

Article

# On the Nuisance Parameter Elimination Principle in Hypothesis Testing

Andrés Felipe Flórez Rivera <sup>\*</sup>, Luis Gustavo Esteves, Víctor Fossaluzza and Carlos Alberto de Bragança Pereira 

Institute of Mathematics and Statistics, University of São Paulo, São Paulo 05508-090, Brazil; lesteves@ime.usp.br (L.G.E.); victorf@ime.usp.br (V.F.); cpereira@ime.usp.br (C.A.d.B.P.)

<sup>\*</sup> Correspondence: afflorezr@gmail.com

**Abstract:** The Non-Informative Nuisance Parameter Principle concerns the problem of how inferences about a parameter of interest should be made in the presence of nuisance parameters. The principle is examined in the context of the hypothesis testing problem. We prove that the mixed test obeys the principle for discrete sample spaces. We also show how adherence of the mixed test to the principle can make performance of the test much easier. These findings are illustrated with new solutions to well-known problems of testing hypotheses for count data.

**Keywords:**  $p$ -values; Bayes factor; likelihood function; hypothesis testing

## 1. Introduction

Principles of Statistical Inference (or Data Reduction) constitute important guidelines on how to draw conclusions from data, especially when performing standard inferential procedures for unknown parameters of interest, like estimation and hypothesis testing. For instance, the Sufficiency Principle (SP) states that any sufficient statistic retains all relevant information about the unknown parameters that should be used to make inferences about them. It precisely recommends that if  $T$  is a sufficient statistic for the statistical model under consideration and  $x_1$  and  $x_2$  are sample points such that  $T(x_1) = T(x_2)$ , then the observation of any of these points should lead to the same conclusions regarding the parameters of interest.

Besides the place of sufficiency in Statistical Inference, these recommendations cover several issues such as the contrast between post-experimental and pre-experimental reasoning and the roles of non-informative stopping rules, censoring mechanisms and nuisance parameters in data analysis. Among the main principles, the Sufficiency Principle is generally recognized as a cornerstone of Statistical Inference. On the other hand, the Likelihood Principle (LP) and its profound consequences are still subjects of intense debate. The reader will find a detailed discussion of the Likelihood Principle in [1–6].

In this work, we examine the Non-Informative Nuisance Parameter Principle (NNPP) introduced by Berger and Wolpert in 1988 in their remarkable book that concerns the problem of the way inferences about a parameter of interest should be made in the presence of nuisance parameters. Nuisance parameters usually affect inferences about the parameter of interest, like in the estimation of the mean of a normal distribution with unknown variance, in the estimation of the parameters of a linear regression model in the presence of unknown variance, and in the determination of  $p$ -values for specific hypotheses in the analysis of  $2 \times 2$  contingency tables ([7]). In a few words, the NNPP states that under suitable conditions, it is irrelevant whether the value of a non-informative nuisance parameter is known or not in order to draw conclusions about the parameter of interest. Despite the importance of the problem for eliminating nuisance parameters in data analysis, the authors have not explored this principle and its consequences in some depth as far as we have reviewed the literature. For this reason, we revisit the NNPP by formally stating it for the problem



**Citation:** Flórez Rivera, A.F.; Esteves, L.G.; Fossaluzza, V.; de Bragança Pereira, C.A. On the Nuisance Parameter Elimination Principle in Hypothesis Testing. *Entropy* **2024**, *26*, 117. <https://doi.org/10.3390/e26020117>

Academic Editor: Geert Verdoolaege

Received: 1 December 2023

Revised: 23 January 2024

Accepted: 24 January 2024

Published: 29 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of hypothesis testing, present decision rules that meet the principle and show how the performance of a particular test in line with the NNPP can then be simplified.

This work is organized as follows: in Section 2, the NNPP for hypothesis testing is stated, discussed and illustrated under a Bayesian perspective. In Section 3, the Bayesian test procedure based on the concept of adaptive significance level and on an alternative  $p$ -value introduced by Pericchi and Pereira in [8], henceforth named the mixed test, is reviewed and is proven to satisfy the NNPP for discrete sample data when the (marginal) null hypothesis regarding the parameter of interest is a singleton (as a matter of fact, the result also holds when such a null hypothesis is specified by a hyperplane). In that section, we also define conditional versions of the adaptive significance level and  $p$ -value based on suitable statistics and prove that under those conditions, the performance of the mixed test is then simply the comparison between these new conditional quantities. These results are of great importance to make it easier to use the mixed test in various situations. In Section 4, we exemplify the main results by presenting new solutions by using the mixed test for well-known problems of test of hypotheses for count data under suitable reparametrizations of the corresponding models: we revisit the problems of comparison of Poisson population means and of testing the hypotheses of independence and symmetry in contingency tables. We make our final comments in Section 5. The proofs of the theorems and the calculations for one example in Section 4 are found in the Appendix A.

## 2. The Non-Informative Nuisance Parameter Principle for Hypothesis Testing

The problem of the elimination of nuisance parameters in statistical inference has a long history and remains a major issue. Proposals to deal with it include the marginalization of the likelihood function by integrating out the nuisance parameter ([9–11]), the construction of partial likelihood functions ([12–14], among others) and the consideration of conditional likelihood functions based on different notions of non-informativeness, sufficiency and ancillarity. Elimination of nuisance parameters and different notions of non-information have also been studied in more detail in [15–18], where, based on suitable statistics, the concepts of B, S and G non-information are presented. The generalized Sufficiency and Conditionality Principles are also discussed in [17]. On the other hand, Bayesian methods for eliminating nuisance parameters based on a suitable statistic  $T$  involve different definitions of sufficiency: for instance, K-Sufficiency, Q-Sufficiency and L-Sufficiency (see for example [17] and references therein).

In this section, the Non-Informative Nuisance Parameter Principle (NNPP) by Berger and Wolpert is discussed and formally defined for the problem of hypothesis testing. As we will see, on the one hand, the NNPP seems to be fair under both the partial and the conditional non-Bayesian approaches mentioned in the previous paragraph; on the other hand, it sounds really reasonable under the Bayesian standpoint. Despite the relevance of the problem of the elimination of nuisance parameters in data analysis, Berger and Wolpert [1] presented the NNPP but has not explored the principle in-depth as far as we have examined in the literature.

Some notation is needed to continue. We denote by  $\theta$  the unknown parameter and by  $X$  the sample to be observed.  $\Theta$  and  $\mathfrak{X}$  represent the parameter and the sample spaces, respectively. The family of discrete probability distributions for  $X$  is denoted by  $\mathcal{P} = \{\mathbb{P}(\cdot|\theta) : \theta \in \Theta\}$ . In addition, for  $x \in \mathfrak{X}$ ,  $L_x(\cdot)$  denotes the likelihood function for  $\theta$  generated by the sample point  $x$ . By an experiment  $\mathcal{E}$ , we mean, as in [1], a triplet  $\mathcal{E} = (X, \theta, \mathcal{P})$ , with  $X, \theta$  and  $\mathcal{P}$  as defined earlier. Finally, for a subset  $\Theta_0$  of  $\Theta$ , we formulate the null hypothesis  $H : \theta \in \Theta_0$  and the alternative one  $A : \theta \notin \Theta_0$ . We recall that a test function (procedure) for the hypotheses  $H$  versus  $A$  is a function  $\varphi : \mathfrak{X} \rightarrow \{0, 1\}$  that takes the value 1 ( $\varphi(x) = 1$ ) if  $H$  is rejected when  $x \in \mathfrak{X}$  is observed and takes the value 0 ( $\varphi(x) = 0$ ) if  $H$  is not rejected when  $x$  is observed. Under the Bayesian perspective, we also consider a continuous prior density function  $\pi(\cdot)$  for  $\theta$  that induces, when combined with the likelihood function  $L_x(\cdot)$ , a continuous posterior density function for  $\theta$  given  $x$ ,  $\pi(\cdot|x)$ .

In [1], Berger and Wolpert presented the following principle on how to make inferences about an unknown parameter of interest  $\theta_1$  in the presence of a nuisance parameter  $\theta_2$ : when a sample observation, say  $x_0$ , separates information concerning  $\theta_1$  from information on  $\theta_2$ , it is irrelevant whether the value of  $\theta_2$  is known or unknown in order to make inferences about  $\theta_1$  based on the observation of  $x_0$ . In other terms, if the conclusions on  $\theta_1$  were to be the same for every possible value of the nuisance parameter, were  $\theta_2$  known, then the same conclusions on  $\theta_1$  should be reached even if  $\theta_2$  is unknown. These authors then consider the following mathematical setup to formalize these ideas.

Let  $\theta = (\theta_1, \theta_2)$ , with  $\theta_1$  and  $\theta_2$  defined as in the previous paragraph. Consider  $\Theta = \Theta_1 \times \Theta_2$ ; that is, the parameter space is variation independent, where  $\Theta_i \subseteq \mathbb{R}^{n_i}$  is the set of values for  $\theta_i$ ,  $n_i \in \mathbb{N}^*$ ,  $i = 1, 2$ . Suppose the experiment  $\mathcal{E} = (X, \theta, \mathcal{P})$  is carried out to learn about  $\theta$ . Let  $\bar{\mathcal{E}} = ((X, \theta_2), \theta_1, \bar{\mathcal{P}})$  be the “thought” experiment in which the pair  $(X, \theta_2)$  is to be observed (instead of observing only  $X$ ), where  $\bar{\mathcal{P}}$  is the family of distributions for  $(X, \theta_2)$  indexed by  $\theta_1$ . Suppose also that under experiment  $\mathcal{E}$ , the likelihood function generated by a specific  $x_0 \in \mathfrak{X}$  for  $\theta$  has the following factored form:

$$L_{x_0}(\theta_1, \theta_2) = L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2) \tag{1}$$

where  $L_{x_0}^i : \Theta_i \rightarrow \mathbb{R}_+$ ,  $i = 1, 2$ ; that is,  $L_{x_0}^i$  depends on  $\theta$  only through  $\theta_i$ .

Berger and Wolpert then states the **Non-Informative Nuisance Parameter Principle (NNPP)**: if  $\mathcal{E}$  and  $x_0 \in \mathfrak{X}$  are such that (1) holds, and if the inference about  $\theta_1$  from the observation of  $(x_0, \theta_2)$  when  $\bar{\mathcal{E}}$  is performed does not depend on  $\theta_2$ , then the inferential statements made for  $\theta_1$  from  $\mathcal{E}$  and  $x_0$  should be the same as (should coincide with) the inferential statements made from  $\bar{\mathcal{E}}$  and  $(x_0, \theta_2)$  for every  $\theta_2 \in \Theta_2$ .

The authors named such a parameter  $\theta_2$  a **Non-Informative Nuisance Parameter (NNP)**, as the conclusions or decisions regarding  $\theta_1$  from  $\bar{\mathcal{E}}$  and  $(x_0, \theta_2)$  do not depend on  $\theta_2$ .

A likelihood function that satisfies (1) is named a likelihood function with separable parameters ([19]). The factored form of the likelihood function in (1) seems to capture the notion of “absence of information about a parameter, say  $\theta_1$ , from the other,  $\theta_2$ , and vice versa” under both Bayesian and non-Bayesian reasoning. Indeed, under the Bayesian paradigm, posterior independence between  $\theta_1$  and  $\theta_2$  (say, given  $x_0$ ) reflects the fact that one’s opinion about the parameter  $\theta_1$  after observing  $x_0$  is not altered by any information about  $\theta_2$ , and consequently, decisions regarding  $\theta_1$  should not depend on  $\theta_2$ . Since posterior independence between  $\theta_1$  and  $\theta_2$  given  $x_0$  is equivalent to the factored form of the likelihood function generated by  $x_0$  under prior independence, condition (1) sounds really reasonable as a mathematical description of separate information about the parameters. Thus, if a Bayesian statistician should make inferences regarding a parameter  $\theta_1$  in the presence of a nuisance parameter  $\theta_2$ , it would be ideal that these parameters are independent a posteriori; that is, the factored form of the likelihood function holds. This last equivalence is proven in the theorem below.

**Theorem 1.** Let  $\mathcal{E} = (X, \theta, \mathcal{P})$  be an experiment and  $\pi(\cdot)$  be the prior probability density function for  $\theta = (\theta_1, \theta_2)$ . Suppose  $\theta_1$  is independent of  $\theta_2$  ( $\theta_1 \perp\!\!\!\perp \theta_2$ ) a priori. Then, for each  $x \in \mathfrak{X}$ ,

$$\theta_1 \perp\!\!\!\perp \theta_2 \mid X = x \iff \exists L_x^i : \Theta_i \rightarrow \mathbb{R}_+, i = 1, 2, \tag{2}$$

such that  $L_x(\theta_1, \theta_2) = L_x^1(\theta_1)L_x^2(\theta_2)$ .

On the other hand, the condition (1) seems to also be a fair representation of non-informativeness of one parameter on another under a non-Bayesian perspective. In fact, such a factored form of the likelihood function arises, for instance, when the sample  $X$  is

conditioned on particular types of statistics that are simple to interpret under non-Bayesian paradigms. Note that for any statistic  $T$ , one can write

$$L_x(\theta_1, \theta_2) = P(X = x|\theta_1, \theta_2) = P(X = x|T(X) = T(x), \theta_1, \theta_2)P(T(X) = T(x)|\theta_1, \theta_2).$$

If, in addition,  $T$  is a statistic such that its distribution given  $\theta$  depends only on  $\theta_1$  and the conditional distribution of  $X$  given  $T(X) = T(x)$ , and  $\theta$  depends only on  $\theta_2$ , the factored form in (1) is easily obtained (such a statistic was named  $p$ -sufficient for  $\theta_1$  by Basu ([17]). In this situation, all the relevant information on  $\theta_1$  is summarized in  $T$ , and one can fully make inferences on  $\theta_1$  taking into account only the conditional distribution of  $T$  given  $\theta$ , which does not involve  $\theta_2$ . Similarly, if  $T$  is a statistic such that its distribution given  $\theta$  depends only on  $\theta_2$  and the conditional distribution of  $X$  given  $T(X) = T(x)$  and  $\theta$  depends only on  $\theta_1$ , the factored form in (1) holds. Such a statistic was named  $s$ -ancillary for  $\theta_1$  by Basu ([17]), and it is somewhat evident that in this case, conclusions on  $\theta_1$  should be drawn exclusively from the distribution of  $X$  given  $T(X)$  and  $\theta$ , which does not depend on  $\theta_2$ . Such a conditional approach to the problem of elimination of nuisance parameters had already been proposed by Basu ([17]) and in a sense is closely related to the NNPP by Berger and Wolpert. The next theorem formally presents such results.

**Theorem 2.** *Let  $\mathcal{E} = (X, \theta, \mathcal{P})$  be an experiment in which  $\theta = (\theta_1, \theta_2)$  and  $\Theta$  is variation independent. Then, if  $\exists T : \mathfrak{X} \rightarrow \mathcal{T}$  such that  $T$  is either  $p$ -sufficient or  $s$ -ancillary for  $\theta_1$ , then for each  $x \in \mathfrak{X}$ , the likelihood function generated by  $x$ ,  $L_x(\cdot)$  can be factored as (1).*

In summary, it seems reasonable that inferences about  $\theta_1$  and  $\theta_2$  can be performed independently under condition (1). Thus, if only  $\theta_1$  is of interest, then it seems sensible under (1) that we reach the same conclusions on  $\theta_1$  when  $x$  is observed either by using the whole likelihood function  $L_x$  or only the factor  $L_x^1$ . That is, it makes sense to disregard the information contained in  $L_x^2$  and focus on  $L_x^1$ . As mentioned by [19], examples of likelihood functions with separable parameters like (1) are rare, but if (1) holds, it would be a useful property for Bayesian and non-Bayesian statisticians to analyze statistical data, especially in the presence of nuisance parameters. This fact will be illustrated in Sections 3 and 4.

We end this section by formally adapting the general NNPP to the special problem of hypothesis testing, in which inference about an unknown parameter consists of deciding whether a statement about the parameter (a statistical hypothesis) should be rejected or accepted by using the observable quantity  $X$ .

As before, let  $\mathcal{E} = (X, \theta, \mathcal{P})$  be an experiment, with  $\Theta = \Theta_1 \times \Theta_2$ . Let  $\bar{\mathcal{E}} = ((X, \theta_2), \theta_1, \bar{\mathcal{P}})$  be the “thought” experiment in which, in addition to  $X$ ,  $\theta_2$  is observed. Then, consider the following definition.

**Definition 1. Non-Informative Nuisance Parameter (NNP):** *Let  $B \subseteq \Theta_1$  and  $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \rightarrow \{0, 1\}$  be a test for the hypotheses*

$$\begin{aligned} \bar{H} : \theta_1 \in B \\ \bar{A} : \theta_1 \notin B \end{aligned} \tag{3}$$

*Then, we say that  $\theta_2$  is a Non-Informative Nuisance Parameter (NNP) for testing  $\bar{H}$  versus  $\bar{A}$  by using  $\bar{\varphi}$  if, for every  $x \in \mathfrak{X}$  such that (1) holds,  $\bar{\varphi}(x, \theta_2)$  does not depend on  $\theta_2$ ; that is, it depends only on  $x$ .*

In a nutshell, Definition 1 tells us something that appears intuitive: if the decision between  $H$  and  $A$  does not depend on  $\theta_2$ , then  $\theta_2$  does not provide any information about  $\theta_1$ . In the following example, we illustrate this idea.

**Example 1.** Consider that  $\Theta_1 = \Theta_2 = \mathbb{R}$  and the experiment  $\mathcal{E} = ((X, \theta_2), \theta_1, \mathcal{P})$ . Let  $B \subset \mathbb{R}$  and  $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \rightarrow \{0, 1\}$  be the test for the hypotheses

$$\begin{aligned} \bar{H} : \theta_1 &\in B \\ \bar{A} : \theta_1 &\notin B, \end{aligned} \tag{4}$$

such that the null hypothesis is rejected when the conditional probability of  $B$  given  $x$  and  $\theta_2$  is small; that is,

$$\bar{\varphi}(x, \theta_2) = 1 \Leftrightarrow P(\theta_1 \in B | x, \theta_2) < \delta, \tag{5}$$

where  $\delta \in (0, 1)$ . Suppose, in addition, that  $\theta_1$  and  $\theta_2$  are independent a priori. Let us verify that  $\theta_2$  is an NNP for testing these hypotheses by means of  $\bar{\varphi}$ . Let  $x_0 \in \mathfrak{X}$  be such that  $L_{x_0}(\theta_1, \theta_2) = L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)$  for specific functions  $L_{x_0}^1$  and  $L_{x_0}^2$ . Then,

$$\begin{aligned} P(\theta_1 \in B | x_0, \theta_2) &= \int_B \pi(\theta_1 | x_0, \theta_2) d\theta_1 = \\ &= \int_B \left[ \frac{P(X = x_0 | \theta_1, \theta_2) \pi(\theta_2 | \theta_1) \pi_1(\theta_1)}{\int_{\Theta_1} P(X = x_0 | \theta'_1, \theta_2) \pi(\theta_2 | \theta'_1) \pi_1(\theta'_1) d\theta'_1} \right] d\theta_1 = \\ &= \frac{\int_B L_{x_0}(\theta_1, \theta_2) \pi_2(\theta_2) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}(\theta_1, \theta_2) \pi_2(\theta_2) \pi_1(\theta_1) d\theta_1} = \\ &= \frac{\int_B L_{x_0}^1(\theta_1) L_{x_0}^2(\theta_2) \pi_2(\theta_2) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}^1(\theta_1) L_{x_0}^2(\theta_2) \pi_2(\theta_2) \pi_1(\theta_1) d\theta_1} \Rightarrow \\ &\Rightarrow P(\theta_1 \in B | x_0, \theta_2) = \frac{\int_B L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}, \end{aligned} \tag{6}$$

where  $\pi_i$  is the prior of  $\theta_i$ ,  $i = 1, 2$ . Thus, we have that

$$\bar{\varphi}(x_0, \theta_2) = 1 \Leftrightarrow \frac{\int_B L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1} < \delta. \tag{7}$$

Note from Equation (7) that  $\bar{\varphi}(x_0, \theta_2)$  does not depend on  $\theta_2$ . Thus,  $\theta_2$  is an NNP for testing  $\bar{H}$  versus  $\bar{A}$  by using  $\bar{\varphi}$ .

After defining an NNP, we formally state the Non-Informative Nuisance Parameter Principle (NNPP) for hypothesis testing.

**Definition 2. Non-Informative Nuisance Parameter Principle (NNPP):** Let the parameter space be variation independent; that is,  $\Theta = \Theta_1 \times \Theta_2$ . Consider the experiments  $\mathcal{E} = (X, \theta, \mathcal{P})$  and  $\mathcal{E} = ((X, \theta_2), \theta_1, \bar{\mathcal{P}})$ . Let  $B \subseteq \Theta_1$  be the subset of  $\Theta_1$  of interest. In addition, let  $\varphi : \mathfrak{X} \rightarrow \{0, 1\}$  and  $\bar{\varphi} : \mathfrak{X} \times \Theta_2 \rightarrow \{0, 1\}$  be tests for the hypotheses

$$\begin{aligned} H : \theta &\in B \times \Theta_2 & \text{and} & & \bar{H} : \theta_1 &\in B \\ A : \theta &\notin B \times \Theta_2, & & & \bar{A} : \theta_1 &\notin B, \end{aligned} \tag{8}$$

respectively.

If  $\theta_2$  is an NNP for testing  $\bar{H}$  versus  $\bar{A}$  by using  $\bar{\varphi}$  and  $x_0 \in \mathfrak{X}$  such that condition (1) holds, then

$$\varphi(x_0) = 1 \Leftrightarrow \bar{\varphi}(x_0, \theta_2) = 1. \tag{9}$$

The NNPP for statistical hypothesis testing says that if one intends to test a hypothesis regarding only the parameter  $\theta_1$ , it is irrelevant whether  $\theta_2$  is known or unknown if it is non-informative for such a decision-making problem. More formally, if one wants to test a hypothesis concerning only  $\theta_1$  and he observes a sample point  $x_0 \in \mathfrak{X}$  that separates information on  $\theta_1$  from information on  $\theta_2$ —that is, (1) holds—then the performances of the tests  $\varphi$  under the original experiment  $\mathcal{E}$  and  $\bar{\varphi}$  under the “thought” experiment  $\bar{\mathcal{E}}$  should yield the same decision on the hypothesis  $\theta_1 \in B$  if  $\theta_2$  is non-informative for that purpose.

We should mention that the NNPP can be adapted to any other inferential procedure. However, in this work, we focus on the principle for the problem of hypothesis testing. We conclude this section by proving that tests based on the posterior probabilities of the hypotheses satisfy the NNPP under prior independence.

**Example 2** (continuation of Example 1). Consider the conditions of Example 1. Consider  $\mathcal{E} = (X, \theta, \mathcal{P})$  and let  $\varphi : \mathfrak{X} \rightarrow \{0, 1\}$  be the test for the hypotheses

$$\begin{aligned} H : & \theta \in B \times \Theta_2 \\ A : & \theta \notin B \times \Theta_2 \end{aligned} \tag{10}$$

that rejects the null hypothesis  $H$  if its posterior probability is small; that is,

$$\varphi(x) = 1 \Leftrightarrow P(\theta \in B \times \Theta_2 | x) < \delta. \tag{11}$$

Let  $x_0 \in \mathfrak{X}$  be such that  $L_{x_0}(\theta_1, \theta_2) = L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)$ . We can write the posterior probability on the right-hand side of (11) as

$$\begin{aligned} P(\theta \in B \times \Theta_2 | x_0) &= \int_{B \times \Theta_2} \pi(\theta | x_0) d\theta = \frac{\int_{B \times \Theta_2} L_{x_0}(\theta) \pi(\theta) d\theta}{\int_{\Theta} L_{x_0}(\theta) \pi(\theta) d\theta} = \\ &= \frac{\int_{B \times \Theta_2} L_{x_0}^1(\theta_1) L_{x_0}^2(\theta_2) \pi_1(\theta_1) \pi_2(\theta_2) d\theta}{\int_{\Theta_1 \times \Theta_2} L_{x_0}^1(\theta_1) L_{x_0}^2(\theta_2) \pi_1(\theta_1) \pi_2(\theta_2) d\theta} = \frac{\int_B L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}, \end{aligned} \tag{12}$$

where the last equality follows from Fubini’s Theorem. Hence,

$$\varphi(x) = 1 \Leftrightarrow \frac{\int_B L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} L_{x_0}^1(\theta_1) \pi_1(\theta_1) d\theta_1} < \delta. \tag{13}$$

From Equations (7) and (13), we have that  $\varphi(x) = 1 \Leftrightarrow \bar{\varphi}(x, \theta_2) = 1$ . Thus, the NNP Principle is met by tests based on posterior probabilities, as in Example 1. This result also holds when  $\Theta_i \subseteq \mathbb{R}^{n_i}$ ,  $n_i \in \mathbb{N}^*$ ,  $i = 1, 2$ .

In the next section, we examine a second test procedure that is in line with the NNPP. We review the mixed test introduced by Pericchi and Pereira ([8]) and prove that such a test meets the NNPP for simple hypotheses concerning the parameter of interest. We also show how the adherence of the mixed test to the NNPP can then simplify its use.

### 3. The Mixed Test Procedure

The mixed test formally introduced in ([8]) is a test procedure that combines elements from both Bayesian and frequentist views. On the one hand, it considers an (intrinsically Bayesian) prior distribution for the parameter from which predictive distributions for the

data under the competing hypotheses and Bayes factors are derived. On the other hand, the performance of the test depends on ordering the sample space by the Bayes factor and on the integration of these predictive distributions over specific subsets of the sample space in a frequentist-like manner. The mixed test is an optimal procedure in the sense that it minimizes linear combinations of averaged (weighted) probabilities of errors of decision. It also meets a few logical requirements for multiple-hypothesis testing and obeys the Likelihood Principle for discrete sample spaces despite the integration over the sample space it involves. In addition, the test overcomes several of the drawbacks fixed-level tests have. However, a difficulty with the mixed test procedure is the need to evaluate the Bayes factor for every sample point to order the sample space, which may involve intensive calculations. Properties of the mixed test and examples of application are examined in detail in [8,20–25].

Next, we review the general procedure for the performance of the mixed test and then show the test satisfies the NNPP when the hypothesis regarding the parameter of interest is a singleton.

First, we determine the predictive distributions for  $X$  under the competing hypotheses  $H$  and  $A$ ,  $f_H$  and  $f_A$ , respectively. For the null hypothesis  $H : \theta \in \Theta_0$ ,  $\Theta_0 \subset \Theta$ ,  $f_H$  is determined as follows: for each  $x \in \mathfrak{X}$ ,

$$f_H(x) = \int_{\Theta_0} L_x(\theta) dP_H(\theta), \quad (14)$$

where  $P_H$  denotes the conditional distribution of  $\theta$  given  $\theta \in \Theta_0$ . That is, for each  $x \in \mathfrak{X}$ ,  $f_H(x)$  is the expected value of the likelihood function generated by  $x$  against  $P_H$ . Similarly, for the alternative hypothesis  $A: \theta \in \Theta_0^c$  we define

$$f_A(x) = \int_{\Theta_0^c} L_x(\theta) dP_A(\theta), \quad (15)$$

where  $P_A$  denotes the conditional distribution of  $\theta$  given  $\theta \in \Theta_0^c$ . From (14) and (15), we obtain the Bayes factor of  $x \in \mathfrak{X}$  for the hypothesis  $H$  over  $A$  as

$$BF(x) = \frac{f_H(x)}{f_A(x)}. \quad (16)$$

Finally, the mixed test  $\varphi^*: \mathfrak{X} \rightarrow \{0, 1\}$  for the hypotheses  $H$  versus  $A$  consists in rejecting  $H$  when  $x \in \mathfrak{X}$  is observed if and only if the Bayes factor  $BF(x)$  is small. That is, for each  $x \in \mathfrak{X}$ ,

$$\varphi^*(x) = 1 \Leftrightarrow BF(x) \leq b/a, \quad (17)$$

where the positive constants  $a$  and  $b$  reflect the decision maker's evaluation of the impact of the errors of the two types or, equivalently, his prior preferences for the competing hypotheses. A detailed discussion on the specification of such constants is found in [8,20–25].

The mixed test can also be defined as a function of a new significance index. That is, (17) can be rewritten as a comparison between such a significance index and a specific cut-off value. These quantities are defined below.

For the mixed test defined in (17), the  $p$ -value of the observation  $x_0 \in \mathfrak{X}$  is the significance index given by

$$p\text{-value}(x_0) = \sum_{x \in \mathfrak{D}(x_0)} f_H(x), \quad (18)$$

where  $\mathfrak{D}(x_0) = \{x \in \mathfrak{X} : BF(x) \leq BF(x_0)\}$ . Also, we define the adaptive type I error probability of  $\varphi^*$  as

$$\alpha^* = \sum_{x \in \mathfrak{D}} f_H(x), \quad (19)$$

where  $\mathfrak{D} = \{x \in \mathfrak{X} : BF(x) \leq b/a\}$ . Alternatively,  $\alpha^*$  is also known as the adaptive significance level of  $\varphi^*$ .

Pereira et al. [21] proved that the mixed test  $\varphi^*$  for the hypotheses

$$\begin{aligned} H : \theta &\in \Theta_0 \\ A : \theta &\in \Theta_0^c \end{aligned} \tag{20}$$

can be written as

$$\varphi^*(x) = 1 \Leftrightarrow BF(x) \leq b/a \Leftrightarrow p\text{-value}(x) \leq \alpha^* . \tag{21}$$

Note that  $\varphi^*$  consists of comparing the *p-value* with the cut-off  $\alpha^*$ , which depends on the specific statistical model under consideration and on the sample size, as opposed to a standard test with a fixed significance level that does not depend on the sample size.

The former does not have a few of the disadvantages of the latter, such as inconsistency ([8,26]) lack of correspondence between practical significance and statistical significance ([8,27]) and absence of logical coherence under multiple-hypothesis testing. We continue with the main results of the manuscript.

### The Mixed Test Obeys the NNPP

In this subsection, we prove that the mixed test meets the NNPP when the hypothesis about the parameter of interest is simple. Next, we examine further the case in which there is a statistic *s-ancillary* for the parameter of interest and show how the introduction of the concepts of a conditional *p-value* and a conditional adaptive significance level can make performance of the mixed test much easier.

**Theorem 3.** *Let  $\theta = (\theta_1, \theta_2)$  and  $\Theta = \Theta_1 \times \Theta_2$  (that is,  $\Theta$  is variation independent). Let  $\mathcal{E} = (X, \theta, \mathcal{P})$  and  $\bar{\mathcal{E}} = ((X, \theta_2), \theta_1, \bar{\mathcal{P}})$  be two experiments as defined in Section 2. Let  $\theta_0 \in \Theta_1$ . In addition, let  $\varphi^* : \mathfrak{X} \rightarrow \{0, 1\}$  and  $\bar{\varphi}^* : \mathfrak{X} \times \Theta_2 \rightarrow \{0, 1\}$  be the mixed tests for the hypotheses*

$$\begin{aligned} H : \theta &\in \{\theta_0\} \times \Theta_2 & \text{and} & & \bar{H} : \theta_1 &= \theta_0 \\ A : \theta &\notin \{\theta_0\} \times \Theta_2 & & & \bar{A} : \theta_1 &\neq \theta_0, \end{aligned} \tag{22}$$

respectively. Assume  $\theta = (\theta_1, \theta_2)$  is absolutely continuous with prior density function  $\pi$ , with  $\theta_1 \perp\!\!\!\perp \theta_2$ . Then,  $\theta_2$  is a Non-Informative Nuisance Parameter for testing  $\bar{H}$  versus  $\bar{A}$  by using  $\bar{\varphi}^*$ , and for every  $x \in \mathfrak{X}$  such that (1) holds,

$$\varphi^*(x) = 1 \Leftrightarrow \bar{\varphi}^*(x, \theta_2) = 1 \tag{23}$$

Theorem 3 tells us that when the likelihood function may be factored as (1), the mixed test obeys the NNPP. That is to say, if one aims to test a simple hypothesis about the parameter of interest  $\theta_1$  in the presence of a non-informative nuisance parameter  $\theta_2$  by means of the mixed test, then he can disregard  $\theta_2$  in the analysis. Under a purely mathematical viewpoint, when  $x \in \mathfrak{X}$  satisfying (1) is observed, the decision between rejecting and accepting the null hypothesis regarding  $\theta_1$  depends on  $L_x$  only through the factor  $L_x^1$ , which is not a function of  $\theta_2$ , as we can see from Equation (A16) in Appendix A. It should be emphasized that Theorem 3 holds for null hypotheses more general than only simple ones. For instance, the Theorem is still valid when the null hypothesis  $H$  is of the form  $H : \theta \in \Theta_0 \times \Theta_2$ , where  $\Theta_0 \subset \Theta_1$  is a hyperplane of  $\Theta_1$ . The proof of this result is quite similar to the proof of Theorem 3 in Appendix A and for this reason is omitted.

The adherence to the NNPP is indeed an advantage of the mixed test. It may bring a considerable reduction in the calculations involved along the procedure of the mixed test, especially under statistical models for which a statistic *s-ancillary* for the parameter of interest can be found. Such cases are examined after Corollary 1, which follows straightforwardly from Theorems 2 and 3.

**Corollary 1.** Assume the same conditions of Theorem 3 and suppose that  $\exists T : \mathfrak{X} \rightarrow \mathcal{T}$  such that  $T$  is  $p$ -sufficient for  $\theta_2$  and  $s$ -ancillary for  $\theta_1$ . Then, for all  $x \in \mathfrak{X}$ ,  $\varphi^*(x) = 1 \Leftrightarrow \bar{\varphi}^*(x, \theta_2) = 1$ .

Now, let us suppose that under experiment  $\mathcal{E} = (X, \theta, \mathcal{P})$ , there is a statistic  $T : \mathfrak{X} \rightarrow \mathcal{T}$  such that  $T$  is  $s$ -ancillary for  $\theta_1$ . Let  $H : \theta \in \{\theta_0\} \times \Theta_2$  be the hypothesis of interest. From the predictive distribution  $f_H$  for  $X$ , we can define for each value  $t \in T(\mathfrak{X}) = \{T(x) : x \in \mathfrak{X}\}$  the conditional probability function for  $X$  given  $T(X) = t$ ,  $f_{H,t}$  by

$$f_{H,t}(x) = \frac{f_H(x)}{\sum_{y \in \mathfrak{X}: T(y)=t} f_H(y)} \tag{24}$$

if  $T(x) = t$ , and  $f_{H,t}(x) = 0$ , otherwise.

Finally, from the conditional distribution in (24), we define two conditional statistics: the conditional  $p$ -value and the conditional adaptive significance level. Such quantities will be of great importance for the performance of the mixed test, as we will see in the next section.

**Definition 3. Conditional  $p$ -value:** Let  $\mathcal{E} = (X, \theta, \mathcal{P})$  be an experiment for which the statistic  $T : \mathfrak{X} \rightarrow \mathcal{T}$  is  $s$ -ancillary for  $\theta_1$ . Let  $H : \theta \in \{\theta_0\} \times \Theta_2$  be the hypothesis of interest, and  $f_{H,t}$ ,  $t \in T(\mathfrak{X})$ , as in (24). We define the  $p$ -value conditional on  $T$ ,  $p_T$ -value :  $\mathfrak{X} \rightarrow [0, 1]$  for each  $x_0 \in \mathfrak{X}$  by

$$p_T\text{-value}(x_0) = \sum_{x \in \mathfrak{D}(x_0)} f_{H,T(x_0)}(x) = \frac{\sum_{x \in \mathfrak{D}_T^*(x_0)} f_H(x)}{\sum_{x \in \mathfrak{D}_T(x_0)} f_H(x)}, \tag{25}$$

where  $\mathfrak{D}_T^*(x_0) = \{x \in \mathfrak{X} : BF(x) \leq BF(x_0), T(x) = T(x_0)\}$  and  $\mathfrak{D}_T(x_0) = \{x \in \mathfrak{X} : T(x) = T(x_0)\}$ . From Equation (A14), the  $P_T$ -value may be rewritten as

$$p_T\text{-value}(x_0) = \frac{\sum_{x \in \mathfrak{D}_T^*(x_0)} \left[ L_x^1(\theta_0) \int_{\Theta_2} L_x^2(\theta_2) \pi_2(\theta_2) d\theta_2 \right]}{\sum_{x \in \mathfrak{D}_T(x_0)} \left[ L_x^1(\theta_0) \int_{\Theta_2} L_x^2(\theta_2) \pi_2(\theta_2) d\theta_2 \right]}, \tag{26}$$

where  $L_x^1(\theta_0) = P(X = x | T(X) = T(x), \theta_0)$  and  $L_x^2(\theta_2) = P(T(X) = T(x) | \theta_2)$  since  $T$  is  $s$ -ancillary for  $\theta_1$ . It follows that

$$\begin{aligned} p_T\text{-value}(x_0) &= \frac{\sum_{x \in \mathfrak{D}_T^*(x_0)} \left[ P(X = x | T(X) = T(x), \theta_0) \int_{\Theta_2} P(T(X) = T(x) | \theta_2) \pi_2(\theta_2) d\theta_2 \right]}{\sum_{x \in \mathfrak{D}_T(x_0)} \left[ P(X = x | T(X) = T(x), \theta_0) \int_{\Theta_2} P(T(X) = T(x) | \theta_2) \pi_2(\theta_2) d\theta_2 \right]} \\ &= \frac{\sum_{x \in \mathfrak{D}_T^*(x_0)} \left[ P(X = x | T(X) = T(x_0), \theta_0) \int_{\Theta_2} P(T(X) = T(x_0) | \theta_2) \pi_2(\theta_2) d\theta_2 \right]}{\sum_{x \in \mathfrak{D}_T(x_0)} \left[ P(X = x | T(X) = T(x_0), \theta_0) \int_{\Theta_2} P(T(X) = T(x_0) | \theta_2) \pi_2(\theta_2) d\theta_2 \right]} = \\ &= \frac{\sum_{x \in \mathfrak{D}_T^*(x_0)} P(X = x | T(X) = T(x_0), \theta_0)}{\sum_{x \in \mathfrak{D}_T(x_0)} P(X = x | T(X) = T(x_0), \theta_0)}; \end{aligned}$$

that is,

$$p_T\text{-value}(x_0) = \sum_{x \in \mathfrak{D}_T^*(x_0)} P(X = x | T(X) = T(x_0), \theta_0). \tag{27}$$

**Definition 4. Conditional adaptive significance level:** Let  $\mathcal{E} = (X, \theta, \mathcal{P})$  be an experiment for which the statistic  $T : \mathfrak{X} \rightarrow \mathcal{T}$  is  $s$ -ancillary for  $\theta_1$ . Let  $H : \theta \in \{\theta_0\} \times \Theta_2$  be the hypothesis of interest and  $f_{H,t}, t \in \mathcal{T}(\mathfrak{X})$  be as in (24). We define the conditional adaptive significance level given  $T, \alpha_T^* : \mathfrak{X} \rightarrow [0, 1]$ , for each  $x_0 \in \mathfrak{X}$  by

$$\alpha_T^*(x_0) = \sum_{x \in \mathfrak{D}} f_{H,T(x_0)}(x) = \frac{\sum_{x \in \mathfrak{D} \cap \mathfrak{D}_T(x_0)} f_H(x)}{\sum_{x \in \mathfrak{D}_T(x_0)} f_H(x)}. \tag{28}$$

The conditional adaptive significance level  $\alpha_T^*$  may be rewritten as

$$\alpha_T^*(x_0) = \sum_{x \in \mathfrak{D} \cap \mathfrak{D}_T(x_0)} P(X = x | T(X) = T(x_0), \theta_0). \tag{29}$$

Definitions 3 and 4 are conditional versions of Definitions in (18) and (19), respectively. While calculation of the unconditional quantities involves the evaluation of the Bayes factor for every  $x \in \mathfrak{X}$ , the determination of the conditional statistics at a specific sample point  $x_0 \in \mathfrak{X}$  depends only on the values of the Bayes factor for the sample points  $x$  such that  $T(x) = T(x_0)$ , which may be much easier to accomplish. Note also that the  $p_T$ -value and  $\alpha_T^*$  can be seen, respectively, as an alternative (conditional) measure of evidence in favor of the null hypothesis  $H$  and an alternative threshold value for testing the competing hypotheses. As a matter of fact, one can substitute the  $p$ -value and the adaptive significance level with their conditional versions in order to perform the mixed test. This is exactly what the next theorem states.

**Theorem 4.** Assume the same conditions as in Corollary 1 and Theorem 3. Then, for all  $x_0 \in \mathfrak{X}$ ,

$$\varphi^*(x_0) = 1 \Leftrightarrow p_T\text{-value}(x_0) \leq \alpha_T^*(x_0).$$

The results of Theorems 3 and 4 and Corollary 1 suggest a way the mixed test may be used without doing so many calculations: when an ancillary statistic for the parameter of interest,  $T$ , is available, one can perform the test by comparing the conditional statistics  $p_T$ -value and  $\alpha_T^*$  instead of comparing the unconditional ones in Definitions (18) and (19). This possibility is illustrated in the next section.

#### 4. Examples

We now revisit three well-known problems of hypothesis testing for count data and present new solutions to them by means of the mixed test. In each problem, we consider a suitable reparametrization of the standard model in order to ensure that

1. There exists a statistic  $T$  that is ancillary to the new parameter of interest;
2. The hypothesis about the new parameter of interest under the reparametrization is a singleton (or a hyperplane);
3. The new parameter of interest is independent of the new nuisance parameter a priori;
4. The distribution of the data  $X$  given any value of the statistic  $T$  is simple enough to render the calculations of the conditional  $p$ -value and the conditional adaptive significance level easy.

##### 4.1. Comparison of Poisson Means

Suppose we are interested in testing the equality between two Poisson means: say  $\theta_1$  and  $\theta_2$ . Let  $\theta = (\theta_1, \theta_2)$ . For this purpose, let  $X = (X_1, X_2)$  be a random vector to be observed such that given  $\theta, X_1$  and  $X_2$  are independent Poisson random variables with

parameters  $n\theta_1$  and  $n\theta_2$ , respectively, where  $n \in \mathbb{N}^* = \{1, 2, 3, \dots\}$  is a known integer. The hypotheses to be tested are

$$\begin{aligned} H &: \theta \in \Theta_0 \\ A &: \theta \in \Theta_0^c, \end{aligned} \tag{30}$$

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

$$L_x(\theta) = \frac{(n\theta_1)^{x_1}}{x_1!} \frac{(n\theta_2)^{x_2}}{x_2!} e^{-\theta_1 n} e^{-\theta_2 n}. \tag{31}$$

Suppose also that  $\theta_1$  and  $\theta_2$  are independent a priori and that  $\theta_i$  is distributed as a Gamma random variable with parameters  $a_i > 0$  and  $c > 0$ ,  $i = 1, 2$ . That is, the prior density function of  $\theta$  is

$$\pi(\theta) = \frac{c^{a_1}}{\Gamma(a_1)} \theta_1^{a_1-1} e^{-c\theta_1} \mathbb{I}_{(0,\infty)}(\theta_1) \frac{c^{a_2}}{\Gamma(a_2)} \theta_2^{a_2-1} e^{-c\theta_2} \mathbb{I}_{(0,\infty)}(\theta_2). \tag{32}$$

Although one can determine an exact expression for the Bayes factor in this case (as a matter of fact, in [25], the authors first presented a solution to the problem of testing the equality of Poisson means by using weighted likelihoods in the context of a production process monitoring procedure), the use of the mixed test under the above parametrization may be computationally disadvantageous, as the sample space is  $\mathbb{N}^2$  and one should determine infinitely many Bayes factors to perform the test. To overcome this difficulty, we next consider the following reparametrization of the model: let  $\lambda = (\lambda_1, \lambda_2)$  be the new parameter, where

$$\lambda_1 = \frac{\theta_1}{\theta_1 + \theta_2} \quad \text{and} \quad \lambda_2 = \theta_1 + \theta_2. \tag{33}$$

The new parameter space is then  $\Lambda = (0, 1) \times \mathbb{R}_+$ . Now, the hypotheses (30) can be rewritten as

$$\begin{aligned} \tilde{H} &: \lambda \in \Lambda_0 \\ \tilde{A} &: \lambda \in \Lambda_0^c, \end{aligned} \tag{34}$$

with  $\Lambda_0 = \{\frac{1}{2}\} \times \mathbb{R}_+$ . Note that the likelihood function (31) can be rewritten by conditioning on the statistic  $T(X) = X_1 + X_2$  as follows:

$$\begin{aligned} L_x(\theta) &= P(X_1 = x_1, X_2 = x_2 | \theta) = P(X_1 = x_1, X_2 = x_2 | X_1 + X_2 = x_1 + x_2, \theta) P(X_1 + X_2 = x_1 + x_2 | \theta) \\ &= \binom{x_1 + x_2}{x_1} \left(\frac{\theta_1}{\theta_1 + \theta_2}\right)^{x_1} \left(\frac{\theta_2}{\theta_1 + \theta_2}\right)^{x_2} \frac{e^{-n(\theta_1 + \theta_2)} [n(\theta_1 + \theta_2)]^{x_1 + x_2}}{(x_1 + x_2)!}. \end{aligned} \tag{35}$$

Hence, the induced likelihood function for  $\lambda$  generated by  $(x_1, x_2)$  may be factored as

$$\tilde{L}_x(\lambda) = \left[ \binom{x_1 + x_2}{x_1} \lambda_1^{x_1} (1 - \lambda_1)^{x_2} \right] \left[ \frac{(n\lambda_2)^{x_1 + x_2}}{(x_1 + x_2)!} e^{-\lambda_2 n} \right] \tag{36}$$

Note that  $T$  is an ancillary statistic for  $\lambda_1$ , as it is distributed as a Poisson random variable with mean  $n\lambda_2$ , and the conditional distribution of  $X$ , given  $T(X) = t$ ,  $t \in \mathbb{N}$ , depends on  $\lambda$  only through  $\lambda_1$ . The prior distribution for  $\lambda$  is given by

$$\tilde{\pi}(\lambda) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \lambda_1^{a_1-1} (1 - \lambda_1)^{a_2-1} \mathbb{I}_{(0,1)}(\lambda_1) \frac{c^{a_1+a_2}}{\Gamma(a_1 + a_2)} \lambda_2^{a_1+a_2-1} e^{-c\lambda_2} \mathbb{I}_{(0,\infty)}(\lambda_2). \tag{37}$$

Now, as (34), (36) and (37) hold, it follows from Theorem 3 that  $\lambda_2$  is an NNP and that the performance of the mixed test for the hypothesis  $\tilde{H} : \lambda \in \{\frac{1}{2}\} \times \mathbb{R}_+$  against  $\tilde{A}$  based on  $\tilde{L}_x$  and the prior  $\tilde{\pi}$  is equivalent to the performance of the mixed test for the simple hypothesis  $\lambda_1 = \frac{1}{2}$  against  $\lambda_1 \neq \frac{1}{2}$  based on the binomial-like factor of  $\tilde{L}_x$  that depends only on  $\theta_1$  and the marginal Beta prior density for  $\lambda_1$  ignoring the NNP  $\lambda_2$ . In addition, Theorem 4 implies that the test for  $\tilde{H}$  versus  $\tilde{A}$  reduces to the comparison of the statistics  $p_T$ -value and  $\alpha_T^*$  at the observed sample point, say  $x_0 = (x_{01}, x_{02})$ . Note that in this case, one does not need to evaluate the Bayes factor for every point of  $\mathbb{N}^2$  but only for those  $x_{01} + x_{02} + 1$  of them for which the sum of the components is  $x_{01} + x_{02}$ . That is, one needs to evaluate the Bayes factor only for the elements of  $\{(u, v) \in \mathbb{N}^2 : T(u, v) = T(x_0) = x_{01} + x_{02}\}$  when  $x_0 = (x_{01}, x_{02})$  is observed.

From Equations (A14) and (A15), one gets the following predictive functions under  $\tilde{H}$  and under  $\tilde{A}$  for  $X$ :

$$f_{\tilde{H}}(\mathbf{x}) = (1/2)^{x_1+x_2} \binom{x_1+x_2}{x_1} \times K \tag{38}$$

and

$$f_{\tilde{A}}(\mathbf{x}) = \binom{x_1+x_2}{x_1} \left[ \frac{\Gamma(x_1+a_1)\Gamma(x_2+a_2)}{\Gamma(x_1+x_2+a_1+a_2)} \right] \left[ \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} \right] \times K, \tag{39}$$

where  $K = \int_0^\infty \left[ \frac{(n\lambda_2)^{x_1+x_2}}{(x_1+x_2)!} e^{-\lambda_2 n} \right] \frac{c^{a_1+a_2}}{\Gamma(a_1+a_2)} \lambda_2^{a_1+a_2-1} e^{-c\lambda_2} d\lambda_2$ . Consequently, the Bayes factor  $BF(x)$  is

$$BF(\mathbf{x}) = (1/2)^{x_1+x_2} \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)} \frac{\Gamma(x_1+x_2+a_1+a_2)}{\Gamma(x_1+a_1)\Gamma(x_2+a_2)}. \tag{40}$$

Finally, it follows from (28) and (30) that for  $x_0 = (x_{01}, x_{02}) \in \mathbb{N}^2$ ,

$$\begin{aligned} p_T\text{-value}(x_0) &= \sum_{(x_1, x_2) \in \mathcal{D}_T^*(x_0)} \binom{x_1+x_2}{x_1} (1/2)^{x_1+x_2} \\ &= \sum_{x_1=0}^{T(x_0)} \binom{T(x_0)}{x_1} (1/2)^{T(x_0)} \mathbb{I}_{\mathcal{D}(x_0)}(x_1, T(x_0) - x_1) \end{aligned} \tag{41}$$

and

$$\begin{aligned} \alpha_T^*(x_0) &= \sum_{(x_1, x_2) \in \mathcal{D} \cap \mathcal{D}_T(x_0)} \binom{x_1+x_2}{x_1} (1/2)^{x_1+x_2} \\ &= \sum_{x_1=0}^{T(x_0)} \binom{T(x_0)}{x_1} (1/2)^{T(x_0)} \mathbb{I}_{\mathcal{D}}(x_1, T(x_0) - x_1). \end{aligned} \tag{42}$$

Note that in this case, the conditional  $p_T$ -value resembles the frequentist  $p$ -value for the simple hypothesis  $\theta = \frac{1}{2}$  under simple random sampling from the Bernoulli model with parameter  $\theta$  (however, for the calculation of the  $p_T$ -value, the sample space is ordered by the Bayes factor instead of the likelihood ratio).

**Example 3 (Comparison of Poisson means).** In [25], the authors consider that a methodology to detect a shift in a production process is to compare the quality index of the current rating period  $P$ ,  $\theta_2$ , with the quality index of the previous rating period,  $\theta_1$ . Suppose that we want to test if a process is under control; that is, if  $\theta_1 = \theta_2$ . For this purpose, two audit samples of size  $n = 10$  are collected at rating periods  $P - 1$  and  $P$ , 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 suppose that  $X_1$  and  $X_2$  are Poisson random variables with parameters  $n\theta_1$  and  $n\theta_2$ , respectively. Let  $X = (X_1, X_2)$ . For simplicity, we consider the hyperparameters in (32) as  $a_1 = a_2 = c = 1$ . Hence, the predictive functions under the competing hypothesis are given by:

$$f_{\tilde{H}}(\mathbf{x}) = (1/2)^{x_1+x_2} \left[ (x_1+x_2+1) \binom{x_1+x_2}{x_1} \right] \left[ \left( \frac{n}{n+1} \right)^{x_1+x_2} \left( \frac{1}{n+1} \right)^2 \right], \tag{43}$$

and

$$f_{\tilde{A}}(\mathbf{x}) = \left( \frac{n}{n+1} \right)^{x_1+x_2} \left( \frac{1}{n+1} \right)^2. \tag{44}$$

Consequently, the Bayes factor at  $x = (x_1, x_2) \in \mathbb{N}^2$  can be expressed by

$$BF(\mathbf{x}) = (1/2)^{x_1+x_2} (x_1+x_2+1) \binom{x_1+x_2}{x_1}. \tag{45}$$

Now, suppose that two defects are found at rating period  $P - 1$  and nine defects are found at period  $P$ . That is, suppose that  $X = (2, 9)$  is observed. In this case, one gets  $BF(2, 9) = 0.322$ .

Considering  $\frac{b}{a} = 1$ , the conditional adaptive significance level and the conditional  $p_T$ -value at  $(2, 9)$  are, respectively,

$$\alpha_T^*((2, 9)) = \sum_{x_1 \in \{0,1,2,3,8,9,10,11\}} \binom{2+9}{x_1} (1/2)^{2+9} = 0.227 \text{ and}$$

$$P_T\text{-value}((2, 9)) = \sum_{x_1 \in \{0,1,2,9,10,11\}} \binom{11}{x_1} (1/2)^{11} = 0.065 .$$

Since  $P_T\text{-value}((2, 9)) < \alpha_T^*((2, 9))$ , the decision is to reject the null hypothesis (34), where  $\lambda = (\frac{\theta_1}{\theta_1 + \theta_2}, \theta_1 + \theta_2)$ .

Note that although the sample size is small, the null hypothesis can be rejected with a conditional  $P_T$  – value of 0.065. Such a value is not compared with standard (fixed) cut-off values such as 0.01 or 0.05 but rather with the conditional adaptive significance level of 0.227 for  $X = (2, 9)$ . Note also that performance of the mixed test by means of the conditional statistics  $p_T$ -value and  $\alpha_T^*$  when  $X = (2, 9)$  is observed requires the calculation of only finitely many Bayes factors (twelve, precisely) even though the sample space is infinite.

#### 4.2. Test of Symmetry

Suppose we want to test the hypothesis of symmetry in an  $r \times r$  two-way contingency table. Several methods have been proposed for testing diagonal symmetry: see, for example, [28–33] and references therein. Here we propose a solution to this problem by using the mixed test and its properties. We present the simplest case  $r = 2$ . The reader will find the general case  $r > 2$  in Appendix A for the sake of readability.

Suppose each element (individual) of a sample of size  $n$  is classified into four mutually exclusive combinations of the two-valued variables  $X_1$  and  $X_2$ . Let Table 1 represent the observed frequencies of the cross-classifications, where  $X_{ij}$  is the number of individuals classified into the  $i$  – th category of  $X_1$  and the  $j$  – th category of  $X_2$ ,  $i, j = 1, 2$ . Let  $\theta = (\theta_{11}, \theta_{12}, \theta_{21})$ , with  $\theta_{ij} \geq 0$  and  $\theta_{11} + \theta_{12} + \theta_{21} \leq 1$ , where  $\theta_{ij}$  denotes the probability of classification into the  $i$  – th category of  $X_1$  and the  $j$  – th category of  $X_2$ ,  $i, j = 1, 2$ .

**Table 1.** Observed frequencies of  $(X_1, X_2)$  in the  $2 \times 2$  case.

		$X_2$		
		$x_{11}$	$x_{12}$	
$X_1$	$x_{21}$	$x_{21}$	$x_{22}$	$n_{2.}$
		$n_{.1}$	$n_{.2}$	$n$

The hypotheses for testing diagonal symmetry are

$$\begin{aligned} H : \theta_{12} &= \theta_{21} \\ A : \theta_{12} &\neq \theta_{21} \end{aligned} \tag{46}$$

We assume that the vector  $X = (X_{11}, X_{12}, X_{21})$  is, given  $\theta$ , a multinomial random vector with parameters  $n$  and  $\theta$ . The likelihood function generated by  $x = (x_{11}, x_{12}, x_{21})$  is then given by

$$L_x(\theta) = \frac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!} \theta_{11}^{x_{11}} \theta_{12}^{x_{12}} \theta_{21}^{x_{21}} \theta_{22}^{x_{22}}, \tag{47}$$

where  $x_{22} = n - x_{11} - x_{12} - x_{21}$  and  $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$ . Assume also a prior Dirichlet distribution with parameter vector  $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}; \alpha_{22})$ ,  $\alpha_{ij} > 0$  for  $\theta$ . That is,

$$\pi(\theta) = \frac{\Gamma(\alpha_{11} + \alpha_{12} + \alpha_{21} + \alpha_{22})}{\Gamma(\alpha_{11})\Gamma(\alpha_{12})\Gamma(\alpha_{21})\Gamma(\alpha_{22})} \theta_{11}^{\alpha_{11}-1} \theta_{12}^{\alpha_{12}-1} \theta_{21}^{\alpha_{21}-1} \theta_{22}^{\alpha_{22}-1} \mathbb{I}_{\Theta}(\theta), \tag{48}$$

where  $\Theta = \{(u_1, u_2, u_3) \in \mathbb{R}_+^3 : u_1 + u_2 + u_3 \leq 1\}$ .

We should note that the determination of the predictive functions is much easier under the following reparametrization of the model: let us define

$$\lambda_{11} = \theta_{11} \quad \lambda_{21} = \theta_{12} + \theta_{21} \quad \text{and} \quad \lambda_{12} = \frac{\theta_{12}}{\theta_{12} + \theta_{21}}.$$

Let  $\lambda = (\lambda_{12}, (\lambda_{11}, \lambda_{21}))$ . Thus, the new parameter space is  $\Lambda = (0, 1) \times S_2$ , where  $S_2 = \{(u, v) \in \mathbb{R}_+^2 : u + v \leq 1\}$ .

Then, we can reformulate the hypotheses (46) as

$$\begin{aligned} \tilde{H} : \lambda &\in \Lambda_0 \\ \tilde{A} : \lambda &\in \Lambda_0^c, \end{aligned} \tag{49}$$

where  $\Lambda_0 = \{\frac{1}{2}\} \times S_2$ . Note that the likelihood function for  $\theta$  generated by  $x = (x_{11}, x_{12}, x_{21}) \in \{(a, b, c) \in \mathbb{N}^3 : a + b + c \leq n\}$  can be rewritten by conditioning on the statistic  $T(X) = (X_{11}, X_{12} + X_{21})$  as

$$\begin{aligned} L_x(\theta) &= P(X = (x_{11}, x_{12}, x_{21}) | T(X) = (x_{11}, x_{12} + x_{21}), \theta)P(T(X) = (x_{11}, x_{12} + x_{21}) | \theta) \\ &= \binom{x_{12} + x_{21}}{x_{12}} \left(\frac{\theta_{12}}{\theta_{12} + \theta_{21}}\right)^{x_{12}} \left(\frac{\theta_{21}}{\theta_{12} + \theta_{21}}\right)^{x_{21}} \frac{n!}{x_{11}!(x_{12} + x_{21})!x_{22}!} \theta_{11}^{x_{11}} (\theta_{12} + \theta_{21})^{x_{12} + x_{21}} \theta_{22}^{x_{22}}. \end{aligned} \tag{50}$$

Hence, the induced likelihood function for  $\lambda$  generated by  $x = (x_{11}, x_{12}, x_{21})$  may be factored as

$$\tilde{L}_x(\lambda) = \left[ \binom{x_{12} + x_{21}}{x_{12}} \lambda_{12}^{x_{12}} (1 - \lambda_{12})^{x_{21}} \right] \left[ \frac{n!}{x_{11}!(x_{12} + x_{21})!x_{22}!} \lambda_{11}^{x_{11}} \lambda_{21}^{x_{12} + x_{21}} (1 - \lambda_{11} - \lambda_{21})^{x_{22}} \right] \tag{51}$$

Note that  $T$  is an ancillary statistic for  $\lambda_{12}$  as it is a multinomial random vector with parameters  $n$  and  $(\lambda_{11}, \lambda_{21})$ , and the conditional distribution of  $X$  given  $T(X) = t$ ,  $t \in \{(u, v) \in \mathbb{N}^2 : u + v \leq n\}$  depends on  $\lambda$  only through  $\lambda_{12}$ . The prior distribution for  $\lambda$  is given by

$$\tilde{\pi}(\lambda) = \frac{\Gamma(\alpha_{12} + \alpha_{21})}{\Gamma(\alpha_{12})\Gamma(\alpha_{21})} \lambda_{12}^{\alpha_{12} - 1} (1 - \lambda_{12})^{\alpha_{21} - 1} \mathbb{I}_{(0,1)}(\lambda_{12}) \tilde{\pi}_{(\lambda_{11}, \lambda_{21})}(\lambda_{11}, \lambda_{21}), \tag{52}$$

where  $\tilde{\pi}_{(\lambda_{11}, \lambda_{21})}$  is the prior Dirichlet distribution for  $(\lambda_{11}, \lambda_{21})$  with parameter vector  $(\alpha_{11}, \alpha_{12} + \alpha_{21}, \alpha_{22})$ .

As in the example of the previous subsection, the results from Section 3 imply that  $(\lambda_{11}, \lambda_{21})$  is an NNP for testing the hypotheses  $\tilde{H}$  versus  $\tilde{A}$  in (50) by using the mixed test. In addition, we only need to compare the conditional  $p_T$ -value with the conditional adaptive significance level to decide between the hypotheses.

From Equations (A14) and (A15), one gets the following predictive functions under  $\tilde{H}$  and under  $\tilde{A}$  for  $X$ :

$$f_{\tilde{H}}(x_{11}, x_{12}, x_{21}) = (1/2)^{x_{12} + x_{21}} \binom{x_{12} + x_{21}}{x_{12}} \times K \tag{53}$$

and

$$f_{\tilde{A}}(x) = \binom{x_{12} + x_{21}}{x_{12}} \left[ \frac{\Gamma(x_{12} + \alpha_{12})\Gamma(x_{21} + \alpha_{21})}{\Gamma(x_{12} + x_{21} + \alpha_{12} + \alpha_{21})} \right] \left[ \frac{\Gamma(\alpha_{12} + \alpha_{21})}{\Gamma(\alpha_{12})\Gamma(\alpha_{21})} \right] \times K, \tag{54}$$

where

$$K = \int_{S_2} \left[ \frac{n!}{x_{11}!(x_{12} + x_{21})!x_{22}!} \lambda_{11}^{x_{11}} \lambda_{21}^{x_{12} + x_{21}} (1 - \lambda_{11} - \lambda_{21})^{x_{22}} \right] \tilde{\pi}_{(\lambda_{11}, \lambda_{21})}(\lambda_{11}, \lambda_{21}) d\lambda_{11} d\lambda_{21}.$$

Consequently, the Bayes factor  $BF(x)$  is

$$BF(x) = (1/2)^{x_{12}+x_{21}} \frac{\Gamma(\alpha_{12})\Gamma(\alpha_{21})}{\Gamma(\alpha_{12} + \alpha_{21})} \frac{\Gamma(x_{12} + x_{21} + \alpha_{12} + \alpha_{21})}{\Gamma(x_{12} + \alpha_{12})\Gamma(x_{21} + \alpha_{21})}. \tag{55}$$

Finally, it follows from (28) and (30) that for  $x = (x_{11}, x_{12}, x_{21})$ ,

$$\begin{aligned} p_T\text{-value}(x) &= \sum_{(y_{11}, y_{12}, y_{21}) \in \mathcal{D}_T^*(x)} \binom{y_{12} + y_{21}}{y_{12}} (1/2)^{y_{12}+y_{21}} \\ &= \sum_{y_{12}=0}^{x_{12}+x_{21}} \binom{x_{12} + x_{21}}{y_{12}} (1/2)^{x_{12}+x_{21}} \mathbb{I}_{\mathcal{D}(x)}(x_{11}, y_{12}, x_{12} + x_{21} - y_{12}) \end{aligned} \tag{56}$$

and

$$\begin{aligned} \alpha_T^*(x) &= \sum_{(y_{11}, y_{12}, y_{21}) \in \mathcal{D} \cap \mathcal{D}_T(x)} \binom{y_{12} + y_{21}}{y_{12}} (1/2)^{y_{12}+y_{21}} \\ &= \sum_{y_{12}=0}^{x_{12}+x_{21}} \binom{x_{12} + x_{21}}{y_{12}} (1/2)^{x_{12}+x_{21}} \mathbb{I}_{\mathcal{D}}(x_{11}, y_{12}, x_{12} + x_{21} - y_{12}) \end{aligned} \tag{57}$$

In this example (as in the previous subsection), the conditional  $p_T$ -value looks like the frequentist  $p$ -value for the simple hypothesis  $\theta = \frac{1}{2}$  regarding an unknown proportion. We should emphasize that for calculation of the  $p_T$ -value, the sample space is ordered by the Bayes factor in place of the likelihood ratio. Note also that the evaluation of this conditional statistic involves ordering at most  $n + 1$  points of the sample space (exactly those for which the statistic  $T$  takes the value  $T(x_0)$  when  $x_0$  is the effectively observed sample point). On the other hand, if one performs the mixed test without using these conditional quantities, he shall order all  $\binom{n+3}{3} = \frac{(n+3)(n+2)(n+1)}{6}$  elements of the sample space.

**Example 4 (Analysis of opinion swing).** Suppose it is of interest to evaluate whether the proportion of individuals that did not support the US President before the State of the Union Address remained unchanged after his address. For this purpose,  $n = 100$  individuals are surveyed with regard to their support for the President before and after his annual message. The survey results are displayed in the following  $2 \times 2$  contingency Table 2:

**Table 2.** Survey results.

		After		
		No	Yes	
Before	No	20	17	37
	Yes	10	53	63
		30	70	100

Let  $X_1$  ( $X_2$ ) be the support—“No” or “Yes”—for the President before (after) the State of the Union Address. Let  $\theta_{ij}$  be the probability that an individual is classified into the  $i$ -th category of  $X_1$  and  $j$ -th category of  $X_2$  (for instance,  $\theta_{11}$  is the probability that an individual does not support the President both before and after his address). The hypothesis that the support for the President remains unchanged is  $\theta_{21} + \theta_{22} = \theta_{12} + \theta_{22}$ . This is equivalent to the hypothesis that the proportion of swings from “Yes” to “No” is equal to the proportion of swings from “No” to “Yes”; that is, this is equivalent to the symmetry hypothesis  $\theta_{12} = \theta_{21}$ . Thus, we can test such a hypothesis by means of the mixed test considering the mathematical setup of this subsection. Suppose  $\alpha = (1, 1, 1, 1)$ . Then, the Bayes factor is given by

$$BF(x_{11}, x_{12}, x_{21}) = (1/2)^{x_{12}+x_{21}} (x_{12} + x_{21} + 1) \binom{x_{12} + x_{21}}{x_{12}}. \tag{58}$$

For the observed data  $x = (20, 17, 10)$ , we obtain the Bayes factor  $BF((20, 17, 10)) = 1.76$ . Considering  $a = b$  (that is  $b/a = 1$ ), we do not reject the null hypothesis since  $BF((20, 17, 10)) > 1$ . In this case, the conditional adaptive significance level and the conditional  $P_T$ -value at the point  $((20, 17, 10))$  are

$$\alpha_T^*((20, 17, 10)) = \sum_{y_{12} \in \{0, \dots, 9\} \cup \{18, \dots, 27\}} \binom{17+10}{y_{12}} (1/2)^{17+10} = 0.122 \text{ and}$$

$$P_T\text{-value}((20, 17, 10)) = \sum_{y_{12} \in \{0, \dots, 10\} \cup \{17, \dots, 27\}} \binom{27}{y_{12}} (1/2)^{27} = 0.248 .$$

Note that  $p_T\text{-value}((20, 17, 10)) > \alpha_T^*((20, 17, 10))$ , as it was expected. Note also that we ordered only 28 elements of the sample space by the Bayes factor to determine the above conditional quantities. To calculate the unconditional ones, we should have ordered all 176,851 points in the sample space.

### 4.3. Test of Independence

Consider the same statistical model as in the previous subsection. However, now we want to evaluate whether there exists (or not) association between the variables  $X_1$  and  $X_2$ . For this purpose, we may test the independence hypothesis between these variables. Consider the joint distribution for  $(X_1, X_2)$  in Table 3 below:

**Table 3.** Joint distribution of  $X_1$  and  $X_2$  given  $\theta$ .

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

The hypotheses to be tested are

$$\begin{aligned} H : \theta_{11} &= (\theta_{11} + \theta_{12})(\theta_{11} + \theta_{21}) \\ A : \theta_{11} &\neq (\theta_{11} + \theta_{12})(\theta_{11} + \theta_{21}) \end{aligned} \tag{59}$$

It is easy to check that hypotheses  $H$  and  $A$  can be rewritten as

$$\begin{aligned} H : \frac{\theta_{11}}{\theta_{11} + \theta_{12}} &= \frac{\theta_{21}}{1 - (\theta_{11} + \theta_{12})} \\ A : \frac{\theta_{11}}{\theta_{11} + \theta_{12}} &\neq \frac{\theta_{21}}{1 - (\theta_{11} + \theta_{12})} \end{aligned} \tag{60}$$

Let us define

$$\lambda_{11} = \frac{\theta_{11}}{\theta_{11} + \theta_{12}} \quad \lambda_{21} = \frac{\theta_{21}}{1 - (\theta_{11} + \theta_{12})} \quad \text{and} \quad \lambda_{12} = \theta_{11} + \theta_{12}$$

and consider the new parameter  $\lambda = ((\lambda_{11}, \lambda_{21}), \lambda_{12})$ , which takes value in  $\Lambda = (0, 1)^2 \times (0, 1)$ . Let  $T(X) = T(X_{11}, X_{12}, X_{21}) = X_{11} + X_{12}$ . Proceeding as in the previous subsections, we obtain the following induced likelihood function for  $\lambda$  generated by  $x = (X_{11}, x_{12}, x_{21})$

$$\tilde{L}_x(\lambda) = \left[ \binom{x_{11} + x_{12}}{x_{11}} \lambda_{11}^{x_{11}} (1 - \lambda_{11})^{x_{12}} \binom{x_{21} + x_{22}}{x_{21}} \lambda_{21}^{x_{21}} (1 - \lambda_{21})^{x_{22}} \right] \left[ \binom{n}{x_{11} + x_{12}} \lambda_{12}^{x_{11} + x_{12}} (1 - \lambda_{12})^{x_{21} + x_{22}} \right] \tag{61}$$

Note that  $T$  is an ancillary statistic for  $(\lambda_{11}, \lambda_{21})$ , as it is a binomial random variable with parameters  $n$  and  $\lambda_{12}$ . In addition, for each possible value  $t$  of the statistic  $T$ , the conditional distribution of  $X$  given  $T(X) = t$  depends on  $\lambda$  only through  $(\lambda_{11}, \lambda_{21})$ .

The prior distribution for  $\lambda$  is such that  $\lambda_{11}, \lambda_{21}$  and  $\lambda_{12}$  are independent Beta random variables with respective parameters  $\alpha_{11}$  and  $\alpha_{12}, \alpha_{21}$  and  $\alpha_{22}$ , and  $\alpha_{11} + \alpha_{12}$  and  $\alpha_{21} + \alpha_{22}$ . Finally, note that under the new parametrization, the independence hypothesis is

$$\begin{aligned} \tilde{H} &: \lambda \in \Lambda_0 \\ \tilde{A} &: \lambda \notin \Lambda_0, \end{aligned} \tag{62}$$

where  $\Lambda_0 = \{(u, v) \in (0, 1)^2 : u = v\} \times (0, 1)$ .

From the results from Section 3, it follows that  $\lambda_{12}$  is an NNP for testing the hypotheses  $\tilde{H}$  versus  $\tilde{A}$  above by using the mixed test. In addition, we only need to compare the conditional  $P_T$  - value with the conditional adaptive significance level to decide between these hypotheses. In a sense, the test for the hypothesis of independence between  $X_1$  and  $X_2$  by using the conditional statistics resembles the test for the hypothesis of homogeneity were the marginal counts  $X_{11} + X_{12}$  and  $n - X_{11} - X_{12}$  fixed beforehand.

Considering as in the previous section  $\alpha = (1, 1, 1, 1)$ , we obtain the following expression for the Bayes factor:

$$BF(x_{11}, x_{12}, x_{21}) = \frac{36 (x_{11} + x_{21} + 1)! (n + 1 - x_{11} - x_{21})!}{x_{11}! x_{12}! x_{21}! x_{22}! (n + 3) \binom{n+2}{x_{11}+x_{21}+1}}.$$

The conditional predictive probability function for  $X$  given  $T(X) = X_{11} + X_{12} = t$ ,  $t = 0, \dots, n$  is given by

$$f_{H,t}(x_{11}, x_{12}, x_{21}) = \frac{6 \binom{t}{x_{11}} \binom{n-t}{x_{21}}}{(n + 3) \binom{n+2}{x_{11}+x_{21}+1}} \mathbb{I}_{T^{-1}(t)}(x_{11}, x_{12}, x_{21}), \tag{63}$$

where  $T^{-1}(t) = \{(y_{11}, y_{12}, y_{21}) \in \mathfrak{X} : T(y_{11}, y_{12}, y_{21}) = y_{11} + y_{12} = t\}$ .

From the above distribution, one may obtain the conditional  $P_T$  - value and the conditional adaptive significance level  $\alpha_T^*$  at each point in the sample space.

**Example 5 (Market’s directional change).** In [34] it is argued that the directional change of the stock market in January signals the directional change of the market for the remainder of the year. Suppose the following Table 4 summarizes the directional changes of the prices of a few stocks in both periods.

**Table 4.** Survey results.

		After January		
		Up	Down	
January change	up	5	0	5
	down	2	8	10
		7	8	15

In this case, the Bayes factor is given by

$$BF(x_{11}, x_{12}, x_{21}) = \frac{36 (x_{11} + x_{21} + 1)! (16 - x_{11} - x_{21})!}{x_{11}! x_{12}! x_{21}! x_{22}! 18 \binom{17}{x_{11}+x_{21}+1}}. \tag{64}$$

For the observed data  $x = (5, 0, 2)$ , we obtain the Bayes factor  $BF((5, 0, 2)) = 0.244$ . Considering  $a = b$  (that is  $b/a = 1$ ), we reject the null hypothesis by using the mixed test since  $BF((5, 0, 2)) < 1$ . That is, the data from only a few stocks reveal that the directional change for

the remainder of the year depends on the directional change in January. Note that although the sample size is small ( $n = 15$ ) and a cell count is equal to zero, the mixed test can be fully performed, as opposed to standard tests for the hypothesis of independence that rely on asymptotic results. In this case, the conditional adaptive significance level and the conditional  $p_T$ -value at the point  $((5, 0, 2))$  are

$$\alpha_T^*((5, 0, 2)) = 0.0109 \quad \text{and} \quad P_T\text{-value}((5, 0, 2)) = 0.0022 .$$

## 5. Discussion

Statistical hypothesis testing is an important quantitative method that may help the daily activity of scientists from different areas of knowledge. However, with recent computational advances, the misuse of standard tests have come to light. Thus, problems with tests of significance and fixed-level tests have brought a growing need for alternative approaches to hypothesis testing that do not have such drawbacks. Among these alternatives, we revisit in this manuscript the mixed test by Pericchi and Pereira, which combines aspects from two opposing viewpoints: the frequentist and the Bayesian. The mixed test satisfies various reasonable properties one desires when performing a test of hypotheses. Here we prove that the mixed test also meets the Non-Informative Nuisance Parameter Principle (NNPP) for simple hypotheses regarding the parameter of interest. The NNPP concerns the question of how to make inferences about a parameter in the presence of Non-Informative Nuisance Parameters: it states that it is irrelevant whether a Non-Informative Nuisance Parameter is known or unknown in order to draw conclusions about a quantity of interest from data. This principle, though important, has not been explored in some depth, and for this reason, we studied it further in hypothesis testing problems. Nuisance parameters typically affect inferences about a parameter of interest: when the variance is unknown, estimation of the mean of a normal distribution and estimation of the parameters of a linear regression model are examples of this.

Adherence of the mixed test to the NNPP allowed for much easier performance of the test, as the calculations involved were significantly reduced. Indeed, decision making between the competing statistical hypotheses was simplified in the three examples we examined: in each situation, conditioning on a suitable statistic and considering conditional versions of the  $p$ -value and the adaptive significance level were revealed to be an advantageous course of action to use the mixed test. The extent to which the adherence of the mixed test to the NNPP is valid and the use of the mixed test can then be made easier remains unanswered in this work. This issue is the goal of future investigation.

**Author Contributions:** All authors have contributed to the conceptualization, formal analysis and writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* [grant 141161/2018-3].

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Proof of Theorem 2.** Suppose there exists a statistic  $T : \mathcal{X} \rightarrow \mathcal{T}$  such that it is  $p$ -sufficient for  $\theta_1$  and  $s$ -ancillary for  $\theta_2$ . Then,

$$\begin{aligned} P(X = x | \theta) &= P(X = x, T(X) = T(x) | \theta) \\ &= P(X = x | T(X) = T(x), \theta_1, \theta_2) P(T(X) = T(x) | \theta_1, \theta_2) \end{aligned} \quad (\text{A1})$$

The result is immediate: as the conditional distribution of  $X$  given  $T(X) = T(x)$  and  $\theta$  depends only on  $\theta_2$ , and the marginal distribution of  $T$  given  $\theta$  depends only on  $\theta_1$ ,

one can write for each  $x \in \mathfrak{X}$ ,  $P(T(X) = T(x) | \theta_1, \theta_2) = L_x^1(\theta_1)$  and  $P(X = x | T(X) = T(x) | \theta_1, \theta_2) = L_x^2(\theta_2)$ . The proof when  $T$  is p-sufficient for  $\theta_2$  and s-ancillary for  $\theta_1$  is analogous.  $\square$

**Proof of Theorem 1.** We first prove the  $(\Rightarrow)$  part of the theorem. Suppose that  $x \in \mathfrak{X}$  is such that  $\theta_1 \perp\!\!\!\perp \theta_2 | X = x$ ; that is, the posterior distribution of  $\theta$  given  $X = x$  can be factored as  $\pi(\theta | x) = \pi_1(\theta_1 | x)\pi_2(\theta_2 | x)$ . Then,

$$\pi(\theta | x) = \pi_1(\theta_1 | x)\pi_2(\theta_2 | x) \Rightarrow \frac{L_x(\theta)\pi(\theta)}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta} = \frac{\int_{\Theta_2} L_x(\theta)\pi(\theta)d\theta_2}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta} \frac{\int_{\Theta_1} L_x(\theta)\pi(\theta)d\theta_1}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta}. \tag{A2}$$

Due to the fact that  $\theta_1 \perp\!\!\!\perp \theta_2$ , it follows from the last equality in (A2) that

$$L_x(\theta)\pi_1(\theta_1)\pi_2(\theta_2) = \pi_1(\theta_1) \int_{\Theta_2} L_x(\theta)\pi_2(\theta_2)d\theta_2 \frac{\pi_2(\theta_2) \int_{\Theta_1} L_x(\theta)\pi_1(\theta_1)d\theta_1}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta}. \tag{A3}$$

The result follows considering, for instance,

$$L_x^1(\theta_1) = \int_{\Theta_2} L_x(\theta)\pi_2(\theta_2)d\theta_2 \quad \text{and} \quad L_x^2(\theta_2) = \frac{\int_{\Theta_1} L_x(\theta)\pi_1(\theta_1)d\theta_1}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta}. \tag{A4}$$

$\square$

Next, we prove the converse. Suppose that the likelihood can be factored as  $L_x(\theta) = L_x^1(\theta_1)L_x^2(\theta_2)$ . Then,

$$\pi(\theta | x) = \frac{L_x(\theta)\pi(\theta)}{\int_{\Theta} L_x(\theta)\pi(\theta)d\theta} = \frac{L_x^1(\theta_1)L_x^2(\theta_2)\pi_1(\theta_1)\pi_2(\theta_2)}{\int_{\Theta_1 \times \Theta_2} L_x^1(\theta_1)L_x^2(\theta_2)\pi_1(\theta_1)\pi_2(\theta_2)d\theta}. \tag{A5}$$

The posterior marginal density of  $\theta_i$  is obtained from (A5) by integrating out the other component. Thus,

$$\pi_i(\theta_i | x) = \frac{L_x^i(\theta_i)\pi_i(\theta_i)}{\int_{\Theta_i} L_x^i(\theta_i)\pi_i(\theta_i)d\theta_i}, \quad i = 1, 2, \tag{A6}$$

and therefore,  $\pi(\theta_1, \theta_2 | x) = \pi_1(\theta_1 | x)\pi_2(\theta_2 | x)$ ; that is,  $\theta_1 \perp\!\!\!\perp \theta_2 | X = x$ .  $\square$

**Proof of Theorem 3.** We first verify that  $\theta_2$  is an NNP for testing  $\bar{H}$  versus  $\bar{A}$  by means of  $\bar{\varphi}^*$ . Recall that

$$\bar{\varphi}^*(x, \theta_2) = 1 \Leftrightarrow \frac{\bar{f}_{\bar{H}}(x, \theta_2)}{\bar{f}_{\bar{A}}(x, \theta_2)} < \frac{b}{a}, \tag{A7}$$

where  $\bar{f}_{\bar{H}}$  ( $\bar{f}_{\bar{A}}$ ) is the predictive distribution for  $(X, \theta_2)$  obtained under  $\bar{H}$  ( $\bar{A}$ ). In this case, the likelihood function generated by  $(x_0, \theta_2)$  for  $\theta_1$  with  $x_0 \in \mathfrak{X}$  such that (1) holds and  $\theta_2 \in \Theta_2$  is

$$\begin{aligned} \bar{L}_{(x_0, \theta_2)}(\theta_1) &= P(X = x_0 | \theta_1, \theta_2)\pi_2(\theta_2 | \theta_1) = L_{x_0}(\theta_1, \theta_2)\pi_2(\theta_2) \\ &= L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)\pi_2(\theta_2). \end{aligned} \tag{A8}$$

Then, the predictive function under the null hypothesis  $\bar{H}$  can be calculated as

$$\bar{f}_{\bar{H}}(x_0, \theta_2) = \int_{\{\theta_0\}} \bar{L}_{(x_0, \theta_2)}(\theta_1)dP_{\bar{H}}(\theta_1), \tag{A9}$$

where  $P_{\bar{H}}$  is degenerate at  $\theta_0$  conditional distribution of  $\theta_1|\theta_1 = \theta_0$ . Thus,

$$\bar{f}_{\bar{H}}(x_0, \theta_2) = L_{x_0}^1(\theta_0)L_{x_0}^2(\theta_2)\pi_2(\theta_2). \tag{A10}$$

In addition, the predictive function under the alternative hypothesis  $\bar{A}$  is given by

$$\begin{aligned} \bar{f}_{\bar{A}}(x_0, \theta_2) &= \int_{\Theta_1} L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)\pi_2(\theta_2)dP_{\bar{H}_1}(\theta_1) \\ &= \int_{\Theta_1} L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)\pi_2(\theta_2)\pi_1(\theta_1)d\theta_1 \\ &= L_{x_0}^2(\theta_2)\pi_2(\theta_2) \int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1. \end{aligned} \tag{A11}$$

Thus, the Bayes factor can be expressed by

$$\begin{aligned} \frac{\bar{f}_{\bar{H}}(x_0, \theta_2)}{\bar{f}_{\bar{A}}(x_0, \theta_2)} &= \frac{L_{x_0}^1(\theta_0)L_{x_0}^2(\theta_2)\pi_2(\theta_2)}{L_{x_0}^2(\theta_2)\pi_2(\theta_2) \int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1} \\ &= \frac{L_{x_0}^1(\theta_0)}{\int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1}. \end{aligned} \tag{A12}$$

Note that Equation (A12) does not depend on  $\theta_2$ . As a result, the test in (A7) does not depend on  $\theta_2$ , and consequently,  $\theta_2$  is an NNP for testing  $\bar{H}$  versus  $\bar{A}$  by means of  $\bar{\varphi}^*$ . Now, we shall determine the test  $\varphi^*$  for  $H$  versus  $A$ . The predictive distribution for  $X$  at  $x_0$  under the null hypothesis is

$$f_H(x_0) = \int_{\Theta_1 \times \Theta_2} L_{x_0}(\theta_1, \theta_2)d\mathbb{P}_H(\theta_1, \theta_2). \tag{A13}$$

It is not difficult to verify that for fixed  $\theta_0 \in \Theta_1$ , the conditional distribution of  $\theta$  given  $\theta_1 = \theta_0$  is such that  $\theta_1$  is degenerate at  $\theta_0$ , and  $\theta_2$  is independent of  $\theta_1$  with density  $\pi_2$ .

Then,

$$\begin{aligned} f_H(x_0) &= \int_{\Theta_1 \times \Theta_2} L_{x_0}(\theta_1, \theta_2)dP_H(\theta_1, \theta_2) \\ &= \int_{\Theta_2} L_{x_0}^1(\theta_0)L_{x_0}^2(\theta_2)\pi_2(\theta_2)d\theta_2 \\ &= L_{x_0}^1(\theta_0) \int_{\Theta_2} L_{x_0}^2(\theta_2)\pi_2(\theta_2)d\theta_2. \end{aligned} \tag{A14}$$

For the alternative hypothesis, we have that

$$\begin{aligned} f_A(x_0) &= \int_{\Theta_1 \times \Theta_2} L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)dP_A(\theta_1, \theta_2) \\ &= \int_{\Theta_1 \times \Theta_2} L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)dP(\theta_1, \theta_2) \\ &= \int_{\Theta_1} \int_{\Theta_2} L_{x_0}^1(\theta_1)L_{x_0}^2(\theta_2)\pi_1(\theta_1)\pi_2(\theta_2)d\theta_1d\theta_2 \\ &= \int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1 \int_{\Theta_2} L_{x_0}^2(\theta_2)\pi_2(\theta_2)d\theta_2. \end{aligned} \tag{A15}$$

Finally,

$$\begin{aligned} \frac{f_H(x_0)}{f_A(x_0)} &= \frac{L_{x_0}^1(\theta_0) \int_{\Theta_2} L_{x_0}^2(\theta_2)\pi_2(\theta_2)d\theta_2}{\int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1 \int_{\Theta_2} L_{x_0}^2(\theta_2)\pi_2(\theta_2)d\theta_2} \\ &= \frac{L_{x_0}^1(\theta_0)}{\int_{\Theta_1} L_{x_0}^1(\theta_1)\pi_1(\theta_1)d\theta_1} = \frac{\bar{f}_{\bar{H}}(x_0, \theta_2)}{\bar{f}_{\bar{A}}(x_0, \theta_2)}. \end{aligned} \tag{A16}$$

Hence,

$$\frac{f_{H_0}(x_0)}{f_{H_1}(x_0)} < \frac{a}{b} \Leftrightarrow \frac{\bar{f}_{H_0}(x_0, \theta_2)}{\bar{f}_{H_1}(x_0, \theta_2)} < \frac{a}{b'} \tag{A17}$$

and consequently,

$$\varphi^*(x_0) = 1 \Leftrightarrow \bar{\varphi}^*(x_0, \theta_2) = 1. \tag{A18}$$

□

**Proof of Corollary 1.** The corollary follows directly from Theorems 2 and 3. □

**Proof of Theorem 4.** From Theorem 3 and Corollary 1, we have that for each  $x_0 \in \mathfrak{X}$ ,

$$\varphi^*(x_0) = 1 \Leftrightarrow BF(x_0) \leq \frac{b}{a}. \tag{A19}$$

Then,

$$\begin{aligned} \varphi^*(x_0) = 1 &\Rightarrow BF(x_0) \leq \frac{b}{a} \Rightarrow \mathfrak{D}(x_0) \subseteq \mathfrak{D} \Rightarrow \\ &\Rightarrow \sum_{x \in \mathfrak{D}(x_0)} f_{H,T(x_0)}(x) \leq \sum_{x \in \mathfrak{D}} f_{H,T(x_0)}(x) \Rightarrow p_T\text{-value}(x_0) \leq \alpha_T^*(x_0). \end{aligned}$$

Thus,

$$\varphi^*(x_0) = 1 \Rightarrow P_T\text{-value}(x_0) \leq \alpha_T^*(x_0). \tag{A20}$$

The converse is proven by the contrapositive.

$$\varphi^*(x_0) = 0 \Rightarrow \frac{b}{a} < BF(x_0) \Rightarrow \mathfrak{D} \cup \{x_0\} \subseteq \mathfrak{D}(x_0).$$

As  $x_0 \notin \mathfrak{D}$  if  $BF(x_0) > \frac{b}{a}$ , we obtain that

$$\varphi^*(x_0) = 0 \Rightarrow \sum_{x \in \mathfrak{D}} f_{H,T(x_0)}(x) + f_{H,T(x_0)}(x_0) \leq \sum_{x \in \mathfrak{D}(x_0)} f_{H,T(x_0)}(x).$$

Since  $x_0 \in \mathfrak{D}_T(x_0)$  and  $BF(x_0) > \frac{b}{a} > 0$ , it follows that  $f_{H,T(x_0)}(x_0) > 0$ . Thus,

$$\varphi^*(x_0) = 0 \Rightarrow \alpha_T^*(x_0) < P_T\text{-value}(x_0), \tag{A21}$$

and consequently,

$$p_T\text{-value}(x_0) \leq \alpha_T^*(x_0) \Rightarrow \varphi^*(x_0) = 1. \tag{A22}$$

From (A20) and (A22), the result follows. □

**Mixed test for symmetry hypothesis for  $r \times r$  contingency tables**

In this case, Table A1 represents the observed frequencies of the cross-classification of  $n$  units by the variables  $X_1$  and  $X_2$ .

**Table A1.** Observed frequencies of  $X_1$  and  $X_2$  in the  $3 \times 3$  case.

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

Let  $X = (X_{ij})$  be the  $(r^2 - 1)$ -dimensional vector of cell counts and  $\theta = (\theta_{ij})$  be the  $(r^2 - 1)$ -dimensional vector of cell probabilities, where  $X_{ij}$  and  $\theta_{ij}$  are self explanatory.

Suppose that  $X$  is a multinomial random vector with parameters  $n$  and  $\theta$ . The likelihood function generated by  $x \in \mathfrak{X}$  for  $\theta$  is given by

$$L_x(\theta) = \frac{n!}{x_{11}! \dots x_{rr}!} \theta_{11}^{x_{11}} \dots \theta_{rr}^{x_{rr}}. \tag{A23}$$

The hypotheses for testing diagonal symmetry are

$$\begin{aligned} H : \theta_{ij} &= \theta_{ji} \quad \forall i \neq j. \\ A : \theta_{ij} &\neq \theta_{ji} \quad \text{for at least one } i \neq j. \end{aligned} \tag{A24}$$

We also assume a prior Dirichlet distribution with parameter  $\alpha = (\alpha_{ij})$  for  $\theta$ . That is,

$$\pi(\theta) = \frac{\Gamma(\alpha_{11} + \dots + \alpha_{rr})}{\Gamma(\alpha_{11}) \dots \Gamma(\alpha_{rr})} \theta_{11}^{\alpha_{11}-1} \dots \theta_{rr}^{\alpha_{rr}-1}. \tag{A25}$$

To perform the mixed test for the symmetry hypothesis, we consider the following reparametrization of the model: we define

$$\begin{aligned} \lambda_{ij} &= \frac{\theta_{ij}}{\theta_{ij} + \theta_{ji}} && \text{for } i < j \\ \lambda_{ij} &= \theta_{ij} + \theta_{ji} && \text{for } i > j \\ \lambda_{ij} &= \theta_{ij} && \text{for } i = j. \end{aligned} \tag{A26}$$

Let  $\lambda = (\lambda_1, \lambda_2)$ , where  $\lambda_1$  is the  $\binom{r^2-r}{2}$ -dimensional vector for which the components are  $\lambda_{ij}$ 's such that  $i < j$ , and  $\lambda_2$  is the  $\binom{r^2-r}{2} + r - 1$ -dimensional vector for which the components are  $\lambda_{ij}$ 's such that  $i \geq j$ . The new parameter space is then  $\Lambda = (0, 1)^{\binom{r^2-r}{2}} \times S_{\binom{r^2-r}{2} + r - 1}$ .

Then, we can rewrite the hypotheses (A24) as

$$\begin{aligned} \tilde{H} : \lambda &\in \Lambda_0 \\ \tilde{A} : \lambda &\in \Lambda_0^c. \end{aligned} \tag{A27}$$

where  $\Lambda_0 = B \times S_{\binom{r^2-r}{2} + r - 1}$ , and  $B$  is the singleton  $B = \{(\frac{1}{2}, \dots, \frac{1}{2})\}$ .

As in previous sections, we consider a statistic  $T$  that is s-ancillary for  $\lambda_1$ :  $T$  is the  $\binom{r^2-r}{2} + r - 1$ -dimensional vector for which the components are the sums  $X_{ij} + X_{ji}$  for  $i < j$  and  $X_{ii}$  for  $i = 1, \dots, r - 1$ . The induced likelihood function for  $\lambda$  generated by  $x$  is

$$\tilde{L}_x(\lambda) = \left[ \prod_{i < j} \binom{x_{ij} + x_{ji}}{x_{ij}} \lambda_{ij}^{x_{ij}} (1 - \lambda_{ij})^{x_{ji}} \right] \left[ \frac{n!}{\prod_i x_{ii}! \prod_{i > j} (x_{ij} + x_{ji})!} \prod_i \lambda_{ii}^{x_{ii}} \prod_{i > j} \lambda_{ij}^{x_{ij} + x_{ji}} \right]. \tag{A28}$$

We can easily see that the likelihood function in (A28) can be factored as  $\tilde{L}_x(\lambda_1, \lambda_2) = \tilde{L}_x^1(\lambda_1) \tilde{L}_x^2(\lambda_2)$ . In addition, the prior distribution for  $\lambda$  is such that  $\lambda_1$  and  $\lambda_2$  are independent:  $\lambda_2$  being a Dirichlet random vector and  $\lambda_1$  a vector of independent Beta random variables. That is,

$$\tilde{\pi}(\lambda) = \tilde{\pi}_1(\lambda_1) \tilde{\pi}_2(\lambda_2) = \prod_{i < j} \frac{\Gamma(\alpha_{ij} + \alpha_{ji})}{\Gamma(\alpha_{ij}) \Gamma(\alpha_{ji})} \lambda_{ij}^{\alpha_{ij}-1} (1 - \lambda_{ij})^{\alpha_{ji}-1} \mathbb{I}_{(0,1)}(\lambda_{ij}) \tilde{\pi}_2(\lambda_2) \tag{A29}$$

From Theorem 3, we have that  $\lambda_2$  is an NNP for testing  $\tilde{H}$  versus  $\tilde{A}$  by means of the mixed test. In addition, the mixed test for  $\tilde{H}$  reduces to the mixed test for the simple hypothesis  $\lambda_1 = (\frac{1}{2}, \dots, \frac{1}{2})$  were  $\lambda_2$  known. From Theorem 4, it follows that we only need to compare the conditional  $P_T$  - value with the conditional adaptive significance level to test  $\tilde{H}$  against  $\tilde{A}$ . From (28), (30) and (A28), we obtain for  $x = (x_{ij}) \in \mathfrak{X}$ :

$$p_{T\text{-value}}(x) = \sum_{y \in \mathfrak{D}_T^*(x)} \prod_{i < j} \binom{y_{ij} + y_{ji}}{y_{ij}} \left(\frac{1}{2}\right)^{y_{ij} + y_{ji}} = \sum_{y \in \mathfrak{D}_T^*(x)} \prod_{i < j} \binom{x_{ij} + x_{ji}}{y_{ij}} \left(\frac{1}{2}\right)^{x_{ij} + x_{ji}} \quad (\text{A30})$$

and

$$\alpha_T^*(x) = \sum_{y \in \mathfrak{D} \cap \mathfrak{D}_T(x)} \prod_{i < j} \binom{y_{ij} + y_{ji}}{y_{ij}} \left(\frac{1}{2}\right)^{y_{ij} + y_{ji}} = \sum_{y \in \mathfrak{D} \cap \mathfrak{D}_T(x)} \prod_{i < j} \binom{x_{ij} + x_{ji}}{y_{ij}} \left(\frac{1}{2}\right)^{x_{ij} + x_{ji}}. \quad (\text{A31})$$

In this case, these conditional quantities are simply determined by the products of binomial-type probabilities.

## References

- Berger, J.; Wolper, R. *The Likelihood Principle*; Institute of Mathematical Statistics: Hayward, CA, USA, 1988.
- Mayo, D. On the Birnbaum Argument for the Strong Likelihood Principle. *Stat. Sci.* **2014**, *29*, 227–239. [[CrossRef](#)]
- Dawid, A. Discussion of “On the Birnbaum Argument for the Strong Likelihood Principle”. *Stat. Sci.* **2014**, *29*, 240–241. [[CrossRef](#)]
- Evans, M. Discussion of “On the Birnbaum Argument for the Strong Likelihood Principle”. *Stat. Sci.* **2014**, *29*, 242–246. [[CrossRef](#)]
- Hannig, J. Discussion of “On the Birnbaum Argument for the Strong Likelihood Principle”. *Stat. Sci.* **2014**, *29*, 254–258. [[CrossRef](#)]
- Bjørnstad, J. Discussion of “On the Birnbaum Argument for the Strong Likelihood Principle”. *Stat. Sci.* **2014**, *29*, 259–260. [[CrossRef](#)]
- Shan, G. *Exact Statistical Inference for Categorical Data*; Academic Press: Cambridge, MA, USA, 2016.
- Pericchi, L.; Pereira, C. Adaptive significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Braz. J. Probab. Stat.* **2016**, *30*, 70–90. [[CrossRef](#)]
- Butler, R.W. Predictive Likelihood Inference with Applications. *J. R. Stat. Soc. Ser. B* **1986**, *48*, 1–38. [[CrossRef](#)]
- Severini, T. Integrated likelihoods for functions of a parameter. *Stat* **2018**, *7*, e212. [[CrossRef](#)]
- Berger, J.; Liseo, B.; Wolpert, R. Integrated Likelihood Methods for Eliminating Nuisance Parameters. *Stat. Sci.* **1999**, *14*, 1–28. [[CrossRef](#)]
- Cox, D.R. Partial likelihood. *Biometrika* **1975**, *62*, 269–276. [[CrossRef](#)]
- Dawid, A.P. On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. R. Stat. Soc. Ser. B* **1975**, *37*, 248–258. [[CrossRef](#)]
- Sprott, D.A. Marginal and conditional sufficiency. *Biometrika* **1975**, *62*, 599–605. [[CrossRef](#)]
- Barndorff-Nielsen, O. Nonformation. *Biometrika* **1976**, *63*, 567–571. [[CrossRef](#)]
- Barndorff-Nielsen, O. *Information and Exponential Families: In Statistical Theory*; Wiley: Chichester, UK, 1978.
- Basu, D. On the Elimination of Nuisance Parameters. *J. Am. Stat. Assoc.* **1977**, *72*, 355–366. [[CrossRef](#)]
- Jørgensen, B. The rules of conditional inference: Is there a universal definition of nonformation? *J. Ital. Stat. Soc.* **1994**, *3*, 355. [[CrossRef](#)]
- Pace, L.; Salvan, A. *Principles of Statistical Inference: From a Neo-Fisherian Perspective*; World Scientific Publishing Company Pte Limited: Singapore, 1997.
- Gannon, M.; Pereira, C.; Polpo, A. Blending Bayesian and Classical Tools to Define Optimal Sample-Size-Dependent Significance Levels. *Am. Stat.* **2019**, *73*, 213–222. [[CrossRef](#)]
- Pereira, C.; Nakano, E.; Fossaluzza, V.; Esteves, L.; Gannon, M.; Polpo, A. Hypothesis Tests for Bernoulli Experiments: Ordering the Sample Space by Bayes Factors and Using Adaptive Significance Levels for Decisions. *Entropy* **2017**, *19*, 696. [[CrossRef](#)]
- Olivera, M. Definição do nível de significância em função do tamanho amostral. Master’s Thesis, IME, Universidade de São Paulo, São Paulo, Brazil, 2014.
- Pereira, B.; Pereira, C. A Likelihood approach to diagnostic test in clinical medicine. *Stat. J.* **2005**, *3*, 77–98.
- Montoya, D.; Irony, T.; Pereira, C.; Whittle, M. An unconditional exact test for the Hardy-Weinberg equilibrium law: Sample space ordering using the Bayes factor. *Genet. Soc. Am.* **2001**, *158*, 875–883.
- Irony, T.; Pereira, C. Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas IME-USP* **1995**, *2*, 27–46.
- DeGroot, M. *Probability and Statistics*; Adson Wesley: Boston, MA, USA, 1986.
- Freeman, P. The role of *p*-values in analysing trial resultss. *Stat. Med.* **1993**, *12*, 15–16. [[CrossRef](#)]
- Bowker, A. A Test for Symmetry in Contingency Tables. *J. Am. Stat. Assoc.* **1948**, *43*, 572–574. [[CrossRef](#)]
- Ireland, C.; Ku, H.; Kullback, S. Symmetry and marginal homogeneity of an  $r \times r$  contingency table. *J. Am. Stat. Assoc.* **1969**, *64*, 1323–1341. [[CrossRef](#)]
- Kullback, S. Marginal Homogeneity of Multidimensional Contingency Tables. *Ann. Math. Stat.* **1971**, *42*, 594–606. [[CrossRef](#)]
- Bernardo, G.; Lauretto, M.; Stern, J. The full Bayesian significance test for symmetry in contingency tables. *AIP Conf. Proc.* **2012**, *1443*, 198–205.
- Agresti, A. *Categorical Data Analysis*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013.

- 
33. Tahata, K.; Tomizawa, S. Symmetry and asymmetry models and decompositions of models for contingency tables. *SUT J. Math.* **2014**, *50*, 131–165. [[CrossRef](#)]
  34. McClave, J.T.; Benson, P.G.; Sincich, T.T. *Statistics for Business and Economics*; Number 519.5; Pearson: London, UK, 2001.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.