Received: 3 December 2021 | Revised: 12 October 2022 | Accepted: 13 October 2022

DOI: 10.1002/asmb.2727

RESEARCH ARTICLE

WILEY

A Bayesian homogeneity test for comparing Poisson populations

Francisco Javier Girón[1] | **María Martel–Escobar**[2] | **Francisco–José Vázquez–Polo**[2]

[1]Real Academia de Ciencias Exactas, Físicas y Naturales de España, Madrid, Spain

[2]Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Canary Islands, Spain

Correspondence
Francisco–José Vázquez–Polo, Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Canary Islands.
Email: fjvpolo@dmc.ulpgc.es

Abstract

For a wide class of daily applications in industrial quality control, there may be interest in comparing several Poisson means. A large catalogue of frequentist procedures for this hypothesis testing problem is available. However, some common drawbacks of them are their low power, interpretation of the *p*-values for multiple comparison, among many others. In this paper, we present a unified Bayesian approach to the problem based on a model selection approach developed using a product partition clustering model. The posterior probabilities for models obtained are derived directly from the associated Bayes factors which are calculated by considering a simple hierarchical prior structure which has a quasi–closed form, easily computed by numerical procedures. This approach constitutes a readily implementable alternative to frequentist multiple testing procedures where uncertainty concerning all possible "types of homogeneity" is ignored. The proposed methodology allows for an intuitive interpretation based directly on posterior probabilities of the partitions involved in the testing problem. We illustrate its performance with three real data sets.

KEYWORDS

Bayesian test, homogeneity, Poisson, hierarchical priors

1 | INTRODUCTION

The use of the Poisson distribution in a large variety of scientific fields has widely been recognized and the literature is full of case studies where the Poisson model is well suited for data modeling where events of interest are counted randomly per unit of space or time. The problem of comparison of counts arising from Poisson distributions is also of interest in the statistical literature[1] as well as in the quality control one.[2] In industrial quality control, the Poisson distribution and its extensions is a common statistical model in many disciplines from CUSUM charts to monitored count–type observations correlated over time where a Poisson marginal is appropriate[3] to reliability,[4] to name just a couple.

Since the seminal paper by Przyborowski and Wilenski[5] introducing a conditional test for testing the difference between two Poisson means, many methods based on conditional binomial distributions has been developed.[6,7] Also, either normal approximations for assessing confidence interval for the parameter difference, $\lambda_1 - \lambda_2$ or numerical procedures for computing confidence intervals for the ratio λ_1/λ_2 are given in Cousins.[8] Based on Anscombe's variance stabilizing transformation, Brown and Zhao[9] proposed an alternative procedure for testing whether a sample observations comes from a simple Poisson distribution when two or more populations are considered. In general, this problem is related to the homogeneity test for different statistical models and the multiple testing options when the null hypothesis of homogeneity is rejected. There are a large number of procedures to multiple testing in the literature, all of them with several concerns and difficulties such as the right interpretation of different *p*-values for multiple comparison, power behavior of the Pearson's χ^2–statistic, among many others.[10] Other alternatives as the parametric bootstrap has also been considered. Chiu and Wang[11] found that common asymptotic tests perform similarly and their powers are close to those of their parametric bootstrap counterparts when the observed counts are large enough. There are just a few available papers using a Bayesian approach to the problem.[12-14] The above references, frequentist and Bayesian ones, show that there is no single procedure that performs better than the others.

In this paper, we address the problem of hypothesis testing about several Poisson means from a different standpoint, considering that it can be regarded as a clustering problem and thus, jointly with the original homogeneity problem, all possible data configurations can be tested using a Bayesian procedure from which the posterior probabilities of each cluster model is computed. We present a unified Bayesian approach that overcomes the drawbacks of other procedures. The following motivating example help us to focus on the problem.

1.1 | Motivating example

Table 1, extracted from Nelson[15] corresponds to an study developed in Poirier et al.[16] on potential carcinogenicity of several chemical compounds, where four groups of mice were treated with different doses of methyl iodide: zero (labeled as $i = 1$), low (corresponds to 0.06 mmoles per kg of mouse and labeled as $i = 2$), medium (0.15 mmoles per kg, $i = 3$) and high (0.31 mmoles per kg, $i = 4$). A Poisson model is assumed for each group. The interest is focused on testing the null hypothesis that all the four Poisson means are equal.

Nelson(1987) conducted a χ^2−test with $4 - 1 = 3$ degrees of freedom, based on the approximation to the multinomial distribution. As Nelson pointed out (example 7, page 177):

> *Therefore, the null hypothesis of equal Poisson processes is not rejected. Three of the four expected values are less than five, but the fear that a small expected value (denominator) will cause a spuriously large contribution to chi–square is of no concern in this example because this result falls far short of being statistically significant.*

Following the Bayesian clustering procedure proposed in this article, the results obtained offer a much more detailed and accurate reading for the Poisson means comparison purposes. The most probable configuration corresponds to the hypothesis that all samples are homogeneous and has posterior probability 0.416 (see Table 3 in Section 3.1). This result is in some accordance with Nelson.[15] However, the structure of the clustering partition induced by the dataset in Table 2 indicates that different forms of homogeneity could be present. To cluster the samples, we follow the product partition model suggested by Barry and Hartigan.[17]

Let $\boldsymbol{x}_i = (n_i, t_i)$ denote the observed data in sample $i = 1, \dots, 4$. There is a one–to–one correspondence between every partition of the parametric space and the partition of the sample data. For instance, the null hypothesis of equal Poisson means $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ or equivalently $\{\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}\}$, corresponds to the single data set $\{\{\boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \boldsymbol{x}_3 \cup \boldsymbol{x}_4\}\}$, which we denote $\{\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$. At the other extreme, when all samples are heterogeneous among them, the configuration $\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$ is equivalent to the multiple hypothesis $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$.

When k Poisson populations are compared, $r \in \{1, \dots, k\}$ type of groupings or clusters are possible (in Table 2 referred to as homogeneity, type 2–heterogeneity, type 3–heterogeneity and heterogeneity, respectively). The total number of possible partitions is given by the Bell number[18] $B_k = \sum_{r=1}^{k} S(k, r)$, where $S(k, r)$ is the Stirling number of the second kind

TABLE 1 Number of tumors (t_i) detected in (n_i) mice

	Doses			
	1	**2**	**3**	**4**
n_i	154	19	20	11
t_i	34	4	6	5

TABLE 2 Cluster configurations for dataset in Nelson[15]

Configurations	**Parameters relationship**
Homogeneity	
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$
Type 2–Homogeneity	
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 \neq \lambda_3 = \lambda_4$
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_3 \neq \lambda_2 = \lambda_4$
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_4\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}\}$	$\lambda_1 = \lambda_4 \neq \lambda_2 = \lambda_3$
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_3 = \lambda_4$
$\{\{\boldsymbol{x}_2\}, \{\boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_2 \neq \lambda_2 = \lambda_3 = \lambda_4$
$\{\{\boldsymbol{x}_3\}, \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4\}\}$	$\lambda_3 \neq \lambda_1 = \lambda_2 = \lambda_4$
$\{\{\boldsymbol{x}_4\}, \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}\}$	$\lambda_4 \neq \lambda_1 = \lambda_2 = \lambda_3$
Type 3–Homogeneity	
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 \neq \lambda_3 \neq \lambda_4$
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_3 \neq \lambda_2 \neq \lambda_4$
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_4\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}\}$	$\lambda_1 = \lambda_4 \neq \lambda_2 \neq \lambda_3$
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_3 \neq \lambda_4$
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_4\}, \{\boldsymbol{x}_3\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_4 \neq \lambda_3$
$\{\{\boldsymbol{x}_2\}, \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_2 \neq \lambda_1 \neq \lambda_3 = \lambda_4$
Heterogeneity	
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$

which indicates the number of (ordered) partitions of the sample of size k into r clusters.[19] Table 2 shows the 15 partitions corresponding to $k = 4$, that is, $B_4 = 15$.

The selection of any one model among all of them is then an alternative to frequentist multiple testing procedures where uncertainty concerning all possible configurations is ignored. The procedure is not affected by the 'size of means', that is, discussions based on whether the rates are small or large as in the frequentist methodology do not arise in our procedure. Including, the case where a preponderance of zeros be present in the sample.

The rest of the paper is organized as follows. Section 2 proposes the novel approach to homogeneity test for several Poisson populations. The third section provides three real–dataset examples where we illustrate the use and versatility of this approach. Section 4 presents the results of numerical simulations on the frequentist evaluation of the Bayesian procedure presented. In Section 5, we derive a natural extensions of the methodology proposed. Finally, Section 6 presents some concluding remarks. In the Supplementary Materials, `Mathematica` and `R` codes can be downloaded, which we have developed in order to make the novel approach more accesible and reproducible.

2 THE MODEL

Following Chang et al.,[20] let us assume that we have independent Poisson observations

$$X_i \sim \text{Po}(\lambda_i), \quad i = 1, 2, \ldots, k(\geq 2), \tag{1}$$

with $\lambda_i = n_i \gamma_i$ and where X_i represent the number of events of type–i taking place over time or space of size n_i. The problem of interest is to test

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_k \quad \text{versus} \quad H_a : \gamma_i \neq \gamma_j \quad \text{for some } \ i \neq j.$$

That is, we are interested in comparing the rate of occurrence of events γ_i's. However in practice, for given population i, we have samples of size n_i from independent $\text{Po}(\gamma_i)$ and by sufficiency one can look only at $S_i = \sum_{j=1}^{n_i} X_{ij}$ which follows a Poisson distribution with mean $n_i\gamma_i$, and therefore the role of X_i in (1) can be played by S_i. For simplicity then, we will use the simplest notation in model (1).

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ be independent samples of k Poisson populations $\text{Po}(\lambda_1), \dots, \text{Po}(\lambda_k)$ with sample sizes $n_1, \dots, n_k$. Also, let us consider $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_m)$ any configuration of length m, with $1 \le m \le k$, that is, $\boldsymbol{z}$ is a partition of sample $\boldsymbol{x}$ in m blocks. For instance, Table 2 contains all configurations $\boldsymbol{z}$ of length $m = 1, \dots, 4$. We denote by $t_i = \sum_{j=1}^{n_i} x_{ij}$ the total number of counts in sample $\boldsymbol{x}_i$, $i = 1, \dots, k$, $n = \sum_{i=1}^{k} n_i$ and $t = \sum_{i=1}^{k} t_i$.

It is apparent that the likelihood function have different expressions on each cluster configuration of the samples $\boldsymbol{z}$. Each model induced by $\boldsymbol{z}$ indicates a different heterogeneity structure of the sampling model, and its posterior probability informs us about the uncertainty for this configuration. The comparison of Poisson means is then equivalent to a homogeneity test with null model $H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda$. Here, we assume that $\lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda$, and the parameters of the competing model (hypothesis) $H_{1\boldsymbol{z}}$ are denoted by $\lambda_1, \dots, \lambda_m$, in order to simplify the notation and keeping in mind that they are different.

In the Bayesian approach, for each configuration structure in a set of candidate models, model uncertainty is quantified based on its posterior model probability, and Bayes factors play a crucial role discriminating between competing models. The Bayes factor for comparing model H_0 vs $H_{1\boldsymbol{z}}$ is given by

$$B_{0\boldsymbol{z}} = \frac{f_0(\boldsymbol{x})}{f_1(\boldsymbol{z})}, \tag{2}$$

where $f_0(\boldsymbol{x})$ and $f_1(\boldsymbol{z})$ are the marginal distributions of the data $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively, given by

$$f_0(\boldsymbol{x}) = \int_0^\infty f_0(\boldsymbol{x}|\lambda)\pi(\lambda)\, d\lambda, \tag{3}$$

and

$$f_1(\boldsymbol{z}) = \int_0^\infty f_1(\boldsymbol{z}|\lambda)\pi(\lambda)\, d\lambda, \tag{4}$$

where $\pi(\lambda)$ is the prior distribution on the hyperparameters λ, and $f_1(\boldsymbol{z}|\lambda)$ is defined below, and it depends on the hierarchical model that relates the $\lambda_1, \dots, \lambda_m$ of partition $\boldsymbol{z}$ with the parameter λ of the null hypothesis.

In general, the best model (equivalently, configuration $\boldsymbol{z}^\star$) is the one that maximizes the posterior probability of the models

$$\Pr(\boldsymbol{z}|\boldsymbol{x}) = \frac{\Pr(\boldsymbol{z})B_{0\boldsymbol{z}}}{\sum_{\boldsymbol{z}\in C_k} \Pr(\boldsymbol{z})B_{0\boldsymbol{z}}}, \quad \forall \boldsymbol{z} \in C_k, \tag{5}$$

where $\Pr(\boldsymbol{z})$ denotes the prior probability of configuration $\boldsymbol{z}$ and C_k the set of all cluster configurations given k samples. A common objective prior for hypothesis testing purposes is the discrete uniform prior $\Pr(\boldsymbol{z}) = \frac{1}{B_k}$, for all $z \in C_k$.

2.1 | The marginals distributions

The likelihood function of the data $\boldsymbol{x}$ under the null hypothesis is

$$f_0(\boldsymbol{x}|\lambda) = \prod_{i=1}^{k} \text{Po}(\boldsymbol{x}_i|\lambda) = \frac{1}{\prod_{i=1}^{k}\prod_{j=1}^{n_i} x_{ij}!} \lambda^{\sum_{i=1}^{k} t_i} \exp(-\lambda \sum_{i=1}^{k} n_i) \propto \lambda^t \exp(-n\lambda), \tag{6}$$

On the other hand, for any configuration $\boldsymbol{z}$ of length m, the likelihood function is

$$f_1(\boldsymbol{z}|\lambda_1, \dots, \lambda_m) = f_1(\boldsymbol{z}_1, \dots, \boldsymbol{z}_m|\lambda_1, \dots, \lambda_m) = \prod_{s=1}^{m} f(\boldsymbol{z}_s|\lambda_s) \propto \prod_{s=1}^{m} \lambda_s^{t_s} \exp(-m_s\lambda_s), \tag{7}$$

where t_s and m_s are the total number of counts and the length of vector of sample $\boldsymbol{z}_s$, respectively. Observe that the proportionally constant in (7) coincides with that in expression (6).

However, the likelihood function $f_1(\boldsymbol{z}|\lambda)$ in (4) cannot be obtained using the information from $\boldsymbol{x}_i$ in sample i, which is related to λ_i but not to λ. Therefore, we need a distribution $\pi(\lambda_i|\lambda)$ to link parameters λ_i and λ. Following Girón,[21] desirable properties for the link distributions $\pi(\lambda_i|\lambda)$ should be: (i) related to that of the statistical model $f(x|\lambda)$, (ii) centered around the parameter λ, and (iii) its variance (or any other measure of dispersion if variance does not exist) accommodates tentative existing differences among the λ_i.

The parameters $\lambda_1, \dots, \lambda_m$ of any partition can be a priori regarded as exchangeable, that is, they are conditionally independent given the parameter of the null model λ, according to the link distribution $\pi(\lambda_i|\lambda)$ and hence, the joint link distribution for any partition of length m is given by

$$\pi(\lambda_1, \dots, \lambda_m|\lambda) = \prod_{s=1}^{m} \pi(\lambda_s|\lambda). \tag{8}$$

Using (8), the likelihood function $f_1(\boldsymbol{z}|\lambda)$ is obtained by integration

$$f_1(\boldsymbol{z}|\lambda) = \int_0^\infty f_1(\boldsymbol{z}_1, \dots, \boldsymbol{z}_m|\lambda_1, \dots, \lambda_m)\pi(\lambda_1, \dots, \lambda_m|\lambda)\, d\lambda_1 \dots d\lambda_m = \prod_{s=1}^{m}\left(\int_0^\infty f(\boldsymbol{z}_s|\lambda_s)\pi(\lambda_s|\lambda)\, d\lambda_s\right). \tag{9}$$

According to the above mentioned properties, a well suited choice of the link distribution relating parameter λ_i with parameter λ is the conjugate Gamma$(\lambda_i|\lambda/2, 1/2)$, i.e.

$$\pi(\lambda_i|\lambda) = \frac{1}{2^{\lambda/2}\Gamma(\lambda/2)}\ \lambda_i^{\lambda/2-1}\exp(-\lambda_i/2), \tag{10}$$

where $\mathbb{E}(\lambda_i|\lambda) = \lambda$ and $\mathbb{V}(\lambda_i|\lambda) = 2\lambda$. Observe that this hierarchical conditional prior satisfy all three desirable properties given above.

From (7) and (10) we obtain

$$f_1(\boldsymbol{z}|\lambda) = \frac{1}{\Gamma(\lambda/2)^m}\prod_{s=1}^{m}\frac{\Gamma(t_s+1/2)}{(m_s+1/2)^{t_s}(2m_s+1)^{\lambda/2}}. \tag{11}$$

Other alternative links can obviously be considered. Moreno et al.[22] considered intrinsic linking distributions in meta–analysis scenarios which also present the above interesting properties. In practice, numerical results obtained from both link distributions are very similar, even for small samples. Hence, we adopt a conjugate gamma link distribution which is very simple and does not require intensive use of numerical procedures for compute the quantities of interest.

Finally, to complete integrals in (3) and (4) we use the non–informative reference prior for the parameter of the null λ,

$$\pi(\lambda) \propto \lambda^{-1/2}, \tag{12}$$

which can be thought of as a limit Gamma$(\lambda|1/2, \varepsilon)$ with $\varepsilon \to 0$. Observe that since $z_j|\lambda_j \sim \text{Po}(\lambda_j)$ and $\lambda_j|\lambda \sim \text{Ga}(\lambda_j|\lambda/2, 1/2)$, the conditional distribution of $z_j|\lambda$ is a Negative Binomial distribution with parameters $\lambda/2$ and $1/3$. The mean and variance of this distribution are λ y 3λ, respectively. After some algebra, its associated Fisher information is given by

$$I(\lambda) = -\mathbb{E}_{Z|\lambda}\left(\frac{\partial^2 \log f(z|\lambda)}{\partial\lambda^2}\right) = \frac{1}{4}\mathbb{E}_{Z|\lambda}\left(\psi^{(1)}(z+\lambda/2) - \psi^{(1)}(z)\right),$$

where $\psi^{(1)}(z)$ represents the first derivative of the digamma function $\psi(z) = \log\Gamma(z)$, and $Z|\lambda \sim \text{NB}(z|\lambda/2, 1/3)$. This quantity need to be computed numerically over the parametric space $\lambda > 0$, and presents a decreasing J–inverted behavior slowly approaching zero as λ grows. Fortunately, such Negative Binomial distribution can be adequately approximated by a Normal $\mathcal{N}(z|\lambda, 3\lambda)$, which Fisher information is given by $(2\lambda+3)/6\lambda^2$, and henceforth an uninformative prior "à la Jeffreys" is $\pi(\lambda) \propto \sqrt{I(\lambda)} \propto \lambda^{-1/2}$, for large values of λ, as we proposed in (12).

Hence,

$$f_0(\boldsymbol{x}) \propto \int_0^\infty \lambda^t \exp(-n\lambda)\lambda^{-1/2}\, d\lambda = \frac{\Gamma(t+1/2)}{n^{t+1/2}}, \tag{13}$$

and

$$f_1(\boldsymbol{z}) \propto \int_0^\infty \frac{\lambda^{-1/2}}{\Gamma(\lambda/2)^m} \prod_{s=1}^{m} \frac{\Gamma(t_s+1/2)}{(m_s+1/2)^{t_s}(2m_s+1)^{\lambda/2}}\, d\lambda. \tag{14}$$

As the arbitrary constant involved in the improper prior is the same in equations (13) and (14), they cancel out in the expression of the Bayes factor in (2),

$$B_{0\boldsymbol{z}} = \frac{\Gamma(t+1/2)}{n^{t+1/2}} \left(\int_0^\infty \frac{\lambda^{-1/2}}{\Gamma(\lambda/2)^m} \prod_{s=1}^{m} \frac{\Gamma(t_s+1/2)}{(m_s+1/2)^{t_s}(2m_s+1)^{\lambda/2}}\, d\lambda \right)^{-1}. \tag{15}$$

Bayes factor in (15) is computed numerically for each sample $\boldsymbol{z}$.

3 | NUMERICAL ILLUSTRATIONS

To illustrate the procedure developed above, we now analyse two real datasets. The `Mathematica` and `R` codes for each case study are available in the Supplementary Material section.

3.1 | Motivating example revisited

We revisit the real dataset from Nelson (1987). Table 3 shows the posterior probabilities of each cluster configuration. The prior distribution on the partitions $\Pr(\boldsymbol{z})$ in (3) plays an important role now. Figures in Table 3 are obtained from a Uniform prior assigned to each configuration which assigns the same probability to every model, that is,

$$\Pr(\boldsymbol{z}) = \frac{1}{B_4} = \frac{1}{15}, \quad \forall \boldsymbol{z} \in \mathcal{C}_4.$$

We can now obtain several quantities of interest. The partition identified as the null hypothesis is the most probable with posteriori probability 0.416. The Type 2–homogeneity cluster concentrates the highest posterior probability (0.471) with the $\lambda_1 = \lambda_2 = \lambda_3 \neq \lambda_4$ configuration being the most probable with probability 0.140. Observe that the homogeneity along with the Type 2–homogeneity clusters accumulate more than 0.887 of the posterior probability, indicating that there is a clearcut structure of homogeneity in this dataset. Furthermore, the configuration which represents the hypothesis that all samples are heterogeneous among them, or equivalently $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$, has a very small posterior probability 0.005 meaning that full heterogeneity is clearly rejected.

On other hand, by summing up the probabilities of all the configurations where each of the samples appears as a single cluster, we obtain their marginal probabilities. In this way, we can observe which population is more different from the others. In particular, population 4 is the more distant with a probability of 0.217. The remaining marginal probabilities are 0.102, 0.105, and 0.107 for populations 1, 2, and 3, respectively.

Similarly, pairwise comparisons are also easy to obtain. Posterior marginal probabilities of each one are: 0.120 (for the pair $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$), 0.073 (for $\{\boldsymbol{x}_1, \boldsymbol{x}_3\}$), 0.045 (for $\{\boldsymbol{x}_1, \boldsymbol{x}_4\}$), 0.055 (for $\{\boldsymbol{x}_2, \boldsymbol{x}_3\}$), 0.059 (for $\{\boldsymbol{x}_2, \boldsymbol{x}_4\}$) and 0.107 (for $\{\boldsymbol{x}_3, \boldsymbol{x}_4\}$).

Posterior distributions of λ_i

The posterior behavior of the parameters of interest λ_i's is also relevant once the clustering procedure is applied to the data, and can help us to compare Poisson populations.

TABLE 3 Posterior probabilities of the partitions in the motivating example

Configurations	Parameters relationship	Post. Prob.
Homogeneity		
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4\}\}$	$\lambda_1=\lambda_2=\lambda_3=\lambda_4$	0.416
Type 2–Homogeneity		
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_2\},\{\boldsymbol{x}_3,\boldsymbol{x}_4\}\}$	$\lambda_1=\lambda_2\neq\lambda_3=\lambda_4$	0.092
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_3\},\{\boldsymbol{x}_2,\boldsymbol{x}_4\}\}$	$\lambda_1=\lambda_3\neq\lambda_2=\lambda_4$	0.049
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_4\},\{\boldsymbol{x}_2,\boldsymbol{x}_3\}\}$	$\lambda_1=\lambda_4\neq\lambda_2=\lambda_3$	0.036
$\{\{\boldsymbol{x}_1\},\{\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4\}\}$	$\lambda_1\neq\lambda_2=\lambda_3=\lambda_4$	0.051
$\{\{\boldsymbol{x}_2\},\{\boldsymbol{x}_1,\boldsymbol{x}_3,\boldsymbol{x}_4\}\}$	$\lambda_2\neq\lambda_1=\lambda_3=\lambda_4$	0.050
$\{\{\boldsymbol{x}_3\},\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_4\}\}$	$\lambda_3\neq\lambda_1=\lambda_2=\lambda_4$	0.053
$\{\{\boldsymbol{x}_4\},\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\}\}$	$\lambda_4\neq\lambda_1=\lambda_2=\lambda_3$	0.140
Type 3–Homogeneity		
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_2\},\{\boldsymbol{x}_3\},\{\boldsymbol{x}_4\}\}$	$\lambda_1=\lambda_2\neq\lambda_3\neq\lambda_4$	0.029
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_3\},\{\boldsymbol{x}_2\},\{\boldsymbol{x}_4\}\}$	$\lambda_1=\lambda_3\neq\lambda_2\neq\lambda_4$	0.024
$\{\{\boldsymbol{x}_1,\boldsymbol{x}_4\},\{\boldsymbol{x}_2\},\{\boldsymbol{x}_3\}\}$	$\lambda_1=\lambda_4\neq\lambda_2\neq\lambda_3$	0.010
$\{\{\boldsymbol{x}_1\},\{\boldsymbol{x}_2,\boldsymbol{x}_3\},\{\boldsymbol{x}_4\}\}$	$\lambda_1\neq\lambda_2=\lambda_3\neq\lambda_4$	0.019
$\{\{\boldsymbol{x}_1\},\{\boldsymbol{x}_2,\boldsymbol{x}_4\},\{\boldsymbol{x}_3\}\}$	$\lambda_1\neq\lambda_2=\lambda_4\neq\lambda_3$	0.010
$\{\{\boldsymbol{x}_1\},\{\boldsymbol{x}_2\},\{\boldsymbol{x}_3,\boldsymbol{x}_4\}\}$	$\lambda_1\neq\lambda_2\neq\lambda_3=\lambda_4$	0.016
Heterogeneity		
$\{\{\boldsymbol{x}_1\},\{\boldsymbol{x}_2\},\{\boldsymbol{x}_3\},\{\boldsymbol{x}_4\}\}$	$\lambda_1\neq\lambda_2\neq\lambda_3\neq\lambda_4$	0.005

We first need compute the posterior distribution of the hyperparameter λ. Equation (12) provides the prior $\pi(\lambda)$, and equation (11) when computed for the homogeneity partition $\boldsymbol{x}$, the so called integrated likelihood of the whole data

$$f(\boldsymbol{x}|\lambda)\propto\frac{1}{\Gamma(\lambda/2)^k}\prod_{i=1}^{k}\frac{\Gamma(t_i+1/2)}{(n_i+1/2)^{t_i}(2n_i+1)^{\lambda/2}}. \tag{16}$$

Therefore, by Bayes theorem, the posterior density of λ is proportional to

$$\pi(\lambda|\boldsymbol{x})\propto\pi(\lambda)f(\boldsymbol{x}|\lambda)=\lambda^{-1/2}\frac{1}{\Gamma(\lambda/2)^k}\prod_{i=1}^{k}\frac{\Gamma(t_i+1/2)}{(n_i+1/2)^{t_i}(2n_i+1)^{\lambda/2}}. \tag{17}$$

where the normalizing constant can be easily obtained by numerical integration.

From the hierarchical model developed in Section 2, we can now derive the posterior distributions of the λ_i in terms of the posterior of λ.

Given λ and $\boldsymbol{x}$, it follows that $\lambda_i|\lambda,\boldsymbol{x}$ are conditionally independent and each follows a Gamma distribution, that is

$$\pi(\lambda_i|\lambda,\boldsymbol{x})=\text{Gamma}(\lambda_i|t_i+\lambda,n_i+1/2), \tag{18}$$

and, consequently

$$\pi(\lambda_i|\boldsymbol{x})=\int_0^{\infty}\text{Gamma}(\lambda_i|t_i+\lambda,n_i+1/2)\ \pi(\lambda|\boldsymbol{x})\ d\lambda. \tag{19}$$

where Gamma$(\lambda_i|t_i+\lambda/2,n_i+1/2)$ denotes the Gamma distribution of the corresponding parameters.

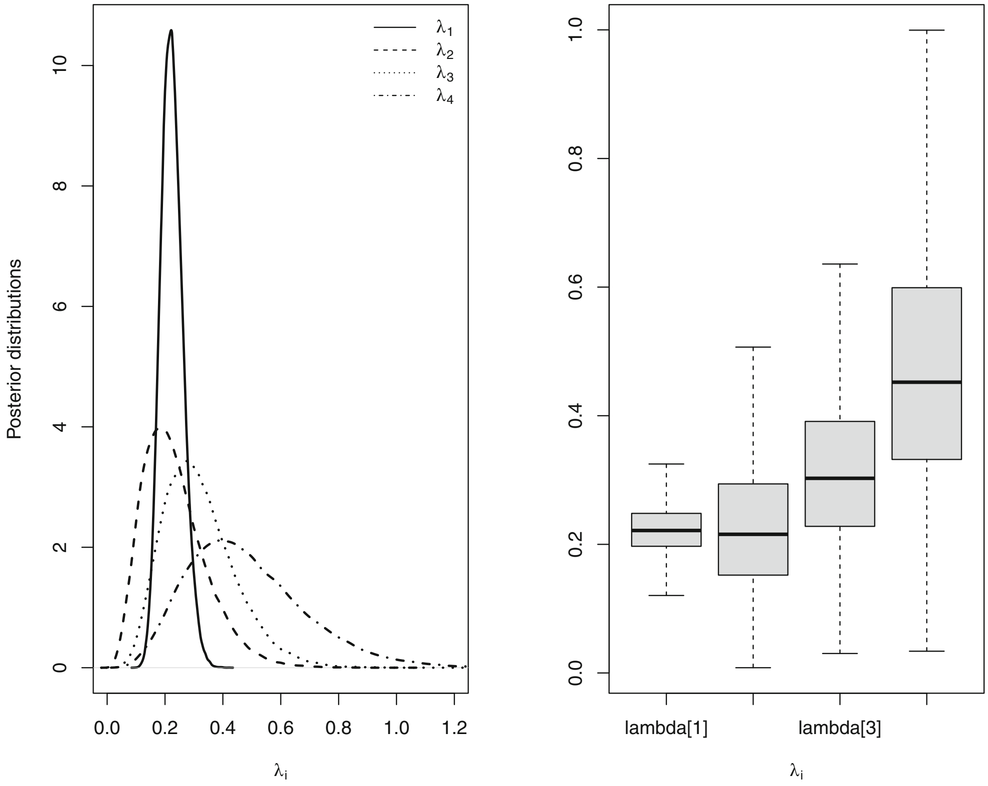

FIGURE 1 Posterior distributions (left panel) and boxplot (right panel) of λ_i, $i = 1, \dots, 4$

TABLE 4 Posterior summaries of λ_i, $i = 1, \dots, 4$

	Post. mean	Post. median	95% HPD interval
λ_1	0.224	0.221	(0.150, 0.298)
λ_2	0.233	0.216	(0.046, 0.448)
λ_3	0.319	0.303	(0.100, 0.568)
λ_4	0.482	0.453	(0.122, 0.883)

Thus, the posterior distribution for λ_i $(i = 1, \dots, k)$ can be either computed by numerical integration or, better still, by Markov chain Monte Carlo (MCMC) simulation. Given the hierarchical structure of the link distribution $\pi(\lambda_i|\lambda)$ a MCMC implementation to find the posterior distributions is particularly easy to code. A `R` code was carried out using `JAGS` (Just Another Gibbs Sampler) through the `rjags` package.[23] See Supplementary Material where we provide codes for posterior distributions and diagnostics of convergence of the corresponding chains generated.

Figure 1 displays the posterior densities of the four Poisson parameters obtained from the hierarchical linking distributions. The densities overlap in a not negligible interval, which points out to the fact that the hypothesis of homogeneity cannot be rejected. The values of the four parameters are small and close to each other (see Table 4) except perhaps the corresponding parameter to the fourth sample λ_4, thus confirming the relatively high probability of marginal $\boldsymbol{x}_4$ obtained before (0.243). Table 4 also contains 95% highest posterior density intervals (HPD intervals).

3.2 | Example 2

With the purpose of comparing a battery of several test statistics for homogeneity in Poisson populations, Chiu and Wang[11] revisited the dataset from Laird and Olivier[24] corresponding to analyze the patient survival after heart valve replacement

TABLE 5 Exposure in months and deaths

(n_i, t_i)	Aortic	Mitral
Young (< 55)	$\boldsymbol{x}_1 = (1259, 4)$	$\boldsymbol{x}_2 = (2082, 1)$
Old (≥ 55)	$\boldsymbol{x}_3 = (1417, 7)$	$\boldsymbol{x}_4 = (1647, 9)$

TABLE 6 Posterior probabilities of the partitions in the motivating example

Configurations	Parameters relationship	Post. Prob.
Homogeneity		
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	**0.037**
Type 2–Homogeneity		
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 \neq \lambda_3 = \lambda_4$	**0.097**
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2, \boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_3 \neq \lambda_2 = \lambda_4$	0.004
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_4\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}\}$	$\lambda_1 = \lambda_4 \neq \lambda_2 = \lambda_3$	0.008
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_3 = \lambda_4$	0.003
$\{\{\boldsymbol{x}_2\}, \{\boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_2 \neq \lambda_1 = \lambda_3 = \lambda_4$	**0.653**
$\{\{\boldsymbol{x}_3\}, \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4\}\}$	$\lambda_3 \neq \lambda_1 = \lambda_2 = \lambda_4$	0.006
$\{\{\boldsymbol{x}_4\}, \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}\}$	$\lambda_4 \neq \lambda_1 = \lambda_2 = \lambda_3$	0.012
Type 3–Homogeneity		
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_2 \neq \lambda_3 \neq \lambda_4$	0.007
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 = \lambda_3 \neq \lambda_2 \neq \lambda_4$	0.049
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_4\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}\}$	$\lambda_1 = \lambda_4 \neq \lambda_2 \neq \lambda_3$	0.043
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_3 \neq \lambda_4$	0.001
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_4\}, \{\boldsymbol{x}_3\}\}$	$\lambda_1 \neq \lambda_2 = \lambda_4 \neq \lambda_3$	0.001
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 \neq \lambda_3 = \lambda_4$	0.074
Heterogeneity		
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$	$\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$	0.005

Note: Bold indicates important values.

operations in a sample of 109 patients along a large number of months of observation. Table 5 shows the observed sample for exposure in months (n_i) and deaths (t_i).

Here $k = 4$, and all counts are less than 10. Such data sample is considered in Chiu and Wang[11] as a situation presenting small k and small λ, and the p-values of all the test considered (the commonly Pearson's χ^2, likelihood ratio and score tests and the U and V test by Potthoff and Whittinghill[25]), either asymptotic or their bootstrap counterpart, range from 0.00 to 0.045. Thus, the null hypothesis of homogeneity is rejected at the 0.05 level of significance, that is, the rates of mortality are concerned by the age (young or old) and/or the valve (aortic or mitral).

However, the Bayesian homogeneity test provides a more detailed analysis of the data. Table 6 shows the results of applying the Bayesian multiple test. As in the above example, a uniform prior distribution over the set of all configurations is considered. Among all 15 possible configurations, we highlight both that accumulate the highest probability and the configuration that represents the null hypothesis of homogeneity (accounting for more than 0.75 together).

The null hypothesis of homogeneity, that is, the homogeneity configuration in Table 6, has a small probability $\Pr(H_0|\boldsymbol{x}) = 0.037$. Hence, the null hypothesis of homogeneity is strongly rejected as well as the hypothesis of full heterogeneity which has also a very small probability 0.005. So, we conclude in the same sense that in Chiu and Wang.[11] However, additional information can be extracted looking at the posterior probabilities in Table 6. Particularly noteworthy is the partition $\{\{\boldsymbol{x}_2\}, \{\boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{x}_4\}\}$ where, with probability 0.653, the hypothesis $\lambda_2 \neq \lambda_1 = \lambda_3 = \lambda_4$ is true. This indicates the presence of homogeneity in three of the samples considered, except for the second one. In fact, from Table 6, we

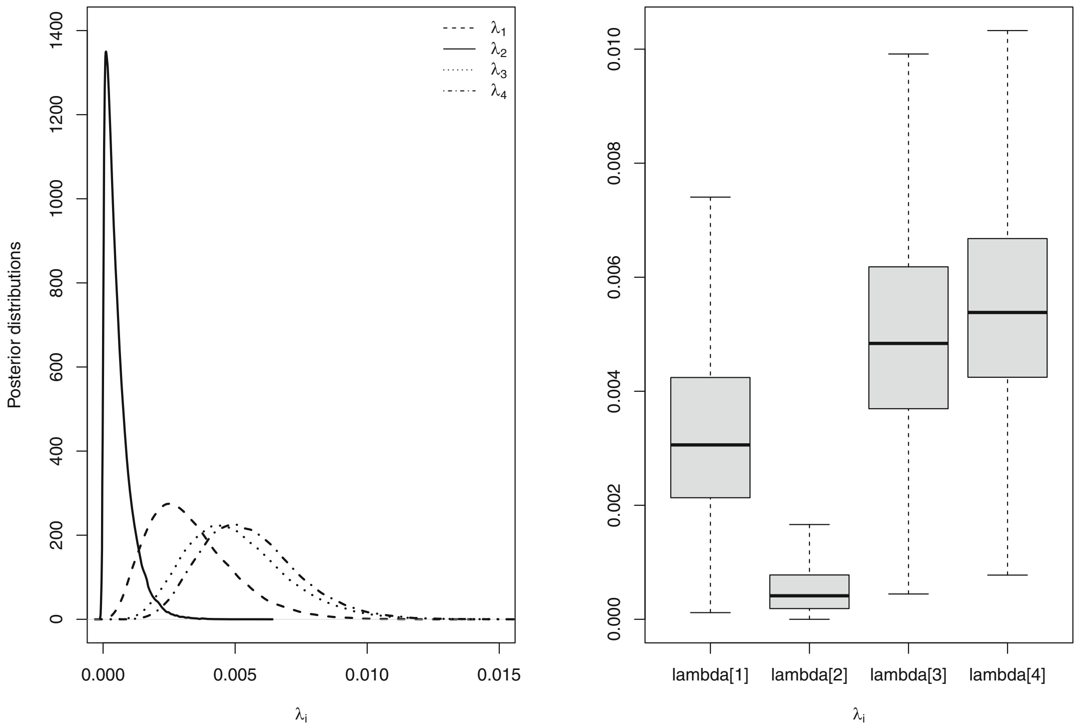

FIGURE 2 Posterior distributions (left panel) and boxplot (right panel) of λ_i, $i = 1, \ldots, 4$

TABLE 7 Posterior summaries of λ_i, $i = 1, \ldots, 4$

	Post. mean	Post. median	95% HPD interval
λ_1	0.0033	0.0030	(0.0006, 0.0065)
λ_2	0.0005	0.0004	(0.00, 0.0016)
λ_3	0.0050	0.0048	(0.0017, 0.0088)
λ_4	0.0056	0.0054	(0.0022, 0.0092)

obtain that the marginal probabilities of each separate population are 0.084, 0.825, 0.062, and 0.074, respectively. That is, population 2 is the most distinct one as can be seen from Figure 2.

Table 7 contains basic summaries of the posterior densities of the four populations considered. All the 95% Highest Posterior Density intervals overlap each other except that from population 2, confirming the above comment.

3.3 | Example 3

As in Example 3 in Nelson,[15] we consider now the case of two clerks who are to be compared for mistakes made on invoices. The observed data were $\boldsymbol{x}_1 = (100, 12)$ and $\boldsymbol{x}_2 = (300, 13)$. This example can be considered a simplified internal control situation in an auditing context. In such situations, we can imagine that a large preponderance of zeros will be present in the samples and only a few of the selected invoices will contain one or more errors. The χ^2 test gives a p-value of 0.011 and the null hypothesis of homogeneity is rejected at 0.05 of significance.

For this case, where $k = 2$, only two clusters are possible in the partition set (homogeneity and heterogeneity, respectively). Table 8 shows the posterior probabilities for each of these configurations. Although certainly the null hypothesis has a lower posterior probability (and consequently, could be rejected) let us note that the posterior odds ratio between heterogeneity and homogeneity is only 1.56, indicating a weak evidence in favor of heterogeneity according to Jeffreys' scale. ("Not worth more than a bare mention", Kass and Raftery[26p. 791]).

TABLE 8 Posterior probabilities of the partitions in the Example 3

Configurations	Parameters relationship	Post. Prob.
Homogeneity		
$\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}\}$	$\lambda_1 = \lambda_2$	0.39
Heterogeneity		
$\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}\}$	$\lambda_1 \neq \lambda_2$	0.61

4 | FREQUENTIST EVALUATION OF THE BAYESIAN TEST

As the Bayes test does not depend on a simple statistic, its frequentist properties (those referred to Type I and Type II errors of the test) have to be evaluated via simulation under the null and the different alternatives or configurations. Type I errors for several number of samples k, different values of the sample size n (assuming that all samples have the same size) and the common parameter λ can be computed by simulation. As an example we set $k = 4$, and consider four different sample sizes $n = 10, 25, 50{,}100$, and four values of the parameter $\lambda = 0.1, 1, 10, 50$.

We first simulate $k \times 1000$ samples from a Poisson distribution with parameter λ and sample size n, and we count the number of times that the true model (in this case, the null hypothesis) has the highest posterior probability, and divide this number by 1000. This provides the frequency of successes of the Bayesian test. Then, the complement to 1 of these quantities is a MonteCarlo estimate of the Type I error. We consider two prior distributions for the configuration: the uniform prior (if we have in mind clustering all configurations), and the prior that assign 1/2 to the null hypothesis and a uniform distribution to the rest of configurations (if we are considering a testing hypothesis problem). The results are given in Tables 9 and 10 respectively.

If we really are in a testing environment, where the null hypothesis has an important or relevant status, we can see from Table 10 that except for $\lambda = 0.1$ and small sample sizes $n = 10$ and 20, all Type I errors are smaller than the usual bound of 0.05. Further, for medium and large values of n and/or λ posterior probabilities tend to be similar and smaller than the usual α–level $= 0.05$.

As we have commented above, Type II errors can also be calculated by simulation for any of the configurations. As this number can be very large, to estimate Type II errors for all configurations is a too formidable task. Nevertheless, we have observed in large scale simulations that when simulated from a certain configuration, and under the cluster environment, the true model or configuration is the one with highest posterior probability and it is therefore chosen by the Bayes procedure.

To end this section, we consider the simplest problem when $k = 2$ of testing $H_0 : \lambda_1 = \lambda_2$ vs $H_1 : \lambda_1 \neq \lambda_2$. For this problem, the two environments give the same prior probability to both hypotheses, 1/2 each. Type II errors depend on the

TABLE 9 Type I error for different choices of n and λ assuming the clustering context

	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$	$\lambda = 50$
$n = 10$	0.66	0.222	0.217	0.243
$n = 25$	0.325	0.131	0.15	0.192
$n = 50$	0.142	0.122	0.11	0.117
$n = 100$	0.097	0.065	0.091	0.079

TABLE 10 Type I error for different choices of n and λ assuming the testing of the null hypothesis

	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$	$\lambda = 50$
$n = 10$	0.104	0.043	0.049	0.043
$n = 25$	0.079	0.033	0.024	0.043
$n = 50$	0.04	0.025	0.02	0.024
$n = 100$	0.017	0.015	0.009	0.017

TABLE 11 Type I errors for different choices of n and $\lambda_1 = \lambda_2 = 1$, estimated via simulation

$n = 10$	$n = 25$	$n = 50$	$n = 100$
0.069	0.053	0.029	0.021

TABLE 12 Type II errors for different choices of n and $\lambda_1 = 1$ and $\lambda_2 = 2, 2.5, 3, 3.5, 4$

(λ_1, λ_2)	$(1, 2)$	$(1, 2.5)$	$(1, 3)$	$(1, 3.5)$	$(1, 4)$
$n = 10$	0.48	0.20	0.06	0.04	0.01
$n = 25$	0.23	0.02	0.00	0.00	0.00
$n = 50$	0.04	0.00	0.00	0.00	0.00
$n = 100$	0.00	0.00	0.000	0.00	0.00

values of parameters λ_1 and λ_2. We present results for a fixed value of λ_1 and different values of λ_2 for different samples sizes.

Type I error for different sample sizes are given in Table 11.

Table 12 shows a simulation of the Type II errors for testing $H_0 : \lambda_1 = \lambda_2$ vs $H_1 : \lambda_1 \neq \lambda_2$ when $\lambda_1 = 1, \lambda_2 = 2, 2.5, 3, 3.5$ and 4 and several sample sizes $n = 10, 25, 50, 100$.

This simulation shows the expected behavior of Type II errors. Type II error vanishes at sample size grows. On the other hand, when the Kullback–Leibler distance between Poisson models increases the Type II error also vanishes.

5 | EXTENSIONS

Example 3 above concludes with the rejection of the null hypothesis and accepting that clerk two is the more accurate,[15] that is, $H_a : \lambda_1 > \lambda_2$ is accepted conditional on the alternative hypothesis that $\lambda_1 \neq \lambda_2$.

The Bayesian procedure also provides a simple solution to testing inequality hypotheses on the parameters. In many occasions it would be interesting to test the null hypothesis $H_0 : \lambda_1 > \lambda_2 > \dots > \lambda_k$ vs H_1 : the ordering of the parameters is any other different permutation.

The Bayesian answer to this testing problem, under the hierarchical model, is the computation of the posterior probability of the null space $\Lambda_0 = \{(\lambda_1, \dots, \lambda_k) : \lambda_1 > \lambda_2 > \dots > \lambda_k\}$ which can be easily obtained from the hierarchical model, and it turns out to be

$$\Pr(H_0|\boldsymbol{n}, \boldsymbol{t}) = \int_0^\infty \int_{\Lambda_0} \prod_{i=1}^k \text{Gamma}(\lambda_i | t_i + \lambda/2, n_i + 1/2)\, d\lambda_i \pi(\lambda|\boldsymbol{x})\, d\lambda, \tag{20}$$

where $\boldsymbol{n} = (n_1, \dots, n_k)$ and $\boldsymbol{t} = (t_1, \dots, t_k)$. For large values of k the numerical computation is better done by using MCMC methods, which take advantage of the posterior conditional independence of the λ_i given λ, that is, simulating first from the posterior of the hyperparameter λ, and then from Gamma distributions.

On the other hand, assuming that the data of the four groups were independent, the Bayesian answer to this test would be the posterior probability of the null space $\Lambda_0 = \{(\lambda_1, \dots, \lambda_k) : \lambda_1 > \lambda_2 > \dots > \lambda_k\}$, i.e.

$$\Pr(H_0|\boldsymbol{n}, \boldsymbol{t}) = \int_{\Lambda_0} \prod_{i=1}^k \text{Gamma}(\lambda_i | t_i + 1/2, n_i)\, d\lambda_i, \tag{21}$$

where $\text{Gamma}(\lambda_i | t_i + 1/2, n_i)$ denotes the posterior densities of λ_i obtained using the reference Jeffreys prior. Same as in the hierarchical setting, this posterior probability has to be evaluated by numerical integration or MonteCarlo simulation. The MCMC computations are, in this case, easier due to the independence of the posterior distributions of the λ_i.

For the data of Example 3 in Section 3.3, $k = 2$ the posterior probability of $\lambda_1 > \lambda_2$, given that $\lambda_1 \neq \lambda_2$, was roughly the same, 0.993, for both formulas.

6 | CONCLUSIONS

This paper provides a simple Bayesian solution to the comparison of the Poisson populations problem. A hierarchical model for testing homogeneity of the means is considered to obtain the Bayes factor for comparing each alternative model with the baseline homogeneity model. Bayesian practitioners base their decisions on the posterior probabilities of all models involved without resorting to sometimes unmeaning p-values as in the frequentist approach. In other words, frequentist methods can be neither used for quantifying evidence for the null hypothesis of homogeneity nor for testing multiple hypotheses simultaneously. The proposed Bayesian approach gives a procedure without these constraints. Furthermore, in contrast with the frequentist approach, no ad–hoc restrictions are needed for particular situations such as a small number of samples k to test their homogeneity, small values of some of the estimates of parameters λ_i, sparse data, and/or small and different sample sizes. The proposed Bayesian procedure is valid under all these situations.

Nevertheless, the prior distribution on the partitions models deserve some attention. In common practical situations, where the number of populations to be compared is small with respect to the samples sizes n_i and no prior information about the models is available, a uniform prior over the set of all configurations C_k seems a reasonable prior distribution. On the other hand, Negrín et al.[27] found a high sensitivity to the prior on the partitions when we focus on estimation and clustering in binary data for meta–analysis.

The proposed Bayesian procedure implies obtaining posterior probabilities for all posible partitions; however, computational difficulties can be arise when the number of Poisson populations to be compared is moderately large ($k > 15$) due to time required for computation. In that case, an efficient stochastic algorithm to search a set of good clusters models is required.[18]

ACKNOWLEDGMENTS

We are grateful to an Associate Editor and to an anonymous referee for their valuable, constructive comments and suggestions, which have greatly helped us improve this paper. MME and FJVP partially funded by grant PID2021–1279890B-I00 (Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, Spain).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

María Martel–Escobar https://orcid.org/0000-0002-7013-4747
Francisco–José Vázquez–Polo https://orcid.org/0000-0002-0632-6138

REFERENCES

1. Kirshnamoorthy K, Thomson J. A more powerful test for comparing two Poisson means. *J Stat Plan Inf*. 2004;119:23-35.
2. Lee YJ. Tests for monotone trend in K Poisson means. *J Qual Tech.* 1985;17:44-49.
3. Weiss CH, Testik MC. Residuals–based CUSUM charts for Poisson INAR(1) processes. *J Qual Tech.* 2015;47:30-42.
4. Ryan KJ, Hamada MS, Reese CS. A Bayesian hierarchical power law process model for multiple repairable systems with an application to supercomputer reliability. *J Qual Tech.* 2011;43:209-223.
5. Przyborowski J, Wilenski H. Homogeneity of results in testing samples from Poisson series. *Biometrika*. 1940;31:313-323.
6. Nelson LS. Power in comparing Poisson means: II Two–sample test. *J Qual Tech.* 1991;23:163-166.
7. Sahai H, Misra SC. Comparing means of two Poisson distributions. *The Mathematical Scientist*. 1992;17:60-67.
8. Cousins RD. Improved central confidence intervals for the ratio of Poisson means. *Nucl Instrum Methods Phys Res, Sect A*. 1998;417:391-399.
9. Brown LD, Zhao LH. A test for the Poisson distribution. *Sankhya*. 2002;64(Series A, 3):611-625.
10. Farcomeni A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat Methods Med Res*. 2008;17:347-388.
11. Chiu SN, Wang L. Homogeneity tests for several Poisson populations. *Comput Stat Data Anal*. 2009;53:4266-4278.
12. Singh B. Bayesian inference for comparing several treatment means under a Poisson model. *Commun Stat Theory Methods*. 1980;9:1137-1145.
13. Zaslavsky BG. Bayesian versus frequentist hypotheses testing in clinical trials with dichotomous and countable outcomes. *J Biopharm Stat*. 2010;20:985-997.
14. Kawasaki Y, Miyaoka E. A Bayesian inference of $P(\lambda_1 < \lambda_2)$ for two Poisson parameters. *J App Stat*. 2012;39:2141-2152.
15. Nelson LS. Comparison of Poisson means: the general case. *J Qual Tech*. 1987;19:173-179.
16. Poirier LA, Stoner GD, Shimkin MB. Bioassay of alkyl halides and nucleotides base analogs by pulmonary tumor response in strain A mice. *Cancer Res*. 1975;35:1411-1415.

17. Barry D, Hartigan JA. Product partition models for change point problems. *Ann Stat*. 1992;20:260-279.
18. Casella G, Moreno E, Girón FJ. Cluster analysis, model selection, and prior distributions on models. *Bayesian Anal*. 2014;9:613-658.
19. Rota GC. The number of partitions of a set. *Amer Math Monthly*. 1964;71:498-504.
20. Chang CH, Pal N, Lin JJ. A note of comparing several Poisson means. *Commun Stat Simul Comput*. 2010;39:1605-1627.
21. Girón FJ. *Bayesian Testing of Statistical Hypotheses Málaga*. Arguval Ed; 2021.
22. Moreno E, Vázquez-Polo FJ, Negrín MA. Objective Bayesian meta–analysis for sparse discrete data. *Stat Med*. 2014;33:3676-3692.
23. Plummer M, Stukalov A, Denwood M. Package rjags. Accessed September, 2021. http://cran.stat.unipd.it/web.packages/rjags/rjags.pdf. 2021.
24. Laird N, Olivier D. Covariance analysis of censored survival data using log–linear analysis techniques. *J Am Stat Assoc*. 1981;76:231-240.
25. Potthoff RF, Whittinghill M. Testing for homogeneity: II The Poisson distribution. *Biometrika*. 1966;53:183-190.
26. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc*. 1985;90:773-795.
27. Negrín MA, Martel M, Vázquez Polo FJ. Bayesian meta–analysis for binary data and prior distribution on models. *Int J Env Res Pub Health*. 2021;18:809.

How to cite this article: Girón FJ, Martel–Escobar M, Vázquez–Polo FJ. A Bayesian homogeneity test for comparing Poisson populations. *Appl Stochastic Models Bus Ind*. 2022;38(6):1158-1171. doi: 10.1002/asmb.2727