mathematics

MDPI

Article

Assessing Multinomial Distributions with a Bayesian Approach

Luai Al-Labadi [1,*], Petru Ciur [1], Milutin Dimovic [1] and Kyuson Lim [2]

[1] Department of Mathematical & Computational Sciences, University of Toronto Mississauga, Toronto, ON L5L 1C6, Canada; petru.ciur@mail.utoronto.ca (P.C.)
[2] Department of Mathematics & Statistics, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada; limk15@mcmaster.ca
* Correspondence: luai.allabadi@utoronto.ca

Abstract: This paper introduces a unified Bayesian approach for testing various hypotheses related to multinomial distributions. The method calculates the Kullback–Leibler divergence between two specified multinomial distributions, followed by comparing the change in distance from the prior to the posterior through the relative belief ratio. A prior elicitation algorithm is used to specify the prior distributions. To demonstrate the effectiveness and practical application of this approach, it has been applied to several examples.

Keywords: dirichlet distribution; hypothesis testing; Kullback–Leibler divergence; multinomial distribution; relative belief inferences

MSC: 62F15; 62F03

check for updates

Citation: Al-Labadi, L.; Ciur, P.; Dimovic, M.; Lim, K. Assessing Multinomial Distributions with a Bayesian Approach. *Mathematics* **2023**, *11*, 3007. https://doi.org/10.3390/math11133007

Academic Editor: Chao Wang

Received: 10 June 2023
Revised: 2 July 2023
Accepted: 4 July 2023
Published: 6 July 2023

1. Introduction

Multinomial distribution tests are a crucial statistical tool in many fields, especially when data can be categorized into multiple groups. These tests were first proposed by Karl Pearson in 1890 and have since been widely used to analyze and make inferences about the probabilities or proportions associated with each category in the multinomial distribution [1].

Let the sample space $\mathcal{A}$ of a random experiment be the union of a finite number k of mutually disjoint sets (categories) $A_1, \ldots, A_k$. Assume that $P(A_j) = \theta_j$, $j = 1, \ldots, k$, where $\sum_{j=1}^{k} \theta_j = 1$. Here θ_j represents the probability that the outcome is an element of the set A_j. The random experiment is to be repeated n independent times. Define the random variables Y_j to be the number of times the outcome is an element of set A_j, $j = 1, \ldots, k$. That is, $Y_1, \ldots, Y_k = n - Y_1 - Y_2 - \cdots - Y_{k-1}$ denote the frequencies with which the outcome belongs to $A_1, \ldots, A_k$, respectively. Then the joint probability mass function (pmf) of $Y_1, \ldots, Y_k$ is the multinomial with parameters $n, \theta_1, \ldots, \theta_k$ [2]. It is desired to test the null hypothesis:

$$H_0^1 : \theta_j = \theta_{j0}, \text{ for } j = 1, \ldots, k \tag{1}$$

against all alternatives, where θ_{j0} are known constants. Within the classical frequentist framework, to test H_0^1, it is common to use the test statistic [3]:

$$\chi^2 = \sum_{j=1}^{k} \frac{(Y_j - n\theta_{j0})^2}{n\theta_{j0}}. \tag{2}$$

It is known that, under H_0^1, the limiting distribution of χ^2 is chi-squared with $k-1$ degrees of freedom. When H_0^1 is true, $n\theta_{j0}$ represents the expected value of Y_j. This implies the observed value χ^2 should not be too large if H_0^1 is true. For a given significance level

α, an approximate test of size α is to reject H_0^1 if the observed $\chi^2 > \chi^2_{k-1}(\alpha)$, where the $\chi^2_{k-1}(\alpha)$ is the $1-\alpha$ quantile of the chi-squared distribution with $k-1$ degrees of freedom; otherwise, fail to reject H_0^1. Other possible tests for H_0^1 include Fisher's exact test and likelihood ratio tests [4].

If there are r independent samples, then the interest is to test whether the r samples come from the same multinomial population or that r multinomial populations are different. Let $A_1, A_2, \ldots, A_k$ denote k possible types of categories in the ith sample, $i = 1, \ldots, k$. Let the probability that an outcome of category A_j will occur for the ith population (or ith sample) be denoted by $\theta_{j|i}$. Note that, $\sum_{j=1}^{k} \theta_{j|i} = 1$ for each $i = 1, \ldots, r$. Moreover, let $Y_{j|i}$ be the number of times the outcome is an element of A_j in sample i. Consider the completely specified hypothesis:

$$H_0^2 : \theta_{j|i} = \theta_{j0|i}, \text{ for } j = 1, 2, \ldots, k. \tag{3}$$

Under H_0^2, the test statistic in (2) can be extended to

$$\chi^2 = \sum_{i=1}^{r} \chi_i^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(Y_{j|i} - n_i\theta_{j0|i})^2}{n_i\theta_{j0|i}}. \tag{4}$$

If H_0^2 is true, then χ^2 in (4) has an approximately chi-squared distribution with $r(k-1)$ degrees of freedom. Likewise, for a given significance level α, an approximate test of size α is to reject H_0^2 if the observed χ^2 is bigger than $\chi^2_{r(k-1)}(\alpha)$; otherwise, fail to reject H_0^2 [5].

A third and more common hypothesis is to test whether the r multinomial populations are the same without specifying the values of the $\theta_{j|i}$. That is, we consider the null hypothesis:

$$H_0^3 : \theta_{j|1} = \theta_{j|2} = \cdots = \theta_{j|r} = \theta_j, \text{ for } j = 1, 2, \ldots, k. \tag{5}$$

The test statistics to test H_0^3 are given by

$$\chi^2 = \sum_{i=1}^{r} \chi_i^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(Y_{j|i} - n_i\hat{\theta}_{j|i})^2}{n_i\hat{\theta}_{j|i}}, \tag{6}$$

where $\hat{\theta}_{j|i} = \frac{\sum_{i=1}^{r} Y_{j|i}}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} Y_{j|i}}{n}$. Here, n_i denotes the sample size of sample i and $\sum_{i=1}^{r} Y_{j|i}$ represents the total in category A_j. Note that $\hat{\theta}_{j|i}$ represents the pooled maximum likelihood estimator (MLE) of θ_j under H_0^3. It is known that the limiting distribution of χ^2 in (6) is a chi-squared distribution with $(r-1)(k-1)$ degrees of freedom. So, for a given significance level α, an approximate test of size α is to reject H_0^3 if the observed $\chi^2 > \chi^2_{(r-1)(k-1)}(\alpha)$; otherwise, fail to reject H_0^3 [5]. It is worth mentioning that several other frequentist methods for testing the multinomial distribution have been proposed, utilizing different distance measures. These methods include the Euclidean distance proposed by [6], the smooth total variation distance introduced by [7], and ϕ-divergences discussed by [8]. These approaches provide alternative ways to assess the goodness-of-fit of the multinomial distribution using distance metrics.

Refs. [9–13] made early advances in Bayesian methods for analyzing categorical data, focusing on smoothing proportions in contingency tables and inference about odds ratios, respectively. These methods typically employed conjugate beta and Dirichlet priors. Ref. [14,15] extended these methods to develop Bayesian analogs of small-sample frequentist tests for 2×2 tables, also using such priors. Ref. [16] recommended the use of the uniform prior for predictive inference, but other priors were also suggested by discussants of his paper. The Jeffreys prior is the most commonly used prior for binomial inference, partially due to its invariance to the scale of measurement for the parameter. Reference priors (see [17]), such as the Jeffreys prior for the binomial parameter (see [18]), are viable

options, but their specification can be computationally complex. Ref. [10] may have been the first to utilize an empirical Bayesian approach with contingency tables, estimating parameters in gamma and log-normal priors for association factors. Empirical Bayes involves estimating the prior distribution from the observed data itself and is particularly useful when dealing with large amounts of data. Refs. [19,20] derived integral expressions for the posterior distributions of the difference, ratio, and odds ratio under independent beta priors. Ref. [19] introduced Bayesian highest posterior density (HPD) confidence intervals for these measures. The HPD approach ensures that the posterior probability matches the desired confidence level, and the posterior density is higher inside the interval than outside. Ref. [21] discussed Bayesian confidence intervals for association parameters in 2×2 tables. They argued that to achieve good coverage performance in the frequentist sense across the entire parameter space, it is advisable to use relatively diffuse priors. Even uniform priors are often too informative, and they recommended the use of the Jeffreys prior. Bayesian methods for analyzing categorical data have been extensively surveyed in the literature, including comprehensive reviews by [22,23] with a focus on contingency table analysis. Refs. [24–26] proposed tests based on Bayesian nonparametric methods using Dirichlet process priors.

We build on the recent work of [27] by extending their Bayesian approach for hypothesis testing on one-sample proportions based on Kullback–Leibler divergence and relative belief ratio, using a uniform (0, 1) prior on binomial proportions, to multinomial distributions. Our goal is to provide a comprehensive Bayesian approach for testing hypotheses H_0^1, H_0^2, and H_0^3. We derive distance formulas and use the Dirichlet distribution as a prior on probabilities. To ensure proper values of the prior's hyperparameters, we employ the elicitation algorithm developed by [28]. The proposed approach offers several advantages, including computational simplicity, ease of interpretation, evidence in favor of the null hypothesis, and no requirement to specify a significance level.

The paper is structured as follows. Section 2 provides an overview of the relative belief ratio inference and KL divergence. Section 3 details the proposed approach, including the formulas and computational algorithms. In Section 4, several examples are presented to illustrate the approach. Finally, Section 5 contains concluding remarks and discussions.

2. Relevant Background

2.1. Inferences Using Relative Belief

Ref. [29] introduced the relative belief ratio, which has become a popular tool in statistical hypothesis testing theory. Several works have employed this approach, including [30–35].

Suppose we have a statistical model with a density function $\{f_\theta(\boldsymbol{y}) : \theta \in \Theta\}$ with respect to the Lebesgue measure on the parameter space Θ. Let $\pi(\theta)$ be a prior on Θ. After observing the data $\boldsymbol{y}$, the posterior distribution of θ can be expressed as

$$\pi(\theta|\boldsymbol{y}) = \frac{f_\theta(\boldsymbol{y})\pi(\theta)}{m(\boldsymbol{y})},$$

where $m(\boldsymbol{y}) = \int_\Theta f_\theta(\boldsymbol{y})\pi(\theta)d\theta$.

Assume that the goal is to draw inferences about the parameter θ. If the prior $\pi(\cdot)$ and the posterior $\pi(\cdot|\boldsymbol{y})$ are continuous at θ, then the relative belief ratio for a hypothesized value θ_0 of θ can be expressed as follows:

$$RB(\theta_0|\boldsymbol{y}) = \frac{\pi(\theta_0|\boldsymbol{y})}{\pi(\theta_0)},$$

the ratio of the posterior density to the prior density at θ_0. In other words, $RB(\theta_0|\boldsymbol{y})$ quantifies how the belief in θ_0 being the true value has changed from prior to posterior. It is worth noting that when $\pi(\cdot)$ and $\pi(\cdot|\boldsymbol{y})$ are discrete, the relative belief ratio is defined through limits, and further details can be found in [29].

The relative belief ratio $RB(\theta_0|\boldsymbol{y})$ provides a measure of evidence for θ_0 being the true value. A value of $RB(\theta_0|\boldsymbol{y}) > 1$ indicates evidence in favor of θ_0 being the true value, whereas $RB(\theta_0|\boldsymbol{y}) < 1$ indicates evidence against θ_0 being the true value. If $RB(\theta_0|\boldsymbol{y}) = 1$, there is no evidence in either direction.

Once the relative belief ratio is calculated, it is important to determine the strength of the evidence in favor of or against $H_0 : \theta = \theta_0$. A common way to quantify this is by computing the tail probability [29]:

$$Str(\theta_0|\boldsymbol{y}) = \Pi(RB(\theta|\boldsymbol{y}) \leq RB(\theta_0|\boldsymbol{y})|\boldsymbol{y}) = \int_{\{\theta\in\Theta: RB(\theta|\boldsymbol{y})\leq RB(\theta_0|\boldsymbol{y})\}} \pi(\cdot|\boldsymbol{y})\, d\theta, \tag{7}$$

where $\Pi(\cdot|\boldsymbol{y})$ in (7) is the posterior cumulative distribution function with posterior density $\pi(\cdot|\boldsymbol{y})$. Therefore, equation (7) represents the posterior probability that the true value of θ has a relative belief ratio no greater than that of the hypothesized value θ_0. When $RB(\theta_0|\boldsymbol{y}) < 1$, there is evidence against θ_0. A small value of $Str(\theta_0|\boldsymbol{y})$ indicates a high posterior probability that the true value has a relative belief ratio greater than $RB(\theta_0|\boldsymbol{y})$, indicating strong evidence against θ_0. Conversely, when $RB(\theta_0|\boldsymbol{y}) > 1$, there is evidence in favor of θ_0. A large value of $Str(\theta_0|\boldsymbol{y})$ indicates a low posterior probability that the true value has a relative belief ratio greater than $RB(\theta_0|\boldsymbol{y})$, indicating strong evidence in favor of θ_0. A small value of $Str(\theta_0|\boldsymbol{y})$ indicates weak evidence in favor of θ_0.

2.2. KL Divergence

The KL divergence, also referred to as relative entropy, is a measure of dissimilarity between two probability distributions that quantifies how far apart they are from each other. It was introduced by Solomon Kullback and Richard Leibler in 1951. Let P and Q be two discrete cumulative distribution functions (cdf's) on the same probability space φ, with corresponding probability mass functions (pmf's) p and q (with respect to the counting measure). The KL divergence between p and q is given by:

$$d(p,q) = \sum_{x\in\varphi} p(x) \log\left(\frac{p(x)}{q(x)}\right).$$

The KL divergence is always non-negative, and it attains its minimum value when $p = q$ almost surely. This property makes it a useful tool in many areas of machine learning and information theory, such as hypothesis testing, model selection, and clustering. One interpretation of the KL divergence is that it measures how much information is lost when using Q to approximate P. It is worth noting that the KL divergence is not symmetric: $d(p,q)$ and $d(q,p)$ are generally not equal. Therefore, it is important to specify which distribution is the "true" or "target" distribution and which is the "approximating" or "predicted" distribution for some applications when using KL divergence in practice, as noted by [36].

The following lemma is essential to the proposed approach.

Lemma 1. *Let $p(y_1, y_2, \ldots, y_r) = p_1(y_1)p_2(y_2)\cdots p_r(y_r)$ and $q(y_1, y_2, \ldots, y_r) = q_1(y_1)q_2(y_2)\cdots q_r(y_r)$, where $p_i(y_i)$ and $q_i(y_i)$ are probability mass functions with supports $y_i = 1, \ldots, n_i, i = 1, \ldots, r$. Then*

$$d(p,q) = \sum_{i=1}^{r} d(p_i, q_i).$$

Proof. We have

$$
\begin{aligned}
d(p,q) &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p(y_1, y_2, \ldots, y_r) \log \frac{p(y_1, y_2, \ldots, y_r)}{q(y_1, y_2, \ldots, y_r)} \\
&= \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) \log \frac{p_1(y_1) \cdots p_r(y_r)}{q_1(y_1) \cdots q_r(y_r)} \\
&= \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) [\log p_1(y_1) + \cdots + \log p_r(y_r) - \log q_1(y_1) - \cdots - \log q_r(y_r)] \\
&= \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) \log p_1(y_1) + \cdots + \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) \log p_r(y_r) \\
&\quad - \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) \log q_1(y_1) - \cdots - \sum_{y_1=0}^{n_1} \cdots \sum_{y_r=0}^{n_r} p_1(y_1) \cdots p_r(y_r) \log q_r(y_r).
\end{aligned}
$$

Since, for $i = 1, \ldots, r$, $\sum_{y_i=1}^{n_i} p_i(y_i) = \sum_{y_i=1}^{n_i} q_i(y_i) = 1$, we have

$$
\begin{aligned}
d(p,q) &= \sum_{y_1=0}^{n_1} p_1(y_1) \log p_1(y_1) + \cdots + \sum_{y_r=0}^{n_r} p_r(y_r) \log p_r(y_r) \\
&\quad - \sum_{y_1=0}^{n_1} p_1(y_1) \log q_1(y_1) - \cdots - \sum_{y_r=0}^{n_r} p_r(y_r) \log q_r(y_r) \\
&= \sum_{y_1=0}^{n_1} p_1(y_1) \log \frac{p_1(y_1)}{q_1(y_1)} + \cdots + \sum_{y_r=0}^{n_r} p_r(y_r) \log \frac{p_r(y_r)}{q_r(y_r)} \\
&= d(p_1, q_1) + \cdots + d(p_r, q_r).
\end{aligned}
$$

□

3. The Approach

3.1. Bayesian One-Sample Multinomial

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_k) \sim \text{multinomial}(n, \theta_1, \ldots, \theta_k)$. The joint pmf of $Y_1, \ldots, Y_k$ is given by

$$
p(y_1, \ldots, y_k) = \binom{n}{y_1, y_2, \ldots, y_k} \prod_{j=1}^{k} \theta_j^{y_j}, \tag{8}
$$

where $\binom{n}{y_1, y_2, \ldots, y_k} = \frac{n!}{y_1! \cdots y_k!}$, $\sum_{j=1}^{k} \theta_j = 1$, and $\sum_{j=1}^{k} y_j = n$.

To test the null hypothesis H_0^1 as defined in (1), we first compute the Kullback–Leibler (KL) divergence between $p(y_1, \ldots, y_k)$ and the pmf under H_0^1 represented by

$$
q(y_1, \ldots, y_k) = \binom{n}{y_1, y_2, \ldots, y_k} \prod_{j=1}^{k} \theta_{j0}^{y_i}. \tag{9}
$$

Here, θ_{j0} denotes the null hypothesis value for θ_j. The following proposition provides the formula for the KL divergence between p and q.

Proposition 1. *Let $p(y_1, \ldots, y_k)$ and $q(y_1, \ldots, y_k)$ be two joint probability mass functions as defined in (8) and (9), respectively. We have,*

$$
d(p,q) = n \sum_{j=1}^{k} \left[\theta_j \log \left(\frac{\theta_j}{\theta_{j0}} \right) \right].
$$

Proof. Let the support of Y_j be $1, 2, \ldots, n_j$, $j = 1, \ldots, k$. We have

$$\begin{aligned} d(p,q) &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \log \frac{p(y_1, \ldots, y_k)}{q(y_1, \ldots, y_k)} \\ &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \log \frac{\binom{n}{y_1, y_2, \ldots, y_k} \prod_{j=1}^{k} \theta_j^{y_j}}{\binom{n}{y_1, y_2, \ldots, y_k} \prod_{j=1}^{k} \theta_{j0}^{y_j}} \\ &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \log \prod_{j=1}^{k} \left[\frac{\theta_j}{\theta_{j0}} \right]^{y_j}. \end{aligned}$$

Using the properties of logarithmic function, we get

$$\begin{aligned} d(p,q) &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \sum_{j=1}^{k} y_j \log \left[\frac{\theta_j}{\theta_{j0}} \right] \\ &= \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \times y_1 \times \log \left[\frac{\theta_1}{\theta_{01}} \right] \\ &\quad \cdots + \sum_{y_1=0}^{n_1} \cdots \sum_{y_k=0}^{n_k} p(y_1, \ldots, y_k) \times y_k \times \left[\frac{\theta_k}{\theta_{k0}} \right] \\ &= \mathbb{E}[Y_1] \times \log \left[\frac{\theta_1}{\theta_{01}} \right] + \cdots + \mathbb{E}[Y_k] \times \log \left[\frac{\theta_k}{\theta_{k0}} \right] \\ &= \sum_{j=1}^{k} \mathbb{E}[Y_j] \log \left[\frac{\theta_j}{\theta_{j0}} \right]. \end{aligned}$$

Since the marginal probability mass function of Y_j, $j = 1, \ldots, k$, is the binomial with parameters n and θ_j, we get

$$d(p,q) = \sum_{j=1}^{k} n\theta_j \log \left[\frac{\theta_j}{\theta_{j0}} \right] = n \sum_{j=1}^{k} \left[\theta_j \log \left(\frac{\theta_j}{\theta_{j0}} \right) \right].$$

□

To connect the distance formula presented in Proposition 1 with the test statistic χ^2 in (2), we use the Taylor series expansion of the function $f(x) = x \log \frac{x}{x_0}$ about x_0. This gives us

$$f(x) = (x - x_0) + 0.5(x - x_0)^2 \frac{1}{x_0} + \cdots$$

If H_0^1 is true and n is large, then we can approximate the distance $d(p,q)$ as

$$d(p,q) \approx n \sum_{j=1}^{k} (\theta_j - \theta_{j0}) + 0.5n \sum_{j=1}^{k} \frac{(\theta_j - \theta_{j0})^2}{\theta_{j0}}. \tag{10}$$

Since the probabilities sum to 1, the first term in (10) equals 0. The second term in (10) can be expressed as

$$0.5 \sum_{j=1}^{k} \frac{(n\theta_j - n\theta_{j0})^2}{n\theta_{j0}} = 0.5 \sum_{j=1}^{k} \frac{(E(Y_j) - n\theta_{j0})^2}{n\theta_{j0}}.$$

This shows a direct connection between the KL divergence and χ^2.

For the proposed Bayesian test, the probabilities $\theta_1, \ldots, \theta_k$ are now random. The suggested prior for the joint probabilities $(\theta_1, \ldots, \theta_k)$ is the Dirichlet distribution with parameters $\alpha_1, \ldots \alpha_k$. That is,

$$(\theta_1, \theta_2, \ldots, \theta_k) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_k). \tag{11}$$

To elicit the prior, we use the elicitation algorithm developed by [28], which requires some domain knowledge to provide a lower bound for each θ_i. For convenience, we have made this algorithm available on Shiny at the following link: https://bayesian-chi-square-test.shinyapps.io/Dirichlet_process_Kyuson_lim/ (accessed on 23 May 2023). For comparison purposes, we also considered the non-informative (uniform) prior, and Jeffreys prior; see Section 4. For the proposed Bayesian approach, when $(\theta_1, \ldots, \theta_k)$ has the prior defined in (11), we put

$$D = n \sum_{j=1}^{k} \left[\theta_j \log \left(\frac{\theta_j}{\theta_{j0}} \right) \right]. \tag{12}$$

We also have that the posterior distribution of $(\theta_1, \ldots, \theta_k)$ given the observed data $\boldsymbol{y} = (y_1, \ldots, y_k)$ is Dirichlet$(\alpha_1 + y_1, \alpha_2 + y_2, \ldots, \alpha_k + y_k)$. We write

$$D_y = n \sum_{j=1}^{k} \left[\theta_j \log \left(\frac{\theta_j}{\theta_{j0}} \right) \right]. \tag{13}$$

Note that,

$$E(\theta_j|\boldsymbol{y}) = \frac{\alpha_j + y_j}{\sum_{j=1}^{k}(\alpha_j + y_j)} = \frac{\alpha_j + y_j}{n + \sum_{j=1}^{k} \alpha_j} = \frac{n}{n + \sum_{j=1}^{k} \alpha_j} \frac{y_j}{n} + \left(1 - \frac{n}{n + \sum_{j=1}^{k} \alpha_j} \right) \frac{\alpha_j}{\sum_{j=1}^{k} \alpha_j},$$

which is a convex combination between the sample proportion (MLE) and the prior mean. As $n \to \infty$, the weak law of large numbers ensures that $E(\theta_j|\boldsymbol{y})$ converges in probability to the true value of θ_j. Hence, if H_0^1 is true, then $D_y \overset{a.s.}{\to} 0$. Conversely, if H_0^1 is false, then $D_y \overset{a.s.}{\to} c$, where $c > 0$. Proposition 1 establishes that $d(p, q) = 0$ if and only if $\theta_j = \theta_{j0}$. Therefore, testing $H_0^1 : \theta_j = \theta_{j0}$ is equivalent to testing $d(p, q) = 0$. It follows that when H_0^1 is true, the distribution of D_y should be more concentrated around 0 than that of D. So, the proposed test involves comparing the distributions of D and D_y around 0 using the relative belief ratio:

$$RB_D(0|\boldsymbol{y}) = \frac{\pi_D(0|\boldsymbol{y})}{\pi_D(0)}, \tag{14}$$

where $\pi_D(0)$ and $\pi_D(0|\boldsymbol{y})$ represent the probability density functions of D and D_y, respectively. If $RBD(0|\boldsymbol{y}) > 1$, it provides evidence in favor of H_0 (since the distribution of D_y is more concentrated around 0 than that of D). If $RB_D(0|\boldsymbol{y}) < 1$, there is evidence against H_0^1 (as the distribution of D_y is less concentrated around 0 than that of D). Additionally, we compute the strength of evidence $Str_D(0|\boldsymbol{y}) = \Pi_D(RB(d|\boldsymbol{y}) \leq RB(0|\boldsymbol{y})|\boldsymbol{y})$, where $\Pi_D(\cdot|\boldsymbol{y})$ is the cumulative distribution function of D_y. As $\pi_D(\cdot|\boldsymbol{y})$ and $\pi_D(\cdot)$ in (14) have no closed forms, $RB_D(0|\boldsymbol{y})$ and $Str_D(0|\boldsymbol{y})$ need to be approximated. The following Algorithm 1 summarizes the steps required to test H_0^1.

Algorithm 1 RB test for H_0^1

(i) Generate $(\theta_1, \theta_2, \ldots, \theta_k)$ from Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_k)$ based on the algorithm of [28] and compute D as defined in (12).

(ii) Repeat step (ii) to obtain a sample of r_1 values of D.

(iii) Generate $(\theta_1, \theta_2, \ldots, \theta_k)$ given the observed data $\boldsymbol{y} = (y_1, \ldots, y_k)$ from Dirichlet$(\alpha_1 + y_1, \alpha_2 + y_2, \ldots, \alpha_k + y_k)$ and compute D_y as defined in (13).

(iv) Repeat step (iii) to obtain a sample of r_2 values of D_y.

(v) Compute the relative belief ratio and the strength as follows:

(a) Let L be a positive number. Let $\hat{F}_D$ denote the empirical cdf of D based on the prior sample in (3) and for $i = 0, \ldots, L$, let $\hat{d}_{i/L}$ be the estimate of $d_{i/L}$, the (i/L)-the prior quantile of D. Here $\hat{d}_0 = 0$, and $\hat{d}_1$ is the largest value of D. Let $\hat{F}_D(\cdot \mid \boldsymbol{y})$ denote the empirical cdf based on D_y. For $d \in [\hat{d}_{i/L}, \hat{d}_{(i+1)/L})$, estimate $RB_D(d \mid \boldsymbol{y}) = \pi_D(d|\boldsymbol{y})/\pi_D(d)$ by

$$\widehat{RB}_D(d \mid \boldsymbol{y}) = L\{\hat{F}_D(\hat{d}_{(i+1)/L} \mid \boldsymbol{y}) - \hat{F}_D(\hat{d}_{i/L} \mid \boldsymbol{y})\}, \tag{15}$$

the ratio of the estimates of the posterior and prior contents of $[\hat{d}_{i/L}, \hat{d}_{(i+1)/L})$. Thus, we estimate $RB_D(0 \mid \boldsymbol{y}) = \pi_D(0|\boldsymbol{y})/\pi_D(0)$ by $\widehat{RB}_D(0 \mid \boldsymbol{y}) = L\hat{F}_D(\hat{d}_{p_0} \mid \boldsymbol{y})$ where $p_0 = i_0/L$ and i_0 are chosen so that i_0/L is not too small (typically $i_0/L \approx 0.05$).

(b) Estimate the strength $\Pi_D(RB_D(d \mid \boldsymbol{y}) \leq RB_D(0 \mid \boldsymbol{y}) \mid \boldsymbol{y})$ by the finite sum

$$\sum_{\{i \geq i_0 : \widehat{RB}_D(\hat{d}_{i/L} \mid \boldsymbol{y}) \leq \widehat{RB}_D(0 \mid \boldsymbol{y})\}} (\hat{F}_D(\hat{d}_{(i+1)/L} \mid \boldsymbol{y}) - \hat{F}_D(\hat{d}_{i/L} \mid \boldsymbol{y})). \tag{16}$$

For fixed L, as $r_1 \to \infty, r_2 \to \infty$, then $\hat{d}_{i/L}$ converges almost surely to $d_{i/L}$ and (15) and (16) converge almost surely to $RB_D(d \mid \boldsymbol{y})$ and $\Pi_D(RB_D(d \mid \boldsymbol{y}) \leq RB_D(0 \mid \boldsymbol{y}) \mid \boldsymbol{y})$, respectively. See [34] for the details.

3.2. Bayesian r-Sample Multinomial Test with a Completely Specified Null Hypothesis

Consider r independent samples $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_r$ where each $\boldsymbol{Y}_i = (Y_{1|i}, \ldots, Y_{k|i})$ follows a multinomial distribution with parameters n_i and $\boldsymbol{\theta}_i = (\theta_{1|i}, \ldots, \theta_{k|i})$, where $\sum_{j=1}^k \theta_{j|i} = 1$ for $i = 1, \ldots, r$. Here, $\theta_{j|i}$ denotes the probability of an outcome falling in category j for the ith sample, and $Y_{j|i}$ represents the number of times the outcome falls in category j in the ith sample. The null hypothesis to be tested is $H_0^2 : \theta_{j|i} = \theta_{j0|i}$ for $j = 1, 2, \ldots, k$, where $\theta_{j0|i}$ are known constants.

Let the joint distribution of $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_r$ be

$$p(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r) = \prod_{i=1}^r p(\boldsymbol{y}_i) = \prod_{i=1}^r \left\{ \binom{n_i}{y_{1|i}, y_{2|i}, \ldots, y_{k|i}} \prod_{j=1}^k \theta_{j|i}^{y_{j|i}} \right\}. \tag{17}$$

The proposed test is based on measuring the KL divergence between p and

$$q(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r) = \prod_{i=1}^r q(\boldsymbol{y}_i) = \prod_{i=1}^r \left\{ \binom{n_i}{y_{1|i}, y_{2|i}, \ldots, y_{k|i}} \prod_{j=1}^k \theta_{j0|i}^{y_{j0|i}} \right\}. \tag{18}$$

The following proposition provides the expression for the KL divergence between p and q. The proof follows directly from Lemma 1 and Proposition 1.

Proposition 2. *Let $p(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r)$ and $q(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r)$ be the two joint probability mass functions as defined in (17) and (18), respectively. Then*

$$d(p,q) \quad = \quad \sum_{i=1}^{r}\left\{ n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{\theta_{j|i}}{\theta_{j0|i}}\right)\right]\right\}.$$

Proposition 2 provides a connection between the KL divergence formula and the test statistic χ^2 in (6). Using a Taylor series expansion, we can approximate the distance $d(p,q)$ as follows:

$$d(p,q) \quad \approx \quad 0.5\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{\left(n_i\theta_{j|i} - n_i\theta_{j0|i}\right)^2}{n_i\theta_{j0|i}} = 0.5\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{\left(E(Y_{j|i}) - n_i\theta_{j0|i}\right)^2}{n_i\theta_{j0|i}},$$

which reveals a close connection to χ^2.

For the proposed Bayesian test of H_0^2, we adopt the prior $(\theta_{1|i}, \ldots, \theta_{k|i}) \sim$ Dirichlet $(\alpha_{1|i}, \alpha_{2|i}, \ldots, \alpha_{k|i})$, and use the algorithm developed in [28] to elicit the hyperparameters $\alpha_{j|i}$ for $j = 1, \ldots, k$ and $i = 1, \ldots, r$. In this case, we define the divergence measure as

$$D = \sum_{i=1}^{r}\left\{ n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{\theta_{j|i}}{\theta_{j0|i}}\right)\right]\right\}, \tag{19}$$

where $\theta_{j0|i}$ is the hypothesized value of $\theta_{j|i}$ under the null hypothesis.

The posterior distribution of $(\theta_{1|i}, \ldots, \theta_{k|i})$ given the observed data $\boldsymbol{y}_i = (y_{1|i}, \ldots, y_{k|i})$ is then Dirichlet$(\alpha_{1|i} + y_{1|i}, \alpha_{2|i} + y_{2|i}, \ldots, \alpha_{k|i} + y_{k|i})$. In this case, the divergence measure becomes

$$D_y = \sum_{i=1}^{r}\left\{ n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{\theta_{j|i}}{\theta_{j0|i}}\right)\right]\right\}. \tag{20}$$

The following algorithm outlines the steps required to test H_0^2 using the proposed Bayesian test (Algorithm 2):

Algorithm 2 RB test for H_0^2

(i) For $i = 1, \ldots, r$, generate $(\theta_{1|i}, \theta_{2|i}, \ldots, \theta_{k|i})$ from Dirichlet$(\alpha_{1|i}, \alpha_{2|i}, \ldots, \alpha_{k|i})$ based on the algorithm of [28] and compute D as defined in (19).

(ii) Repeat step (i) to obtain a sample of r_1 values of D.

(iii) For $i = 1, \ldots, r$, generate $(\theta_{1|i}, \theta_{2|i}, \ldots, \theta_{k|i})$ given the observed data $\boldsymbol{y}_i = (y_{1|i}, \ldots, y_{k|i})$ from Dirichlet$(\alpha_{1|i} + y_{1|i}, \alpha_{2|i} + y_{2|i}, \ldots, \alpha_{k|i} + y_{k|i})$ and compute D_y as defined in (20).

(iv) Repeat step (iii) to obtain a sample of r_2 values of D_y.

(v) Compute the relative belief ratio and strength as described in Algorithm 1.

3.3. Bayesian Test for Homogeneity in r-Sample Multinomial Data

Consider r independent samples $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_r$, where for $i = 1, \ldots, r$, $\boldsymbol{Y}_i = (Y_{1|i}, \ldots, Y_{k|i})$ $\sim$ multinomial$(n_i, \theta_{1|i}, \ldots, \theta_{k|i})$ with $\sum_{j=1}^{k}\theta_{j|i} = 1$. To test the null hypothesis H_0^3 as defined in (5), it is required to measure the KL divergence between p and q as defined in (17) and (18) with $\theta_{j0|i}$ is replaced by θ_j. This requirement is offered in the following proposition.

Proposition 3. *Consider the probability mass functions p and q as defined in (17) and (18) with $\theta_{j0|i}$ is replaced by θ_j. Then*

$$d(p,q) = \sum_{i=1}^{r}\left\{n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{\theta_{j|i}}{\theta_j}\right)\right]\right\} \tag{21}$$

and

$$\theta_j^\star = \arg\min_{\theta_j} d(p,q) = \frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}. \tag{22}$$

Proof. (21) follows directly from Lemma 2 by setting $\theta_{j|i} = \theta_j$. To prove (22), we use we use Lagrange multiplier with the constraint $\sum_{j=1}^{r}\theta_j = 1$:

$$L = L(\theta_j,\lambda) = \sum_{i=1}^{r}\left\{n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{\theta_{j|i}}{\theta_j}\right)\right]\right\} + \lambda\left(\sum_{j=1}^{r}\theta_j - 1\right).$$

Now,

$$\frac{\partial L}{\partial \theta_j} = -\sum_{i=1}^{r} n_i\frac{\theta_{j|i}}{\theta_j} + \lambda, \;\; j = 1,\ldots,k.$$

Setting $\frac{\partial L}{\partial \theta_j} = 0$ gives

$$\theta_j = \frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{\lambda}, \;\; j = 1,\ldots,k.$$

Summing over both sides and applying the constraint gives

$$\lambda = \sum_{j=1}^{k}\sum_{i=1}^{r} n_i\theta_{j|i} = \sum_{i=1}^{r} n_i \sum_{j=1}^{k}\theta_{j|i} = \sum_{i=1}^{r} n_i = n.$$

Hence,

$$\theta_j^\star = \frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}.$$

□

Note that $\theta_j^\star$ represents the weighted average of $\theta_{j|i}$. Substituting $\theta_j^\star$ into (21), we get

$$\widehat{d}(p,q) = \sum_{i=1}^{r}\left\{n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{n\theta_{j|i}}{\sum_{i=1}^{r} n_i\theta_{j|i}}\right)\right]\right\}, \tag{23}$$

which is equal to 0 under H_0^3. We can also establish a connection between (23) and the test statistic χ^2 in (6). By Taylor series expansion, when n is large and under H_0^3, we have

$$\begin{aligned}
\widehat{d}(p,q) &\approx 0.5\sum_{i=1}^{r} n_i \sum_{j=1}^{k} \frac{\left(\theta_{j|i} - \frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}\right)^2}{\frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}} \\
&= 0.5\sum_{i=1}^{r}\sum_{j=1}^{k} \frac{\left(n_i\theta_{j|i} - n_i\frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}\right)^2}{n_i\frac{\sum_{i=1}^{r} n_i\theta_{j|i}}{n}} \\
&= 0.5\sum_{i=1}^{r}\sum_{j=1}^{k} \frac{\left(E(Y_{j|i}) - n_i\frac{\sum_{i=1}^{r} E(Y_{j|i})}{n}\right)^2}{n_i\frac{\sum_{i=1}^{r} E(Y_{j|i})}{n}},
\end{aligned}$$

which is closely linked to χ^2. The proposed Bayesian test for H_0^3 uses the prior described in Section 3.1. We write

$$\widehat{D} = \sum_{i=1}^{r}\left\{n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{n\theta_{j|i}}{\sum_{i=1}^{r} n_i\theta_{j|i}}\right)\right]\right\}. \tag{24}$$

Moreover, for the posterior distribution of $(\theta_{1|i},\ldots,\theta_{k|i})$ given the observed data $\boldsymbol{y}_i = (y_{1|i},\ldots,y_{k|i})$, we write

$$\widehat{D}_y = \sum_{i=1}^{r}\left\{n_i \sum_{j=1}^{k}\left[\theta_{j|i}\log\left(\frac{n\theta_{j|i}}{\sum_{i=1}^{r} n_i\theta_{j|i}}\right)\right]\right\}. \tag{25}$$

The following algorithm is used to test H_0^3 (Algorithm 3).

Algorithm 3 RB test for H_0^3

(i) For $i = 1,\ldots,r$, generate $(\theta_{1|i},\theta_{2|i},\ldots,\theta_{k|i})$ from Dirichlet$(\alpha_{1|i},\alpha_{2|i},\ldots,\alpha_{k|i})$ based on the algorithm of [28] and compute $\widehat{D}$ as defined in (24).

(ii) Repeat step (i) to obtain a sample of r_1 values of $\widehat{D}$.

(iii) For $i = 1,\ldots,r$, generate $(\theta_{1|i},\theta_{2|i},\ldots,\theta_{k|i})$ given the observed data $\boldsymbol{y}_i = (y_{1|i},\ldots,y_{k|i})$ from Dirichlet$(\alpha_{1|i}+y_{1|i},\alpha_{2|i}+y_{2|i},\ldots,\alpha_{k|i}+y_{k|i})$ and compute $\widehat{D}_y$ as defined in (25).

(iv) Repeat step (iii) to obtain a sample of r_2 values of $\widehat{D}_y$.

(v) Compute the relative belief ratio and strength using Algorithm 1, but replace D and D_y with $\widehat{D}$ and $\widehat{D}_y$, respectively.

4. Examples

This section presents three examples that demonstrate the effectiveness of our approach in evaluating H_0^1, H_0^2, and H_0^3. We use Algorithms 1–3, with fixed values of $L = 20$, $i_0 = 1$, and $r_1 = r_2 = 10^4$. To further investigate the efficacy of our approach, we consider three different prior distributions: uniform prior, Jeffreys prior, and an elicited prior based [28]. Additionally, we compute the p-values using the test statistics discussed in Section of this paper. The approach was implemented using R (version 4.2.1), and the code is available upon request from the corresponding author.

Example 1 (*Rolling Die*; [5])**.** *We roll a die 60 times and seek to test whether it is unbiased, that is, whether $H_0^1 : \theta_j = 1/6$ for $j = 1,\ldots,k$. The Table 1 below presents the recorded data:*

Table 1. Data of Example 1.

	1	2	3	4	5	6	Total
Observed	8	11	5	12	15	6	60

We will use a Bayesian approach to address this problem. We employ three priors: the uniform prior represented by Dirichlet $(1,1,1,1,1,1)$, Jeffreys prior represented by Dirichlet $(0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$, and the elicited prior Dirichlet (5.83, 5.83, 5.83, 5.83, 5.83, 5.83) obtained using the algorithm proposed by [28], with a lower bound of 0.05 applied to all probabilities. It is worth noting that setting the lower bound in [28] to 0 yields the uniform prior. Additionally, we will include the p-value for the corresponding frequentist test as a reference. The results of our analysis are presented in Table 2. Clearly, both the proposed Bayesian approach, considering the three priors, and the frequentist approach lead to the same conclusion. It should be noted that the uniform prior and the Jefferey prior have a wider spread around zero compared to the elicited prior. As a result, they have higher relative belief ratios in this example. However, this is not practically significant in our case as we calibrate the relative belief ratio through the strength. See Figure 1.

Table 2. The RB and its strength (Str) for Example 1.

Prior	RB (Strength)	Decision
Uniform	15.125 (1)	Strong evidence in favor of H_0^1
Jeffreys	19.824 (1)	Strong evidence in favor of H_0^1
Evan et al.	1.900 (1)	Strong evidence in favor of H_0^1
p-value	0.3027	Fail to reject H_0^1 at $\alpha = 0.05$

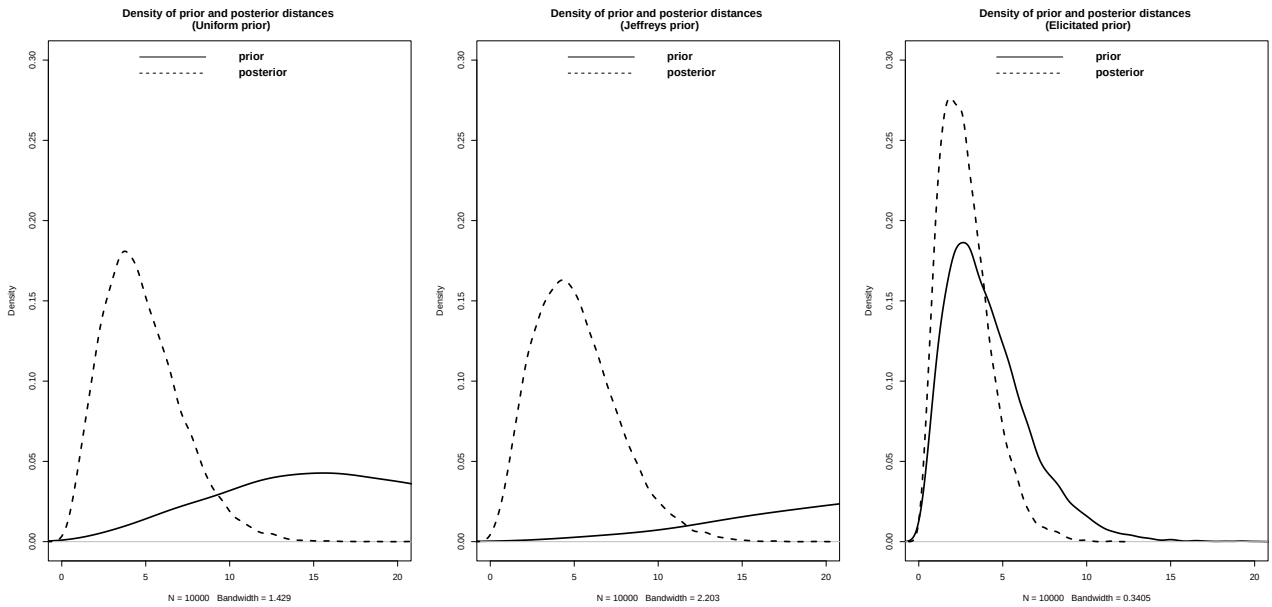

Figure 1. Density plot of distances in Example 1.

Example 2 (*Operation Trial* [5])**.** *In a system consisting of four independent components, let $\theta_{j|i}$ denote the probability of successful operation of the ith component, $i = 1,2,3,4$. We will test the null hypothesis $H_0^2 : \theta_{1|1} = 0.9, \theta_{2|1} = 0.1, \theta_{1|2} = 0.9, \theta_{2|2} = 0.1, \theta_{1|3} = 0.8, \theta_{2|3} = 0.2$, $\theta_{1|4} = 0.8$, $\theta_{2|4} = 0.2$, given that in 50 trials, the components operated as follows (Table 3):*

Table 3. Data of Example 2.

Component	Successful	Failure
1	40	10
2	48	2
3	45	5
4	40	10

We use the priors: Dirichlet $(1,1)$, Dirichlet $(0.5,0.5)$, and Dirichlet $(33.38,5.62)$. We obtain the latter prior using algorithm of [28], with lower bounds of $\theta_{1|i} = 0.7$ and $\theta_{2|i} = 0.1$ for all $i = 1,2,3,4$. Table 4 displays the results of our analysis. As in Example 1, both the uniform prior and the Jefferey prior exhibit less concentration around zero when compared to the elicited prior. This, in turn, leads to a notably different conclusion than that of the elicited prior and the p-value calculated using the chi-square test. See also Figure 2.

Table 4. The RB and its strength (Str) for Example 2.

Prior	RB (Strength)	Decision
Uniform	20(1)	Strong evidence in favor of H_0^2
Jeffreys	19.998 (0.000)	Weak evidence in favor of H_0^2
Evan et al.	0.592 (0.047)	Strong evidence against H_0^2
p-value	0.030	Fail to reject H_0^1 at $\alpha = 0.05$

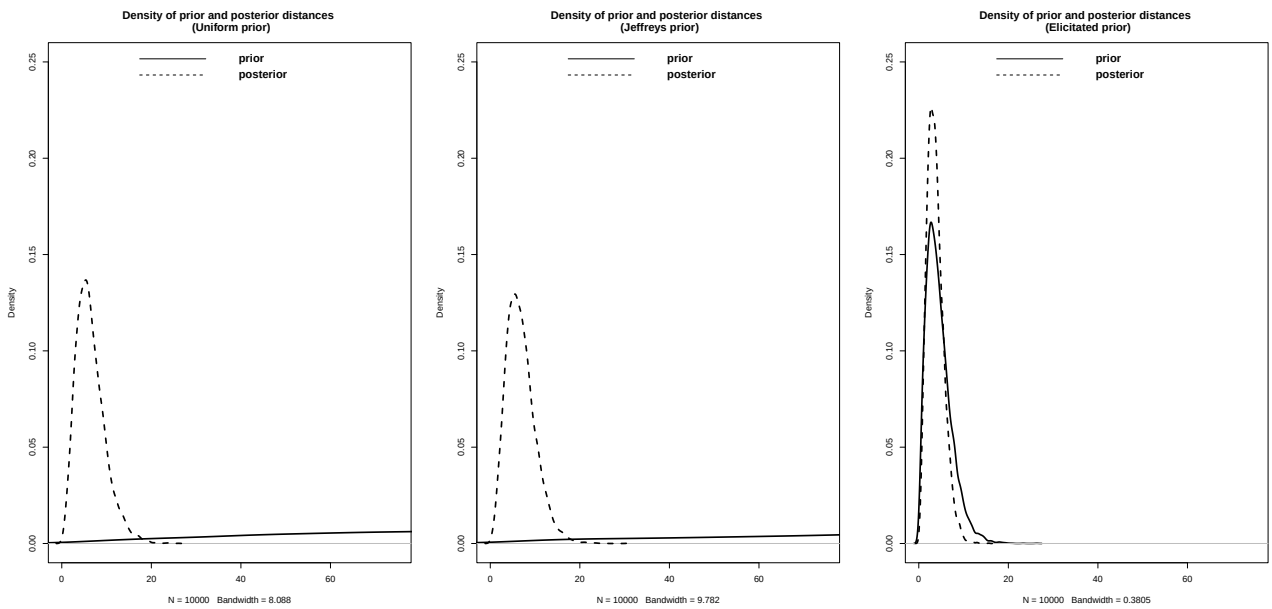

Figure 2. Density plot of distances in Example 2.

Example 3 (*Clinical Trial*; [37])**.** *A study was performed to determine whether the type of cancer differed between blue-collar, white-collar, and unemployed workers. A sample of 100 of each type of worker diagnosed as having cancer was categorized into one of three types of cancer. The results are shown in Table 5. See also Table 12.6 of [37]. The hypothesis to be tested is that the proportions of the three cancer types are the same for all three occupation groups. That is,* $H_0^3 : \theta_{j|1} = \theta_{j|2} = \theta_{j|3}$ *for all j (types of cancer), where* $\theta_{j|i}$ *is the probability of occupation i having cancer type j.*

Table 5. Data of Example 3.

Occupation	Type of Cancer			
	Lung	**Stomach**	**Other**	**Total**
Blue collar	53	17	30	100
White Collar	10	67	23	100
Unemployed	30	30	40	100

Similar to the previous two examples, we utilize the uniform prior Dirichlet $(1,1,1)$, Jeffreys prior Dirichlet $(0.5,0.5,0.5)$, and the elicited prior Dirichlet $(3,3,3)$. We obtained the elicited prior using the algorithm of [28] by setting a lower bound of 0.05 for all probabilities. Table 6 summarizes the results of our analysis. Similar to the previous examples, Jeffreys prior is not sufficiently concentrated around zero, which makes it inefficient when there is evidence against H_0. See Figure 3.

Table 6. The RB and its strength (Str) for Example 3.

Prior	RB (Strength)	Decision
Uniform	0.102 (0.010)	Strong evidence against H_0
Jeffreys	1.874 (0.129)	Weak evidence in favor of H_0
Evan et al.	0.012 (0.000)	Strong evidence against H_0
p-value	0.000	Reject H_0^1 at $\alpha = 0.05$

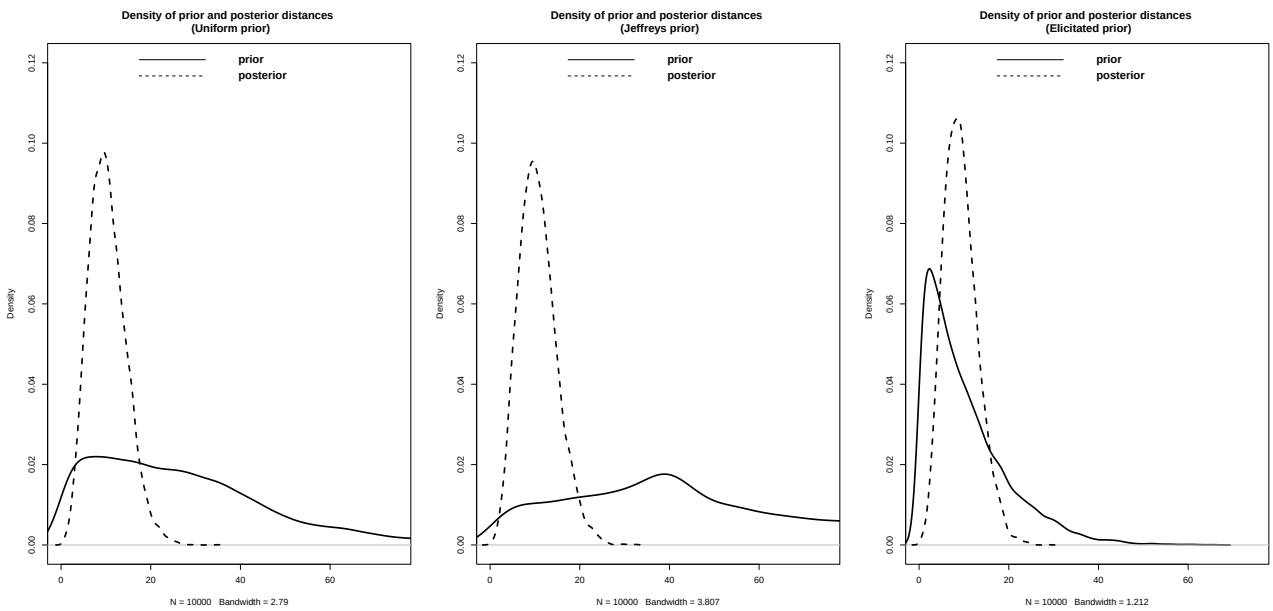

Figure 3. Density plot of distances in Example 3.

5. Concluding Remarks

This study presents a Bayesian method for testing hypotheses related to multinomial distributions. Our approach involves calculating the Kullback–Leibler divergence between two multinomial distributions and comparing the change in distance from the prior to the posterior through the relative belief ratio. To specify the prior distributions, we employ a prior elicitation algorithm. We recommend avoiding the use of Jeffreys prior or the uniform prior unless there is a valid reason to use them. Through several examples, we

demonstrated the effectiveness of our approach. Future research may expand our approach to include testing for independence and other related cases.

Author Contributions: Methodology, L.A.-L. and P.C.; Software, P.C., M.D. and K.L.; Resources, K.L.; Writing—review & editing, M.D.; Supervision, L.A.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: There are no data associated with this paper.

Acknowledgments: We would like to express our sincere gratitude to the Editor and the anonymous referees for their valuable and constructive comments. Their insightful feedback and suggestions have greatly contributed to the improvement of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Agresti, A. *An Introduction to Categorical Data Analysis*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2007.
2. Hogg, R.V.; McKean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*, 8th ed.; Person: Boston, MA, USA, 2019.
3. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **1900**, *50*, 157–175. [CrossRef]
4. Rice, J.A. *Mathematical Statistics and Data Analysis*, 3rd ed.; Brookes/Cole: Belmont, MA, USA, 2007.
5. Bain, L.J.; Engelhardt, M. *Introduction to Probability and Mathematical Statistics*; Duxbury Press: North Scituate, MA, USA, 1992.
6. Frey, J. An exact multinomial test for equivalence. *Can. J. Stat.* **2009**, *37*, 47–59.
7. Ostrovski, V. Testing equivalence of multinomial distributions. *Stat. Probab. Lett.* **2017**, *124*, 77–82.
8. Alba-Fernández, M.V.; Jiménez-Gamero, M.D. Equivalence Tests for Multinomial Data Based on ϕ-Divergences. In *Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision and Control*; Balakrishnan, N., Gil, M.Á., Martín, N., Morales, D., Pardo, M.d.C., Eds.; Springer: Cham, Switzerland, 2023; Volume 445. [CrossRef]
9. Good, I.J. The population frequencies of species and the estimation of population parameters. *Biometrika* **1953**, *40*, 237–264. [CrossRef]
10. Good, I.J. On the estimation of small frequencies in contingency tables. *J. R. Stat. Soc. Ser. B* **1956**, *18*, 113–124. [CrossRef]
11. Good, I.J. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*; MIT Press: Cambridge, MA, USA, 1965. [CrossRef]
12. Good, I.J. A Bayesian significance test for multinomial distributions (with Discussion). *J. R. Stat. Soc. Ser. B* **1967**, *29*, 399–431. [CrossRef]
13. Lindley, D.V. The Bayesian analysis of contingency tables. *Ann. Math. Stat.* **1964**, *35*, 1622–1643. [CrossRef]
14. Altham, P.M.E. Exact Bayesian analysis of a 2×2 contingency table, and Fisher's "exact" significance test. *J. R. Stat. Soc. Ser.* **1969**, *31*, 261–269.
15. Altham, P.M.E. The analysis of matched proportions. *Biometrika* **1971**, *58*, 561–576.
16. Geisser, S. On prior distributions for binary trials. *Am. Stat.* **1984**, *38*, 244–247. [CrossRef]
17. Bernardo J.M.; Ramón, J.M. An introduction to Bayesian reference analysis: Inference on the ratio of multinomial parameters. *Statistician* **1998**, *47*, 101–135. [CrossRef]
18. Bernardo J.M.; Smith, A.F.M. *Bayesian Theory*; Wiley: Hoboken, NJ, USA, 1994. [CrossRef]
19. Hashemi, L.; Nandram, B.; Goldberg, R. Bayesian analysis for a single 2×2 table. *Stat. Med.* **1997**, *16*, 1311–1328. [CrossRef]
20. Nurminen, M.; Mutanen, P. Exact Bayesian analysis of two proportions. *Scand. J. Stat.* **1987**, *14*, 67–77. [CrossRef]
21. Agresti, A.; Min, Y. Frequentist performance of Bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics* **2005**, *61*, 515–523. [CrossRef]
22. Agresti, A.; Hitchcock, D.B. Bayesian inference for categorical data analysis. *Stat. Methods Appl.* **2005**, *14*, 297–330. [CrossRef]
23. Leonard, T.; Hsu, J.S.J. The Bayesian Analysis of Categorical Data—A Selective Review. In *Aspects of Uncertainty*; Freeman, P.R., Smith, A.F.M., Eds.; A Tribute to D. V. Lindley; Wiley: New York, NY, USA, 1994; pp. 283–310. [CrossRef]
24. Carota, C. A family of power-divergence diagnostics for goodness-of-fit. *Can. J. Stat.* **2007**, *35*, 549–561.
25. Kim, M.; Nandram, B.; Kim, D.H. Nonparametric Bayesian test of homogeneity using a discretization approach. *J. Korean Data Inf. Sci. Soc.* **2018**, *29*, 303–311. [CrossRef]
26. Quintana, F.A. Nonparametric Bayesian analysis for assessing homogeneity in $k \times l$ contingency tables with fixed right margin totals. *J. Am. Stat. Assoc.* **1998**, *93*, 1140–1149.
27. Al-Labadi, L.; Cheng, Y.; Fazeli-Asl, F.; Lim, K.; Weng, Y. A Bayesian one-sample test for proportion. *Stats* **2022**, *5*, 1242–1253. [CrossRef]
28. Evans, M.; Guttman, I.; Li, P. Prior elicitation, assessment and inference with a Dirichlet prior. *Entropy* **2017**, *19*, 564. [CrossRef]

29. Evans, M. *Measuring Statistical Evidence Using Relative Belief*; Monographs on Statistics and Applied Probability 144; Taylor & Francis Group, CRC Press: Boca Raton, RL, USA, 2015. [CrossRef] [PubMed]
30. Abdelrazeq, I.; Al-Labadi, L.; Alzaatreh, A. On one-sample Bayesian tests for the mean. *Statistics* **2020**, *54*, 424–440. [CrossRef]
31. Al-Labadi, L. The two-sample problem via relative belief ratio. *Comput. Stat.* **2021**, *36*, 1791–1808.
32. Al-Labadi, L.; Berry, S. Bayesian estimation of extropy and goodness of fit tests. *J. Appl. Stat.* **2020**, *49*, 357–370. [CrossRef]
33. Al-Labadi, L.; Evans, M. Optimal robustness results for relative belief inferences and the relationship to prior-data conflict. *Bayesian Anal.* **2017**, *12*, 705–728. [CrossRef]
34. Al-Labadi, L.; Evans, M. Prior-based model checking. *Can. J. Stat.* **2018**, *46*, 380–398. [CrossRef]
35. Al-Labadi, L.; Patel, V.; Vakiloroayaei, K.; Wan, C. Kullback–Leibler divergence for Bayesian nonparametric model checking. *J. Korean Stat. Soc.* **2020**, *50*, 272–289. [CrossRef]
36. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006. [CrossRef]
37. Freund, R.J.; Wilson, W.J.; Mohr, D.L. *Statistical Methods*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2010.