})}{\Gamma(\alpha_{11})\Gamma(\alpha_{12})}\lambda_1^{\alpha_{11}-1}(1-\lambda_1)^{\alpha_{12}-1}\right]\left[\frac{\Gamma(\alpha_{21}+\alpha_{22})}{\Gamma(\alpha_{21})\Gamma(\alpha_{22})}\lambda_2^{\alpha_{21}-1}(1-\lambda_2)^{\alpha_{22}-1}\right] \\ &\times\left[\frac{\Gamma(\sum_{i,j=1}^{2}\alpha_{ij})}{\Gamma(\alpha_{11}+\alpha_{12})\Gamma(\alpha_{21}+\alpha_{22})}\lambda_3^{\alpha_{11}+\alpha_{12}-2}(1-\lambda_3)^{\alpha_{21}+\alpha_{22}-2}\right]. \end{aligned} \tag{A-44}$$

Note that the prior distribution (A-44) can be seen as a product of Beta distributions. Thus, the predictive function under the null hypothesis (3-21) can be written as

$$
\begin{aligned}
f_{\tilde{H}}(x) &= \int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
&= \frac{\oint_{\Lambda} L(\boldsymbol{\lambda}|\,\boldsymbol{x})\pi(\boldsymbol{\lambda})|J| d\mathbb{P}_H(\lambda)}{\oint_{\Lambda} \pi(\boldsymbol{\lambda})|J| d\mathbb{P}_H(\lambda)} \\
&= \frac{n!}{x_{11}! \ldots x_{22}!} \left[\frac{\int_0^1 t^{x_{11}+x_{21}+\alpha_{11}+\alpha_{21}-2}(1-t)^{x_{12}+x_{22}+\alpha_{12}+\alpha_{22}-2} dt \times \int_0^1 \lambda_3^{x_{11}+x_{12}+\alpha_{11}+\alpha_{12}-2}(1-\lambda_3)^{x_{21}+x_{22}+\alpha_{21}+\alpha_{22}-2}|J|\sqrt{\Delta}\, d\lambda_3}{\int_0^1 t^{\alpha_{11}+\alpha_{21}-2}(1-t)^{\alpha_{12}+\alpha_{22}-2} dt \times \int_0^1 \lambda_3^{\alpha_{11}+\alpha_{12}-2}(1-\lambda_3)^{\alpha_{21}+\alpha_{22}-2}|J|\sqrt{\Delta}\, d\lambda_3} \right],
\end{aligned}
\tag{A-45}
$$

where $|J| = \lambda_3(1-\lambda_3)$ is the Jacobian of the transformation and $\Delta = \left(\frac{\partial \lambda_1}{\partial t}\right)^2 + \left(\frac{\partial \lambda_2}{\partial t}\right)^2$. Hence, the arch $\sqrt{\Delta}$ that represents the null hypothesis H is given by the equations

$$
\lambda_1 = \lambda_2 = t \quad \text{then,} \quad \frac{\partial \lambda_1}{\partial t} = \frac{\partial \lambda_2}{\partial t} = 1, \tag{A-46}
$$

thus,

$$
\begin{aligned}
\sqrt{\Delta} &= \sqrt{\left(\frac{\partial \lambda_1}{\partial t}\right)^2 + \left(\frac{\partial \lambda_2}{\partial t}\right)^2} \\
&= \sqrt{(1)^2 + (1)^2} \\
&= \sqrt{2},
\end{aligned}
\tag{A-47}
$$

and consequently,

$$
\begin{aligned}
f_{\tilde{H}}(x) &= \frac{n!}{x_{11}! \ldots x_{22}!} \left[\frac{\int_0^1 t^{x_{11}+x_{21}+\alpha_{11}+\alpha_{21}-2}(1-t)^{x_{12}+x_{22}+\alpha_{12}+\alpha_{22}-2} dt \times \int_0^1 \lambda_3^{x_{11}+x_{12}+\alpha_{11}+\alpha_{12}-1}(1-\lambda_3)^{x_{21}+x_{22}+\alpha_{21}+\alpha_{22}-1} d\lambda_3}{\int_0^1 t^{\alpha_{11}+\alpha_{21}-2}(1-t)^{\alpha_{12}+\alpha_{22}-2} dt \times \int_0^1 \lambda_3^{\alpha_{11}+\alpha_{12}-1}(1-\lambda_3)^{\alpha_{21}+\alpha_{22}-1} d\lambda_3} \right] \\
&= \frac{n!}{x_{11}! \ldots x_{22}!} \left[\frac{\Gamma(n_{.1}+\alpha_{11}+\alpha_{21}-1)\Gamma(n_{.2}+\alpha_{12}+\alpha_{22}-1)\Gamma(n_{1.}+\alpha_{11}+\alpha_{12}) \times \Gamma(n_{2.}+\alpha_{21}+\alpha_{22})\Gamma(\sum_{i,j}^2 \alpha_{ij}-2)\Gamma(\sum_{i,j}^2 \alpha_{ij})}{\Gamma(n+\sum_{i,j=1}^2 \alpha_{ij}-2)\Gamma(n+\sum_{i,j=1}^2 \alpha_{ij})\Gamma(\alpha_{11}+\alpha_{21}-1) \times \Gamma(\alpha_{12}+\alpha_{22}-1)\Gamma(\alpha_{11}+\alpha_{22})\Gamma(\alpha_{21}+\alpha_{22})} \right].
\end{aligned}
\tag{A-48}
$$

where, $n_{.j} = \sum_{i=1}^2 x_{ij}$ and $n_{i.} = \sum_{j=1}^2 x_{ij}$, with $i, j = 1, 2$. Now, for the predictive function under the alternative hypothesis A we have that

$$
\begin{aligned}
f_{\bar{A}}(x) =& \int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\bar{A}}(\boldsymbol{\lambda}) \\
=& \frac{n!}{x_{11}!\ldots x_{22}!}\frac{\Gamma(\sum_{i,j=1}^{2}\alpha_{ij})}{\Gamma(\alpha_{11})\ldots\Gamma(\alpha_{22})} \int_{(0,1)^2} \lambda_1^{x_{11}+\alpha_{11}-1}(1-\lambda_1)^{x_{12}+\alpha_{12}-1}\lambda_2^{x_{21}+\alpha_{21}-1} \\
& \times (1-\lambda_2)^{x_{22}+\alpha_{22}-1} d\lambda_1 d\lambda_2 \int_0^1 \lambda_3^{\sum_{j=1}^{2}(x_{1j}+\alpha_{1j}-1)}(1-\lambda_3)^{\sum_{j=1}^{2}(x_{2j}+\alpha_{2j}-1)}|J| d\lambda_3 \\
=& \frac{n!}{x_{11}!\ldots x_{22}!}\frac{\Gamma(\sum_{i,j=1}^{2}\alpha_{ij})}{\Gamma(\alpha_{11})\ldots\Gamma(\alpha_{22})} \int_0^1 \lambda_1^{x_{11}+\alpha_{11}-1}(1-\lambda_1)^{x_{12}+\alpha_{12}-1} d\lambda_1 \int_0^1 \lambda_2^{x_{21}+\alpha_{21}-1} \\
& \times (1-\lambda_2)^{x_{22}+\alpha_{22}-1} d\lambda_2 \int_0^1 \lambda_3^{x_{11}+x_{12}+\alpha_{11}+\alpha_{12}-1}(1-\lambda_3)^{x_{21}+x_{22}+\alpha_{21}+\alpha_{22}-1} d\lambda_3 \\
=& \frac{n!}{x_{11}!\ldots x_{22}!}\left[\frac{\Gamma(\sum_{i,j=1}^{2}\alpha_{ij})\Gamma(x_{11}+\alpha_{11})\Gamma(x_{12}+\alpha_{12})\Gamma(x_{21}+\alpha_{21})\Gamma(x_{22}+\alpha_{22})}{\Gamma(\sum_{i,j=1}^{2}x_{ij}+\alpha_{ij})\prod_{i,j=1}^{2}\Gamma(\alpha_{ij})}\right].
\end{aligned}
\tag{A-49}
$$

Note that the likelihood function (A-43) comes factorized as (2-31) and the prior distribution of $\boldsymbol{\lambda}$ can be written as a product of distributions, in this case, Beta distributions. Then, by Theorem 4, the predictive functions can be calculated disregarding λ_3, that is

$$
f_H(\boldsymbol{x}|T=x_{i.}) = \binom{x_{11}+x_{12}}{x_{11}}\binom{x_{21}+x_{22}}{x_{21}} \times \left[\frac{\Gamma(n_{.1}+\alpha_{11}+\alpha_{21}-1)\Gamma(n_{.2}+\alpha_{12}+\alpha_{22}-1)\Gamma(\sum_{i,j}^{2}\alpha_{ij}-2)}{\Gamma(n+\sum_{i,j=1}^{2}\alpha_{ij}-2)\Gamma(\alpha_{11}+\alpha_{21}-1)\Gamma(\alpha_{12}+\alpha_{22}-1)}\right], \tag{A-50}
$$

and

$$
f_A(\boldsymbol{x}|T=x_{i.}) = \binom{x_{11}+x_{12}}{x_{11}}\binom{x_{21}+x_{22}}{x_{21}} \times \left[\frac{\prod_{i=1}^{2}\Gamma(\alpha_{i1}+\alpha_{i2})\Gamma(x_{11}+\alpha_{11})\Gamma(x_{12}+\alpha_{12})\Gamma(x_{21}+\alpha_{21})\Gamma(x_{22}+\alpha_{22})}{\prod_{i,j=1}^{2}\Gamma(\alpha_{ij})\Gamma(x_{11}+x_{12}+\alpha_{11}+\alpha_{12})\Gamma(x_{21}+x_{22}+\alpha_{21}+\alpha_{22})}\right], \tag{A-51}
$$

where $x_{i.} = \sum_{j=1}^{2} x_{ij}$. As a result, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed as

$$
Bf(\boldsymbol{x}) = \left[\frac{\Gamma(n_{.1}+\alpha_{11}+\alpha_{21}-1)\Gamma(n_{.2}+\alpha_{12}+\alpha_{22}-1)\Gamma(\sum_{i,j=1}^{2}\alpha_{ij}-2) \times \Gamma(n_{1.}+\alpha_{11}+\alpha_{12})\Gamma(n_{2.}+\alpha_{21}+\alpha_{22})\prod_{i,j=1}^{2}\Gamma(\alpha_{ij})}{\Gamma(n+\sum_{i,j=1}^{2}\alpha_{ij}-2)\Gamma(\alpha_{11}+\alpha_{21}-1)\Gamma(\alpha_{12}+\alpha_{22}-1)\Gamma(x_{11}+\alpha_{11}) \times \Gamma(x_{12}+\alpha_{12})\Gamma(x_{21}+\alpha_{21})\Gamma(x_{22}+\alpha_{22})\prod_{i=1}^{2}\Gamma(\alpha_{i1}+\alpha_{i2})}\right]. \tag{A-52}
$$

Independence hypothesis for $r \times l$ contingency tables

As in 2×2 case, under the null hypothesis in (3-26) the predictive function (3-29) is not easy to compute. Then, one more time, we use a suitable reparameterization that satisfies the conditions of Theorem 4. Then, consider the following reparameterization

$$\begin{array}{llll}\lambda_{11}=\frac{\theta_{11}}{\theta_{1.}}; & \lambda_{12}=\frac{\theta_{12}}{\theta_{1.}}; & \ldots & ;\lambda_{1(l-1)}=\frac{\theta_{1(l-1)}}{\theta_{1.}}\\ \lambda_{21}=\frac{\theta_{21}}{\theta_{2.}}; & \lambda_{22}=\frac{\theta_{22}}{\theta_{2.}}; & \ldots & ;\lambda_{2(l-1)}=\frac{\theta_{2(l-1)}}{\theta_{2.}}\\ \vdots & \vdots & \ldots & \vdots\\ \lambda_{r1}=\frac{\theta_{r1}}{\theta_{r.}}; & \lambda_{r2}=\frac{\theta_{r2}}{\theta_{r.}}; & \ldots & ;\lambda_{r(l-1)}=\frac{\theta_{r(l-1)}}{\theta_{r.}}\\ \eta_1=\theta_{1.}; & \eta_2=\theta_{2.}; & \ldots & ;\eta_{r-1}=\theta_{(r-1).}.\end{array} \tag{A-53}$$

Thus, by using equation (A-53) the likelihood function is defined as

$$L(\lambda|\boldsymbol{x})=\prod_{i=1}^{r}\left[\frac{(\sum_{j=1}^{l}x_{ij})!}{\prod_{j=1}^{l}x_{ij}!}\prod_{j=1}^{l}\lambda_{ij}^{x_{ij}}\right]\left[\frac{n!}{\prod_{i=1}^{r}(\sum_{j=1}^{l}x_{ij})!}\right]\prod_{i=1}^{r}\eta_i^{\sum_{j=1}^{l}x_{ij}}. \tag{A-54}$$

and the prior distribution is given by

$$\pi(\boldsymbol{\lambda},\boldsymbol{\eta})=\prod_{i=1}^{r}\left[\frac{\Gamma(\sum_{j=1}^{l}\alpha_{ij})}{\prod_{j=1}^{l}\Gamma(\alpha_{ij})}\prod_{j=1}^{l}\lambda_{ij}^{\alpha_{ij}-1}\right]\frac{\Gamma(\sum_{i=1}^{r}\sum_{j=1}^{l}\alpha_{ij})}{\prod_{i=1}^{r}\Gamma(\sum_{j=1}^{l}\alpha_{ij})}\prod_{i=1}^{r}\eta_i^{\sum_{j=1}^{l}\alpha_{ij}-1}. \tag{A-55}$$

Therefore, the predictive function under each hypothesis (3-21) can be written as:

$$\begin{aligned} f_{\tilde{H}}(x)=&\int_{\Lambda}L(\boldsymbol{\lambda}|\boldsymbol{x})d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda})\\ =&\frac{\oint_{\Lambda}L(\boldsymbol{\lambda}|\boldsymbol{x})\pi(\boldsymbol{\lambda})|J|d\mathbb{P}_H(\lambda)}{\oint_{\Lambda}\pi(\boldsymbol{\lambda})|J|d\mathbb{P}_H(\lambda)}\\ =&\frac{n!}{\prod_{i=1}^{r}\prod_{j=1}^{l}x_{ij}!}\left[\frac{\int_{(0,1)^{l-1}}\prod_{j=1}^{l}t_j^{\sum_{i=1}^{r}(x_{ij}+\alpha_{ij}-1)}\sqrt{\Delta}\,dt\int_{(0,1)^{r-1}}\prod_{i=1}^{r}\eta_i^{\sum_{j=1}^{l}(x_{ij}+\alpha_{ij}-1)}|J|\,d\boldsymbol{\eta}}{\int_{(0,1)^{l-1}}\prod_{j=1}^{l}t_j^{\sum_{i=1}^{r}(\alpha_{ij}-1)}\sqrt{\Delta}\,dt\int_{(0,1)^{r-1}}\prod_{i=1}^{r}\eta_i^{\sum_{j=1}^{l}(\alpha_{ij}-1)}|J|\,d\boldsymbol{\eta}}\right],\end{aligned} \tag{A-56}$$

where $|J|=\prod_{i=1}^{r}\eta_i^{r-1}$ is the Jacobian of the transformation and $\Delta=\sum_{i=1}^{m}D_i$, where D_i represents the determinant of the i-th squared sub-matrix of order $(l-1)$ of the partial matrix of derivatives (A-39). Note that the new parameterization converted the independence test into one of homogeneity, thereby, in this case, the equations to the arch $\sqrt{\Delta}$ also are given by (A-38). Consequently,

$$\begin{aligned} f_{\tilde{H}}(x)=&C_2\times\left[\frac{\int_{(0,1)^{l-1}}\prod_{j=1}^{l}t_j^{\alpha_j^*-r}\sqrt{\sum_{i=1}^{m}D_i}\,dt\int_{(0,1)^{r-1}}\prod_{i=1}^{r}\eta_i^{\alpha_i^*-l+(r-1)}\,d\boldsymbol{\eta}}{\int_{(0,1)^{l-1}}\prod_{j=1}^{l}t_j^{\alpha_j^{**}-r}\sqrt{\sum_{i=1}^{m}D_i}\,dt\int_{(0,1)^{r-1}}\prod_{i=1}^{r}\eta_i^{\alpha_i^{**}-l+(r-1)}\,d\boldsymbol{\eta}}\right]\\ =&C_2\times\left[\frac{\begin{array}{r}\prod_{j=1}^{l}\Gamma\left(\alpha_j^*-r+1\right)\prod_{i=1}^{r}\Gamma\left(\alpha_i^*-l+r\right)\\ \times\Gamma\left(\sum_{j=1}^{l}(\alpha_j^{**}-r+1)\right)\Gamma\left(\sum_{i=1}^{r}\alpha_i^{**}-l+r\right)\end{array}}{\begin{array}{r}\Gamma\left(\sum_{j=1}^{l}(\alpha_j^*-r+1)\right)\Gamma\left(\sum_{i=1}^{r}(\alpha_i^*-l+r)\right)\\ \times\prod_{j=1}^{l}\Gamma\left(\alpha_j^{**}-r+1\right)\prod_{i=1}^{r}\Gamma\left(\alpha_i^{**}-l+r\right)\end{array}}\right],\end{aligned} \tag{A-57}$$

and

$$
\begin{aligned}
f_{\bar{A}}(x) =& \int_\Lambda L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\bar{A}}(\boldsymbol{\lambda}) \\
=& C_2 \times \left[\begin{array}{r} \prod_{i=1}^r \left[\frac{\Gamma(\sum_{j=1}^l \alpha_{ij})}{\prod_{j=1}^l \Gamma(\alpha_{ij})} \right] \int_{(0,1)^{l-1}} \prod_{i=1}^r \left[\prod_{j=1}^l \lambda_{ij}^{x_{ij}+\alpha_{ij}-1} \right] d\boldsymbol{\lambda} \\ \times \int_{(0,1)^{r-1}} \prod_{i=1}^r \eta_i^{\sum_{j=1}^l (x_{ij}+\alpha_{ij}-1)} |J| \, d\boldsymbol{\eta} \end{array} \right] \\
=& C_2 \times \prod_{i=1}^r \left[\frac{\Gamma(\alpha_i^{**})}{\prod_{j=1}^l \Gamma(\alpha_{ij})} \int_{(0,1)^{l-1}} \prod_{j=1}^l \lambda_{ij}^{x_{ij}+\alpha_{ij}-1} d\boldsymbol{\lambda_i} \right] \int_{(0,1)^{r-1}} \prod_{i=1}^r \eta_i^{\alpha_i^* - l + (r-1)} d\boldsymbol{\eta} \\
=& C_2 \times \left[\frac{\prod_{i=1}^r \left[\prod_{j=1}^l \Gamma(x_{ij}+\alpha_{ij}) \Gamma(\alpha_i^{**}) \right] \prod_{i=1}^r \Gamma(\alpha_i^* - l + r) \Gamma(\sum_{i=1}^r (\alpha_i^{**} - l + r))}{\prod_{i=1}^r \left[\Gamma(\alpha_i^*) \prod_{j=1}^l \Gamma(\alpha_{ij}) \right] \Gamma(\sum_{i=1}^r (\alpha_i^* - l + r)) \prod_{i=1}^r \Gamma(\alpha_i^{**} - l + r)} \right],
\end{aligned} \tag{A-58}
$$

where $C_2 = \frac{n!}{\prod_{i=1}^r \prod_{j=1}^l x_{ij}!}$, $\alpha_i^* = \sum_{j=1}^l x_{ij} + \alpha_{ij}$, $\alpha_j^* = \sum_{i=1}^r x_{ij} + \alpha_{ij}$, $\alpha_i^{**} = \sum_{j=1}^l \alpha_{ij}$ and $\alpha_j^{**} = \sum_{i=1}^r \alpha_{ij}$. Analogously to the 2×2 case, the likelihood function (A-54) is factorized as (2-31) and the prior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ can be written as a product of distributions, now, Dirichlet densities. Then, by Theorem 4, the predictive functions can be calculated independently of $\boldsymbol{\eta}$, that is

$$
f_{\bar{H}}(\boldsymbol{x}|\, T = x_{i.}) = \prod_{i=1}^r \left[\frac{(\sum_{j=1}^l x_{ij})!}{\prod_{j=1}^l x_{ij}!} \right] \left[\frac{\prod_{j=1}^l \Gamma(\alpha_j^* - r + 1) \Gamma\left(\sum_{j=1}^l (\alpha_j^{**} - r + 1) \right)}{\Gamma\left(\sum_{j=1}^l (\alpha_j^* - r + 1) \right) \prod_{j=1}^l \Gamma(\alpha_j^{**} - r + 1)} \right], \tag{A-59}
$$

and

$$
f_{\bar{A}}(\boldsymbol{x}|\, T = x_{i.}) = \prod_{i=1}^r \left[\frac{(\sum_{j=1}^l x_{ij})!}{\prod_{j=1}^l x_{ij}!} \right] \left[\frac{\prod_{i=1}^r \left[\prod_{j=1}^l \Gamma(x_{ij}+\alpha_{ij}) \Gamma(\alpha_i^{**}) \right]}{\prod_{i=1}^r \left[\Gamma(\alpha_i^*) \prod_{j=1}^l \Gamma(\alpha_{ij}) \right]} \right]. \tag{A-60}
$$

where $x_{i.} = \sum_{j=1}^r x_{ij}$. As a result, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed by

$$
Bf(\boldsymbol{x}) = \left[\frac{\prod_{j=1}^l \Gamma\left(\alpha_j^* - r + 1 \right) \Gamma\left(\sum_{j=1}^l (\alpha_j^{**} - r + 1) \right) \prod_{i=1}^r \left[\Gamma(\alpha_i^*) \prod_{j=1}^l \Gamma(\alpha_{ij}) \right]}{\Gamma\left(\sum_{j=1}^l (\alpha_j^* - r + 1) \right) \prod_{j=1}^l \Gamma\left(\alpha_j^{**} - r + 1 \right) \prod_{i=1}^r \left[\prod_{j=1}^l \Gamma(x_{ij}+\alpha_{ij}) \Gamma(\alpha_i^{**}) \right]} \right]. \tag{A-61}
$$

A.5 Diagonal Symmetry

A.5.1 Diagonal Symmetry for 3×3 case

Under the null hypothesis in (3-39) we have that $\theta_{ij} = \theta_{ji} \, \forall \, i < j$, then, using the reparameterization (3-42) we have:

$$
\begin{aligned}
\theta_{12} = \theta_{21} \quad &\Leftrightarrow \quad \eta_{12}\lambda_{12} = \eta_{12}(1 - \lambda_{12}) \\
\theta_{13} = \theta_{31} \quad &\Leftrightarrow \quad \eta_{13}\lambda_{13} = \eta_{13}(1 - \lambda_{13}) \\
\theta_{23} = \theta_{32} \quad &\Leftrightarrow \quad \eta_{23}\lambda_{23} = \eta_{23}(1 - \lambda_{23}).
\end{aligned} \tag{A-62}
$$

Consequently,

$$\begin{aligned}\theta_{12} = \theta_{21} \quad &\Leftrightarrow \quad \lambda_{12} = 1/2 \\ \theta_{13} = \theta_{31} \quad &\Leftrightarrow \quad \lambda_{13} = 1/2 \\ \theta_{23} = \theta_{32} \quad &\Leftrightarrow \quad \lambda_{23} = 1/2.\end{aligned} \tag{A-63}$$

Hence, the hypotheses (3-39) can be rewritten as:

$$\begin{aligned}\tilde{H}: \quad & \boldsymbol{\lambda} \in \Lambda_0 \\ \tilde{A}: \quad & \boldsymbol{\lambda} \in \Lambda_0^c,\end{aligned} \tag{A-64}$$

where,

$$\Psi = \{(i,j) : i < j \text{ for } i,j = 1,2,3.\}, \tag{A-65}$$

then, $\Lambda_0 = B \times \Lambda^*$, with $B = \{\lambda_{ij} : \lambda_{ij} = 1/2, \ \forall (i,j) \in \Psi.\}$ and $\Lambda^* = \{(\eta_{12}, \ldots, \eta_{22}) \in (0,1)^5 : \sum \eta_{ij} \leq 1\}$. Thus, the new likelihood function for $\boldsymbol{\lambda}$ generated by $x = (x_{11}, \ldots, x_{33})$ is given by

$$L(\boldsymbol{\lambda}|\boldsymbol{x}) = \left[\prod_{\Psi} \binom{x_{ij} + x_{ji}}{x_{ij}} \lambda_{ij}^{x_{ij}} (1 - \lambda_{ij})^{x_{ji}}\right] \left[\frac{n!}{\prod_{\Psi}(x_{ij} + x_{ji})! \prod_{i=1}^{3} x_{ii}!} \prod_{\Psi} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{3} \eta_{ii}^{x_{ii}}\right], \tag{A-66}$$

and the prior distribution is given by

$$\pi(\boldsymbol{\lambda}) = \left[\prod_{\Psi} \frac{\Gamma(\alpha_{ij} + \alpha_{ji})}{\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})} \lambda_{ij}^{\alpha_{ij}-1} (1 - \lambda_{ij})^{\alpha_{ji}-1}\right] \left[\frac{\Gamma(\sum_{i,j=1}^{3} \alpha_{ij})}{\prod_{i=1}^{3} \Gamma(\alpha_{ii}) \prod_{\Psi} \Gamma(\alpha_{ij} + \alpha_{ji})} \prod_{\Psi} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{3} \eta_{ii}^{\alpha_{ii}-1}\right]. \tag{A-67}$$

Therefore, the predictive function under null hypothesis for the 3×3 case is defined by

$$\begin{aligned} f_{\tilde{H}}(x) &= \int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ &= \frac{n!}{\prod_{i,j=1}^{3} x_{ij}!} \int_{\Lambda} \left[\prod_{\Psi} \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}}\right] \left[\prod_{\Psi} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{3} \eta_{ii}^{x_{ii}}\right] d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ &= \frac{n!}{\prod_{i,j=1}^{3} x_{ij}!} \left[\frac{\int_B \left[\prod_{\Psi} \lambda_{ij}^{x_{ij}+\alpha_{ij}-1} (1-\lambda_{ij})^{x_{ji}+\alpha_{ji}-1}\right] d\boldsymbol{\lambda} \int_{\Lambda^*} \left[\prod_{\Psi} \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-2} \prod_{i=1}^{3} \eta_{ii}^{x_{ii}+\alpha_{ii}-1}\right] |J| d\boldsymbol{\eta}}{\int_B \left[\prod_{\Psi} \lambda_{ij}^{\alpha_{ij}-1} (1-\lambda_{ij})^{\alpha_{ji}-1}\right] d\boldsymbol{\lambda} \int_{\Lambda^*} \left[\prod_{\Psi} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-2} \prod_{i=1}^{3} \eta_{ii}^{\alpha_{ii}-1}\right] |J| d\boldsymbol{\eta}}\right], \end{aligned} \tag{A-68}$$

where $|J|$ represents the jacobian of the transformation, given by

$$|J| = \eta_{12}\eta_{13}\eta_{23}.$$

The predictive function under null hypothesis of (A-64) can be written as

$$
\begin{aligned}
f_{\tilde{H}}(\boldsymbol{x}) =& \frac{n! \times (1/2)^{\sum_\Psi x_{ij}+x_{ji}}}{\prod_{i,j=1}^{r} x_{ij}!} \left[\frac{\displaystyle\int_{\Lambda^*} \prod_{\Psi} \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{3} \eta_{ii}^{x_{ii}+\alpha_{ii}-1} d\boldsymbol{\eta}}{\displaystyle\int_{\Lambda^*} \prod_{\Psi} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{3} \eta_{ii}^{\alpha_{ii}-1} d\boldsymbol{\eta}} \right] \\
=& \left[\frac{\begin{array}{r} n! \times (1/2)^{\sum_\Psi x_{ij}+x_{ji}} \prod_\Psi \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^3 \Gamma(x_{ii}+\alpha_{ii}) \\ \times \Gamma\left(\sum_\Psi \alpha_{ij}+\alpha_{ji}+\sum_{i=1}^3 \alpha_{ii}\right) \end{array}}{\begin{array}{r} \prod_{i,j=1}^3 x_{ij}! \Gamma\left(\sum_\Psi x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}+\sum_{i=1}^3 x_{ii}+\alpha_{ii}\right) \\ \times \prod_\Psi \Gamma(\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^3 \Gamma(\alpha_{ii}) \end{array}} \right].
\end{aligned} \tag{A-69}
$$

Now, the predictive function under the alternative hypothesis for the 3×3 case is defined by

$$
\begin{aligned}
f_{\tilde{A}}(x) =& \int_\Lambda L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
=& \frac{n!}{\prod_{i,j=1}^3 x_{ij}!} \int_\Lambda \left[\prod_\Psi \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}}\right] \left[\prod_\Psi \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^3 \eta_{ii}^{x_{ii}}\right] d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
=& C_3 \times \left[\int_{B^c} \prod_\Psi \lambda_{ij}^{x_{ij}+\alpha_{ij}-1} (1-\lambda_{ij})^{x_{ji}+\alpha_{ji}-1} d\boldsymbol{\lambda} \int_{\Lambda^*} \prod_\Psi \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^3 \eta_{ii}^{x_{ii}+\alpha_{ii}-1} d\boldsymbol{\eta} \right], \\
=& C_3 \times \left[\frac{\prod_\Psi \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_\Psi \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})} \frac{\prod_\Psi \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^3 \Gamma(x_{ii}+\alpha_{ii})}{\Gamma\left(\sum_\Psi x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}+\sum_{i=1}^3 x_{ii}+\alpha_{ii}\right)} \right],
\end{aligned} \tag{A-70}
$$

with $C_3 = \frac{n!}{\prod_{i,j=1}^3 x_{ij}!} \frac{\Gamma(\sum_{i,j=1}^3 \alpha_{ij})}{\prod_{i,j=1}^3 \Gamma(\alpha_{ij})}$. Note that, the likelihood function (A-66) comes factorized as (2-31) and the prior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ can be written as a product of distributions, one more time, Beta distributions. Then, by Theorem 4, the predictive functions can be calculated without considering $\boldsymbol{\eta}$, that is

$$
f_{\tilde{H}}(\boldsymbol{x}|T = x_\Psi) = \left[(1/2)^{\sum_\Psi x_{ij}+x_{ji}} \prod_\Psi \binom{x_{ij}+x_{ji}}{x_{ij}} \right], \tag{A-71}
$$

and

$$
f_{\tilde{A}}(\boldsymbol{x}|T = x_\Psi) = \left[\prod_\Psi \binom{x_{ij}+x_{ji}}{x_{ij}} \frac{\prod_\Psi \Gamma(\alpha_{ij}+\alpha_{ji})}{\prod_\Psi \Gamma(\alpha_{ij})\Gamma(\alpha_{ji})} \frac{\prod_\Psi \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_\Psi \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})} \right], \tag{A-72}
$$

where $x_\Psi = \sum_\Psi x_{ij} + x_{ji}$. As a result, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed as

$$
\begin{aligned}
Bf(\boldsymbol{x}) =& \frac{f_{\tilde{H}}(\boldsymbol{x})}{f_{\tilde{A}}(\boldsymbol{x})} \\
=& (1/2)^{\sum_\Psi x_{ij}+x_{ji}} \left[\frac{\prod_\Psi \Gamma(\alpha_{ij})\Gamma(\alpha_{ji})}{\prod_\Psi \Gamma(\alpha_{ij}+\alpha_{ji})} \frac{\prod_\Psi \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}{\prod_\Psi \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})} \right].
\end{aligned} \tag{A-73}
$$

A.5.2 Diagonal Symmetry for $r \times r$ case

For the general case the hypothesis of symmetry (A-64) has the same form. Then, proceeding as in the previous section, we have

$$
\begin{array}{ccc}
\theta_{12} = \theta_{21} & \Leftrightarrow & \eta_{12}\lambda_{12} = \eta_{12}(1-\lambda_{12}) \\
\vdots & \vdots & \vdots \\
\theta_{1r} = \theta_{r1} & \Leftrightarrow & \eta_{1r}\lambda_{1r} = \eta_{1r}(1-\lambda_{1r}) \\
\theta_{23} = \theta_{32} & \Leftrightarrow & \eta_{23}\lambda_{23} = \eta_{23}(1-\lambda_{23}) \\
\vdots & \vdots & \vdots \\
\theta_{2r} = \theta_{r2} & \Leftrightarrow & \eta_{2r}\lambda_{2r} = \eta_{2r}(1-\lambda_{2r}) \\
\vdots & \vdots & \vdots \\
\theta_{(r-1)r} = \theta_{r(r-1)} & \Leftrightarrow & \eta_{(r-1)r}\lambda_{(r-1)r} = \eta_{(r-1)r}(1-\lambda_{(r-1)r}).
\end{array}
\tag{A-74}
$$

Consequently,

$$
\begin{array}{ccc}
\theta_{12} = \theta_{21} & \Leftrightarrow & \lambda_{12} = 1/2 \\
\vdots & \vdots & \vdots \\
\theta_{1r} = \theta_{r1} & \Leftrightarrow & \lambda_{1r} = 1/2 \\
\theta_{23} = \theta_{32} & \Leftrightarrow & \lambda_{23} = 1/2 \\
\vdots & \vdots & \vdots \\
\theta_{2r} = \theta_{r2} & \Leftrightarrow & \lambda_{2r} = 1/2 \\
\vdots & \vdots & \vdots \\
\theta_{(r-1)r} = \theta_{r(r-1)} & \Leftrightarrow & \lambda_{(r-1)r} = 1/2.
\end{array}
\tag{A-75}
$$

Hence, the hypotheses (A-64) can be generalized for the $r \times r$ case as

$$
\begin{aligned}
\tilde{H}: &\quad \boldsymbol{\lambda} \in \Lambda_0 \\
\tilde{A}: &\quad \boldsymbol{\lambda} \in \Lambda_0^c,
\end{aligned}
\tag{A-76}
$$

where,

$$
\Psi^* = \{i, j : i < j \ \text{ for } \ i, j = 1, 2, \ldots, r.\},
\tag{A-77}
$$

then, the space Λ_0 is $\Lambda_0 = B^* \times \Lambda^*$, where $B^* = \{\lambda_{ij} : \lambda_{ij} = 1/2, \ \forall\,(i,j) \in \Psi^*\}$ and $\Lambda^* = \{\forall \eta_{ij} \in (0,1)^{(r^2+r-4)/2} : \sum \eta_{ij} \leq 1\}$. Thus, the new likelihood function for $\boldsymbol{\lambda}$ generated by $x = (x_{11}, \ldots, x_{33})$ can be generalized as

$$
L(\boldsymbol{\lambda}|\boldsymbol{x}) = \frac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \left[\prod_{\Psi^*} \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}} \right] \left[\prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}} \right],
\tag{A-78}
$$

and the prior distribution is generalized as

$$
\pi(\boldsymbol{\lambda}) = \left[\prod_{\Psi^*} \frac{\Gamma(\alpha_{ij}+\alpha_{ji})}{\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})} \lambda_{ij}^{\alpha_{ij}-1} (1-\lambda_{ij})^{\alpha_{ji}-1} \right] \left[\frac{\Gamma(\sum_{i,j=1}^{r} \alpha_{ij})}{\prod_{i=1}^{r} \Gamma(\alpha_{ii}) \prod_{\Psi} \Gamma(\alpha_{ij}+\alpha_{ji})} \prod_{\Psi^*} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{r} \eta_{ii}^{\alpha_{ii}-1} \right],
\tag{A-79}
$$

Hence, the predictive function under the null hypothesis H is defined by

$$
\begin{aligned}
f_{\tilde{H}}(x) =& \int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
=& \tfrac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \int_{\Lambda} \left[\textstyle\prod_{\Psi^*} \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}}\right] \left[\textstyle\prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}}\right] d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\
=& \frac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \left[\frac{\int_{B^*} \left[\prod_{\Psi^*} \lambda_{ij}^{x_{ij}+\alpha_{ij}-1}(1-\lambda_{ij})^{x_{ji}+\alpha_{ji}-1}\right] d\boldsymbol{\lambda} \int_{\Lambda^*} \left[\prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-2} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}+\alpha_{ii}-1}\right] |J| d\boldsymbol{\eta}}{\int_{B^*} \left[\prod_{\Psi^*} \lambda_{ij}^{\alpha_{ij}-1}(1-\lambda_{ij})^{\alpha_{ji}-1}\right] d\boldsymbol{\lambda} \int_{\Lambda^*} \left[\prod_{\Psi^*} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-2} \prod_{i=1}^{r} \eta_{ii}^{\alpha_{ii}-1}\right] |J| d\boldsymbol{\eta}}\right],
\end{aligned}
\tag{A-80}
$$

where $|J|$ represents the jacobian of the transformation, it is given by:

$$
|J| = \prod_{\Psi^*} \eta_{ij}.
$$

Thus, the predictive function under null hypothesis of (A-76) is given by

$$
\begin{aligned}
f_{\tilde{H}}(\boldsymbol{x}) =& \frac{n! \times (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}}}{\prod_{i,j=1}^{r} x_{ij}!} \left[\frac{\displaystyle\int_{\Lambda^*} \prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}+\alpha_{ii}-1} d\boldsymbol{\eta}}{\displaystyle\int_{\Lambda^*} \prod_{\Psi^*} \eta_{ij}^{\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{r} \eta_{ii}^{\alpha_{ii}-1} d\boldsymbol{\eta}}\right] \\
=& \left[\frac{\begin{array}{r} n! \times (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}} \prod_{\Psi^*} \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^{r} \Gamma(x_{ii}+\alpha_{ii}) \\ \times \Gamma\left(\sum_{\Psi^*} \alpha_{ij}+\alpha_{ji}+\sum_{i=1}^{r} \alpha_{ii}\right) \end{array}}{\begin{array}{r} \prod_{i,j=1}^{r} x_{ij}! \Gamma\left(\sum_{\Psi^*} x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}+\sum_{i=1}^{r} x_{ii}+\alpha_{ii}\right) \\ \times \prod_{\Psi^*} \Gamma(\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^{r} \Gamma(\alpha_{ii}) \end{array}}\right].
\end{aligned}
\tag{A-81}
$$

Now, the predictive function under the alternative hypothesis for the $r \times r$ case is:

$$
\begin{aligned}
f_{\tilde{A}}(x) =& \int_{\Lambda} L(\boldsymbol{\lambda}|\boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
=& \tfrac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \int_{\Lambda} \left[\textstyle\prod_{\Psi^*} \lambda_{ij}^{x_{ij}} (1-\lambda_{ij})^{x_{ji}}\right] \left[\textstyle\prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}}\right] d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\
=& C_4 \times \left[\int_{B^{*c}} \textstyle\prod_{\Psi^*} \lambda_{ij}^{x_{ij}+\alpha_{ij}-1}(1-\lambda_{ij})^{x_{ji}+\alpha_{ji}-1} d\boldsymbol{\lambda} \int_{\Lambda^*} \prod_{\Psi^*} \eta_{ij}^{x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}-1} \prod_{i=1}^{r} \eta_{ii}^{x_{ii}+\alpha_{ii}-1} d\boldsymbol{\eta}\right] \\
=& C_4 \times \left[\frac{\prod_{\Psi^*} \Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_{\Psi^*} \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})} \frac{\prod_{\Psi^*} \Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}) \prod_{i=1}^{r} \Gamma(x_{ii}+\alpha_{ii})}{\Gamma\left(\sum_{\Psi^*} x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji}+\sum_{i=1}^{r} x_{ii}+\alpha_{ii}\right)}\right],
\end{aligned}
\tag{A-82}
$$

with $C_4 = \frac{n!}{\prod_{i,j=1}^{r} x_{ij}!} \frac{\Gamma(\sum_{i,j=1}^{r} \alpha_{ij})}{\prod_{i,j=1}^{r} \Gamma(\alpha_{ij})}$. As in the 3×3 case, the likelihood function (A-78) comes factorized as (2-31) and the prior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ can be written as a product of distributions, in this case, Beta distributions. Then, by Theorem 4, the predictive functions can be calculated without considering $\boldsymbol{\eta}$, that is

$$
f_{\tilde{H}}(\boldsymbol{x}|\, T = x_{\Psi^*}) = (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}} \prod_{\Psi^*} \binom{x_{ij}+x_{ji}}{x_{ij}},
\tag{A-83}
$$

and

$$f_{\tilde{A}}(\boldsymbol{x}|T = x_{\Psi^*}) = \left[\prod_{\Psi^*}\binom{x_{ij}+x_{ji}}{x_{ij}}\frac{\prod_{\Psi^*}\Gamma(\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi^*}\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})}\frac{\prod_{\Psi^*}\Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}{\prod_{\Psi^*}\Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}\right]. \tag{A-84}$$

where $x_{\Psi^*} = \sum_{\Psi^*} x_{ij} + x_{ji}$. As a result, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed

$$\begin{aligned} Bf(\boldsymbol{x}) =& \frac{f_{\tilde{H}}(\boldsymbol{x})}{f_{\tilde{A}}(\boldsymbol{x})} \\ =& (1/2)^{\sum_{\Psi^*} x_{ij}+x_{ji}}\left[\frac{\prod_{\Psi^*}\Gamma(\alpha_{ij})\Gamma(\alpha_{ji})}{\prod_{\Psi^*}\Gamma(\alpha_{ij}+\alpha_{ji})}\frac{\prod_{\Psi^*}\Gamma(x_{ij}+x_{ji}+\alpha_{ij}+\alpha_{ji})}{\prod_{\Psi^*}\Gamma(x_{ij}+\alpha_{ij})\Gamma(x_{ji}+\alpha_{ji})}\right]. \end{aligned} \tag{A-85}$$

A.6 Poisson means comparison

Let $\mathbf{X} = (X_1, X_2)$ be the observations following a Poisson distribution with parameters $m\theta_1$ and $n\theta_2$ respectively, with $m, n \in \mathbb{N}$. Then, the likelihood function is given by

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{(m\theta_1)^{x_1}}{x_1!}\frac{(n\theta_2)^{x_2}}{x_2!}e^{-\theta_1 m}e^{-\theta_2 n}. \tag{A-86}$$

Assuming that $\boldsymbol{\theta}$ has prior distribution as

$$\pi(\boldsymbol{\theta}) = \frac{b^a}{\Gamma(a)}\frac{d^c}{\Gamma(c)}\theta_1^{a-1}\theta_2^{c-1}e^{-\theta_1 b}e^{-\theta_2 d}, \tag{A-87}$$

that is, $\theta_1 \in \mathbb{R}_+$ and $\theta_2 \in \mathbb{R}_+$ are distributed as a Gamma distribution with parameters $(a, b) \in \mathbb{R}_+$ and $(c, d) \in \mathbb{R}_+$ respectively and $\theta_1 \perp\!\!\!\perp \theta_2$. Then, the predictive function under the null hypothesis H of (3-56) can be expressed by

$$\begin{aligned} f_{\tilde{H}}(x) =& \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_{\tilde{H}}(\boldsymbol{\Theta}) \\ =& \frac{\oint_\Theta L(\boldsymbol{\theta}|\,\boldsymbol{x})\pi(\boldsymbol{\theta})d\mathbb{P}_H(\theta)}{\oint_\Theta \pi(\boldsymbol{\theta})d\mathbb{P}_H(\theta)} \\ =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\frac{\int_0^\infty t^{(x_1+x_2+a+c-1)-1}e^{-t(b+d+m+n)}\sqrt{\Delta}\,dt}{\int_0^\infty t^{(a+c-1)-1}e^{-t(b+d)}\sqrt{\Delta}\,dt}\right], \end{aligned} \tag{A-88}$$

where $\Delta = \left(\frac{\partial\theta_1}{\partial t}\right)^2 + \left(\frac{\partial\theta_2}{\partial t}\right)^2$. Hence, the arch $\sqrt{\Delta}$ that represents the null hypothesis H is given by the equations $\theta_1 = \theta_2 = t$, then

$$\begin{aligned} \Delta =& \left(\frac{\partial\theta_1}{\partial t}\right)^2 + \left(\frac{\partial\theta_2}{\partial t}\right)^2 \\ =& 2, \end{aligned} \tag{A-89}$$

consequently,

$$\begin{aligned} f_{\tilde{H}}(x) =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\frac{\displaystyle\int_0^{\infty} t^{(x_1+x_2+a+c-1)-1}e^{-t(b+d+m+n)}\,dt}{\displaystyle\int_0^{\infty} t^{(a+c-1)-1}e^{-t(b+d)}\,dt}\right] \\ =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\frac{\Gamma(x_1+x_2+a+c-1)(b+d)^{(a+c-1)}}{(b+d+m+n)^{(x_1+x_2+a+c-1)}\Gamma(a+c-1)}\right]. \end{aligned} \tag{A-90}$$

Analogously, the predictive function under the alternative hypothesis A is given by

$$\begin{aligned} f_A(\boldsymbol{x}) =& \int_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_A(\boldsymbol{\theta}) \\ =& \frac{b^a}{\Gamma(a)}\frac{d^c}{\Gamma(c)}\frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\int_0^{\infty}\theta_1^{x_1+a-1}e^{-\theta_1(m+b)}d\theta_1\int_0^{\infty}\theta_2^{x_2+c-1}e^{-\theta_2(n+d)}d\theta_2\right] \\ =& \frac{b^a}{\Gamma(a)}\frac{d^c}{\Gamma(c)}\frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\frac{\Gamma(x_1+a)}{(m+b)^{(x_1+a)}}\frac{\Gamma(x_2+c)}{(n+d)^{(x_2+c)}}\right]. \end{aligned} \tag{A-91}$$

Note that in this case there not exists line integral because the alternative hypothesis A of (3-56) only excludes the line where $\theta_1 = \theta_2$. Hence, Bayes factor $Bf(\boldsymbol{x})$ is given by

$$Bf(\boldsymbol{x}) = \frac{\Gamma(a)\Gamma(c)}{b^a d^c}\left[\frac{\Gamma(x_1+x_2+a+c-1)(b+d)^{(a+c-1)}}{(b+d+m+n)^{(x_1+x_2+a+c-1)}\Gamma(a+c-1)}\frac{(m+b)^{(x_1+a)}(n+d)^{(x_2+c)}}{\Gamma(x_1+a)\Gamma(x_2+c)}\right]. \tag{A-92}$$

Next, we shall show that assuming equals sample sizes and equals rate parameters, i.e.,

$$n = m \qquad \text{and} \qquad b = d, \tag{A-93}$$

and by using a suitable reparameterization in order to obtain a factored likelihood, the P-P test procedure meets the condition of Theorem 4, and consequently, this problem will be reduced to a “simple” sharp hypothesis testing problem. Thus, the new variables can be written

$$\lambda_1 = \frac{\theta_1}{\theta_1+\theta_2} \quad \text{and} \quad \lambda_2 = \theta_1+\theta_2, \tag{A-94}$$

where, the new parametric space is $\Lambda = [0,1]\times\mathbb{R}_+$. Now, let the hypotheses (3-56) be rewritten as

$$\begin{aligned} \tilde{H} &: \lambda_1 \in \Lambda_0 \\ \tilde{A} &: \lambda_1 \in \Lambda_0^c, \end{aligned} \tag{A-95}$$

with $\Lambda_0 = B\times\Lambda^*$, where $B = \{\lambda_1 : \lambda_1 = \{1/2\}\}$ and $\Lambda^* = (0,\infty)$. Note that, since X_1 and X_2 are independent the likelihood function (3-57) can be expressed as

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2\,|\,\boldsymbol{\theta}) =& P(X_1 = x_1\,|\,X_1+X_2 = x_1+x_2,\boldsymbol{\theta})P(X_1+X_2 = x_1+x_2\,|\,\boldsymbol{\theta}) \\ =& n^{x_1+x_2}\binom{x_1+x_2}{x_1}\left(\frac{\theta_1}{\theta_1+\theta_2}\right)^{x_1}\left(\frac{\theta_2}{\theta_1+\theta_2}\right)^{x_2}\frac{e^{-n(\theta_1+\theta_2)}(\theta_1+\theta_2)^{x_1+x_2}}{(x_1+x_2)!}. \end{aligned} \tag{A-96}$$

Hence, the new likelihood function for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ generated by $\boldsymbol{x} = (x_1, x_2)$ can be written as

$$L(\boldsymbol{\lambda} \mid \boldsymbol{x}) = n^{x_1+x_2} \binom{x_1+x_2}{x_1} \lambda_1^{x_1} (1-\lambda_1)^{x_2} \frac{e^{-\lambda_2}\lambda_2^{x_1+x_2}}{(x_1+x_2)!}. \tag{A-97}$$

and the prior distribution is given by

$$\pi(\boldsymbol{\lambda}) = \left[\frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)}\lambda_1^{a-1}(1-\lambda_1)^{c-a}\right]\left[\frac{b^a d^c}{\Gamma(a+c)}\lambda_2^{a+c-2}e^{-\lambda_2\lambda_1(b-d)-\lambda_2 d}\right]. \tag{A-98}$$

Thus, the predictive function under the null hypothesis $\tilde{H}$ can be expressed by

$$\begin{aligned} f_{\tilde{H}}(\boldsymbol{x}) =& \int_\Lambda L(\boldsymbol{\lambda} \mid \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\int_\Lambda \lambda_1^{x_1}(1-\lambda_1)^{x_2} e^{-\lambda_2}\lambda_2^{x_1+x_2} d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\left[\frac{(1/2)^{x_1+x_2}\displaystyle\int_0^\infty \lambda_2^{x_1+x_2+a+c-1}e^{-\lambda_2\frac{(m+n+b+d)}{2}}d\lambda_2}{\displaystyle\int_0^\infty \lambda_2^{a+c-1}e^{-\lambda_2\frac{(b+d)}{2}}d\lambda_2}\right], \end{aligned} \tag{A-99}$$

and the predictive function under the alternative hypothesis $\tilde{A}$ as:

$$\begin{aligned} f_{\tilde{A}}(\boldsymbol{x}) =& \int_\Lambda L(\boldsymbol{\lambda} \mid \boldsymbol{x}) d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\ =& \frac{m^{x_1}n^{x_2}}{x_1!x_2!}\int_\Lambda \lambda_1^{x_1}(1-\lambda_1)^{x_2} e^{-\lambda_2}\lambda_2^{x_1+x_2} d\mathbb{P}_{\tilde{A}}(\boldsymbol{\lambda}) \\ =& C_5 \times \left[\int_0^1\int_0^\infty \lambda_1^{x_1+a-1}(1-\lambda_1)^{x_2+c-1}\lambda_2^{x_1+x_2+a+c-1}e^{-(\lambda_1\lambda_2)(m+b)}e^{-\lambda_2(1-\lambda_1)(n+d)}d\lambda_2\lambda_1\right] \\ =& C_5 \times \left[\int_0^1\int_0^\infty \lambda_1^{x_1+a-1}(1-\lambda_1)^{x_2+c-1}\lambda_2^{x_1+x_2+a+c-1}e^{-\lambda_2(\lambda_1(m+b-n-d)+n+d)}d\lambda_2\lambda_1\right], \end{aligned} \tag{A-100}$$

where $C_5 = \frac{b^a}{\Gamma(a)}\frac{d^c}{\Gamma(c)}\frac{m^{x_1}n^{x_2}}{x_1!x_2!}$. Now, note that assuming $n = m$ and $b = d$ the predictive functions are reduced to

$$\begin{aligned} f_{\tilde{H}}(\boldsymbol{x}) =& \frac{n^{x_1+x_2}}{x_1!x_2!}\left[\frac{(1/2)^{x_1+x_2}\displaystyle\int_0^\infty \lambda_2^{x_1+x_2+a+c-1}e^{-\lambda_2(n+b)}d\lambda_2}{\displaystyle\int_0^\infty \lambda_2^{a+c-1}e^{-\lambda_2 b}d\lambda_2}\right] \\ =& (1/2)^{x_1+x_2}\left[\frac{n^{x_1+x_2}}{x_1!x_2!}\right]\left[\frac{\Gamma(x_1+x_2+a+c)\,b^{a+c}}{\Gamma(a+c)(n+b)^{x_1+x_2+a+c}}\right], \end{aligned} \tag{A-101}$$

and

$$\begin{aligned} f_{\tilde{A}}(\boldsymbol{x}) =& C_5 \times \left[\int_0^1\int_0^\infty \lambda_1^{x_1+a-1}(1-\lambda_1)^{x_2+c-1}\lambda_2^{x_1+x_2+a+c-1}e^{-\lambda_2(n+b)}d\lambda_2\lambda_1\right] \\ =& C_5 \times \left[\int_0^1 \lambda_1^{x_1+a-1}(1-\lambda_1)^{x_2+c-1}d\lambda_1\int_0^\infty \lambda_2^{x_1+x_2+a+c-1}e^{-\lambda_2(n+b)}d\lambda_2\right], \\ =& \frac{n^{x_1+x_2}}{x_1!x_2!}\left[\frac{\Gamma(x_1+a)\Gamma(x_2+c)}{(n+b)^{(x_1+x_2+a+c)}}\right]\left[\frac{b^{a+c}}{\Gamma(a)\Gamma(c)}\right]. \end{aligned} \tag{A-102}$$

As in the previous cases, the likelihood function (A-97) comes factorized as (2-31) and the prior distribution of $\boldsymbol{\lambda}$ can be written as a product of distributions, in this case, of Gamma distributions. Then, by Theorem 4, the predictive functions can be calculated unregarding λ_2, in this way

$$f_{\tilde{H}}(\boldsymbol{x}|T = x_1 + x_2) = (1/2)^{x_1+x_2}\binom{x_1 + x_2}{x_1}, \tag{A-103}$$

and

$$f_{\tilde{A}}(\boldsymbol{x}|T = x_1 + x_2) = \binom{x_1 + x_2}{x_1}\left[\frac{\Gamma(x_1 + a)\Gamma(x_2 + c)}{\Gamma(x_1 + x_2 + a + c)}\right]\left[\frac{\Gamma(a + c)}{\Gamma(a)\Gamma(c)}\right]. \tag{A-104}$$

Consequently, the Bayes factor $Bf(\boldsymbol{x})$ can be expressed as

$$Bf(\boldsymbol{x}) = (1/2)^{x_1+x_2}\frac{\Gamma(a)\Gamma(c)}{\Gamma(a + c)}\frac{\Gamma(x_1 + x_2 + a + c)}{\Gamma(x_1 + a)\Gamma(x_2 + c)}. \tag{A-105}$$

A.7 Hardy–Weinberg Equilibrium

For testing (3-78), we consider that the sample size n is known and sample elements of genotypes frequencies are obtained independently in such a way that the vector $\boldsymbol{X} = (X_1, X_2, X_3)$ follows a Multinomial distribution with parameters n and $\boldsymbol{\theta}$. Thus, the Likelihood function for $\boldsymbol{\theta}$ generated by $\boldsymbol{x} = (x_1, x_2, x_3)$ is defined by

$$L(\boldsymbol{\theta}\,|\,\boldsymbol{x}) = \frac{n!}{\prod_{i=1}^{3} x_i!}\prod_{i=1}^{3}\theta_i^{x_i} \tag{A-106}$$

In addition, we assume that the parameter vector $\boldsymbol{\theta}$ follows a Dirichlet distribution, with parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ for $\alpha_i > 0$, that is

$$\pi(\boldsymbol{\theta}) = \frac{\Gamma(\sum_{i=1}^{3}\alpha_i)}{\prod_{i=1}^{3}\Gamma(\alpha_i)}\prod_{i=1}^{3}\theta_i^{\alpha_i - 1} \tag{A-107}$$

Then, the predictive function under the null hypothesis of (3-78) is given by

$$\begin{aligned}
f_H(x) &= \int_\Theta L(\boldsymbol{\theta}|\boldsymbol{x})d\mathbb{P}_{\tilde{H}}(\boldsymbol{\theta}) \\
&= \frac{n!}{\prod_{i=1}^{3} x_i!}\int_\Theta \prod_{i=1}^{3}\theta_i^{x_i}d\mathbb{P}_{\tilde{H}}(\boldsymbol{\theta}) \\
&= \frac{\oint_\Theta L(\boldsymbol{\theta}|\,\boldsymbol{x})\pi(\boldsymbol{\theta})|J|d\mathbb{P}_H(\theta)}{\oint_\Theta \pi(\boldsymbol{\theta})|J|d\mathbb{P}_H(\theta)} \\
&= 2^{x_2}\frac{n!}{\prod_{i=1}^{3} x_i!}\frac{\int_0^1\left(\sqrt{1 - 3p(1-p)}\right)p^{2A_1 + A_2 - 2}(1-p)^{2A_3 + A_2 - 2}dp}{\int_0^1\left(\sqrt{1 - 3p(1-p)}\right)p^{2\alpha_1 + \alpha_2 - 2}(1-p)^{2\alpha_3 + \alpha_2 - 2}dp},
\end{aligned} \tag{A-108}$$

where $A_i = x_i + \alpha_i$ for $i = 1, 2, 3$. Note that the exact calculation of (A-108) is not feasible. Thereby, as in the previous cases, we are going to apply a suitable reparameterization that makes the task of obtaining (A-108) easier. For this, consider the new variables

$$\lambda_1 = \frac{\sqrt{\theta_1}}{\sqrt{\theta_1} + \sqrt{\theta_3}}; \qquad \lambda_2 = \sqrt{\theta_1} + \sqrt{\theta_3}. \tag{A-109}$$

Then, the new parametric space Λ is given by $\Lambda = [0, 1] \times \left[0, (\lambda_1^2 + (1 - \lambda_1)^2)^{-1/2}\right]$ and the new hypotheses can be expressed as

$$\begin{aligned} \tilde{H} &: \boldsymbol{\lambda} \in \Lambda_0 \\ \tilde{A} &: \boldsymbol{\lambda} \in \Lambda_0^c, \end{aligned} \tag{A-110}$$

with $\Lambda_0 = B \times \{1\}$ and $B = [0, 1]$. Then, the new predictive function under the null hypothesis $\tilde{H}$ is given by

$$\begin{aligned} f_{\tilde{H}}(x) &= \int_\Lambda L(\boldsymbol{\lambda} \,|\, \boldsymbol{x}) d\mathbb{P}_{\tilde{H}}(\boldsymbol{\lambda}) \\ &= \frac{\oint_\Lambda L(\boldsymbol{\lambda}|\boldsymbol{x})\pi(\boldsymbol{\lambda})|J| d\mathbb{P}_H(\lambda)}{\oint_\Lambda \pi(\boldsymbol{\lambda})|J| d\mathbb{P}_H(\lambda)} \\ &= \frac{n!}{\prod_{i=1}^3 x_i!} \frac{\int_0^1 t^{2(A_1-1)+1}(1-t^2-(1-t)^2)^{A_2-1}(1-t)^{2(A_3-1)+1}\sqrt{\Delta}\, dt}{\int_0^1 t^{2(\alpha_1-1)+1}(1-t^2-(1-t)^2)^{\alpha_2-1}(1-t)^{2(\alpha_3-1)+1}\sqrt{\Delta}\, dt} \\ &= 2^{x_2} \frac{n!}{\prod_{i=1}^3 x_i!} \frac{\int_0^1 t^{2(A_1-1)+A_2}(1-t)^{2(A_3-1)+A_2}\sqrt{\Delta}\, dt}{\int_0^1 t^{2(\alpha_1-1)+\alpha_2}(1-t)^{2(\alpha_3-1)+\alpha_2}\sqrt{\Delta}\, dt}, \end{aligned} \tag{A-111}$$

where $|J|$ is the jacobian of the transformation (A-109) given by $|J| = 4\lambda_1(1 - \lambda_1)\lambda_2^3$ and $\Delta = \left(\frac{\partial\lambda_1}{\partial t}\right)^2 + \left(\frac{\partial\lambda_2}{\partial t}\right)^2$. Hence, the arch $\sqrt{\Delta}$ that represent the null hypothesis H is given by the equations $\lambda_1 = t$ and $\lambda_2 = 1$, then,

$$\begin{aligned} \Delta &= \left(\frac{\partial\lambda_1}{\partial t}\right)^2 + \left(\frac{\partial\lambda_2}{\partial t}\right)^2 \\ &= (1)^2 + (0)^2 . \end{aligned} \tag{A-112}$$

Consequently, $f_{\tilde{H}}(\boldsymbol{x})$ can be computed as

$$f_{\tilde{H}}(x) = 2^{x_2} \frac{n!}{\prod_{i=1}^3 x_i!} \left[\frac{\Gamma(2A_1 + A_2 - 1)\Gamma(2A_3 + A_2 - 1)\Gamma(2\sum_{i=1}^3 \alpha_i - 2)}{\Gamma(2\alpha_1 + \alpha_2 - 1)\Gamma(2\alpha_3 + \alpha_2 - 1)\Gamma(2\sum_{i=1}^3 A_i - 2)}\right]. \tag{A-113}$$

Now, the predictive function under the alternative hypothesis H is calculated in its original parameterization; it is given by

$$
\begin{aligned}
f_A(x) &= \int_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{x}) d\mathbb{P}_A(\boldsymbol{\theta}) \\
&= \left[\frac{\Gamma(\sum_{i=1}^{3}\alpha_i)}{\prod_{i=1}^{3}\Gamma(\alpha_i)} \frac{n!}{\prod_{i=1}^{3} x_i!}\right] \int_{(0,1)^2} \theta_1^{A_1-1}\theta_2^{A_2-1}\theta_3^{A_3-1} d\boldsymbol{\theta} \\
&= \left[\frac{\Gamma(\sum_{i=1}^{3}\alpha_i)}{\prod_{i=1}^{3}\Gamma(\alpha_i)} \frac{n!}{\prod_{i=1}^{3} x_i!}\right] \frac{\prod_{i=1}^{3}\Gamma(A_i)}{\Gamma(\sum_{i=1}^{3} A_i)}
\end{aligned} \tag{A-114}
$$

As a result, the Bayes factor $Bf(\boldsymbol{x})$ is given by

$$
Bf(\boldsymbol{x}) = 2^{x_2}\left[\frac{\Gamma(2A_1+A_2-1)\Gamma(2A_3+A_2-1)\Gamma(2\sum_{i=1}^{3}\alpha_i-2)}{\Gamma(2\alpha_1+\alpha_2-1)\Gamma(2\alpha_3+\alpha_2-1)\Gamma(2\sum_{i=1}^{3}A_i-2)} \frac{\prod_{i=1}^{3}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{3}\alpha_i)} \frac{\Gamma(\sum_{i=1}^{3}A_i)}{\prod_{i=1}^{3}\Gamma(A_i)}\right]. \tag{A-115}
$$

Bibliography

Agresti (2007) A Agresti. *An Introduction to Categorical Data Analysis.* Wiley. From page 32

Agresti (2013) A Agresti. *Categorical data analysis. 3rd.* From page 4, 33, 37, 44

Agresti and David (2005) A Agresti and H David. Bayesian inference for categorical data analysis. *Statistical Methods and Applications*, 14(3):297–330. From page 59

Andersen (2012) E.B. Andersen. *The Statistical Analysis of Categorical Data.* Springer Berlin Heidelberg. From page 32

Barber and Ogle (2014) Jarrett J. Barber and Kiona Ogle. To p or not to p? *Ecology*, 95(3):621–626. From page 2

Barndorff-Nielsen (1976) O Barndorff-Nielsen. Nonformation. *Biometrika*, 63(3):567–571. From page 17

Barndorff-Nielsen (1978) Ole Barndorff-Nielsen. *Information and exponential families: in statistical theory.* Chichester: Wiley. From page 17

Barry (1996) J Barry. *Probabilidade: um curso em nível intermediário.* Number 519.2. From page 39

Basu (1977) D Basu. On the elimination of nuisance parameters. *Journal of the American Statistical Association*, 72(358):355–366. From page 17, 18

Berger (1985) J Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer. From page 12

Berger and Delampady (1987) J Berger and M Delampady. Testing precise hypotheses. *Statistical Science*, 2(3):317–335. From page 1, 2, 3

Berger and Wolper (1988) J Berger and R Wolper. *The Likelihood Principle.* Institute of Mathematical Statistics, Hayward, California. From page 11, 16, 17, 18

Bernardo *et al.* (2012) G Bernardo, M Lauretto and J Stern. The full bayesian significance test for symmetry in contingency tables. *AIP Conference Proceedings*, 1443(1):198–205. From page 44

Bowker (1948) A Bowker. A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43(244):572–574. From page 44

Cenipa (2019) Cenipa. Ocorrências aeronáuticas na aviação civil brasileira. URL http://dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira. From page 56

Cornfield (1966) J Cornfield. Sequential trials, sequential analysis and the likelihood principle. *The American Statistician*, 20(2):18–23. From page 2, 8

Council *et al.* **(1996)** National Research Council *et al.* *The evaluation of forensic DNA evidence*. National Academies Press. From page 52

Courant and Fritz (1974) R Courant and J Fritz. *Introduction to calculus and analysis*, volume 2. Springer Science & Business Media. From page 39

DeGroot (1986) M DeGroot. *Probability and Statistics*. Adson Wesley: Boston, MA, USA. From page 3, 8

Demidenko (2016) E Demidenko. The p-value you can't buy. *The American Statistician*, 70(1):33–38. From page 1

Dickey and Lientz (1970) J Dickey and B Lientz. The weighted likelihood ratio, sharp hypotheses, about chances, the order of a markov chain. *The annals of Mathematical Statistics*, 41(1):214–226. From page 2

Eaton (1989) M Eaton. Group invariance applications in statistics. *Regional conference series in Probability and Statistics*, pages 133. From page 12

Etz and Wagenmakers (2017) A Etz and E Wagenmakers. J. b. s. haldane's contribution to the bayes factor hypothesis test. *Institute of Mathematical Statistics*, 32(2):313–329. From page 2, 8

Fisher (1935) R Fisher. The fiducial argument in statistical inference. *Annals of eugenics*, 6(4):391–398. From page 12

Flórez (2015) A Flórez. Elicitación de un a distribución subjetiva del vector de parámetros π de la distribución multinomial. *Master dissertation: Universidad Nacional de Colombia*. From page 3

Fraser (1968) D Fraser. *The structure of inference*. Wiley. From page 12

Fraser (1979) D.A.S. Fraser. *Inference and linear models*. McGraw-Hill Companies. From page 12

Gabriel (1969) K. R. Gabriel. Simultaneous test procedures–some theory of multiple comparisons. *The Annals of Mathematical Statistics*, 40. From page 2

Gannon *et al.* **(2019)** M Gannon, C Pereira and A Polpo. Blending bayesian and classical tools to define optimal sample-size-dependent significance levels. *The American Statistician*, 73(sup1):213–222. From page 11

Gill (1999) J Gill. The insignificance of null hypothesis significance testing. *Political Research Quarterly*, 52(3):647–674. From page 3

Good (1950) I Good. *Probability and the weighting of evidence.* Hafner, New York. From page 2

Good (1965) I Good. *The estimation of probabilities,.* MIT Press, Cambridge. From page 2

Graffelman (2019) J Graffelman. Exploring diallelic genetic markers: The hardyweinberg package. *https://cran.r-project.org/web/packages/HardyWeinberg/vignettes/HardyWeinberg.pdf*, pages 29. From page 51, 52

Graffelman (2015) J Graffelman. Exploring diallelelic genetic markers: the hardyweinberg package. *Journal of Statistical Software*, 64(3):1–23. From page 55

Graziadei (2015) H Graziadei. Testes bayesianos para homogeneidade marginal em tabelas de contingência. *Dissertação de mestrado, IME, USP.* From page 4, 33

Hacking (1965) I. Hacking. *Logic of Statistical Inference.* Cambridge University Press. From page 1, 7

Head *et al.* **(2015)** M Head, L Holman, R Lanfear, A Kahn and MJennions. The extent and consequences of p-hacking in science. *PLoS biology*, 13(3). From page 4

Hotelling (1936) H Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4): 321–377. From page 12

Hunt and Stein (1946) G Hunt and C Stein. Most stringent tests of statistical hypotheses. *never published (Take from Lehmann and Romano (2008)).* From page 12

Ireland *et al.* **(1969)** CT Ireland, HH Ku and S Kullback. Symmetry and marginal homogeneity of an r× r contingency table. *Journal of the American Statistical Association*, 64 (328):1323–1341. From page 44

Irony and Pereira (1995) T Irony and C Pereira. Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas IME-USP*, 2(1): 27–46. From page 2, 4, 11, 49, 51, 59

Jeffreys (1998) H. Jeffreys. *The Theory of Probability.* OUP Oxford. From page 2

Jeffreys (1935) H Jeffreys. Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(2):203—-222. From page 2, 6, 7

Jeffreys (1936) H Jeffreys. Further significance tests. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):416–445. From page 2, 7

Johnstone *et al.* **(1986)** D Johnstone, G Barnard and D Lindley. Tests of significance in theory and practice. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 35(5):491–504. From page 1, 2

Jørgensen (1994) B Jørgensen. The rules of conditional inference: Is there a universal definition of nonformation? *Journal of the Italian Statistical Society*, 3(3):355. From page 17

Kass and Raftery (1995) R Kass and A Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795. From page 2, 3, 8

Klein and Linton (2013) M Klein and P Linton. On a comparison of tests of homogeneity of binomial proportions. *Journal of Statistical Theory and Applications*, 12(3):208–224. From page 4, 33

Kullback (1971) S Kullback. Marginal homogeneity of multidimensional contingency tables. *The Annals of Mathematical Statistics*, 42(2):594–606. From page 37, 44

Lavine and Schervish (1999) M Lavine and M Schervish. Bayes factors: What they are and what they are not. *The American Statistician*, 53(2):119–122. From page 8

Lehmann (2012) E Lehmann. Some principles of the theory of testing hypotheses. *Springer*, pages 139–164. From page 12

Lehmann and Romano (2008) E Lehmann and J Romano. *Testing statistical hypotheses*. Springer Science & Business Media. From page 12

Lehmann and Romano (2006) E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer New York. From page 3, 8

Lindley (1957) D Lindley. A statistical paradox. *Biometrika*, 44:187–192. From page 1, 4, 11, 33

Lindley (1968) D Lindley. The choice of variables in multiple regression. *The Royal Statistical Society*, B(30):31–66. From page 2

Lindley and Smith (1972) D Lindley and A Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1):1–18. From page 59

Lindley *et al.* **(1979)** D Lindley, A Tversky and R Brown. On the reconciliation of probability assessments. *Journal of the Royal Statistical Society*, 142(2):146–180. From page 7

Menéndez *et al.* **(2005)** M Menéndez, J Pardo, L Pardo and K Zografos. On tests of symmetry, marginal homogeneity and quasi-symmetry in two-way contingency tables based on minimum φ-divergence estimator with constraints. *Journal of Statistical Computation and Simulation*, 75(7):555–580. From page 37

Montoya *et al.* **(2001)** D Montoya, T Irony, C Pereira and M Whittle. An unconditional exact test for the hardy-weinberg equilibrium law: Sample space ordering using the bayes factor. *The Genetics Society of America*, 158:875–883. From page 2, 11, 52, 55

Neyman (1960) J Neyman. *First course in probability and statistics.* Holt, Rinehart and Winston. From page 1

Neyman and Pearson (1957) J Neyman and E Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A:263–294. From page 4, 33, 55

Olivera (2014) M Olivera. Definição do nivel de significancia em função do tamanho amostral. *Dissertação de mestrado, IME, Universidade de São Paulo.* From page 11

Pace and Salvan (1997) L. Pace and A. Salvan. *Principles of Statistical Inference: From a Neo-Fisherian Perspective.* World Scientific Publishing Company Pte Limited. From page 5, 12, 16

Pereira and Pereira (2005) B Pereira and C Pereira. A likelihood approach to diagnostic test in clinical medicine. *Statistical Journal*, 3(1):77–98. From page 4, 11, 59

Pereira (1985) C Pereira. Testing hypotheses of different dimensions: Bayesian view and classical interpretation. *Professor Thesis, Institute Mathematics and Statistics, USP.* From page 4

Pereira and Stern (1999) C Pereira and J Stern. Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110. From page 7, 21

Pereira and Wechsler (1993) C Pereira and S Wechsler. On the concept of p-value. *Revista Brasileira de Probabilidade e Estatística*, 7:159–177. From page 2

Pereira *et al.* **(2017)** C Pereira, E Nakano, V Fossaluza, L Esteves, M Gannon and A Polpo. Hypothesis tests for bernoulli experiments: Ordering the sample space by bayes factors and using adaptive significance levels for decisions. *Entropy*, 19:696. From page vii, 2, 3, 4, 7, 8, 9, 10, 11, 12, 33, 59, 65

Pericchi and Pereira (2016) L Pericchi and C Pereira. Adaptive significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Brazilian Journal of Probability and Statistics*, 30(1):70–90. From page 3, 4, 8, 11, 48, 59

Pitman (1939) EJG Pitman. Tests of hypotheses concerning location and scale parameters. *Biometrika*, 31(1/2):200–215. From page 12

Pérez and Pericchi (2014) M Pérez and L Pericchi. Changing statistical significance with the amount of information: The adaptive α significance level. *Statistics & probability letters*, 85:20–24. From page 3

R Core Team (2017) R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/. From page 55

Rosnow and Rosenthal (1992) R Rosnow and R Rosenthal. Statistical procedures and the justification of knowledge in psychological science. From page 3

Rubin (1984) D Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172. From page 7

Savage (1961) L Savage. The subjective basis of statistical practice. *Unpublished, University of Mich. Notes*. From page 2

Schervish (1996a) M Schervish. P values: What they are and what they are not. *The American Statistician*, 50(3):203–206. From page 1, 2

Schervish (1996b) M Schervish. *Theory of Statistics*. Springer Series in Statistics. From page 5, 6

Shoemaker *et al.* (1998) J Shoemaker, L Painter and B Weir. A bayesian characterization of hardy-weinberg disequilibrium. *Genetics*, 149(4):2079–2088. From page 52

Sulewski (2019) Piotr Sulewski. The lms for testing independence in two-way contingency tables. *Biometrical Letters*, 56(1):17–43. From page 58

Tahata and Tomizawa (2014) K Tahata and S Tomizawa. Symmetry and asymmetry models and decompositions of models for contingency tables. *SUT journal of Mathematics*, 50(2):131–165. From page 44

Thomson *et al.* (2009) G Thomson, H Maldonado, A Lancaster, J Hollenbach, L Barcellos, S Mack and R Single. Hardy weinberg proportions methods manual version 0.1.2. pages 1–9. From page 52

Trafimow and Marks (2015) D Trafimow and M Marks. Editorial. *Basic and Applied Social Psychology*, 37:1–2. From page 1, 3

Upton (1982) G Upton. A comparison of alternative tests for the 2x2 comparative trial. *Journal of the Royal Statistical Society*, 145:86–105. From page 4, 32, 33

Wald (1947) A Wald. *Sequential Analysis*. Wiley, New York. From page 2

Wasserstein and Lazar (2016) R Wasserstein and N Lazar. The asa statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133. From page 3

Wigginton *et al.* **(2005)** J Wigginton, D Cutler and G Abecasis. A note on exact tests of hardy-weinberg equilibrium. *The American Journal of Human Genetics*, 76(5):887–893. From page 55

Wilks (1935) S Wilks. The likelihood test of independence in contingency tables. *The Annals of Mathematical Statistics*, 6:190–196. From page 4, 33, 55

Winkler (1967) R Winkler. The assessment of prior distributions in bayesian analysis. *Journal of the American Statistical Association*, 63(139):776–800. From page 7

Wrinch and Jeffreys (1919) D Wrinch and H Jeffreys. On some aspects of the theory of probability. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 38(228):715–731. From page 2

Wrinch and Jeffreys (1923a) D Wrinch and H Jeffreys. On certain fundamental principles of scientific inquiry. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 45(266):368–374. From page 2

Wrinch and Jeffreys (1923b) D Wrinch and H Jeffreys. On certain fundamental principles of scientific inquiry. (second paper). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 45(266):368–374. From page 2

Yates (1984) Frank Yates. Tests of significance for 2× 2 contingency tables. *Journal of the Royal Statistical Society: Series A (General)*, 147(3):426–449. From page 3