Conditional Goodness-of-Fit Tests for Discrete Distributions

Rasmus Erlemann · Bo Henry Lindqvist

Accepted: 24 December 2021
© Grace Scientific Publishing 2022, corrected publication 2023

Abstract

In this paper, we address the problem of testing goodness-of-fit for discrete distributions, where we focus on the geometric distribution. We define new likelihood-based goodness-of-fit tests using the beta-geometric distribution and the type I discrete Weibull distribution as alternative distributions. The tests are compared in a simulation study, where also the classical goodness-of-fit tests are considered for comparison. Throughout the paper we consider conditional testing given a minimal sufficient statistic under the null hypothesis, which enables the calculation of exact $p$ values. For this purpose, a new method is developed for drawing conditional samples from the geometric distribution and the negative binomial distribution. We also explain briefly how the conditional approach can be modified for the binomial, negative binomial and Poisson distributions. It is finally noted that the simulation method may be extended to other discrete distributions having the same sufficient statistic, by using the Metropolis–Hastings algorithm.

Keywords Goodness-of-fit · Conditional distribution · Geometric distribution · Monte Carlo simulations · Sufficient statistic

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $F$. Goodness-of-fit testing is concerned with how well a family of distributions $\mathcal{F}$ fits the data as a probability model. The null hypothesis is $F \in \mathcal{F}$ and the alternative hypothesis is $F \notin \mathcal{F}$. The in-depth literature on this topic includes [4]. In the literature of goodness-of-fit testing, most of the work has been focused on continuous distributions, i.e., $\mathcal{F}$ is a family of continuous distributions. For discrete distributions, the main interest has been in the Poisson distribution which plays a special role in probability theory. Goodness-of-fit
tests for the Poisson distribution go at least back to [6,15]. More recent studies of the Poisson distribution are [16,18].

Common alternatives to the Poisson distribution are the negative binomial distribution and its special case, the geometric distribution. The latter distribution is of particular interest since it is the discrete counterpart of the exponential distribution and is hence an important distribution with various applications, for example in survival analysis, reliability analysis and queuing theory. Bracquemond et al. [3] presented a comprehensive study of different goodness-of-fit test statistics for the geometric distribution and a comparison between them in a simulation study. Another paper considering tests for the geometric distribution is [12]. The present paper will mainly be concerned with goodness-of-fit testing for the geometric distribution, although several ideas considered can easily be modified to cover other discrete distributions.

Traditional goodness-of-fit tests, both in the continuous and discrete distribution cases, are the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling tests, see, for example, [4]. Various methods are used for finding critical values, typically based on standard asymptotic techniques or parametric bootstrapping. There are, however, also other methods or tricks available, often used to tailor goodness-of-fit testing for specific models.

One such “trick”, which will be the main tool in the present paper, is to condition on sufficient statistics under the null hypothesis model to be tested. Such approaches go back to the 1950s. More specifically, [6] obtained in this way exact versions of the chi-squared test and an alternative test based on the dispersion for the Poisson distribution, using the fact that the sum of the observations is a sufficient statistic in this case. As a follow-up, [15] used the same idea to derive an exact test for the Poisson case based on a likelihood ratio statistic (see Sect. 3.3.1). Conditioning on sufficient statistics has also been used recently in [2,14]. While the just cited papers have considered models with one unknown parameter under the null hypothesis, [8] did goodness-of-fit testing for the two-parameter negative binomial distribution, assuming both parameters unknown. Then she conditioned on the sum of the observations in order to eliminate the probability parameter and then using an asymptotic approach having only one unknown parameter.

The most important ingredient of a goodness-of-fit test is of course the test statistic. The three standard tests, the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling test, are already mentioned. These are examples of tests based on the empirical distribution function of the data. While the Kolmogorov–Smirnov statistic considers the maximal difference between the null model and the empirical distribution of the data, the two other tests are based on the corresponding integrated squared difference. A well-known fact is that the Anderson–Darling statistic differs from the Cramér–von Mises statistic in that it gives more weight to extreme values of the observations. There are in the literature also considered other test statistics that are known to be large when the null hypothesis model does not hold, but without connection to particular alternative models. Examples are chi-squared tests and tests based on Fisher’s index of dispersion, which is the ratio of the variance to the mean, and is well known to be 1 for the Poisson distribution. Closely related to these tests are the tests derived by [9] based on characterizations of the Poisson, binomial and negative binomial distributions by their power-series representations (see Sect. 3.3.2).
The above tests are essentially not tailored for specific alternative distributions. There might in applications be of importance, however, to have tests that are particularly powerful for given alternative distributions. One purpose of the present paper is to investigate how well the standard goodness-of-fit tests for the geometric distribution will do compared to tests tailored for specific alternatives.

As a final comment on the use of conditional testing, one might ask what is possibly lost in power by such an approach when compared to unconditional ones. We have not pursued this problem, but refer to [10] who concluded from a particular study that calculated p values from conditional tests are highly correlated with p values found by parametric bootstrapping. An apparent advantage with the conditional tests is, moreover, that these tests are exact, while the bootstrap based tests are not exact (albeit almost so).

In the second section, we introduce how conditional tests are used in goodness-of-fit testing with discrete null hypothesis. We also cover how the p values and powers are calculated with Monte Carlo methods in that setting. In the third section we present our method for drawing conditional simulations from the geometric distribution. It is based on the so-called stars and bars representation, introduced by [5]. In the same section, we also introduce some classical test statistics and define new likelihood-based tests. In the end of the section, there are two examples where the data is simulated from the beta-geometric and the discrete Weibull distribution of type I and we calculate the conditional p values for both cases. The fourth section consists of the power study. In the fifth section we consider a real-life data set and use previously mentioned methods to test if the geometric distribution fits the data. The sixth section consists of conclusions and a brief outline for possible future work in this subject. The last section is “Appendix”. There are proofs, algorithm descriptions and a method for the negative binomial. References are at the very end of this paper.

In the following we shall let \( \mathbb{N} = \{1, 2, \ldots, \} \) and \( \mathbb{N}_0 = \{0, 1, 2, \ldots\} \). A sample of random variables \( Y_1, Y_2, \ldots, Y_n \) is denoted shortly in its vector form by a bold letter, \( \mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \). Bold \( \mathbf{P} \) is reserved for the probability function to differentiate it from other functions.

### 2 Conditional Tests in Goodness-of-Fit Testing for Discrete Distributions

#### 2.1 Calculation of Conditional P Values by Monte Carlo Simulation

For illustration we focus on goodness-of-fit testing for the geometric distribution. Suppose \( \mathbf{X} = (X_1, \ldots, X_n) \) is a random sample from a population with values in \( \mathbb{N}_0 \). The null and alternative hypotheses are as follows.

- \( H_0 \) : The random sample is from a population which has the geometric distribution,
- \( H_1 \) : The random sample is not from a population which has the geometric distribution.

Let \( D = D(\mathbf{X}) \) be a test statistic such that large values of \( D \) are supposed to indicate deviations from \( H_0 \). Suppose that, under the null hypothesis, \( T = T(\mathbf{X}) \) is a
sufficient statistic. Algorithm 2 in “Appendix” calculates, by Monte Carlo simulation, the conditional $p$ value of the test from the formula

$$p_{\text{cond}} = P(D(X) \geq D(x_{\text{obs}}) \mid T(X) = t)$$

where $x_{\text{obs}}$ is the observed value of $X$ and $t$ is the observed value of $T(X)$. For the Monte Carlo simulation, one therefore needs a way of simulating from the conditional distribution of $X$ given $T(X) = t$. In the next section we show how this can be done in the case of the geometric distribution.

2.2 Calculation of Test Power of Conditional Tests

To calculate the power of a goodness-of-fit test for a given alternative distribution, for a given significance level $\alpha$, we proceed as follows. Draw a large number $M$ data sets from the alternative distribution. For the $i$th such set, calculate the conditional $p$ value, $p_{\text{cond}}^i$ by Algorithm 2, $i = 1, 2, \ldots, M$. The simulated power of the test can then be calculated by

$$\beta(\alpha) \approx \frac{\sum_{i=1}^{M} I(p_{\text{cond}}^i \leq \alpha)}{M}.$$ 

Power calculations require a large number of iterations. Let $K$ be the number of iterations used to calculate each conditional $p$ value. If $M$ is the number of data sets drawn to approximate the power, then in total we are doing $M$ times $K$ iterations. The number of data sets $M$ is chosen to be as large as possible depending on computational capabilities. We chose $M = 1000$.

It should be noted that since we are dealing with discrete distributions, for a given data set $x$ with $T(x) = t$, there are only finitely many possible data sets in the conditional distribution of $X$ given $T(X) = t$. This means that, although we fix a significance level $\alpha$ and are guaranteed a size of the test that is $\leq \alpha$, the size may be strictly less than $\alpha$. This problem is of course of less concern if $n$ is large.

3 Goodness-of-fit Testing in the Geometric Distribution

3.1 Conditional Sampling from the Geometric Distribution

Let $X_1, X_2, \ldots, X_n$ be iid random variables, such that $X_i \sim \text{Geom}(p)$ for all $i = 1, 2, \ldots, n$, i.e.,

$$P(X_i = x) = p(1 - p)^x \text{ for } x = 0, 1, 2, \ldots.$$ 

Then $T(X) = \sum_{i=1}^{n} X_i = t$ is a sufficient statistic. The conditional distribution of $X$ given $T(X) = t$ is calculated as

$$P(X = x \mid T(X) = t) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \sum_{i=1}^{n} X_i = t).$$
Fig. 1 Stars and bars representation

\[ P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, \sum_{i=1}^{n} X_i = t) \]

\[ = \begin{cases} 
\frac{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{P(\sum_{i=1}^{n} X_i = t)}, & \text{if } \sum_{i=1}^{n} X_i = t \\
0, & \text{if } \sum_{i=1}^{n} X_i \neq t
\end{cases} \]

If we restrict the support to be \( S = \{(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = t, x_1, \ldots, x_n \in \mathbb{N}_0\} \), we get

\[ P(X = x \mid T(X) = t) = \frac{p^n(1 - p)^{\sum_{i=1}^{n} x_i}}{p^n(1 - p)^{t(n-1)}} = \frac{1}{\binom{t+n-1}{n-1}}. \] (1)

As the conditional probability (1) does not depend on \( x \in S \), the distribution of \( X \mid T(X) = t \) is uniform on \( S \), i.e., on the set of all possible ways that \( n \) nonnegative integers sum to \( t \). [5] introduced a representation of such sums through the so called “stars and bars” framework. To construct one such sum, we lay down \( t \) stars and put \( n - 1 \) bars between them. The sum is constructed by counting the number of stars between the bars, letting the first element be the number of stars in front of the first bar, and letting the last one be the number to the right of the last bar. For example, for \( t = 8 \) and \( n = 4 \), Fig. 1 represents the sum \( 1 + 0 + 3 + 4 = 8 \).

The following lemma states that the representation gives rise to any possible sum and conversely. The proof is given in “Appendix”.

**Lemma 1** For \( t, n \in \mathbb{N} \), let

\[ L_1 = \left\{ (x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = t, x_1, \ldots, x_n \in \mathbb{N}_0 \right\} \]

and

\[ L_2 = \{(k_1, \ldots, k_{n-1}) : k_1 < k_2 < \ldots < k_{n-1} < t + n, k_1, \ldots, k_{n-1} \in \mathbb{N}\}. \]

Define a transformation \( \phi : L_2 \rightarrow L_1 \), such that

\[ \phi(k_1, \ldots, k_{n-1}) = (k_1 - 1, (k_2 - 2) - (k_1 - 1), \ldots, (k_{n-1} - (n - 1) - (k_{n-2} - (n - 2)), t - (k_{n-1} - (n - 1))). \]
Then $\phi$ is a bijection between the sets $L_2$ and $L_1$.

**Example 1** The $k_j$ in $L_2$ are the positions of the bars in the stars and bars representation. For example, the stars and bars representation of $n = 4$ and $t = 1 + 0 + 3 + 4 = 8$ in Fig. 1, correspond to

$$k_1 = 2, \ k_2 = 3, \ k_3 = 7.$$  

**Lemma 2** Let $t, n \in \mathbb{N}$. The set

$$\left\{ (x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = t, \ x_1, \ldots, x_n \in \mathbb{N}_0 \right\}$$

has $\binom{t+n-1}{n-1}$ elements.

The proof of Lemma 2 can be found in [5]. In fact, it also follows from (1).

Algorithm RandomKSubsets from [20] is a method for drawing the so-called bars $k_1, \ldots, k_{n-1}$ uniformly from $L_2$. We have modified it by making recursive calls into iterative ones. This allows the algorithm to be used with large values of $n$ and $t$ more efficiently and there are no issues with recursion depth limitations. It is described by Algorithm 1 in “Appendix”, where a proof of its correctness is also given (Lemma 3).

Algorithm 1 gives us a sample $k_1, \ldots, k_{n-1}$ and the last step is to transform it back into an element from $L_1$. We use the previously defined function $\phi$ for it and

$$\phi(k_1, \ldots, k_{n-1}) \sim X \mid T(X) = t.$$  

### 3.2 Standard Goodness-of-fit Test Statistics for Discrete Distributions

The following is a general setup for calculation of test statistics for the most common goodness-of-fit tests, with focus on the geometric distribution. The setup essentially follows the one of [18] who, in particular, studied the performance for the Poisson distribution.

Let $x_1, \ldots, x_n$ be the observed sample and $t = \sum_{i=1}^{n} x_i$ is the sufficient statistic. Maximum likelihood estimator for the geometric distribution is given by

$$\hat{p} = \frac{n}{t + n}.$$  

In order to avoid trivial cases, we will assume $t > 0$ and $0 < \hat{p} < 1$. Now, define for $j = 0, 1, 2, \ldots$,

- $o_j = \#\{i : x_i = j\}$ = observed number of values $j$ for the sample
- $\hat{p}_j = \hat{p}(1 - \hat{p})^j$ = probability of value $j$ in geometric distribution
- $\hat{e}_j = n\hat{p}_j$ = estimated expected number of values $j$ for the sample
From this define, for $k = 0, 1, 2, \ldots$,

$$
\hat{Z}_k = \sum_{j=0}^{k} (o_j - \hat{e}_j) \equiv O_k - \hat{E}_k
$$

$$
\hat{H}_k = \sum_{j=0}^{k} \hat{p}_j
$$

where $O_k = \sum_{j=0}^{k} o_j$ is the observed number of values $\leq k$ in the sample and $\hat{E}_k = n \hat{H}_k$ is its estimated expected value.

Further, define

$$
M_{0}^{u} = \min \{ j : o_{j'} = 0 \text{ for all } j' > j \}
$$

$$
M_{1}^{u} = \min \{ j : p_{j'} < 10^{-3}/n \text{ for all } j' > j \}
$$

$$
M^{u} = \max \{ M_{0}, M_{1} \}
$$

$$
M_{0}^{l} = \max \{ j : o_{j'} = 0 \text{ for all } j' < j \}
$$

$$
M_{1}^{l} = \max \{ j : p_{j'} < 10^{-3}/n \text{ for all } j' < j \}
$$

$$
M^{l} = \min \{ M_{0}, M_{1} \}
$$

### 3.2.1 The Cramér–von Mises Test

The Cramér–von Mises test statistic is defined by

$$
W^{2} = \frac{1}{n} \sum_{M^{l}} \hat{Z}_i^2 \hat{p}_i.
$$

### 3.2.2 The Anderson–Darling Test

The Anderson–Darling test statistic is defined by

$$
A^{2} = \frac{1}{n} \sum_{M^{l}} \frac{\hat{Z}_i^2 \hat{p}_i}{\hat{H}_i(1 - \hat{H}_i)}.
$$

### 3.2.3 The Kolmogorov–Smirnov Test

The Kolmogorov–Smirnov test statistic should ideally be defined as $\max_{k=1,2,\ldots} |Z_k|$. As shown by [3], the maximum will always occur for a $k \leq M_{0}^{u}$, so that we define

$$
KS = \max_{k=0,1,2,\ldots,M_{0}^{u}} |Z_k|.
$$
To see this, recall that $\hat{Z}_k = O_k - \hat{E}_k$. Now $O_k = n$ for $k \geq M_0^u$, while $\hat{E}_k < n$ and $\hat{E}_k$ is increasing in $k$. Hence, for $k \geq M_0$, $|Z_k| = Z_k$ and is decreasing.

### 3.3 Likelihood Based Tests

In the present subsection we study tests that are derived with the aim of having high power against given alternative distributions. The main tool is here to consider likelihood functions.

#### 3.3.1 Test Versus Heterogeneous Geometric Observations

In this subsection we follow the idea of [15], who considered the Poisson distribution where we consider the geometric distribution.

Suppose $X_1, \ldots, X_n$ are independent and geometrically distributed, but with different parameters $p_i$. The log likelihood for data $x_1, \ldots, x_n$ would then be

$$
\ell(p_1, \ldots, p_n) = \sum_{i=1}^{n} \left( \ln(p_i) + x_i \ln(1 - p_i) \right),
$$

which is maximized by $\hat{p}_i = 1/(1 + x_i)$ for $i = 1, \ldots, n$. The relevant null hypothesis is now

$$
H_0 : p_1 = p_2 = \ldots = p_n = p.
$$

The log likelihood under the null hypothesis is then $\ell(p, \ldots, p) = n \ln(p) + t \ln(1 - p)$ where $t = \sum_{i=1}^{n} x_i$, which is maximized by $\hat{p} = n/(n + t)$.

The likelihood ratio statistic can therefore be written

$$
\ell(\hat{p}_1, \ldots, \hat{p}_n) - \ell(\hat{p}, \ldots, \hat{p})
= \sum_{i=1}^{n} \left[ \ln \left( \frac{1}{1 + x_i} \right) + x_i \ln \left( \frac{x_i}{1 + x_i} \right) \right] - n \ln \left( \frac{n}{n + t} \right) - t \ln \left( \frac{t}{n + t} \right)
= \sum_{i=1}^{n} \left[ x_i \ln(x_i) - (x_i + 1) \ln(x_i + 1) \right] - n \ln \left( \frac{n}{n + t} \right) - t \ln \left( \frac{t}{n + t} \right)
$$

Since we consider conditional tests given $\sum_{i=1}^{n} X_i = t$, we may exclude the last terms above, which after rewriting the first sum gives the test statistic

$$
CR = \sum_{i=1}^{n} [x_i \ln(x_i) - (x_i + 1) \ln(x_i + 1)] = \sum_{j=M_0^u}^{n} o_j \left( j \ln j - (j + 1) \ln(j + 1) \right)
$$

where we use $0 \ln 0 = 0$. 

\copyright  Springer
3.3.2 The Beta-Geometric Distribution

In the previous subsection we considered the alternative hypothesis that the observations were geometrically distributed, but with possibly different parameters $p_i$. Suppose now that these $p_i$ are drawn independently from the beta distribution.

Thus, for a single observation $X$ we assume that it is geometrically distributed with parameter $p$, where $p$ is generated from the beta-distribution with parameters $\alpha > 0$ and $\beta > 0$. Let $B(\alpha, \beta)$ be the Beta function, defined by

$$B(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$  

Then it is seen that the unconditional distribution of $X$ is what has been named the beta-geometric distribution,

$$P(X = x) = \int_0^1 p(1-p)^x \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}dp = \frac{B(\alpha + 1, \beta + x)}{B(\alpha, \beta)} \quad (2)$$

for $x = 0, 1, 2, \ldots$ As suggested by [13], a useful reparametrization is via

$$\pi = \frac{\alpha}{\alpha + \beta}, \quad \theta = \frac{1}{\alpha + \beta}.$$  \hspace{1cm} (3)

With this parametrization it is seen that $\theta = 0$ corresponds to the geometric distribution with $p = \pi$. Tests for the geometric distribution null hypothesis can then be derived by testing

$$H_0 : \theta = 0 \text{ vs. } \theta > 0.$$  

Using the reparametrization (3), we find from (2), noting that $\alpha = \pi/\theta, \beta = (1-\pi)/\theta$ and using properties of the gamma function,

$$P(X = x) = \frac{\alpha \prod_{j=1}^x (\beta + x - j)}{\prod_{j=0}^x (\alpha + \beta + x - j)} = \frac{\pi \prod_{j=0}^{x-1} (1-\pi + j\theta)}{\prod_{j=0}^x (1+j\theta)} \quad (4)$$

for $x = 0, 1, \ldots$ (Note that the formula also holds for $x = 0$, giving $P(X = 0) = \pi$, since an empty product by convention equals 1.)

The log likelihood for the data $x_1, \ldots, x_n$ with values in $\mathbb{N}_0$ is hence, see also [13],

$$\ell(x; \pi, \theta) = n \ln \pi + \sum_{i=1}^n \sum_{j=0}^{x_i-1} \ln(1-\pi + j\theta) - \sum_{i=1}^n \sum_{j=0}^{x_i} \ln(1+j\theta) \quad (5)$$
Differentiating with respect to \( \theta \) and \( \pi \), give, respectively,

\[
\frac{\partial \ell}{\partial \theta} = \frac{n \sum_{i=1}^{x_i} j}{1 - \pi + j \theta} - \frac{n \sum_{j=0}^{x_i} j}{1 + j \theta},
\]

\[
\frac{\partial \ell}{\partial \pi} = \frac{n}{\pi} - \frac{n \sum_{i=1}^{x_i} \frac{1}{1 - \pi + j \theta}}.
\]

Letting \( \theta = 0 \) in the expression for \( \frac{\partial \ell}{\partial \theta} \) gives the score statistic for testing \( H_0 : \theta = 0 \), namely

\[
S = \sum_{i=1}^{n} \sum_{j=0}^{x_i-1} \frac{j}{1 - \pi} - \sum_{i=1}^{n} \sum_{j=0}^{x_i} j = \pi \frac{\sum_{i=1}^{n} x_i^2 - (2-\pi) \sum_{i=1}^{n} x_i}{2(1-\pi)}. \tag{6}
\]

The score test in general rejects \( H_0 : \theta = 0 \) for large values of \( |S| \). Indeed, it can be shown from the rightmost expression in (6) that, under \( H_0 \) where the \( X_i \) are geometrically distributed with probability \( p = \pi \), we have \( E(S) = 0 \). Paul [13] replaced \( \pi \) by the maximum likelihood estimate \( \hat{\pi} \) under \( H_0 \), and divided the expression in (6) by an estimate of the standard deviation of \( S \) under \( H_0 \), which is \( \sqrt{n/\hat{\pi}} \). The resulting statistic then has an asymptotically standard normal distribution under \( H_0 \).

Let now \( m_1 = (1/n) \sum_{i=1}^{n} x_i \) and \( m_2 = (1/n) \sum_{i=1}^{n} x_i^2 \) be the first and second empirical moments, respectively, from the data \( x \). Replacing \( \pi \) by \( \hat{\pi} = n/(t+n) = 1/(1+m_1) \) we can write the right-hand side of (6) as

\[
\frac{m_2}{n \frac{1}{1+m_1} - \left(2 - \frac{1}{1+m_1}\right) \frac{m_1}{2 \left(1 - \frac{1}{1+m_1}\right)} = \frac{n}{2} \frac{m_2 - m_1 - 2m_1^2}{m_1}. \tag{7}
\]

Since we consider conditional testing given \( t \), or equivalently given \( \hat{\pi} \) or given \( m_1 \), we will use the numerator of (6) as our statistic, i.e.,

\[
SB = m_2 - m_1 - 2m_1^2. \tag{7}
\]

Actually, we might also have deleted the terms involving \( m_1 \). We keep them, however, due to the fact that \( SB \) is of some importance, as explained below.

Not now that, since \( \theta \geq 0 \) is a model restriction, the maximum of (5) may occur at a boundary where \( \theta = 0 \). But for \( \theta = 0 \), (5) is simply the log likelihood of the geometric distribution and is hence maximized by \( \pi = \hat{\pi} \). Thus if \( SB > 0 \), then we know that the maximum of (5) is not at a boundary point with \( \theta = 0 \), and must hence be at a point \( (\pi, \theta) \) with \( \theta > 0 \). This point may hence presumably be found by using the partial derivatives derived below. If, instead, \( SB < 0 \), then the maximum likelihood estimate is likely to be at the point \( (\hat{\pi}, 0) \).

It follows from the above that if \( SB < 0 \), then the numerical value is uninteresting, because it corresponds to parameter values outside of the parameter set and intuitively...
to parameters for which we would reject the null hypothesis. We therefore suggest to replace $SB$ by $SB_0 = \max(0, SB)$ and call this the score test statistic for the null hypothesis $\theta = 0$.

Paul [13] considered the score test and in addition the likelihood ratio test based on the log likelihood (5) and standard asymptotics (taking into account the fact that the null hypothesis is on the boundary of the parameter space). He further noted that the likelihood ratio test, as well as the score test, is rather liberal (non-conservative) in regard to size. He therefore found that a bootstrap test might be preferable.

Singh et al. [17] considered both maximum likelihood estimation and moment estimation of $\alpha$ and $\beta$. The moment estimators are obtained as follows. First, define

$$
\mu_1 \equiv E(X) = \frac{\beta}{\alpha - 1} \quad \text{for } \alpha > 1
$$

$$
\mu_2 \equiv E(X^2) = \frac{\beta(\alpha + 2\beta)}{(\alpha - 1)(\alpha - 2)} \quad \text{for } \alpha > 2
$$

Solving for $\alpha$ and $\beta$ we get

$$
\alpha = \frac{2(\mu_2 - \mu_1^2)}{\mu_2 - \mu_1 - 2\mu_1^2}
$$

$$
\beta = \mu_1(\alpha - 1)
$$

The moment estimators $\tilde{\alpha}$ and $\tilde{\beta}$ for $\alpha$ and $\beta$ are obtained by substituting the empirical moments $m_1$ and $m_2$ for $\mu_1$ and $\mu_2$. This leads to an estimator for the parameter $\theta$ which can be expressed by

$$
\tilde{\theta} = (\tilde{\alpha} + \tilde{\beta})^{-1} = \frac{m_2 - m_1 - 2m_1^2}{2m_2 - m_1^2 + m_1m_2}
$$

It is noticeable that the numerator of $\tilde{\theta}$ equals $SB$ (see (7)). Thus $\tilde{\theta}$ and $SB$ have the same sign (since both denominators above is always positive). As already noted, this sign is of importance for maximum likelihood estimation based on (5). Note, on the other hand, that in a conditional test using $\hat{\theta}$ we cannot ignore the denominator of $\tilde{\theta}$, since it contains $m_2$.

As a final note in this subsection, the test statistic $SB_0$ appears to be essentially identical to the one for the geometric distribution which is derived in [9]. These authors derive test statistics from characterizations of distributions given by certain powerseries distribution laws. These include Poisson, binomial and the negative binomial distribution. Their general test statistic is

$$
\hat{c} = \frac{\frac{1}{n} \sum_{i=1}^n X_i (X_i - 1)}{\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2} = \frac{m_2 - m_1}{m_1^2}
$$

(8)
and the test for the geometric distribution rejects the null hypothesis if a normalized version of
\[ |\hat{c} - 2| = \frac{|m_2 - m_1 - 2m_1^2|}{m_1^2} \]
is large, where the normalization leads to an asymptotically standard normal distribution under \( H_0 \). Since the normalization is a function of \( m_1 \) only, and the denominator of (8) can be deleted, we in fact end up with the test statistic \( |SB| \).

Example

Suppose we have observed the data in Table 1, which are simulated from a beta-geometric distribution with \( n = 100, \pi = 0.4, \theta = 0.125 \).

Using the described test statistics for testing the null hypothesis of a geometric distribution, we obtained the conditional \( p \) values given in Table 2. We note that also the three standard tests are able to detect the deviation from the geometric distribution here, while \( p \) values are remarkably lower for the tests derived above that are tailored for detecting deviations in the direction of a beta-geometric distribution. In fact, the same low \( p \) values are obtained for tests versus the discrete Weibull distribution that will be studied below.

Maximum likelihood estimates for \( \pi \) and \( \theta \) in the beta-geometric distribution can be calculated using the R-package VGAM, giving \( \hat{\pi} = 0.4274, \hat{\theta} = 0.1166 \). These estimates are used to calculate the estimated expected counts in Table 1. It is remarkable that these are much closer to the observed values than the ones estimated from the geometric distribution.

3.3.3 The Discrete Weibull Distribution of Type I

Let for \( x = 0, 1, 2, \ldots, \)
\[
P(X = x) = q^x\beta - q^{(x+1)\beta} \tag{9}
\]
where \( 0 < q < 1 \) and \( \beta > 0 \). This is the probability distribution of the type I Weibull distribution, which was introduced by [11]. We denote it by \( \mathcal{W}(q, \beta) \). The geometric distribution with parameter \( p \) is now a special case obtained when \( q = 1 - p \) and \( \beta = 1 \).

The R-package DiscreteWeibull contains routines for this distribution, including simulation of data and estimation of parameters.

The discrete hazard rate of a random variable with values in the (nonnegative) integers can be defined by [1] \( \lambda(x) = P(X = x|X \geq x) \). From (9) we get
\[
\lambda(x) = \frac{P(X = x)}{P(X \geq x)} = \frac{q^x\beta - q^{(x+1)\beta}}{q^x\beta} = 1 - q^{(x+1)\beta} - x^\beta
\]
Table 1  Data simulated from the beta-geometric distribution with $n = 100, \pi = 0.4, \theta = 0.125$.

| $j$ | $o_j$ | $\hat{e}_j^g$ | $\hat{e}_j^b$ |
|-----|-------|---------------|---------------|
| 0   | 42    | 35.4          | 42.7          |
| 1   | 24    | 22.8          | 21.9          |
| 2   | 11    | 14.8          | 12.2          |
| 3   | 8     | 9.5           | 7.3           |
| 4   | 4     | 6.2           | 4.6           |
| 5   | 4     | 4.0           | 3.0           |
| 6   | 0     | 2.6           | 2.1           |
| 7   | 1     | 1.7           | 1.4           |
| 8   | 0     | 1.1           | 1.0           |
| 9   | 2     | 0.7           | 0.8           |
| 10  | 2     | 0.4           | 0.6           |
| 11  | 0     | 0.3           | 0.4           |
| 12  | 0     | 0.2           | 0.3           |
| 13  | 0     | 0.1           | 0.3           |
| 14  | 0     | 0.1           | 0.2           |
| 15  | 1     | 0.0           | 0.2           |
| 16  | 1     | 0.0           | 0.1           |

The column $o_j$ gives the number of observations $x_i$ that resulted in $x_i = j$. The two last columns give the estimated expected frequencies under a geometric distribution and beta-geometric distribution, respectively.

Table 2  Conditional $p$ values obtained by simulating 10000 data sets from the conditional distribution.

| Statistic | $W^2$ | $A^2$ | $KS$ | $CR$ | $SB$ | $SB_0$ | $\hat{\theta}$ | $|SW|$ | $SW_L$ | $SW_U$ |
|-----------|-------|-------|------|------|------|--------|------------------|-------|--------|--------|
| $p_{\text{cond}}$ | 0.034 | 0.028 | 0.059 | 0.009 | 0.004 | 0.004 | 0.004 | 0.005 | 0.004 | 0.996 |

Here, $W^2$, $A^2$, $KS$ denote the standard tests displayed in Sect. 3.2, respectively, the Cramér–von Mises, the Anderson–Darling and the Kolmogorov–Smirnov tests. Further, $CR$ denotes the likelihood ratio test versus heterogeneous geometric distributions (Sect. 3.3.1); $SB$, $SB_0$, $\hat{\theta}$ are variations of score tests versus the beta-geometric distribution (Sect. 3.3.2); and finally the $|SW|$, $SW_L$ and $SW_U$ are variations of score tests versus the Weibull distribution of type I (Sect. 3.3.3).

which is seen to be increasing in $x$ if $\beta > 1$ and decreasing in $x$ if $\beta < 1$, and constant equal to $p$ when $\beta = 1$, which corresponds to the geometric distribution.

Suppose now we have data $x_1, \ldots, x_n$ with values in $\mathbb{N}_0$. Testing the null hypothesis that the data come from the geometric distribution is now equivalent to testing $H_0 : \beta = 1$ vs. $H_1 : \beta \neq 1$, or possibly the one-sided versions of the alternative. The testing can be done by a likelihood ratio test. It follows from (9) that the log-likelihood for the sample $x_1, \ldots, x_n$ from the type I Weibull distribution is given by

$$\ell(q, \beta) = \sum_{i=1}^{n} \ln \left( q^{x_i^\beta} - q^{(x_i+1)^\beta} \right).$$
The likelihood ratio test statistic can be computed by calculating the maximum likelihood estimates of \( q \) and \( \beta \), and of \( p \), which is the parameter under the null hypothesis model. Details are given by [19], while computations can be done using the R-package DiscreteWeibull.

A score test can be derived in a way similar to what we did in Sect. 3.3.2 for the beta-geometric distribution. First, the partial derivative with respect to \( \beta \) of the log-likelihood function \( \ell \), is given by

\[
\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} q_i^\beta x_i^\beta \ln(q) \ln(x_i) - q^{(x_i+1)^\beta} (x_i + 1)^\beta \ln(q) \ln(x_i + 1)
\]

The score statistic of \( H_0 \) can then be found by letting \( \beta = 1 \), which leads to

\[
\frac{\partial \ell}{\partial \beta} \bigg|_{\beta=1} = \ln(q) \sum_{i=1}^{n} (x_i \ln(x_i) - q (x_i + 1) \ln(x_i + 1)).
\]  

(10)

It can now be checked that if the \( x_i \) are from the geometric distribution with parameter \( p \), the expected value of (10) is 0 (noting that \( q = 1 - p \)). The standard approach is now to estimate \( q \) by \( 1 - \hat{p} \) (from the geometric distribution) and divide (10) by the estimated standard deviation, in order to obtain a test statistic which is standard normally distributed under the null hypothesis. We shall, however, consider conditional testing, conditioning on \( \sum_{i=1}^{n} X_i \) or, equivalently, on \( \hat{p} \), and we may hence use the test statistic

\[
SW = \sum_{i=1}^{n} [(1 - \hat{p}) (x_i + 1) \ln(x_i + 1) - x_i \ln(x_i)]
\]

where \( \hat{p} = n/(n + \sum_i x_i) \). If \( x_i = 0 \), we shall let \( x_i \ln(x_i) = 0 \). Note that we have changed the order of the terms inside the sum as compared to (10). This is because \( \ln(q) < 0 \) and will lead to a statistic \( SW \) with the same sign as \( \frac{\partial \ell}{\partial \beta} \bigg|_{\beta=1} \). Then for the two-sided alternative, \( \beta \neq 1 \), we should use the statistic \( |SW| \) as the test statistic. A more powerful test can then be defined for the two one-sided alternatives, by using \( SW_U = SW \) if the alternative is \( \beta > 1 \) and \( SW_L = -SW \) for the alternative \( \beta < 1 \), and reject in both cases for high values of the test statistic.

The resemblance between the statistics \( SW \) and \( CR \) is striking. In fact, \( CR \) is obtained from \( SW \) by letting \( \hat{p} = 0 \), and switching the sign. Simulations and \( p \) value calculations in the following will indicate the possible difference between their merits.

**Example**

Suppose we have observed the data in Table 3, which are simulated from a type I discrete Weibull distribution with \( n = 50, q = 0.8, \beta = 1.4 \) using the R-package DiscreteWeibull.
Table 3  Data simulated from a type I Weibull distribution with $n = 50$, $q = 0.8$, $\beta = 1.4$

| $j$ | $o_j$ | $\hat{e}_g$ | $\hat{e}_b$ |
|-----|-------|-------------|-------------|
| 0   | 13    | 18.0        | 13.8        |
| 1   | 14    | 11.5        | 13.2        |
| 2   | 10    | 7.4         | 9.3         |
| 3   | 8     | 4.7         | 5.9         |
| 4   | 1     | 3.0         | 3.5         |
| 5   | 1     | 1.9         | 2.0         |
| 6   | 0     | 1.2         | 1.1         |
| 7   | 2     | 0.8         | 0.6         |
| 8   | 1     | 0.5         | 0.3         |

The column $o_j$ gives the number of observations $x_j$ that resulted in $x_j = j$. The two last columns give the estimated expected frequencies under a geometric distribution and type I Weibull distribution, respectively.

Table 4  Conditional $p$ values obtained by simulating 10000 data sets from the conditional distribution

| Statistic | $W^2$ | $A^2$ | $KS$ | $CR$ | $SB$ | $SB_0$ | $\hat{\theta}$ | $|SW|$ | $SWL$ | $SWU$ |
|-----------|-------|-------|------|------|------|--------|----------------|-------|-------|-------|
| $p_{\text{cond}}$ | 0.072 | 0.078 | 0.124 | 0.962 | 0.890 | 1.0 | 0.890 | 0.083 | 0.956 | 0.044 |

For abbreviations, see caption of Table 2 and definitions of test statistics in Sects. 3.2 and 3.3. Two-sided (one-sided) testing using $|SW|$ ($SWU$) means that the alternative hypothesis is $\beta \neq 1$ ($\beta > 1$). The statistics $SB$ and $SB_0$ can only be used to detect an increased variance compared to the geometric distribution. Here the test statistic $SB$ is negative, which makes these tests meaningless.

Using all the tests considered so far in the paper, we obtained the conditional $p$ values given in Table 4. We note that also the three standard tests are able to detect the departure from the discrete Weibull distribution here, while $p$ values are remarkably lower for the tests $|SW|$ and $SWL$, which are tailored for detecting deviations in the direction of a discrete Weibull distribution. It should be noted, however, that the test $CR$ as well as the tests versus the beta-geometric distribution are useless for these data. The reason for this last fact is that the beta-geometric distribution always increases the variance of the data as compared to the geometric distribution, while the discrete Weibull with $\beta > 1$ decreases the variance (a property well known for the continuous Weibull distribution). Thus a beta-geometric distribution would have difficulties fitting these data.

Maximum likelihood estimates for $\pi$ and $\theta$ are calculated as $\hat{\pi} = 0.7239$, $\hat{\beta} = 1.267$ using the R-package DiscreteWeibull. These estimates are used to calculate the estimated expected counts in Table 3. Again, these are much closer to the observed values than the ones estimated from the geometric distribution.

### 4 Computer Simulations

#### 4.1 Power Study

We did a power study of the previously defined test statistics, using various sample sizes, alternative distributions, fixing significance level to $\alpha = 0.1$. Note that we
left out $\hat{\theta}$ and $SB$ from the power study, because they had identical powers for all alternatives (see discussion in Sect. 3.3.2 on why $SB_0$ should be preferred over $SB$). In some cases we disregarded sample sizes $n = 5$ or $n = 100$ if the powers were too close to the significance level or 1.

As usual for a wide choice of alternative distributions, there is no best test against all alternatives. From standard tests, $W^2$ has slightly higher powers with small sample sizes. For larger sample sizes, $A^2$ and $W^2$ are almost identical. Maximal type test $KS$ has slightly lower powers than the other standard tests. Tests $CR$, $SB_0$, $SWL$ and $SWU$ are sensitive to the alternative distribution the data comes from. This makes them situational and they lack the versatility of the standard tests. For example, $CR$, $SB_0$ and $SWL$ outperform the standard tests when the data comes from BG or $\mathcal{W}$ distributions. $SWU$ outperforms other tests for Pois, Bin and NB distributions. $|SW|$ is more versatile and has almost identical powers to the standard quadratic tests. Likelihood ratio tests need a versatile comparative alternative distribution to perform well. Type I Weibull distribution fits this role, as we can see from $|SW|$ powers. Test $SWL$ is for the case where $\beta < 1$ and $SWU$ for $\beta > 1$. Under those conditions, they outperform $|SW|$.

Table 5 Conditional power calculations with significance level $\alpha = 0.1$

| Alternative | Sample size | $W^2$ | $A^2$ | $KS$ | $CR$ | $SB_0$ | $|SW|$ | $SWL$ | $SWU$ |
|-------------|-------------|-------|-------|------|------|--------|--------|-------|-------|
| Pois(0.5)   | $n = 25$    | 0.231 | 0.238 | 0.208| 0.002| 0.001  | 0.225  | 0.002 | 0.329 |
|             | $n = 100$   | 0.736 | 0.734 | 0.705| 0.000| 0.000  | 0.763  | 0.000 | 0.851 |
| Pois(1)     | $n = 5$     | 0.119 | 0.119 | 0.110| 0.008| 0.003  | 0.122  | 0.008 | 0.134 |
|             | $n = 25$    | 0.613 | 0.605 | 0.543| 0.000| 0.001  | 0.618  | 0.000 | 0.730 |
|             | $n = 100$   | 0.996 | 0.996 | 0.992| 0.000| 0.000  | 0.998  | 0.000 | 0.999 |
| Pois(2)     | $n = 5$     | 0.332 | 0.321 | 0.163| 0.003| 0.003  | 0.339  | 0.003 | 0.395 |
|             | $n = 25$    | 0.963 | 0.965 | 0.914| 0.000| 0.000  | 0.966  | 0.000 | 0.985 |
| Bin(5, 0.3) | $n = 5$     | 0.418 | 0.410 | 0.294| 0.000| 0.000  | 0.403  | 0.000 | 0.432 |
|             | $n = 25$    | 0.986 | 0.985 | 0.972| 0.000| 0.000  | 0.990  | 0.000 | 0.996 |
| NB(5, 0.5)  | $n = 5$     | 0.531 | 0.466 | 0.434| 0.000| 0.001  | 0.538  | 0.000 | 0.672 |
|             | $n = 25$    | 0.997 | 0.998 | 0.986| 0.000| 0.000  | 1.000  | 0.000 | 1.000 |
| NB(3, 0.7)  | $n = 25$    | 0.395 | 0.393 | 0.339| 0.001| 0.002  | 0.407  | 0.001 | 0.526 |
|             | $n = 100$   | 0.875 | 0.875 | 0.834| 0.000| 0.000  | 0.906  | 0.000 | 0.945 |
| BG(2, 5)    | $n = 5$     | 0.161 | 0.170 | 0.137| 0.267| 0.269  | 0.158  | 0.274 | 0.025 |
|             | $n = 100$   | 0.565 | 0.565 | 0.514| 0.676| 0.705  | 0.608  | 0.706 | 0.007 |
| BG(2, 2)    | $n = 5$     | 0.122 | 0.122 | 0.108| 0.205| 0.201  | 0.125  | 0.207 | 0.018 |
|             | $n = 25$    | 0.558 | 0.570 | 0.504| 0.705| 0.688  | 0.611  | 0.717 | 0.001 |
| $\mathcal{W}(0.7, 0.8)$ | $n = 25$   | 0.320 | 0.338 | 0.282| 0.503| 0.432  | 0.353  | 0.492 | 0.006 |
|             | $n = 100$   | 0.749 | 0.759 | 0.679| 0.874| 0.792  | 0.792  | 0.882 | 0.000 |
| $\mathcal{W}(0.5, 0.5)$ | $n = 5$    | 0.273 | 0.300 | 0.241| 0.443| 0.406  | 0.299  | 0.446 | 0.005 |
|             | $n = 25$    | 0.935 | 0.939 | 0.905| 0.972| 0.947  | 0.943  | 0.964 | 0.000 |

For abbreviations, see caption of Table 2.
Table 6  Real-life data. Observed and estimated expected frequencies for three different models

| Value | 0 | 1 | 2 | 3 | 4 | ≥ 5 |
|-------|---|---|---|---|---|-----|
| Observed frequency | 6 | 4 | 3 | 3 | 2 | 10 |
| Expected frequency, geometric | 3.9 | 3.3 | 2.9 | 2.5 | 2.1 | 13.3 |
| Expected frequency, beta-geometric | 5.0 | 3.9 | 3.1 | 2.5 | 2.0 | 11.5 |
| Expected frequency, discrete Weibull | 6.0 | 3.6 | 2.7 | 2.2 | 1.8 | 11.7 |

Fig. 2  Observed data plotted against fitted distributions. Black line with dots is the observed data, blue line is the geometric, red line the beta-geometric and green line the discrete Weibull distribution

Table 7  Conditional $p$ values obtained by simulating 10000 data sets from the conditional distribution

| Statistic | $W^2$ | $A^2$ | $KS$ | $CR$ | $SB$ | $SB_0$ | $\hat{\theta}$ | $|SW|$ | $SWL$ | $SWU$ |
|-----------|-------|-------|------|------|------|--------|-----------|-------|-------|-------|
| $p_{\text{cond}}$ | 0.107 | 0.117 | 0.315 | 0.042 | 0.134 | 0.134 | 0.110 | 0.047 | 0.953 |

For abbreviations, see caption of Table 2

5 Real-Life Data

In this section we use real-life data from [3]. The data consist of numbers of inspections between discovery of defects in an industrial process. Conditional samples are used to calculate the distribution of goodness-of-fit test statistics following the recipe from Sect. 2. Conditional $p$ values are reported to decide if we should reject or not reject the null hypothesis that the data comes from the geometric distribution.

In order to have the data on the form considered in this paper, we have subtracted 1 from each observation.

The data are given in Table 6. Note that the data $x_i \geq 5$ are lumped together in the table for illustrative purposes. The observed values for these 10 observations are 6, 8, 10, 12, 13, 16, 17, 25, 28, and are used in the simulations and calculations (Fig. 2).

Conditional $p$ values for the various tests are given in Table 7, calculated with 10000 Monte Carlo samples from the conditional distribution.

The standard tests as well as the tests versus beta-geometric distribution still indicate the possibility of a geometric distribution, having $p$ values $> 0.10$, while the hypothesis
of geometric distribution is in fact rejected at 5% significance level by the CR test and the one-sided test versus the type I discrete Weibull distribution with $\beta < 1$. This possibility of the Weibull distribution is also indicated by the fitted expected frequencies as shown in Table 6.

### 6 Conclusion and Future Work

In this paper we studied goodness-of-fit tests for discrete distributions obtained by conditioning on the sufficient statistic under the null hypothesis. We developed, in particular, a method to draw conditional samples from the geometric distribution. These samples are used for calculation of $p$ values for various goodness-of-fit tests. In addition to considering standard goodness-of-fit test, we derived new likelihood-based test statistics for testing of the geometric distribution versus heterogeneity, as well as versus discrete Weibull distributions with both increasing and decreasing hazard.

A power study was conducted to check how the tests perform against data from different alternative distributions. Our simulations suggested that the two-sided test versus the type I discrete Weibull distributions was able to detect bad fit for data from various alternative distributions. The power results for this test, $|SW|$, were in fact generally similar to the ones obtained for the standard quadratic goodness-of-fit tests.

Real-life data from [3] were considered and it was tested whether the geometric distribution is a good fit. The calculated $p$ values suggested that the geometric distribution might not be a good fit according to some of the tests.

For further work, a general method could be described for the case where $T(X) = \sum_{i=1}^{n} X_i$ is sufficient for the family of distributions under the null hypothesis. This is the case for the power series distributions where the probability distribution is of the form, see [9] or [7],

$$P(X = x) = \frac{a(x)\theta^x}{\eta(\theta)} \text{ for } x = 0, 1, 2, \ldots \ (11)$$

where $a(x) \geq 0, \theta > 0, \eta(\theta) = \sum_{y=0}^{\infty} a(y)\theta^y$. The Poisson, binomial, negative binomial and geometric distributions are of this kind. It can be shown from (11) [7] that for samples $X = (X_1, \ldots, X_n)$ from this distribution, we have

$$P(X = x|T(X) = t) \propto \prod_{i=1}^{n} a(x_i) \text{ when } \sum_{i=1}^{n} x_i = t.$$ 

Conditional samples with a given sum $t$ can hence be obtained by the Metropolis algorithm using samples from the conditional geometric distribution as proposals.

### Declarations

**Conflict of interest** On behalf of all the authors, the corresponding author states that there is no conflict of interest.
Appendix

Proof of Lemma 1 The function is defined correctly, because for any \((k_1, k_2, \ldots, k_{n-1}) \in L_2\)
\[
k_1 - 1 + (k_2 - 2) - (k_1 - 1) + \ldots + (k_{n-1} - (n - 1)) = t
\]
and \(k_1 - 1, (k_2 - 2) - (k_1 - 1), \ldots, t - (k_{n-1} - (n - 1)) \in \mathbb{N}_0\), as \(0 < k_1 < k_2 < \ldots < k_{n-1} < t + n\).
Let us assume that \((k_1, k_2, \ldots, k_{n-1}), (v_1, v_2, \ldots, v_{n-1}) \in L_2\), such that
\[
\phi(k_1, k_2, \ldots, k_{n-1}) = \phi(v_1, v_2, \ldots, v_{n-1}).
\]
This implies that
\[
k_1 + 1 = v_1 + 1 \Rightarrow k_1 = v_1,
(2 - k_1 - 1) = (v_2 - 2) - (v_1 - 1) \Rightarrow k_2 = v_2,
\ldots
(k_{n-1} - (n - 1)) - (k_{n-2} - (n - 2)) = (v_{n-1} - (n - 1)) - (v_{n-2} - (n - 2)) \Rightarrow k_{n-1} = v_{n-1},
\]
and we can conclude that \(\phi\) is injective.
Let us fix an element \((x_1, x_2, \ldots, x_n) \in L_1\), then
\[
\phi(x_1 + 1, x_1 + x_2 + 2, \ldots, \sum_{i=1}^{n-1} x_i + (n - 1)) = (x_1, x_2, \ldots, x_n)
\]
and
\[
0 < x_1 + 1 < x_1 + x_2 + 2 < \ldots < \sum_{i=1}^{n-1} x_i + n - 1 < t + n,
\]
which implies that \(\phi\) is surjective. Injectivity and surjectivity imply that \(\phi\) is bijective.
\[\square\]

Lemma 3 Let \(n, t \in \mathbb{N}\) and \((k_1, \ldots, k_{n-1}) \in L_2\) be an arbitrary sample drawn according to Algorithm 1. Then it is drawn uniformly, i.e.,
\[
P(k_1, \ldots, k_{n-1}) = \frac{1}{\binom{n+t-1}{n-1}}
\]
for each \((k_1, \ldots, k_{n-1}) \in L_2\). The probability follows from Lemma 2.
The task is to distribute $t$ stars and $n - 1$ bars randomly on the positions $1, 2, \ldots, t + n - 1$. We start from the right. Then the probability of placing a bar in position $t + n - 1$ is $\frac{n - 1}{t + n - 1}$. Then we proceed conditionally to the left and multiply probabilities of placing bars or stars in order to calculate probabilities of a given configuration.

More precisely, let $(k_1, k_2, \ldots, k_{n-1})$ be an arbitrary sample drawn according to Algorithm 1. We want to calculate

$$P(k_1, \ldots, k_{n-1}). \quad (12)$$

We know that the algorithm accepted $n - 1$ integers (i.e., placements $k_j$ of the bars) in the process. Also, let $V$ denote the number of integers that were not accepted. In total the algorithm ran $V + n - 1$ iterations. The probability $P$ is a product of $V + n - 1$ probabilities. Let us look at the denominator and nominator separately. In the denominator we have

$$(t + n - 1) \cdot (t + n - 2) \cdot \ldots \cdot (t - V - 1). \quad (13)$$

In the numerator we have

$$(n - 1) \cdot (n - 2) \cdot \ldots \cdot 2 \cdot 1 = (n - 1)! \quad (14)$$

from the accepted integers. In the numerator there is also

$$t \cdot (t - 1) \cdot \ldots \cdot (t - V - 1) \quad (15)$$

from the integers that were not accepted. Combining (13), (14) and (15) we get

$$P(k_1, k_2, \ldots, k_{n-1}) = \frac{(n - 1)!t(t - 1) \ldots (t - V - 1)}{(t + n - 1) \ldots (t - V - 1)}$$

$$= \frac{(n - 1)!t(t - 1) \ldots (t - V - 1)(t - V - 2) \ldots 2 \cdot 1}{(t + n - 1) \ldots (t - V - 1)(t - V - 2) \ldots 2 \cdot 1}$$

$$= \frac{(n - 1)! \cdot t!}{(t + n - 1)!}$$

$$= \frac{1}{\binom{t + n - 1}{n - 1}}.$$

$\blacksquare$
Algorithm 1: This algorithm picks \( n - 1 \) bar positions \( k_1, k_2, \ldots, k_{n-1} \) randomly among \( t + n - 1 \) possible positions. After applying the transformation \( \phi \) to the selected bar positions, we get a conditional geometric sample.

\[
\text{Data: } t \text{ and } n \\
\text{Result: } k_1, \ldots, k_{n-1} \\
\text{initialization;} \\
N = 0; \quad \text{// number of accepted integers} \\
V = 0; \quad \text{// number of not accepted integers} \\
I = t + n - 1; \quad \text{// integer to consider} \\
\text{while } N < n - 1 \text{ do} \\
\quad \text{Draw } p \sim U[0, 1]; \\
\quad \text{if } p < (n - 1 - N)/(t + n - 1 - N - V) \text{ then} \\
\qquad k_{n-1-N} = I; \quad \text{// integer } I \text{ was accepted} \\
\qquad N = N + 1; \\
\qquad I = I - 1; \\
\qquad \text{Continue} \\
\quad \text{end} \\
\quad \text{if } p \geq (n - 1 - N)/(t + n - 1 - N - V) \text{ then} \\
\qquad V = V + 1; \quad \text{// integer } I \text{ was not accepted} \\
\qquad I = I - 1; \\
\qquad \text{Continue} \\
\quad \text{end} \\
\text{end}
\]

Algorithm 2: This algorithm draws a (large) number \( V \) of samples \( y \) from the conditional distribution \( X | T(X) = t \) and calculates the relative frequency \( p_{\text{cond}} \) of samples giving a test statistic \( D(y) \) exceeding the observed \( D(x) \).

\[
\text{Data: Integer } V, \text{ Data set } x, \text{ sufficient statistic } T(x) = t \text{ and test statistic } D \\
\text{Result: } p_{\text{cond}} \\
\text{initialization;} \\
\text{count} = 0; \\
\text{for } 1 \text{ to } V \text{ do} \\
\quad \text{Draw } y \sim X | T(X) = t; \\
\quad \text{if } D(y) \geq D(x) \text{ then} \\
\qquad \text{count} = \text{count} + 1; \\
\quad \text{end} \\
\text{end} \\
p_{\text{cond}} = \text{count}/V
\]

Conditional Sampling from the Negative Binomial Distribution

Let \( Y_1 \sim \text{NB}(r_1, p), Y_2 \sim \text{NB}(r_2, p), \ldots, Y_n \sim \text{NB}(r_n, p) \) be independent random variables, where the parameters \( r_1, \ldots, r_n \) are assumed to be known. Then \( T(Y) = \sum_{i=1}^{n} Y_i \) is a sufficient statistic for \( p \). In this subsection, we will show how the algorithm
for the geometric distribution in Sect. 2.1 can be used to draw samples from the conditional distribution $Y \mid T(Y) = t$.

Note first that an argument like the one leading to (1) gives the following expression:

$$P(Y_1 = y_1, \ldots, Y_n = y_n \mid \sum_{i=1}^{n} Y_i = t) = \prod_{i=1}^{n} \left( \frac{y_i - r_i - 1}{y_i - 1} \right) \left( \frac{t + R - 1}{t} \right)$$

where $R = \sum_{i=1}^{n} r_i$.

The following shows that we can sample from this conditional distribution by using the algorithm for the geometric distribution. Note first that we can write for $i = 1, \ldots, n$,

$$Y_i = \sum_{j=1}^{r_i} X_{ij}$$

where $X_{ij}, i = 1, \ldots, n; j = 1, \ldots, r_i$ are i.i.d. from $\text{Geom}(p)$. Then

$$P(Y = y \mid T(Y) = t) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid T(Y_1, Y_2, \ldots, Y_n) = t) = P\left( \sum_{j=1}^{r_1} X_{1j} = y_1, \ldots, \sum_{j=1}^{r_n} X_{nj} = y_n \mid \sum_{i=1}^{n} \sum_{j=1}^{r_i} X_{ij} = t \right)$$

$$= \sum_{(x_{ij}) : \sum_{j} x_{ij} = y_1, \ldots, \sum_{j} x_{nj} = y_n} P\left( X_{ij} = x_{ij} : i = 1, \ldots, n, j = 1, \ldots, r_i \mid \sum_{i,j} X_{ij} = t \right).$$

It follows from this that we can use the method for drawing samples from the conditional geometric distribution to draw conditional samples in the negative binomial case. More precisely, we can first draw a sample $x_1, x_2, \ldots, x_R$, where $R = r_1 + r_2 + \ldots + r_n$, from the conditional distribution of $X_1, X_2, \ldots, X_R \mid T(X) = t$, where $X_i \sim \text{Geom}(p)$ for $i = 1, 2, \ldots, R$ and let

$$y_1 = \sum_{i=1}^{r_1} x_i, \quad y_2 = \sum_{i=r_1+1}^{r_1+r_2} x_i, \ldots, y_n = \sum_{i=r_1+\ldots+r_{n-1}+1}^{R} x_i.$$

We end up with a sample $y$ from the desired conditional distribution $Y \mid T(Y) = t$.

For simulation in practice, notice that Algorithm 1 with input $t$ and $R$ gives numbers $k_1, k_2, \ldots, k_{R-1}$. Using the transformation $\phi$ in Lemma 1, it is then seen that we have
\begin{align*}
y_1 &= kr_1 - r_1 \\
y_2 &= kr_1 + r_2 - k r_1 - r_2 \\
&\vdots \\
y_i &= kr_1 + \ldots + r_i - k r_1 + \ldots + r_{i-1} - r_i \\
&\vdots \\
y_{n-1} &= kr_1 + \ldots + r_{n-1} - k r_1 + \ldots + r_{n-2} - r_{n-1} \\
y_n &= t - k r_1 + \ldots + r_{n-1} + R - r_n.
\end{align*}

References

1. Barlow RE, Marshall AW, Proschan F et al (1963) Properties of probability distributions with monotone hazard rate. Ann Math Stat 34(2):375–389
2. Beltrán-Beltrán JI, O’Reilly FJ (2019) On goodness of fit tests for the Poisson, negative binomial and binomial distributions. Stat Pap 60(1):1–18
3. Bracquemond C, Crétois E, Gaudoin O (2002) A comparative study of goodness-of-fit tests for the geometric distribution and application to discrete time reliability. Laboratoire Jean Kuntzmann, Applied Mathematics and Computer Science, Technical Report
4. D’Agostino RB, Stephens MA (eds) (1986) Goodness-of-fit Techniques. Marcel Dekker Inc, New York
5. Feller W (1968) An introduction to probability theory and its applications. Wiley, New York
6. Fisher RA (1950) The significance of deviations from expectation in a Poisson series. Biometrics 6(1):17–24
7. González-Barrios JM, O’Reilly F, Rueda R (2006) Goodness of fit for discrete random variables using the conditional density. Metrika 64(1):77–94
8. Heller B (1986) A goodness-of-fit test for the negative binomial distribution applicable to large sets of small samples. In: Developments in water science, vol 27. Elsevier, pp 215–220
9. Kyriakoussis A, Li G, Papadopoulos A (1998) On characterization and goodness-of-fit test of some discrete distribution families. J Stat Plan Inference 74(2):215–228
10. Lockhart RA, O’reilly F, Stephens M (2009) Exact conditional tests and approximate bootstrap tests for the von Mises distribution. J Stat Theory Pract 3(3):543–554
11. Nakagawa T, Osaki S (1975) The discrete weibull distribution. IEEE Trans Reliab 24(5):300–301
12. Ozonur D, Gökpinar E, Gökpinar F, Bayrak H (2013) Comparisons of the goodness of fit tests for the geometric distribution. Gazi Univ J Sci 26(3):369–375
13. Paul SR (2005) Testing goodness of fit of the geometric distribution: an application to human fecundability data. J Mod Appl Stat Methods 4(2):8
14. Puig P, Weiß CH (2020) Some goodness-of-fit tests for the Poisson distribution with applications in biodosimetry. Comput Stat Data Anal 144:106878
15. Rao CR, Chakravarti I (1956) Some small sample tests of significance for a Poisson distribution. Biometrics 12(3):264–282
16. Rueda R, Reilly FO, Perez-Abreu V (1991) Goodness of fit for the Poisson distribution based on the probability generating function. Commun Stat Theory Methods 20(10):3093–3110
17. Singh B, Pudir P, Maheshwari S (2014) Parameter estimation of beta-geometric model with application to human fecundability data. arXiv preprint arXiv:1405.6392
18. Spinelli JJ, Stephens MA (1997) Cramér-von Mises tests of fit for the Poisson distribution. Can J Stat La Revue Canadienne de Statistique 25(2):257–268
19. Vila R, Nakano EY, Saulo H (2019) Theoretical results on the discrete Weibull distribution of Nakagawa and Osaki. Statistics 53(2):339–363
20. Wilf HS (1999) East side, west side . . . - an introduction to combinatorial families-with maple programming. https://www.math.upenn.edu/~wilf/lecnotes.html

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.