A Choice Between Poisson and Geometric Distributions

Biswabrata Pradhan *& Debasis Kundu †

Abstract

Both Poisson and geometric distributions can be used quite effectively in modeling count data. In this paper we compare Bayesian and frequentists criteria to choose between Poisson and geometric distributions. In the frequentist choice, we use the ratio of maximized likelihood in discriminating between the two distributions. Asymptotic distributions of the ratio of maximized likelihoods are also obtained, and they can be used to compute the minimum sample size needed to discriminate between the two distributions. We further investigate the Bayesian model selection criterion in choosing the correct model, under a fairly general set of priors. We perform some simulation experiments to see the effectiveness of the proposed methods and the analysis of one data set is performed for illustrative purposes.

Key Words and Phrases: Likelihood ratio test; asymptotic distribution; Bayes factor, model selection, probability of correct selection.

*SQC & OR Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, Pin 700108, India. Corresponding author, e-mail: bis@isical.ac.in

†Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. Corresponding author, e-mail: kundu@iitk.ac.in.

1 Introduction

We address the following problem in this paper. Suppose, an experimenter has a sample of count data, say $\{x_1, \cdots, x_n\}$, and she wants to use the Poisson or geometric distribution to analyze this data set, which one is preferable or it should not make any difference?

Both Poisson and geometric distributions can be used quite effectively to analyze count data. Both the distributions have some physical interpretations also. Therefore, if the data follow one of those physical processes, it is easier to fit that specific distributions, otherwise the problem might be more difficult. It may be noted that even if both the models fit the data reasonably well, it might be very important to choose the correct model, since the inferences based on model will often involve tail probabilities, where the affect of model assumptions is very critical. Therefore, even if we have a small or moderate samples, it is still very important to make the best possible decision based on whatever data are available at hand.

The problem of testing, whether some given observations follow one of the two probability distributions is quite old. Cox (1961, 1962) first proposed to use the ratio of maximized likelihood in discriminating two separate probability distributions. Since then, extensive work has been done to discriminate between two continuous families of distributions, see for example Dey and Kundu (2012), and the references cited therein for an extensive list of references in this topic available till today.

Interestingly, although extensive work has been done in discriminating between two continuous distribution functions, not much attention has been paid in discriminating between two discrete distribution functions. Vu and Maller (1996) proposed a likelihood ratio test of testing H_0: Poisson *vs.* H_1: Binomial, see also Dauxois, Druilhet and Pommeret (2006) in this respect.

In this paper, we consider the problem of discriminating between Poisson and geometric distributions given a sample of size n. The main reasons to choose these two distributions are that both of them have the same support, and also both these distributions have the same number of unknown parameters. Hence, the usual AIC or BIC method will not work in this case. We consider the problem both from the frequentist and Bayesian view points. First we consider the difference of the maximized log-likelihood (DMLL) values to discriminate between these two distributions. We obtain the asymptotic distributions of DMLL under both the null distributions, and it is observed that DMLL is asymptotically normally distributed. The asymptotic distribution can be used to compute the probability of correct selection (PCS), and it is observed in the simulation study that the asymptotic distribution works quite well even for small sample sizes. The asymptotic distribution can be used to obtain the critical regions of the corresponding testing of hypothesis problem also.

We also find the minimum sample size needed in discriminating between the two distribution functions for a given PCS. Using the asymptotic distribution of DMLL, we obtain the minimum sample size needed to discriminate between the two distribution functions for a given user specified protection level, *i.e.* the PCS. The asymptotic distribution of DMLL can be used for testing purposes also, and we have provided the details in Section 4.

For completeness purposes, we further consider the Bayesian model choice between Poisson and geometric distributions under fairly general priors based on Bayes factor approach. It is observed that the Bayes factor can be expressed in explicit form. One data analysis is performed for illustrative purposes, and it is observed that different methods provide the same results in this case.

Rest of the paper is organized as follows. In Section 2, we provide the test criterion, and the corresponding asymptotic distribution is provided in Section 3. Sample size determination and testing of hypotheses problems have been considered in Section 4. Simulation results are presented in Section 5. Bayesian model selection has been presented in Section 6. The analysis of one data set has been presented in Section 7. Finally conclusions appear in Section 8.

2 Selection Criterion

Suppose $X_1, \cdots, X_n$ are independent and identically distributed (*i.i.d.*) random variables from any one of the two distribution functions. The probability mass function (PMF) of a Poisson distribution with the parameter λ is given by

$$p_{PO}(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, 2, \cdots \tag{1}$$

The PMF of geometric distribution is given by

$$p_{GE}(x;\theta) = \theta(1-\theta)^x; \quad x = 0, 1, 2, \cdots, \tag{2}$$

here X denotes the number of failures before the first success. From now on, a Poisson distribution with parameter λ and a geometric distribution with parameter θ will be denoted by $P(\lambda)$ and $G(\theta)$ respectively.

Let us denote the log-likelihood functions of the observations $\{X_1, \cdots, X_n\}$, that they are coming from a $P(\lambda)$ or from a $G(\theta)$, respectively as

$$l_P(\lambda) = -n\lambda + \sum_{i=1}^{n} X_i \ln\lambda - \sum_{i=1}^{n} \ln X_i!, \tag{3}$$

$$l_G(\theta) = n\ln\theta + \ln(1-\theta)\sum_{i=1}^{n} X_i. \tag{4}$$

The logarithm of the ratio of maximized likelihood functions or the difference of the maximized log-likelihood functions is $T = l_p(\widehat{\lambda}) - l_G(\widehat{\theta})$, where $\widehat{\lambda} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\widehat{\theta} = \frac{1}{1+\overline{X}_n}$ are the maximum likelihood estimators (MLEs) of λ and θ, under the assumptions that the data are coming from $P(\lambda)$ and $G(\theta)$ respectively.

The statistic T can be written as

$$T = l_p(\widehat{\lambda}) - l_G(\widehat{\theta}) = n(1+\overline{X}_n)\ln(1+\overline{X}_n) - n\overline{X}_n - \sum_{i=1}^{n} \ln X_i!. \qquad (5)$$

The natural model selection criterion will be to choose Poisson distribution, if $T > 0$, otherwise, choose geometric distribution. In the next section we derive the asymptotic distribution of T.

3 Asymptotic Properties of T

In this section we derive the asymptotic distribution of T for two different cases, namely when the data are coming from $P(\lambda)$ and from $G(\theta)$ respectively. We use the following notations. Almost sure convergence will be denoted by $a.s.$. For any Borel measurable function, $h(\cdot)$, $E_P(h(U))$ and $V_p(h(U))$ will denote the mean and variance of $h(U)$ under the assumptions that U follows $P(\lambda)$. Similarly, we define $E_G(h(U))$ and $V_G(h(U))$ as mean and variance of $h(U)$ under the assumption that U follows $G(\theta)$. Moreover, if $g(\cdot)$ and $h(\cdot)$ are two Borel measurable functions, we define $Cov_P(g(U), h(U)) = E_P(g(U)h(U)) - E_P(g(U))E_P(h(U))$, and $Cov_G(g(U), h(U)) = E_G(g(U)h(U)) - E_G(g(U))E_P(h(U))$,

3.1 Case I: The Data follow Poisson Distribution

In this case we have the following main result.

Theorem 1: Under the assumption that the data follow $P(\lambda)$, the distribution of T is approximately normally distributed with mean $E_P(T)$ and variance $V_P(T)$. The explicit expressions of $E_P(T)$ and $V_P(T)$ will be provided later.

To prove Theorem 1, we need the following lemma.

Lemma 1: Suppose the data are from $P(\lambda)$, then as $n \to \infty$, we have the following;

(i) $\widehat{\lambda} \to \lambda \quad a.s.$

(ii) $\widehat{\theta} \to \widetilde{\theta}, \quad a.s.,$ where

$$E_P[\ln(p_{GE}(X;\widetilde{\theta}))] = \max_{\theta} E_P[\ln(P_{GE}(X;\theta))]$$

(iii) If $T^* = l_P(\lambda) - l_G(\widetilde{\theta})$, then $n^{-1/2}(T - E_P(T))$ is asymptotically equivalent to $n^{-1/2}(T^* - E_P(T^*))$

Proof of Lemma 1: It follows using the similar argument as of White (1982, Theorem 1), and therefore it is omitted. ∎

Proof of theorem 1: Using the central Limit Theorem (CLT), it easily follows that $n^{-1/2}(T^* - E_P(T^*))$ is asymptotically normally distributed. Therefore, proof immediately follows from part (iii) of Lemma 1 and using CLT. ∎

Now we provide the explicit expressions of $\widetilde{\theta}$, $E_P(T)$ and $V_P(T)$. To compute $\widetilde{\theta}$, let us define

$$g(\theta) = E_P[\ln(P_{GE}(X;\theta))] = \ln\theta + \lambda\ln(1-\theta).$$

Therefore, it is immediate that the maximum of $g(\theta)$ occurs at

$$\widetilde{\theta} = \frac{1}{1+\lambda}.$$

Now we provide the expressions for $E_P(T)$ and $V_P(T)$. First observe that $\lim_{n\to\infty} E_P(T)/n$ and $\lim_{n\to\infty} V_P(T)/n$ exist. Suppose we denote $\lim_{n\to\infty} E_P(T)/n = AM_P$ and $\lim_{n\to\infty} V_P(T)/n = AV_P$. Therefore, for large n,

$$\begin{aligned} \frac{E_P(T)}{n} \approx AM_p &= E_P(\ln p_{PO}(X;\lambda) - \ln p_{GE}(X;\widetilde{\theta})) \\ &= (1+\lambda)\ln(1+\lambda) - \lambda - H_1(\lambda), \end{aligned} \tag{6}$$

here $H_1(\lambda) = E(\ln X!) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\ln(n!)\lambda^n}{n!}$. Similarly, for large n,

$$\frac{V_P(T)}{n} \approx AV_p = V_P(\ln p_{PO}(X;\lambda) - \ln p_{GE}(X;\widetilde{\theta}))$$

$$
\begin{aligned}
&= V_P(\ln(1+\lambda)X - \ln X!) \\
&= \lambda(\ln(1+\lambda))^2 + V_P(\ln X!) - 2\ln(1+\lambda)Cov_P(X, \ln X!) \\
&= \lambda(\ln(1+\lambda))^2 + H_2(\lambda) - H_1^2(\lambda) - 2\ln(1+\lambda)(H_3(\lambda) - \lambda H_1(\lambda)), \quad (7)
\end{aligned}
$$

where $H_1(\lambda)$ is same as defined before, and

$$
H_2(\lambda) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}(\ln n!)^2 \lambda^n}{n!}, \quad H_3(\lambda) = \sum_{n=1}^{\infty} \frac{e^{-\lambda}(\ln n!)\lambda^n}{(n-1)!}.
$$

3.2 Case II: The Data follow Geometric Distribution

Along the same line as Theorem 1, in this case we have the following main result.

Theorem 2: Under the assumption that the data follow $G(\theta)$, the distribution of T is approximately normally distributed with mean $E_G(T)$ and variance $V_G(T)$. The explicit expressions of $E_G(T)$ and $V_G(T)$ will be provided later.

To prove Theorem 2, we need the following lemma, similarly as Lemma 1.

Lemma 2: Suppose the data are from $G(\theta)$, then as $n \to \infty$, we have the following;

(i) $\widehat{\theta} \to \theta \quad a.s.$

(ii) $\widehat{\lambda} \to \widetilde{\lambda}, \quad a.s., \quad$ where

$$
E_G[\ln(p_{PO}(X; \widetilde{\lambda}))] = \max_{\lambda} E_G[\ln(P_{PO}(X; \lambda))]
$$

(iii) If $T_* = l_P(\widetilde{\lambda}) - l_G(\theta)$, then $n^{-1/2}(T - E_G(T))$ is asymptotically equivalent to $n^{-1/2}(T_* - E_P(T_*))$

Proof of Lemma 1: It follows using the similar argument as of White (1982, Theorem 1), and therefore it is omitted. ∎

Proof of theorem 2: It follows along the same line as Theorem 1. ∎

Now to obtain the explicit expressions of $\widetilde{\lambda}$, $E_G(T)$ and $V_G(T)$, first let us define

$$h(\lambda) = E_G[\ln(P_{PO}(X;\lambda))] = -\ln\lambda + \frac{1-\theta}{\theta}\ln\lambda - E(\ln X!).$$

Therefore, it is immediate that the maximum of $h(\lambda)$ occurs at

$$\widetilde{\lambda} = \frac{1-\theta}{\theta}.$$

The expressions for $E_G(T)$ and $V_G(T)$ can be obtained similarly as before. First observe that $\lim_{n\to\infty} E_G(T)/n$ and $\lim_{n\to\infty} V_G(T)/n$ exist. Suppose we denote $\lim_{n\to\infty} E_G(T)/n = AM_G$ and $\lim_{n\to\infty} V_G(T)/n = AV_G$. Therefore, for large n,

$$\begin{aligned}\frac{E_G(T)}{n} \approx AM_G &= E_G(\ln p_{PO}(X;\widetilde{\lambda}) - \ln p_{GE}(X;\theta) \\ &= \frac{\theta - 1 - \ln\theta}{\theta} - H_4(\theta), \end{aligned} \qquad (8)$$

here $H_4(\theta) = E_G(\ln X!) = \sum_{n=0}^{\infty} \theta(1-\theta)^n \ln(n!)$. Similarly, for large n,

$$\begin{aligned}\frac{V_G(T)}{n} \approx AV_G &= V_G(\ln p_{PO}(X;\widetilde{\lambda}) - \ln p_{GE}(X;\theta)) \\ &= V_G(X\ln\theta + \ln X!) \\ &= \frac{(\ln\theta)^2(1-\theta)}{\theta^2} + H_5(\theta) - H_4^2(\theta) + 2\ln\theta Cov_G(X, \ln X!) \\ &= \frac{(\ln\theta)^2(1-\theta)}{\theta^2} + H_5(\theta) - H_4^2(\theta) + 2\ln\theta\left(H_6(\theta) - \frac{H_4(\theta)(1-\theta)}{\theta}\right) \end{aligned} \qquad (9)$$

where $H_4(\theta)$ is same as defined before, and

$$H_5(\theta) = E_G(\ln X!)^2 = \sum_{n=0}^{\infty} \theta(1-\theta)^n (\ln(n!))^2, \quad H_6(\lambda) = E_G(X\ln X!) = \sum_{n=0}^{\infty} \theta(1-\theta)^n n\ln(n!)$$

4 Determination of Sample Size & Testing

4.1 Minimum Sample Size Determination

In this section, similarly as in Gupta and Kundu (2003) it is assumed that the data are coming from one of the two distribution functions, and we propose a method to determine the minimum sample size needed to choose between the geometric and a Poisson distributions, for a given user specified probability correct selection (PCS). It is expected the user knows the PCS, or provides a range of PCSs. It is intuitively clear that if the PCS is more, larger sample size is needed to choose the correct model. Note that as $\lambda \to 0$ and $\theta \to 1$, then both PO(λ) and GE(θ) converge to the same degenerate distribution, namely degenerating at 0. Therefore, in this case it is not possible to discriminate between the two.

If the two distribution functions are very close to each other, it is quite natural that one needs a very large sample size to discriminate between the two distribution functions. It is also true that if the two distribution functions are very close to each other, one may not need to discriminate the two distribution functions from any practical point of view. Therefore, it is assumed that that the user will specify before hand the PCS and also the tolerance limit in terms of the distance between two distribution functions. The tolerance limit simply indicates that one does not want to discriminate between two distribution functions if the distance between them is less that the tolerance limit.

There are several ways to define the distance measures between the two distribution functions. In this paper we mainly take the Kolmogorov-Smirnov (KS) distance measure between the two distribution functions. Based on the PCS, the tolerance limit and using the KS distance measure the minimum sample size can be obtained from the asymptotic distribution of T. The details will be explained below. But first we present the KS distance measure between PO(λ) and GE($\widetilde{\theta}$) for different values of λ and similarly between GE(θ) and PO($\widetilde{\lambda}$) for different values of θ.

Table 1: KS distances between PO(λ) and GE($\widetilde{\theta}$) for different values of λ.

λ	0.100	0.250	0.500	1.000	2.000
KS	0.004	0.021	0.060	0.132	0.198

Table 2: KS distances between GE(θ) and PO($\widetilde{\lambda}$) for different values of θ.

θ	0.300	0.400	0.500	0.600	0.700	0.800
KS	0.203	0.177	0.132	0.087	0.049	0.021

It is clear from the above two tables that the two distribution functions become closer as θ increases or λ decreases. Now from the asymptotic distribution of T, and using the KS distance measures between the two distribution functions, the minimum required sample size can be obtained, so that the PCS achieves a certain protection level p^* for a given tolerance level D^*. The tolerance level D^* is such that we do not want to discriminate between the two distribution functions if the distance between the two is less than D^*. We explain the procedure when the data are coming from the Poisson distribution, but the other case follows along the same line.

From Theorem 1, it is known that T is asymptotically normally distributed with mean $E_P(T)$ and variance $V_P(T)$. Therefore, the PCS can be written approximately as

$$PCS(\lambda) = P(T > 0|\lambda) \approx \Phi\left(\frac{E_P(T)}{\sqrt{V_P(T)}}\right) = \Phi\left(\frac{\sqrt{n}\text{AM}_P(\lambda)}{\sqrt{\text{AV}_P(\lambda)}}\right). \tag{10}$$

Here $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Now to determine the sample size needed to achieve a protection level p^*, equate

$$\Phi\left(\frac{\sqrt{n}\text{AM}_P(\lambda)}{\sqrt{\text{AV}_P(\lambda)}}\right) = p^*, \tag{11}$$

and solve for n, *i.e.*

$$n = \frac{z_{p^*}^2 \text{AV}_P(\lambda)}{(\text{AM}_P(\lambda))^2}, \tag{12}$$

here z_{p^*} is the $100p^*$ percentile point of a standard normal distribution. Similarly, if the data are coming from a geometric distribution, the minimum sample size needed to achieve a protection level p^* is

$$n = \frac{z_{p^*}^2 \text{AV}_G(\theta)}{(\text{AM}_G(\theta))^2}, \tag{13}$$

From (12) and (13), it is clear that the minimum sample size needed to discriminate between a Poisson and a geometric distribution depends on the unknown parameters.

In Table 3 we provide $\text{AM}_P(\lambda)$, $\text{AV}_P(\lambda)$ and $R_1(\lambda) = \dfrac{\text{AV}_P(\lambda)}{(\text{AM}_P(\lambda))^2}$ for different values of λ.

Table 3: The values of AM_P and AV_P for different values of λ.

	λ				
	0.25	0.50	1.00	2.00	3.00
$\text{AM}_P(\lambda)$	0.00799	0.02713	0.08145	0.20466	0.31787
$\text{AV}_P(\lambda)$	0.01327	0.04557	0.13155	0.28312	0.37311
$R_1(\lambda)$	207.9	61.9	19.8	6.8	3.7

In Table 4 we provide $\text{AM}_G(\theta)$, $\text{AV}_G(\theta)$ and $R_2(\theta) = \dfrac{\text{AV}_G(\theta)}{(\text{AM}_G(\theta))^2}$ for different values of θ.

Table 4: The values of AM_G and AV_G for different values of θ.

	θ					
	0.30	0.40	0.50	0.60	0.70	0.80
$\text{AM}_G(\theta)$	-0.45307	-0.23279	-0.12153	-0.06135	-0.02815	-0.01046
$\text{AV}_G(\theta)$	2.58845	1.07245	0.47760	0.21262	0.08799	0.02998
$R_2(\theta)$	12.6	19.8	32.3	56.5	111.0	273.9

From Table 3 (4) it is clear that R_1 (R_2) increases as λ (θ) decreases (increases) as expected. The tolerance level D^* provides the lower limit of limit of λ and the upper limit of θ.

Now we will explain how we can determine the minimum sample size required for a given user specified protection level p^* and the tolerance level D^*, to discriminate between a Poisson and a geometric distribution functions. Suppose $p^* = 0.90$ and $D^* = 0.021$ in terms of KS distance. Here $D^* = 0.021$ means we do not want to discriminate between a Poisson and a geometric distribution functions if the KS distance between the two is less than 0.021. From Tables 1 and 2 it is observed that for $D^* = 0.021$, $\lambda \geq 0.250$ and $\theta \leq 0.80$. Therefore, if the data are coming from a Poisson distribution, then one needs $n = z_{0.9} \times R_1(0.25) =$

$1.28 \times 207.9 \approx 267$. Similarly, if the data are coming from a geometric distribution, one needs $n = z_{0.9} \times R_2(0.80) = 1.28 \times 273.9 \approx 351$. Therefore, for the given tolerance level D^* $= 0.021$, one needs at least $n = 351$ (max(267,351)) to meet the protection level $p^* = 0.9$ simultaneously for both the cases.

4.2 Testing of Hypotheses

In this subsection we present the discrimination problem as a testing of hypotheses problem, as it was considered by Dumonceaux and Antle (1973) or Kundu and Manglick (2004). Let us consider the following two testing of hypothesis problems:

Problem 1: H_0: Poisson vs. H_1 Geometric

Problem 2: H_0: Geometric vs. H_1: Poisson.

Our asymptotic results derived in the previous section can be used to test the above two hypotheses as follows:

Problem 1: Reject the null hypothesis H_0 at the $\alpha\%$ level of significance if

$$T < n \times \text{AM}_P(\widehat{\lambda}) - z_\alpha \times \sqrt{n \times \text{AV}_P(\widehat{\lambda})},$$

and accept otherwise.

Problem 2: Reject the null hypothesis H_0 at the $\alpha\%$ level of significance if

$$T > -n \times \text{AM}_G(\widehat{\theta}) + z_\alpha \times \sqrt{n \times \text{AV}_G(\widehat{\theta})},$$

and accept otherwise.

5 Simulation Results

5.1 Simulation Results

In this section we present some simulation results mainly to see how the proposed criterion behave for different sample sizes and for different parameter values. We have also reported the theoretical PCS based on the asymptotic distributions derived in the previous section. We have taken different sample sizes and different parameter values. When the null distribution is Poisson, we have taken different λ values namely λ = 0.025, 0.50, 1.00, 2.00, 3.00, and when the null distribution is geometric, then we have taken different θ values namely $\theta = 0.3$, 0.4, 0.5, 0.6 and 0.7. In both the cases we have taken n = 20, 40, 60, 80 and 100. The results are presented in Tables 5 & 6. The PCSs are computed based on 10000 replications. For comparison purposes in each case in brackets below, we have also reported the theoretical PCSs based on asymptotic distributions.

Some of the points are quite clear from the two tables. As sample size increases the PCSs increase as expected in both the cases for all parameter values. For small θ values or for large λ values, the PCSs are quite high. It is not very surprising, as θ decreases or λ increases the K-S distance between the two increases. Theoretical PCSs based on asymptotic distributions match very well with the corresponding empirical values even for moderate sample sizes. It is also observed that for large values of λ or for small values of θ the match is quite well even for small sample sizes.

6 Bayesian Model Selection

In this section we consider the Bayesian model selection criterion to choose between the two distribution functions. In the Bayesian frame work the model selection is performed using

Table 5: The probability of correct selection based on Monte Carlo Simulations when the data are coming from Poisson.

	n				
λ	20	40	60	80	100
0.25	0.73 (0.63)	0.75 (0.67)	0.77 (0.71)	0.79 (0.73)	0.80 (0.76)
0.50	0.83 (0.74)	0.86 (0.79)	0.87 (0.84)	0.90 (0.87)	0.91 (0.90)
1.00	0.88 (0.84)	0.95 (0.92)	0.96 (0.96)	0.98 (0.98)	0.99 (0.99)
2.00	0.97 (0.96)	0.99 (0.99)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
3.00	0.99 (0.99)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)

the Bayes factor approach, see for example Kass and Raftery (1995).

In case of Poisson distribution for the parameter λ we assume the conjugate gamma prior with the scale and shape parameters as c and d respectively, with the following probability density function;

$$\pi_1(\lambda; c, d) = \frac{c^d}{\Gamma(d)} \lambda^{d-1} e^{-c\lambda} \tag{14}$$

The joint posterior of λ and $data$ is given by

$$\pi_p(\lambda, data) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!} \times \frac{c^d}{\Gamma(d)} \lambda^{d-1} e^{-c\lambda} = \frac{c^d e^{-(n+c)\lambda} \lambda^{d+\sum x_i - 1}}{\Gamma(d) \prod x_i!} \tag{15}$$

In case of geometric distribution, for the parameter θ we assume the conjugate beta prior with the parameters a and b respectively, with the following probability density function;

$$\pi_2(\theta; a, b) = \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \theta^{a-1} (1-\theta)^{b-1} \tag{16}$$

The joint posterior density of θ and $data$ is given by

$$\pi_g(\theta, data) = \theta^n (1-\theta)^{\sum x_i} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{n+a-1} (1-\theta)^{b+\sum x_i - 1}. \tag{17}$$

Table 6: The probability of correct selection based on Monte Carlo Simulations when the data are coming from Geometric distribution.

θ	n				
	20	40	60	80	100
0.3	0.92 (0.89)	0.98 (0.96)	0.99 (0.99)	1.00 (0.99)	1.00 (1.00)
0.4	0.83 (0.84)	0.94 (0.92)	0.98 (0.96)	0.99 (0.98)	1.00 (0.99)
0.5	0.73 (0.78)	0.87 (0.87)	0.93 (0.91)	0.96 (0.94)	0.98 (0.96)
0.6	0.62 (0.72)	0.77 (0.80)	0.84 (0.85)	0.88 (0.88)	0.92 (0.91)
0.7	0.51 (0.66)	0.64 (0.73)	0.72 (0.77)	0.77 (0.80)	0.82 (0.83)

Therefore, the Bayes Factor is given by

$$\text{BF} = \frac{\int_0^\infty \pi_p(\lambda, data)d\lambda}{\int_0^1 \pi_g(\theta, data)d\theta} = \frac{c^d\Gamma(d+\sum x_i)\Gamma(a)\Gamma(b)\Gamma(n+\sum x_i+a+b)}{\Gamma(d)\prod x_i!\Gamma(a+b)\Gamma(n+a)\Gamma(\sum x_i+b)(n+c)^{d+\sum x_i}}. \tag{18}$$

In case of non-informative priors for both the cases, *i.e.* when c=d= 0 and a=b=1, the Bayes factor (18) takes the following form

$$\text{BF} = \frac{\Gamma(\sum x_i + n + 2)}{(\prod x_i!)(\sum x_i)\Gamma(n+1)n^{\sum x_i}}.$$

If BF > 1, then Poisson is selected otherwise, geometric is preferred.

7 Data Analysis

Data Set: It is a famous data set and it has been taken from Dauxois, Druilhet and Pommeret (2006). It consists of the frequencies of Prussian killed by horse kicks. We want to see which one between geometric or Poisson distribution fits better? In this case the MLEs of λ and θ are 0.70 and 0.59 respectively, and T = 8.33. Therefore, we prefer Poisson distribution compared to the geometric distribution. We obtain the observed and the expected

Table 7: Prussian Horse Kick Data Set

Number of Deaths	0	1	2	3	4	5
Frequencies	144	91	32	11	2	0

Table 8: Observed and expected frequencies for Bortkewisch's Data Set

Frequencies	144	91	32	11	2	0
Poisson	139.0	97.3	34.1	7.9	1.4	0.3
Geometric	165.2	67.7	27.8	11.4	4.7	3.2

frequencies based on $\widehat{\lambda}$ and $\widehat{\theta}$ for both the cases, and they are presented below; The χ^2 values corresponding to Poisson and geometric distributions are 2.5 and 16.1 respectively. Therefore, clearly Poisson provides a better goodness of fit compared to the geometric distribution. Based on non-parametric bootstrapping it is observed that the PCS is 0.99. In this case log(BF) = 1107.7, and this also indicates that Poisson is the preferred model.

Finally we want to perform the following test:

$$H_0 : \text{geometric distribution} \quad vs. \quad H_1 : \text{Poisson distribution} \tag{19}$$

Based on the asymptotic distribution of T, the p value becomes 0.002, and we therefore reject the null hypothesis.

8 Conclusions

In this paper we consider discriminating between two one parameter discrete distributions namely Poisson and geometric distributions. We propose to use the ratio of maximized likelihoods to discriminate between the two distribution functions. It is observed that the for certain ranges of the parameter values the method works very well but in the limiting case the method does not work. It may not be very surprising because in the limiting case the two distributions coincide.

We obtain the asymptotic distribution of the test statistic, and based on the asymptotic distribution we compute the probability of correct selection. It has be used for testing purposes also. Based on extensive simulation experiments it is observed that the asymptotic distribution works quite well even for moderate sample sizes. For completeness purposes we have presented the Bayes factor approach also to discriminate between the two distribution functions. One data analysis has been presented for illustrative purposes, and it is observed that different methods agree with each other in this case.

References

[1] Cox, D.R. (1961), "Tests of separate families of hypotheses", *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley, University of California Press, 105-123.

[2] Cox, D.R. (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society*, Ser. B, vol. 24, 406-424.

[3] Dauxois, J-Y, Druilhet, P. and Pommeret, D. (2006), "A Bayesian choice between Poisson, binomial and negative binomial models", *Test*, vol. 15, 423 - 432.

[4] Dumonceaux, R. and Antle, C.E. (1973), "Discriminating between the log-normal and Weibull distribution", *Technometrics*, vol. 15, 923 - 926.

[5] Dey, A.K. and Kundu, D. (2012), "Discriminating between the Weibull and Log-normal distributions for type-II censored data", *Statistics*, vol. 46, 197 - 214.

[6] Gupta, R.D. and Kundu, D. (2003), "Discriminating between Weibull and generalized exponential distributions", *Computational Statistics and Data Analysis*, vol. 43, 179 - 196.

[7] Kass, R.E. and Raftery, A. (1995), "Bayes factor", *Journal of the American Statistical Association*, vol. 90, 769 - 781.

[8] Kundu, D. and Manglick, A. (2004), "Discriminating between the Weibull and log-normal distributions", *Naval Research Logistics*, vol. 51, 893 - 905.

[9] Vu, H.T.V. and Muller, R.A. (1996), "The likelihood ratio test for Poisson versus binomial distributions", *Journal of the American Statistical Association*, vol. 91, 818-824.

[10] White, H. (1982), "Maximum likelihood estimation of mis-specified models", *Econometrica*, vol. 50, 1-25.

[11] White, H. (1982), "Regularity conditions for Cox's test of non-nested hypotheses", *Econometrics*, vol. 19, 301 - 318.