Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (1) (2025), 319 – 335
DOI : 10.15672/hujms.1292359

Research Article

Intrinsic priors for comparing zero-inflation parameters in Poisson models

Kipum Kim[1], Hyeon Jun Jeong[2], Yongdai Kim[3], Seong W. Kim*[1]

[1] *Department of Mathematical Data Science, Hanyang University, Ansan, 15588, South Korea*
[2] *Department of Economics, Hanyang University, Ansan, 15588, South Korea*
[3] *Department of Statistics, Seoul National University, Seoul, 08826, South Korea*

Abstract

Prior elicitation is an important issue in both objective and subjective Bayesian inferences. In hypothesis testing and model selection, choosing appropriate prior distributions becomes significantly more critical. In an objective Bayesian analysis, one utilizes noninformative priors such as Jeffreys priors or reference priors for hypothesis testing which are often improper, making unspecified constants to be contained in the Bayes factor. Thus, the resulting Bayes factor should be adjusted. In this paper, we consider default Bayes procedures for testing zero-inflation parameters in a zero-inflated Poisson distribution. In particular, we derive a set of intrinsic priors based on an approximation procedure. Extensive simulations and analyses of two real datasets are performed to support the methodology developed in the paper. It is shown that the proposed Bayesian and frequentist approaches yield similar comparable results.

Mathematics Subject Classification (2020). 62F03, 62F15

Keywords. Fractional Bayes factor, intrinsic Bayes factor, intrinsic prior, training sample, zero inflation

1. Introduction

Analysis of discrete data is conducted in various fields such as natural sciences, social sciences, public health, and geology. There are several types of distributions in utilizing discrete data. Xiao et al. [35] employed the geometric distribution in a regression model to perform Bayesian inference. Azexedo et al. [2] used various distributions, including Poisson, negative binomial, COM-Poisson, and generalized Poisson distributions to analyze tuberculosis data. The Poisson distribution is commonly used when a random variable of interest is the number of events occurring in a given time interval. For instance, it would be interesting to observe how many earthquakes will occur in one year; or to see how many home runs can be produced by a baseball batter in each game. More often than not, these count data possess an excessive number of zeros, hindering analysis with the regular Poisson distribution. Under these circumstances with excessive zero patterns,

*Corresponding Author.
Email addresses: rainbowlion@hanyang.ac.kr (K. Kim), su970211@hanyang.ac.kr (H. J. Jeong), ydkim903@snu.ac.kr (Y. Kim), seong@hanyang.ac.kr (S. W. Kim)
Received: 06.05.2023; Accepted: 05.01.2025

zero-inflated models would be a remedy to circumvent loss of information or tendencies of biased estimators.

Research on the analysis of zero-inflated outcomes was started by [10] and was further developed by [28]. Lambert [21] proposed the zero-inflated Poisson (ZIP) regression model, where a Poisson-Bernoulli mixture structure is proposed to deal with two sources of excessive zero values. Later, a considerable amount of work was performed in analyzing zero-inflated count data through ZIP models. Random effects were incorporated into ZIP models [14, 26, 37], and marginalized ZIP regression models were proposed and extensively analyzed by [22] and [24]. On the other hand, ZIP regression mixture models were proposed by [23], and latent factor ZIP models were suggested and utilized by [29].

Prior elicitation has been one of the major issues in both objective and subjective Bayesian inferences in which the prior distribution should account for uncertainties and beliefs about unknown parameters before data are observed. As pointed out by [4], it is often not durable to properly and subjectively impose prior distributions due to time constraints or resources. Consequently, default Bayesian procedures were proposed and developed with an objective perspective. According to more related work on prior elicitation under an objective Bayesian context performed by [12], it is noted that the selection of prior distributions is crucial when dealing with model selection or hypothesis testing. It must be performed with caution in the use of noninformative priors such as Jeffreys priors [16] or reference priors of [5]. Note that the Jeffreys prior is derived by taking the square root of the determinant of the Fisher information matrix. On the other hand, reference priors are based on the Kullback-Leibler divergence and divide the parameter space into 'parameter of interest' and nuisance parameters. More often than not, these noninformative priors are usually improper, and the resulting pdf associated with these priors does not have a finite integral. Ultimately, the marginal distribution calculated with the improper prior involves an arbitrary constant, which hinders the resulting Bayes factor from being well-defined due to the ratio of two unspecified arbitrary constants. Thus, it is indispensable to properly impose objective and default prior specifications.

To overcome this arbitrariness, Berger and Pericchi [7] introduced a new model selection criterion called the intrinsic Bayes factor (IBF) using a data-splitting idea. A part of the full data often called a training sample, is utilized to remove the arbitrariness of improper priors, producing a well-defined Bayes factor. The IBF has been successful in producing more stable results under various settings and problems in the model selection and hypothesis testing context. Although there has been a considerable amount of work that utilizes the IBF as a model selection tool, we only state a few recent papers. Wang and Pericchi [34] proposed the geometric IBF in conjunction with choosing training sample sizes to come up with stable values. Almodóvar-Rivera and Pericchi-Guerra [1] used IBF methodologies to deal with hypothesis testing problems associated with normal means for two independent populations. Clare [9] proposed a new universal and robust boundary value of the IBF by taking a comprehensive reformulation into account.

However, the IBF approach requires a higher computational cost when either the size of the training sample is large or non-nested model comparisons are performed. On the other hand, O'Hagan [30] proposed another model selection criterion called the fractional Bayes factor (FBF), which provides an adjustment to the ordinary Bayes factor by using a fraction on the likelihood. The FBF methodology is more computationally feasible than the IBF approach simply because it is not necessary to conduct a heavy computation caused by training samples. However, the FBF is often sensitive to the choice of fraction [6, 11]. To circumvent the cons of the FBF approach [13] proposed the approximated adaptive fractional Bayes Factor (AAFBF), which achieves faster convergence by modifying the mean of the prior distribution in the FBF. Ultimately, it turned out that the AAFBF is adaptable to a wide range of statistical models through the use of approximations.

Due to some shortcomings encountered in the IBF and FBF approaches, it is intractable to apply these approaches in some situations, such as when drawing inferences on non-linear models or time series analysis. Eliciting a proper prior would be one remedy under these circumstances to avoid heavy computation or the choice of a fraction. This induced [7] to suggest a (possibly) proper prior that may be a plausible alternative in justifying the full likelihood. This prior is called an *intrinsic prior*, and the resulting ordinary Bayes factor calculated with full samples approaches to the IBF or FBF at least asymptotically.

There has been some work on finding intrinsic priors in the Bayesian hypothesis context, for which most are limited to continuous distributions. Further, it is well known that finding intrinsic priors is not an easy task in many hypothesis testing scenarios, due to the inherent and deeply rooted difficulties that exist in the problem itself. Kim and Sun [20] conducted hypothesis testing for exponential distributions and the power law process to derive intrinsic priors. Moreno [27] utilized half-normal distributions to conduct default Bayesian tests with intrinsic priors. A considerable amount of research was conducted on ZIP models in accordance with default Bayesian procedures. Xie and Goh [36] conducted an objective Bayesian analysis through default priors on zero-inflation and Poisson count parameters in the ZIP. However, limited research has been conducted on analyzing discrete data with default Bayes factors and intrinsic priors. Bayarri et al. [3] proposed an objective Bayesian approach for testing the zero-inflation parameter in the ZIP without attempting to find intrinsic priors. However, soon after, Sivaganesan and Jiang [33] derived intrinsic priors for testing the point null hypothesis associated with the mean of the Poisson distribution. Recently, Han et al. [15] conducted a hypothesis testing on the Poisson count parameter of the ZIP distribution to derive a couple of intrinsic priors when the zero inflation parameter is treated as a nuisance parameter.

A number of work has been done for testing zero inflation parameters based on the frequentist approach using the likelihood ratio test [25, 32]. Some papers deal with two sample tests for deriving intrinsic priors. Kim [18] conducted testing on two independent exponential means to derive a general class of intrinsic priors. Kim and Kim [19] derived intrinsic priors for testing two normal means with the intrinsic approach. In this article, we focus on two independent populations that both follow the ZIP distribution with the same count parameter but different zero-inflation parameters. Under this setup, default Bayesian testing procedures are presented and intrinsic priors will be derived through a reasonable approximation. To the best of our knowledge, aside from those mentioned in this section, there are no other recently published papers that offer value for critique in this field.

The rest of the paper is organized as follows: In Section 2, we present default Bayesian procedures for hypothesis testing and model selection including the IBF and FBF methodologies. We consider the zero-inflated Poisson distribution to present default Bayes factors and the main results for deriving intrinsic priors in testing the equality of zero-inflation parameters. In Section 3, an extensive Monte Carlo simulation study was carried out to evaluate the performance of the proposed procedures. Two real datasets are analyzed to illustrate the proposed methodologies in Section 4. Finally, we finish this article with concluding remarks in Section 5.

2. Default bayesian testing and intrinsic priors for the ZIP distribution

2.1. Bayesian testing and Bayes factors

Suppose that two competing hypotheses, H_0 (null hypothesis) and H_1 (alternative hypothesis), are considered. For data $\boldsymbol{Z}$, model H_j has density $f_j(\boldsymbol{z}|\boldsymbol{\theta}_j)$, where $\boldsymbol{\theta}_j$ are unknown model parameters for $j = 0, 1$. Bayesian model selection proceeds by choosing a prior distribution $\pi_j(\boldsymbol{\theta}_j)$ for $\boldsymbol{\theta}_j$ under model H_j. Let $m_j(\boldsymbol{z})$ denote the marginal of

predictive density under model M_j. That is,

$$m_j(\boldsymbol{z}) = \int_{\Theta_j} f_j(\boldsymbol{z}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j,\ j = 0, 1,$$

where Θ_j is the parameter space for $\boldsymbol{\theta}_j$. Before the data are observed we assign the prior model probability of model H_j being true, denoted by $p(H_j)$ so that $p(H_0) + p(H_1) = 1$. Then the posterior probability that H_j is the true model can be calculated as

$$Pr(H_0|\boldsymbol{z}) = \left[1 + \frac{p(H_1)}{p(H_0)}B_{10}\right]^{-1}, \tag{2.1}$$

where B_{10} is the Bayes factor of model H_1 to model H_0 defined as

$$B_{10}(\boldsymbol{z}) = \frac{m_1(\boldsymbol{z})}{m_0(\boldsymbol{z})} = \frac{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int_{\Theta_0} f_i(\boldsymbol{z}|\boldsymbol{\theta}_0)\pi_0(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0}. \tag{2.2}$$

It is usual or conventional that we assume the same prior model probability of 1/2 each, yielding

$$P(H_0|\boldsymbol{z}) < P(H_1|\boldsymbol{z}) \text{ if and only if } B_{10} > 1.$$

However, when multiple model comparisons are conducted with several prior model probabilities, different probabilities can be assigned based on expert beliefs or justifications with appropriate rationales [8, 31].

A Bayesian model selection criterion often selects model H_1 if $B_{10} > 1$. Kass and Raftery [17] suggested the following interpretations of Bayes factors for evidence against H_0 provided in Table 1.

Table 1. Interpretations for the Bayes factor

Value of B_{10}	Interpretation
1 – 3.2	Not worth more than a bare mention
3.2 – 10	Substantial
10 – 100	Strong
>100	Decisive

As mentioned in Section 1, limited information on model parameters often requires the use of noninformative priors that are typically improper in most cases. For instance, let $\pi_j^N(\boldsymbol{\theta}_j)$ $(j = 0, 1)$ be the improper prior density, then the Bayes factor in (2.2) can be expressed as

$$B_{10}^N(\boldsymbol{z}) = \frac{m_1^N(\boldsymbol{z})}{m_0^N(\boldsymbol{z})} = \frac{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^N(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int_{\Theta_0} f_i(\boldsymbol{z}|\boldsymbol{\theta}_0)\pi_0^N(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0}. \tag{2.3}$$

Since $\pi_j^N(\boldsymbol{\theta}_j)$ is improper, it is defined only up to an arbitrary constant c_j, resulting in an indeterminate Bayes factor. This is a motivation why one needs to use default Bayes factors with the following form:

$$B_{10}^D(\boldsymbol{z}) = B_{10}^N(\boldsymbol{z}) \cdot CF_{01}. \tag{2.4}$$

Here, $B_{10}^N(\boldsymbol{z})$ is defined in (2.3) and should be calculated with the full data $\boldsymbol{z}$ along with improper priors π_0^N and π_1^N. Notice that the correction factor is used to remove arbitrary constants and allow the Bayes factor to be well-defined.

Two methods in conjunction with the correction factor have been proposed and frequently utilized to serve as default Bayes factors. To circumvent this indeterminacy problem, Berger and Pericchi [7] proposed to use a part of the data, often called a training sample. Specifically, let $z(\ell)$ be a minimal training sample for which both marginals

$m_0(z(\ell))$ and $m_1(z(\ell))$ are finite, and no subset of $z(\ell)$ provides finite marginals. Therefore, the correction factor associated with the arithmetic intrinsic Bayes factor (AIBF) can be defined as

$$CF_{01}^{I} = \frac{1}{L}\sum_{\ell=1}^{L} B_{01}^{N}(z(\ell)), \tag{2.5}$$

where L is the total number of all possible minimal samples. Meanwhile, the correction factor related to the fractional Bayes factor (FBF) of [30] is defined as

$$CF_{01}^{F} = \frac{\int_{\Theta_0} f_0(\boldsymbol{z}|\boldsymbol{\theta}_0)^{\delta}\pi_0^N(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)^{\delta}\pi_1^N(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}, \tag{2.6}$$

where fraction δ is chosen arbitrarily but should be chosen properly. Subsequently, two default Bayes factors can be obtained through (2.4) using two different correction factors given by (2.5) and (2.6), respectively.

Remark 2.1. Note that the choice of fraction δ relies upon both sample sizes and number of model parameters [30]. Further, several works have focused on choice of fraction that influences the consistency of Bayes factors [11]. We used a common choice of fraction in both simulation studies and real data analysis that are presented in Sections 3 and 4, respectively.

2.2. Testing for zero inflation parameters in the ZIP

Consider a random variable Z having a zero-inflated Poisson distribution with the following probability mass function:

$$f(z|\lambda,\omega) = \begin{cases} \omega + (1-\omega)e^{-\lambda}, & \text{for } z = 0, \\ (1-\omega)\frac{e^{-\lambda}\lambda^z}{z!}, & \text{for } z = 1, 2, \ldots, \end{cases} \tag{2.7}$$

where ω is often called the *zero-inflation* parameter. We denote the distribution in (2.7) as $\text{ZIP}(\lambda,\omega)$ in short for notation convenience.

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ be a random sample from $\text{ZIP}(\lambda,\omega_1)$, and independently we observe a random sample $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ from $\text{ZIP}(\lambda,\omega_2)$. We are interested in testing

$$H_0 : \omega_1 = \omega_2 \text{ versus } H_1 : \omega_1 \neq \omega_2.$$

Let ω be the common value of ω_1 and ω_2. Then we have $\theta_0 = (\lambda,\omega)$ and $\theta_1 = (\lambda,\omega_1,\omega_2)$, where θ_0 and θ_1 are generic expressions for the parameters under H_0 and H_1, respectively. Let $N = m + n$, and let α_x and α_y denote the numbers of zero observations from each of two populations, i.e., $\alpha_x = \sum_{i=1}^{m} I(X_i = 0)$ and $\alpha_y = \sum_{i=1}^{n} I(Y_i = 0)$. Further, let $s_x = \sum_{i=1}^{m} X_i$ and $s_y = \sum_{i=1}^{n} Y_i$ denote the sums of total observations from each population respectively. For observed data $\boldsymbol{x}$ and $\boldsymbol{y}$, the likelihood functions under H_0 and H_1 are given respectively by

$$\begin{aligned} L_0(\boldsymbol{x},\boldsymbol{y}|\lambda,\omega) &\propto \left[\omega + (1-\omega)e^{-\lambda}\right]^{\alpha_x+\alpha_y}(1-\omega)^{N-\alpha_x-\alpha_y}e^{-(N-\alpha_x-\alpha_y)\lambda}\lambda^{s_x+s_y}, \\ L_1(\boldsymbol{x},\boldsymbol{y}|\lambda,\omega_1,\omega_2) &\propto \left[\omega_1 + (1-\omega_1)e^{-\lambda}\right]^{\alpha_x}\left[\omega_2 + (1-\omega_2)e^{-\lambda}\right]^{\alpha_y} \\ &\quad (1-\omega_1)^{m-\alpha_x}(1-\omega_2)^{n-\alpha_y}e^{-(N-\alpha_x-\alpha_y)\lambda}\lambda^{s_x+s_y}. \end{aligned} \tag{2.8}$$

We consider noninformative priors for both H_0 and H_1 as starting priors under *independent a priori*. That is,

$$\begin{cases} \pi_0^N(\lambda,\omega) = \lambda^{-1/2} \text{ for } \lambda > 0 \text{ and } 0 < \omega < 1, \\ \pi_1^N(\lambda,\omega_1,\omega_2) = \lambda^{-1/2} \text{ for } \lambda > 0,\ 0 < \omega_1 < 1,\ \text{and } 0 < \omega_2 < 1. \end{cases} \tag{2.9}$$

Based on the full data $(\boldsymbol{x}, \boldsymbol{y})$, the marginal under H_0 is

$$m_0(\boldsymbol{x}, \boldsymbol{y}) = \frac{\Gamma(\alpha+1)\Gamma(N-\alpha+1)\Gamma(s+1/2)}{(N-\alpha)^{s+1/2}} \sum_{j=0}^{\alpha} \frac{\Gamma(\alpha+1)}{\Gamma(N+j-\alpha+2)}, \tag{2.10}$$

where $\alpha = \alpha_x + \alpha_y$ and $s = s_x + s_y$. The marginal $m_1(\boldsymbol{x}, \boldsymbol{y})$ under H_1 can also be calculated based on the likelihood L_1 in (2.8) along with the prior π_1^N in (2.9). We tried to obtain a closed form of $m_1(\boldsymbol{x}, \boldsymbol{y})$ using two binomial expansions on the terms inside the brackets in the likelihood. However, no closed form with double summations existed. Thus, the marginal under H_1 is calculated through a direct three-dimensional integration.

When calculating the correction factor, we take a zero observation and a nonzero observation from each of the two ZIP distributions as a training sample. That is, the set of the training sample is $z(l) = \{(x_l, 0), (y_l, 0)\}$. Note that $z(l)$ is not minimal in the sense that both marginals under H_0 and H_1 are finite with $\{x_l, y_l\}$ only. This is not congruent with the statements mentioned in Section 2.1 regarding the definition of *minimal training sample*, since $\{x_l, y_l\}$ is a subset of $z(l)$. However, if we exclude two zero observations in the training sample, the part of the likelihood contributed by zero observations is ignored. This results in identical processes of extracting the intrinsic prior from the ZIP distributions and of deriving the intrinsic prior from the regular Poisson distributions cf. [15]. Thus, we use $z(l)$ as the training sample to proceed for subsequent analysis even though it is not minimal.

Proposition 2.2. *Let $z(l)$ be the training sample. Then, the Bayes factor based on $z(l)$ is*

$$B_{01}(z(l)) = \frac{6}{5} \cdot \frac{12^\kappa + 3 \cdot 8^\kappa + 6 \cdot 6^\kappa}{12^\kappa + 4 \cdot 8^\kappa + 4 \cdot 6^\kappa}, \tag{2.11}$$

where $\kappa = x_l + y_l + 0.5$.

Proof: Since $\alpha_x = \alpha_y = 1$, $s_x = x_l$, and $s_y = y_l$ in (2.8), the marginal of $z(l)$ under H_0 is

$$\begin{aligned} m_0(z(l)) &\propto \int_0^\infty \int_0^1 \left[\omega^2 + 2\omega(1-\omega)e^{-\lambda} + (1-\omega)^2 e^{-2\lambda}\right](1-\omega)^2 e^{-2\lambda} \lambda^{x_l+y_l-0.5} d\omega d\lambda \\ &= \Gamma(\kappa)\left\{\frac{1}{30 \cdot 2^\kappa} + \frac{1}{10 \cdot 3^\kappa} + \frac{1}{5 \cdot 4^\kappa}\right\} \\ &= \frac{\Gamma(\kappa)}{30} \cdot \frac{12^\kappa + 3 \cdot 8^\kappa + 6 \cdot 6^\kappa}{24^\kappa}. \end{aligned}$$

On the other hand, the marginal under H_1 is

$$\begin{aligned} m_1(z(l)) &\propto \int_0^\infty \int_0^1 \int_0^1 \left[\omega_1\omega_2 + \omega_1(1-\omega_2)e^{-\lambda} + (1-\omega_1)\omega_2 e^{-\lambda} + (1-\omega_1)(1-\omega_2)\right] \\ &\times (1-\omega_1)(1-\omega_2)e^{-2\lambda}\lambda^{x_l+y_l-0.5} d\omega_1 d\omega_2 d\lambda \\ &= \frac{\Gamma(\kappa)}{36}\left\{\frac{1}{2^\kappa} + \frac{4}{3^\kappa} + \frac{4}{4^\kappa}\right\} \\ &= \frac{\Gamma(\kappa)}{36} \cdot \frac{12^\kappa + 4 \cdot 8^\kappa + 4 \cdot 6^\kappa}{24^\kappa}. \end{aligned}$$

Thus, the Bayes factor with the training sample is readily available. □

2.3. Intrinsic priors for testing zero-inflation parameters

It is beneficial to find reasonable priors to make use of the full likelihood when justifying default Bayesian procedures for hypothesis testing. As mentioned in subsection 2.1, the use of two default Bayes factors has some cons. The AIBF could require computational cost, whereas the choice of fraction in the FBF could be sensitive. Once reasonable (possibly

proper) priors are available, the correction factor in (2.4) would not be necessary. Rather, we only need to calculate the (ordinary) Bayes factor with intrinsic priors based on the full data $\boldsymbol{z}$. Ultimately, asymptotic equivalence would be obtained in which this Bayes factor gets close to the Bayes factor in (2.4) as the sample size increases.

Under the regularity conditions in [7], the intrinsic prior for H_1 based on the AIBF approach, denoted by π_1^I, is calculated as

$$\pi_1^I(\boldsymbol{\theta}) = E[B_{01}(Z(\ell))|\boldsymbol{\theta}]\pi_1^N(\boldsymbol{\theta}), \tag{2.12}$$

where the expectation is taken with the probability distribution of $Z(\ell)$ under model H_1. On the other hand, the intrinsic prior based on the FBF approach, denoted by π_1^F, is

$$\pi_1^F(\boldsymbol{\theta}) = CF_{01}^{F*}\pi_1^N(\boldsymbol{\theta}). \tag{2.13}$$

Here, $CF_{01}^{F*} = \lim_{n\to\infty} CF_{01}^F$, where CF_{01}^F is given in Eq. (2.6) and n is a given sample size.

Based on the Bayes factor with the training sample $z(l)$ in (2.11), we can derive an intrinsic prior defined in (2.12). The following theorem provides the intrinsic prior using an approximation method when testing the equality of zero-inflation parameters.

Theorem 2.3. *The intrinsic prior for* $(\lambda, \omega_1, \omega_2)$ *under* H_1 *based on the AIBF approach in (2.12) is*

$$\pi_1^I(\lambda,\omega_1,\omega_2) \propto \frac{e^{-2\lambda}}{\sqrt{\lambda}(1-e^{-\lambda})^2}\left[\left\{\sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}\cdot\frac{2\cdot 6^\alpha - 8^\alpha}{12^\alpha+4\cdot 8^\alpha+4\cdot 6^\alpha}\right\}+(e^\lambda-1)^2\right], \tag{2.14}$$

where $\alpha = k + 0.5$.

Proof: Note that each of $x_l \equiv x$ and $y_l \equiv y$ follows a zero-truncated Poisson distribution with parameter λ, and they are independent. Let

$$g(\kappa) = \frac{12^\kappa + 3\cdot 8^\kappa + 6\cdot 6^\kappa}{12^\kappa + 4\cdot 8^\kappa + 4\cdot 6^\kappa}. \tag{2.15}$$

Then, from (2.11), we have

$$\begin{aligned} E[B_{01}^N(z(l))] &= \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} B_{01}^N(z(l))\frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}\frac{e^{-\lambda}\lambda^y}{y!(1-e^{-\lambda})} \\ &= \frac{6}{5}\frac{e^{-2\lambda}}{(1-e^{-\lambda})^2}\sum_{x=1}^{\infty}\sum_{y=1}^{\infty}\frac{\lambda^{x+y}}{x!y!}g(\kappa). \end{aligned} \tag{2.16}$$

Let

$$q(\lambda) \equiv \sum_{x=1}^{\infty}\sum_{y=1}^{\infty}\frac{\lambda^{x+y}}{x!y!}g(\kappa).$$

Note that

$$\begin{aligned}
q(\lambda) &= \left(\frac{\lambda^2}{1!1!}g(2.5)+\frac{\lambda^3}{1!2!}g(3.5)+\frac{\lambda^4}{1!3!}g(4.5)+\frac{\lambda^5}{1!4!}g(5.5)+\cdots\right)\\
&+ \left(\frac{\lambda^3}{2!1!}g(3.5)+\frac{\lambda^4}{2!2!}g(4.5)+\frac{\lambda^5}{2!3!}g(5.5)+\frac{\lambda^6}{2!4!}g(6.5)+\cdots\right)\\
&+ \left(\frac{\lambda^4}{3!1!}g(4.5)+\frac{\lambda^5}{3!2!}g(5.5)+\frac{\lambda^6}{3!3!}g(6.5)+\frac{\lambda^7}{3!4!}g(7.5)+\cdots\right)\\
&\vdots\\
&= \lambda^2 g(2.5)\left[\frac{1}{1!1!}\right]+\lambda^3 g(3.5)\left[\frac{1}{1!2!}+\frac{1}{2!1!}\right]+\lambda^4 g(4.5)\left[\frac{1}{1!3!}+\frac{1}{2!2!}+\frac{1}{3!1!}\right]+\cdots\\
&= \sum_{k=2}^{\infty}\lambda^k g(k+0.5)\sum_{\ell=1}^{k-1}\frac{1}{\ell!(k-\ell)!}.
\end{aligned}$$

On the other hand, from the binomial expansion we have

$$\sum_{\ell=1}^{k-1}\frac{1}{\ell!(k-\ell)!}=\frac{1}{k!}\sum_{\ell=1}^{k-1}\frac{k!}{\ell!(k-\ell)!}=\frac{1}{k!}\sum_{\ell=1}^{k-1}\binom{k}{\ell}=\frac{2^k-2}{k!}.$$

Thus, it follows that

$$q(\lambda)\equiv\sum_{x=1}^{\infty}\sum_{y=1}^{\infty}\frac{\lambda^{x+y}}{x!y!}g(\kappa)=\sum_{k=2}^{\infty}\frac{\lambda^k(2^k-2)}{k!}g(k+0.5). \tag{2.17}$$

Note that $q(\lambda)$ in (2.17) is a non-linear function of λ, and does not have a closed form in terms of λ. Thus, we manipulate the summand in (2.17) to obtain a closed form through an approximation in the following manner. Rewrite $q(\lambda)$ as

$$\begin{aligned}
q(\lambda) &= \sum_{k=2}^{\infty}\frac{\lambda^k(2^k-2)}{k!}g(k+0.5)\\
&= \sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}g(k+0.5)+\sum_{k=\tau+1}^{\infty}\frac{\lambda^k(2^k-2)}{k!}g(k+0.5)\\
&\approx \sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}\frac{12^\alpha+3\cdot 8^\alpha+6\cdot 6^\alpha}{12^\alpha+4\cdot 8^\alpha+4\cdot 6^\alpha}+\left[\sum_{k=2}^{\infty}\frac{\lambda^k(2^k-2)}{k!}-\sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}\right]\\
&= \sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}\left[\frac{12^\alpha+3\cdot 8^\alpha+6\cdot 6^\alpha}{12^\alpha+4\cdot 8^\alpha+4\cdot 6^\alpha}-1\right]+\sum_{k=2}^{\infty}\frac{(2\lambda)^k-2\lambda^k}{k!}\\
&= \left[\sum_{k=2}^{\tau}\frac{\lambda^k(2^k-2)}{k!}\cdot\frac{2\cdot 6^\alpha-8^\alpha}{12^\alpha+4\cdot 8^\alpha+4\cdot 6^\alpha}\right]+(e^\lambda-1)^2.
\end{aligned} \tag{2.18}$$

The last part of (2.18) is readily achieved by the Maclaurin series expansion on e^x. □

To validate the use of approximation in Theorem 2.3, we present empirical calculations on the two functions associated with the summand in (2.18). Specifically, let

$$r(k)=\frac{\lambda^k(2^k-2)}{k!}g(k+0.5),$$

where $g(\cdot)$ is defined in (2.15), and let

$$r^*(k)=\frac{\lambda^k(2^k-2)}{k!}$$

be a counterfeit version of $r(k)$ by excluding $g(k+0.5)$ from $r(k)$. Figure 1 depicts the two functions $r(k)$ and $r^*(k)$ when $\lambda=2$. We observe a considerable difference

between the values of $r(k)$ and $r^*(k)$ around $k = 4$, and the difference decreases as k increases. Ultimately, it vanishes when k exceeds 10. We calculate the differences between $r(k)$ and $r^*(k)$ with several values of λ when $k = 20(2)30$ to validate the plausibility of approximation. From Table 2, we see that the difference decreases as the value of k increases for a fixed value of λ. We set $\tau = 20$ in our computations associated with the intrinsic prior in Theorem 2.3 for both simulation studies and real data analyses.

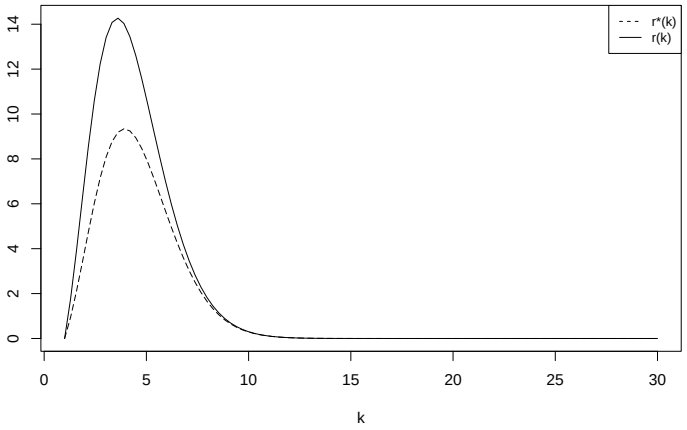

Figure 1. Comparison of $r(k)$ and $r^*(k)$ when $\lambda = 2$ showing the plausibility of approximation

Table 2. The differences of $r(k)$ and $r^*(k)$ with different values of λ

λ	$r(k) - r^*(k)$					
	$k = 20$	$k = 22$	$k = 24$	$k = 26$	$k = 28$	$k = 30$
2	3.34×10^{-10}	5.13×10^{-12}	6.60×10^{-14}	7.22×10^{-16}	6.79×10^{-18}	5.55×10^{-20}
3	1.11×10^{-6}	3.84×10^{-8}	1.11×10^{-9}	2.74×10^{-11}	5.79×10^{-13}	1.06×10^{-14}
4	3.50×10^{-4}	2.15×10^{-5}	1.11×10^{-6}	4.85×10^{-8}	1.82×10^{-9}	5.96×10^{-11}
5	3.03×10^{-2}	2.92×10^{-3}	2.35×10^{-4}	1.60×10^{-5}	9.43×10^{-7}	4.83×10^{-8}

Note that the joint prior in (2.14) does not depend on ω_1 and ω_2. Since the support of each of ω_1 and ω_2 is defined on the interval $(0, 1)$, the marginal intrinsic prior for λ is proportional to $\pi_1^I(\lambda, \omega_1, \omega_2)$. Figure 2 shows the marginal intrinsic prior $\pi_1^I(\lambda)$ when $\tau = 20$. The prior density is unimodal and skewed to the right.

3. Simulation studies

In this section, we perform Monte Carlo simulation studies to evaluate the performance of default Bayes factors for testing the null hypothesis $H_0 : \omega_1 = \omega_2$ against the alternative $H_1 : \omega_1 \neq \omega_2$ when each of two populations independently follows a ZIP distribution with the same λ and different zero inflation parameters ω_1 and ω_2. We generate data with four configurations of (ω_1, ω_2): $(0.2, 0.2)$, $(0.2, 0.3)$, $(0.2, 0.5)$, and $(0.2, 0.7)$. We use two values of $\lambda = 4$ and $\lambda = 6$ to see if there is an effect of λ on the performance of the proposed approach. Regarding sample sizes, we consider $m = n = 10$, 20, and 30 to investigate the performance of the proposed procedures in an asymptotic sense. We calculated two default Bayes factors; the intrinsic and fractional Bayes factors denoted by B_{10}^I and B_{10}^F, respectively. Additionally, the ordinary Bayes factor using the intrinsic prior in (2.14) is

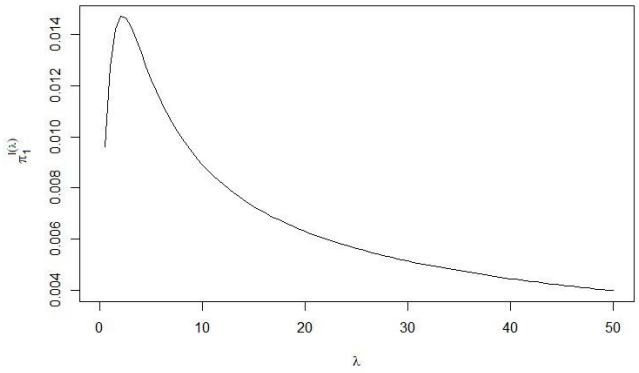

Figure 2. The marginal intrinsic prior $\pi_1^I(\lambda)$ for testing $H_0 : \omega_1 = \omega_2$ vs. $H_1 : \omega_1 \neq \omega_2$

calculated from the full data. This Bayes factor is denoted by B_{10}^{I*} in Tables 3 and 4. A total of 1,000 replications was used to carry out this simulation study.

In Tables 3 and 4, we provide the simulated averages and standard deviations (in parenthesis) of the three Bayes factors based on 1,000 simulated data. As mentioned in Remark 2.1, the common choice of fraction δ would be the size of the (minimal) training sample divided by the whole sample size, i.e., $\delta = 4/N$. First, when simulated data are generated from H_0, i.e., $(\omega_1, \omega_2) = (0.2, 0.2)$, all three Bayes factors decrease with an increase in sample sizes, indicating improved precision. On the other hand, when the data are generated from H_1, all three Bayes factors increase as the sample size increases These trends are perfectly observed despite the value of λ, as is expected from a theoretical point of view.

Second, when $(\omega_1, \omega_2) = (0.2, 0.2)$ the relative differences between B_{10}^I and B_{10}^F, on average, vary between 1.36 and 5.88%. This implies that the choice of fraction δ seems to be adequate for equivalence between the IBF and FBF. However, these relative differences increase as the differential between ω values increases. Regarding asymptotic equivalence between B_{10}^I and B_{10}^{I*}, the relative differences between these two Bayes factors, on average, vary between 14 and 20% when the differential between ω values is less than or equal to 0.1. Moreover, it turned out that the performance was better for $\lambda = 6$ with fixed values of ω. However, these relative differences also increase as the differential between ω values increases. Third, these relative differences decrease as the sample size increases for fixed values of ω and λ in most cases. However, these phenomena are not perfectly consistent.

To compare the default Bayesian approach conducted here with the frequentist approach, we performed a likelihood ratio test (LRT). We note that the maximum likelihood estimates (MLE) do not exist in closed forms, and the Newton-Raphson method was adopted to calculate the MLEs and the resulting P-values, denoted by p_{LR}. It can be obtained in a straightforward manner, and thus the details are omitted. The medians of the P-values calculated based on 1,000 replications in each configuration of parameters are reported in Tables 3 and 4. Even though there were not large differences in the P-values when the data were generated from H_0, we observed small discrepancies between the P-values. However, when the data were generated from H_1, all median P-values decreased as the sample size increased, except for cases with $(\omega_1, \omega_2) = (0.2, 0.3)$ and $\lambda = 4$.

Next, we present the proportion of $B_{10}^I > 1$ (i.e., supporting H_1 based on the Bayes factor), the proportion of $p_{LR} < 0.05$ (i.e., supporting H_1 based on the LRT), the proportion of both Bayesian and frequentist approaches supporting the true model (denoted by $P1$), and the proportion of both Bayesian and frequentist approaches yielding the same conclusion (denoted by $P2$). We see that the proportion of identifying the true model increases as the sample size increases when the differential between ω values is equal to or greater than 0.3. In particular, the proposed model effectively identifies the true model when $(\omega_1, \omega_2) = (0.2, 0.7)$ even with small sample sizes. The proposed approach based on the Bayes factors provides comparable results with the frequentist approach based on the LRT in identifying the true model. The frequentist and Bayesian approaches agree with each other at least 86% of the time in all cases considered here when using a moderate sample size of 30 from each population. The proposed approach based on Bayes factors has higher success rates in perceiving the true model than does the frequentist approach especially when the simulated data are generated under the alternative.

Table 3. The average values and standard deviations (in parenthesis) of the three Bayes factors based on 1,000 replications. The following proportions are provided: $B_{10}^I > 1$, $p_{LR} < 0.05$, $P1$ = Both B_{10}^I and LRT support the true model, and $P2$ = Both B_{10}^I and LRT yield the same conclusion

(ω_1, ω_2)	λ	n	B_{10}^I	B_{10}^F	B_{10}^{I*}	P-value	Proportions			
							$B_{10}^I > 1$	$p_{LR} < 0.05$	$P1$	$P2$
$(0.2, 0.2)$	4	10	1.026 (2.189)	1.012 (1.519)	0.818 (1.759)	0.528 (0.325)	16.5	7.7	76.6	77.4
		20	0.859 (1.879)	0.909 (1.751)	0.688 (1.514)	0.464 (0.302)	16.7	3.8	83.3	87.1
		30	0.817 (2.726)	0.865 (2.468)	0.654 (2.181)	0.528 (0.302)	11.8	5.3	88.2	93.5
	6	10	1.027 (2.247)	1.007 (1.530)	0.849 (1.853)	0.527 (0.339)	16.9	7.8	83.1	90.9
		20	0.820 (1.674)	0.865 (1.542)	0.680 (1.388)	0.528 (0.305)	16.6	4.3	85.9	86.7
		30	0.819 (2.363)	0.859 (2.221)	0.678 (1.956)	0.518 (0.306)	11.4	6.5	88.6	95.1
$(0.2, 0.3)$	4	10	1.444 (3.905)	1.297 (2.433)	1.145 (3.051)	0.228 (0.326)	23.6	10.3	10.3	86.7
		20	2.399 (15.291)	2.158 (11.906)	1.915 (12.116)	0.427 (0.317)	27.8	11.0	11.0	86.2
		30	2.695 (19.055)	2.474 (15.313)	2.154 (15.225)	0.370 (0.310)	23.2	11.4	11.4	91.2
	6	10	1.549 (4.440)	1.346 (2.663)	1.279 (3.649)	0.528 (0.344)	23.4	10.1	10.1	86.7
		20	2.099 (15.746)	1.907 (12.175)	1.784 (13.056)	0.427 (0.325)	28.2	10.0	10.0	88.8
		30	2.743 (29.744)	2.430 (21.895)	2.272 (24.642)	0.345 (0.311)	23.6	14.0	14.0	90.3

4. Real data analysis

In this section, we illustrate the proposed methodologies under the ZIP model using two real datasets that contain excessive zeros observed in time intervals.

Table 4. Continuation of Table 3

(ω_1, ω_2)	λ	n	B_{10}^I	B_{10}^F	B_{10}^{I*}	P-value	Proportions			
							$B_{10}^I > 1$	$p_{LR} < 0.05$	$P1$	$P2$
$(0.2, 0.5)$	4	10	13.339	7.059	10.616	0.154	57.6	31.6	31.6	74.0
			(1.15×10^2)	(44.170)	(91.076)	(0.309)				
		20	63.209	42.561	50.697	0.044	75.0	50.5	50.5	77.5
			(5.92×10^2)	(3.64×10^2)	(4.78×10^2)	(0.208)				
		30	4.71×10^3	2.66×10^3	3.77×10^3	0.001	81.8	67.7	67.7	85.9
			(1.36×10^5)	(7.53×10^4)	(1.09×10^5)	(0.169)				
	6	10	16.444	8.044	13.614	0.155	58.9	33.8	30.7	68.7
			(1.52×10^2)	(53.500)	(1.26×10^2)	(0.384)				
		20	81.987	52.537	67.842	0.044	76.1	50.8	50.8	74.7
			(7.74×10^2)	(4.52×10^2)	(6.41×10^2)	(0.247)				
		30	1.54×10^3	9.18×10^2	1.27×10^3	0.014	82.8	69.7	69.7	86.9
			(3.41×10^4)	(1.93×10^4)	(2.83×10^4)	(0.186)				
$(0.2, 0.7)$	4	10	87.155	37.213	69.838	0.021	89.2	62.6	62.5	73.2
			(3.80×10^2)	(1.38×10^2)	(3.07×10^2)	(0.187)				
		20	5.85×10^4	2.23×10^4	4.69×10^4	0.001	97.3	89.8	89.8	92.5
			(1.40×10^6)	(5.05×10^5)	(1.12×10^6)	(0.067)				
		30	6.30×10^5	3.18×10^6	5.01×10^5	< 0.001	99.5	98.5	98.5	99.0
			(6.00×10^6)	(2.93×10^6)	(4.76×10^6)	(< 0.001)				
	6	10	97.927	39.717	81.022	0.015	90.3	60.4	60.4	70.1
			(4.37×10^2)	(1.50×10^2)	(3.62×10^2)	(0.291)				
		20	8.91×10^4	3.20×10^4	7.37×10^4	0.001	97.7	86.6	86.6	88.9
			(2.01×10^6)	(1.66×10^6)	(6.77×10^5)	(0.221)				
		30	7.84×10^5	3.78×10^5	6.48×10^5	< 0.001	99.6	94.8	94.8	95.2
			(6.86×10^6)	(3.20×10^5)	(5.66×10^6)	(< 0.001)				

4.1. Yellow dust storms data

We consider a dataset related to the number of yellow dust storms that have been observed in South Korea over the last two decades. Sand and dust particles mainly originate from deserts in China and Mongolia and are carried from the Yangtze River through the westerly winds, especially in the Spring. Thus, the frequency of yellow dust occurrence is higher in the first half of the year when the Yangtze River basin is heavily affected. We selected two metropolitan cities of South Korea with a population more than one million: Incheon and Gwangju. The distance between the two cities is about 256 kilo meters, and their locations along with a small part of the eastern seaboard of China are depicted in Figure 3. We collected the number of yellow dust storms observed from the first half of 2003 to the second half of 2022 on a semi-annual basis. Thus, the samples sizes are $m = n = 40$, among which there are 8 and 10 zero observations for Incheon and Gwangju, respectively. The data are accessed on the Korea Meteorological Administration website. At the initial stage, the MLEs for λ were calculated to be 5.01 for Incheon and 4.56 for Gwangju assuming that each of the two populations independently follows a ZIP distribution.

Table 5 shows the results for testing $H_0 : \omega_1 = \omega_2$ based on the data of the two selected cities. We reported three Bayes factors, which were denoted as in the simulation study. Assuming that a default Bayes factor, B_{10}^I is true, the other default Bayes factor, B_{10}^F is slightly overestimated, although there seems to be no large difference between them. On the other hand, B_{10}^{I*} is slightly underestimated with a value of 0.2558. However, all three Bayes factors are less than one, indicating no support for H_1. We also report the P-value

based on the LRT, yielding a value of 0.5921. Thus, we conclude that there is no strong evidence to support the alternative hypothesis.

Table 5. Results for testing $H_0 : \omega_1 = \omega_2$ versus $H_1 : \omega_1 \neq \omega_2$ for yellow dust data

Cities	B_{10}^{I}	B_{10}^{F}	B_{10}^{I*}	P-value
Incheon vs. Gwangju	0.3119	0.3480	0.2558	0.5921

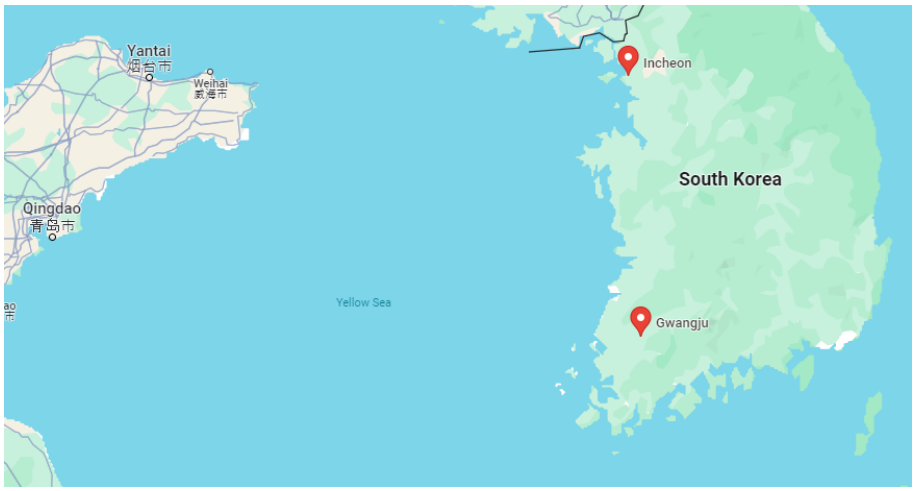

Figure 3. Two cities in South Korea affected by yellow dusts

4.2. Book reading data

In this subsection, we utilized a reading dataset to demonstrate the proposed methodologies as an illustration. National reading surveys have been conducted annually by the Ministry of Culture, Sports, and Tourism since 1993. The survey targets 6,000 adults older than 19 years who live in South Korea, and 3,320 elementary, middle, or high school students nationwide. In this survey, the data were collected based on the answers to the question: "How many paper books did you read from September 2020 to August 2021, excluding textbooks, reference books, and test preparation books?" For illustrative and computational purposes, we only considered adults who live in Gyeonggi Province and read fewer than 50 books per year, resulting in a total of 788 samples. We also limited the sample of students to those who live in Gyeonggi Province and are attending high school. Regarding the selection of adults, we restricted the annual reading quantity to fewer than 50 books. A total of 270 samples was obtained from the student side.

To produce a setup similar to that of the earthquake data experiment, we divided the adult reading dataset into two groups by gender. The number of males was 359 with 53.7% of them being zero observations, whereas the number of females was 429, with 52% of them being zero observations. A plausible explanation for observing such zero observations is the decrease in the market share of paper books due to the recent developments of e-books and audio-books. Another reason would be Korea's fixed pricing system in the paper books market, resulting in relatively high prices for book and limiting purchases.

Assuming that each group (population) independently follows a ZIP distribution, the MLEs of λ turned out to be 6.422 for men and 6.515 for women respectively, showing no big difference in λ values between the two groups. Thus, we proceeded to test H_0 :

$\omega_1 = \omega_2$ with the same λ for the two ZIP distributions after we randomly extracted 30 observations from each of the two groups to avoid heavy computation. Table 6 presents two default Bayes factors and the ordinary Bayes factor computed with the intrinsic priors for analyzing book reading data. We also report the P-value based on the LRT described in Section 3. While we see that two default Bayes factors, B_{10}^I and B_{10}^F are close to each other, there is little difference between the two default Bayes factors and B_{10}^{I*}. All the Bayes factors are less than one, showing no prominent evidence to support H_1. Moreover, the P-value based on the LRT statistic is 0.6054, which does not show significant evidence to support the alternative hypothesis.

Next, we carried out the same test based on the data for adult men versus high school students from Gyeonggi Province. The MLE of λ for the high school student group is 6.411, showing a small difference in λ values between adult men and high school students. This analysis produced very different magnitudes of the Bayes factors, with all being very large values, supporting H_1. Again, there is not much difference between B_{10}^I and B_{10}^F, and there is little difference between the default Bayes factors and B_{10}^{I*}. A very small P-value of 0.0039 showed conformity to the results obtained through the three Bayes factors. Since reading is a mandatory part of the learning curriculum for students in South Korea, there is a considerable difference in the two proportions of zero inflations between adults and students.

Table 6. Results for testing $H_0 : \omega_1 = \omega_2$ versus $H_1 : \omega_1 \neq \omega_2$ for book reading data

	B_{10}^I	B_{10}^F	B_{10}^{I*}	P-value
Men vs. Women	0.4204	0.4221	0.3497	0.6054
Men vs. High school	17.522	16.205	14.253	0.0039

5. Concluding remarks

In this paper, we performed Bayesian hypothesis testing for the zero-inflation parameters when two underlying distributions independently follow ZIP distributions. Intrinsic and fractional approaches were used to calculate the default Bayes factors. An intrinsic prior associated with the intrinsic approach was derived through a reasonable approximation. The proposed Bayesian approach and the existing frequentist approach based on the LRT provided comparable results under the ZIP distribution.

There are some drawbacks to the proposed method. As mentioned in Introduction, we expected asymptotic equivalence properties for the two Bayes factors; however, increasing sample sizes did not significantly reduce the relative differences between the two Bayes factors. The numerical integration blew up when calculating the marginal distribution, revealing limitations to computing for large sample sizes. Finally, we observed relatively poorer results when using smaller values of λ.

Finding intrinsic priors based on the exact method is virtually impossible unless a closed form is available on the expected value of the Bayes factor. This forces us to approximately calculate the expected value. We did not attempt to derive intrinsic priors based on the fractional approach by limiting the correction factor. This is mainly due to the fact that the correction factor requires a complex calculation. However, some of the issues could be resolved if we were to utilize the approach proposed by Gu et al. [13]. This project may present considerable challenges for forthcoming research. Research in this direction is in progress, and we hope to report its results in a future paper.

Acknowledgments

The authors are grateful for constructive feedbacks of the Editors and four anonymous reviewers, leading to considerable improvements from the previous version.

Author contributions. K. Kim developed the theoretical foundations for intrinsic priors and designed the experimental setup for the simulations. H. J. Jeong performed computational work in collaboration with K. Kim and collected real datasets to validate the proposed methodologies. Y. Kim contributed to the interpretation of the experimental results. S. W. Kim organized the manuscript structure and wrote its main content. Additionally, S. W. Kim discussed the theoretical methodologies with Y. Kim. All authors have read and approved the final manuscript.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. Yongdai Kim's work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)]. Seong W. Kim's research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A2C1005271).

Data availability. Data were collected on the Korea Meteorological Administration website (`https://www.weather.go.kr`) and Korea Ministry of Culture, Sports, and Tourism website (`https://mdis.kostat.go.kr/index.do`).

References

[1] I. A. Almodóvar-Rivera and L. R. Pericchi-Guerra, *An objective and robust Bayes factor for the hypothesis test one sample and two population means*, Entropy **26** (1), 1-25, 2024.

[2] A. M. Azexedo, *Í*. J. Silva, M. C. Nery, H. P. Rocha and R. A. Santana, *Counting models for overdispersed data: A review with application to tuberculosis data*, Braz. J. Biometrics **41** (3), 274-286, 2023.

[3] M. J. Bayarri, J. O. Berger and G. S. Datta, *Objective Bayes testing of Poisson versus inflated Poisson models*, IMS Collect. **3**, 105-121, 2008.

[4] J. O. Berger, *The case for objective Bayesian analysis*, Bayesian Anal. **1** (3), 385-402, 2006.

[5] J. O. Berger and J. M. Bernardo, *Estimating a product of means: Bayesian analysis with reference priors*, J. Am. Stat. Assoc. **84** (405), 200-207, 1989.

[6] J. O. Berger and J. Moreta, *Default Bayes factors for nonnested hypothesis testing*, J. Am. Stat. Assoc. **94** (446), 542-554, 1999.

[7] J. O. Berger and L. Pericchi, *The intrinsic Bayes factor for model selection and prediction*, J. Am. Stat. Assoc. **91** (433), 109-122, 1996.

[8] S. Chen, Y. Li, J. Kim and S. W. Kim, *Bayesian change point analysis for extreme daily precipitation*, Int. J. Climatol. **37** (7), 3123-3137, 2017.

[9] R. Clare, *A universal robust bound for the intrinsic Bayes factor*, Ph.D. dissertation, Univ. Puerto Rico, 2024.

[10] A. C. Cohen, *Estimation in mixtures of discrete distributions*, in Proc. Int. Symp. Discrete Distrib., Montreal, 373-378, 1963.

[11] C. Conigliani and A. O'Hagan, *Sensitivity of the fractional Bayes factor to prior distributions*, Can. J. Stat. **28** (2), 343-352, 2000.
[12] G. Consonni, D. Fouskakis, B. Liseo and I. Ntzoufras, *Prior distributions for objective Bayesian analysis*, Bayesian Anal. **13** (2), 627-679, 2018.
[13] X. Gu, J. Mulder and H. Hoijtink, *Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses*, Br. J. Math. Stat. Psychol. **71** (2), 229-261, 2018.
[14] D. B. Hall, *Zero-inflated Poisson and binomial regression with random effects: A case study*, Biometrics **56** (4), 1030-1039, 2000.
[15] Y. Han, H. Hwang, H. K. T. Ng and S. W. Kim, *Default Bayesian testing for the Zero-inflated Poisson distribution*, Stat. Interface **17** (4), 623-634, 2024.
[16] H. Jeffreys, *Theory of Probability*, 3rd ed., Oxford Univ. Press, 1961.
[17] R. E. Kass and A. E. Raftery, *Bayes factors*, J. Am. Stat. Assoc. **90** (430), 773-795, 1995.
[18] S. W. Kim, *Intrinsic priors for testing exponential means*, Stat. Probab. Lett. **46** (2), 195-201, 2000.
[19] S. W. Kim and D. Kim, *Intrinsic priors for two-sample tests in normal populations*, Commun. Stat.-Theory Methods **31** (7), 1091-1105, 2002.
[20] S. W. Kim and D. Sun, *Intrinsic priors for model selection using an encompassing model with applications to censored failure time data*, Lifetime Data Anal. **6**, 251-269, 2000.
[21] D. Lambert, *Zero-inflated Poisson regression, with an application to defects in manufacturing*, Technometrics **34** (1), 1-14, 1992.
[22] K. Lee, Y. Joo, J. J. Song and D. W. Harper, *Analysis of zero-inflated clustered count data: A marginalized model approach*, Comput. Stat. Data Anal. **55** (1), 824-837, 2011.
[23] H. K. Lim, W. K. Li and P. L. H. Yu, *Zero-inflated Poisson regression mixture model*, Comput. Stat. Data Anal. **71**, 151-158, 2014.
[24] D. L. Long, J. S. Preisser, A. H. Herring and C. E. Golin, *A marginalized zero-inflated Poisson regression model with random effects*, J. R. Stat. Soc. C **64** (5), 815-830, 2015.
[25] K. Mahmood and F. Havva, *Inferences for the inflation parameter in the zip distributions: The method of moments*, Stat. Methodol. **8** (4), 377-388, 2011.
[26] Y. Min and A. Agresti, *Random effect models for repeated measures of zero-inflated count data*, Stat. Modell. **5** (1), 1-19, 2005.
[27] E. Moreno, *Objective Bayesian methods for one-sided testing*, Test **14** (1), 181-198, 2005.
[28] J. Mullahy, *Specification and testing of some modified count data models*, J. Econom. **33** (3), 341-365, 1986.
[29] B. Neelon and D. Chung, *The LZIP: A Bayesian latent factor model for correlated zero-inflated counts*, Biometrics **73** (1), 185-196, 2017.
[30] A. O'Hagan, *Fractional Bayes factors for model comparison*, J. R. Stat. Soc. B **57** (1), 99-118, 1995.
[31] L. Perreault, J. Bernier, B. Bobée and E. Parent, *Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting*, J. Hydrol. **235** (3-4), 242-263, 2000.
[32] J. Schwartz and D. Giles, *Bias-reduced maximum likelihood estimation of the zero-inflated Poisson distribution*, Commun. Stat.-Theory Methods **45** (2), 465-478, 2016.
[33] S. Sivaganesan and D. Jiang, *Objective Bayesian testing of a Poisson mean*, Commun. Stat.-Theory Methods **39** (11), 1887-1897, 2010.
[34] Y. Wang and L. Pericchi, A bridge between cross-validation Bayes factors and geometric intrinsic Bayes factors, arXiv: 2006.06495v1.

[35] X. Xiao, Y. Tang, A. Xu and G. Wang, *Bayesian inference for zero-and-one-inflated geometric distribution regression model using Polya-gamma latent variables*, Commun. Stat.-Theory Methods **49** (15), 3730-3743, 2020.

[36] H. Xu, M. Xie and T. N. Goh, *Objective Bayes analysis of zero-inflated Poisson distribution with application to healthcare data*, IIE Trans. **46** (8), 843-852, 2014.

[37] K. K. Yau and A. H. Lee, *Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme*, Stat. Med. **20** (19), 2907-2920, 2001.