Statistics and Probability Letters 216 (2025) 110275

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Waiting time representation of discrete distributions

Tomasz J. Kozubowski [a], Dorota Młynarczyk [b,*], Anna K. Panorska [a]

[a] Department of Mathematics & Statistics, Mail Stop 084, University of Nevada, Reno, NV 89557, USA
[b] Departament de Matemàtiques, Universitat Autònoma de Barcelona, Edifici C, 08193 Bellaterra (Barcelona), Spain

ARTICLE INFO

Keywords:
Bivariate Bernoulli distribution
Discrete hazard rate
Distribution theory
Geometric distribution
Multivariate discrete distribution

ABSTRACT

We discuss a representation of any probability distribution on the set of non-negative integers as a waiting time distribution in a sequence of independent Bernoulli trials. Several associated results are derived and illustrated by examples. Multivariate extensions are briefly treated as well.

1. Introduction

The concept of a discrete random variable representing the waiting time for the first success in a sequence of Bernoulli trials finds applications in various fields, including engineering and science. While the geometric distribution serves as the most common example when the trials are independent and identically distributed (IID), there are also many examples of distributions that arise in the Bernoulli trials scheme when the trials are independent but not necessarily identically distributed. An important example of current interest is the case of the "harmonic" probability of success of the underlying sequence $\{B_n\}$ of independent Bernoulli trials:

$$\mathbb{P}(B_n = 1) = \frac{w_1}{w_1 + w_2 + n - 1}, \quad n \in \mathbb{N} = \{1, 2, \dots\}, w_1 > 0, w_2 \geq 0, \tag{1}$$

as recently studied by Huillet and Möhle (2023). As discussed therein, the one-parameter model with $w_2 = 0$ is closely related to the Ewens sampling formula (Ewens, 1972), the Ewens fragmentation process (Gnedin and Pitman, 2007), as well as to cycles of permutations and Poisson spacings (Najnudel and Pitman, 2020). The general model (1) has been studied extensively, with applications in reliability theory, ecology, biology, random records theory and models for random walks with disasters (see, e.g., Chen and Liu, 1997; Hoshino, 2001; Kozubowski and Podgórski, 2018; Pitman and Yor, 1997; Sibuya, 2014). Other important special cases of (1) include one where $w_1 \in (0, 1)$ and $w_2 = 1 - w_1$, and another where $w_2 = 1$, leading to *Sibuya* and *Yule–Simon* distribution of N, respectively, where

$$N = \inf\{n : B_n = 1\}. \tag{2}$$

Moreover, for $w_1 = w_2 = 1$, in which case we have $\mathbb{P}(B_n = 1) = 1/(1 + n)$, the distribution of N is given by

$$\mathbb{P}(N = n) = \frac{1}{1 + (n-1)} - \frac{1}{1+n} = \frac{1}{n(n+1)}, \quad n \in \mathbb{N}, \tag{3}$$

and provides the waiting time for the first record (the value greater than all the previous values) connected with a sequence of IID continuous random variables (Rényi, 1976). For an excellent summary of the historical development, applications, and new results on Bernoulli sequence-based models of the form (1) we refer to Huillet and Möhle (2023).

* Corresponding author.
E-mail address: dorotaanna.mlynarczyk@uab.cat (D. Młynarczyk).

https://doi.org/10.1016/j.spl.2024.110275
Received 19 December 2023; Received in revised form 22 May 2024; Accepted 10 September 2024
Available online 12 September 2024

The main goal of this paper is to show that, in fact, every discrete distribution supported on the set of non-negative integers $\mathbb{N}_0 = \{0, 1, \ldots\}$ is uniquely represented as the waiting time for the first success in a sequence of independent Bernoulli trials. This interpretation provides an insight into the structure of the particular distribution, which may be crucial in applications. For instance, if the Bernoulli success probability is of the form

$$\mathbb{P}(B_n = 1) = S(g(n)), \quad n \in \mathbb{N}, \tag{4}$$

where $S(x) = \mathbb{P}(X > x)$ is the survival function (SF) of a random variable X and $g(\cdot)$ is a function on $\mathbb{N}$, then the variable N defined by (2) provides the waiting time for the event $\{X_n > g(n)\}$ connected with a sequence $\{X_n\}$ of IID copies of X. Applications where such random variables are of great interest include scenarios in engineering where N represents the waiting time for the process values X_n to get "out of control". Similar applications arise when the success probabilities are of the form $\mathbb{P}(B_n = 1) = F(g(n))$, $n \in \mathbb{N}$, where this time $F(x) = \mathbb{P}(X \le x)$ is the cumulative distribution function (CDF) of X.

The ideas and results presented in this paper have several important consequences, including a new method for constructing bivariate (or multivariate) discrete distributions of a random vector $\mathbf{N} = (N_1, \ldots, N_k)$ *with any given marginals* and nontrivial dependence structures. Indeed, one can work with a k-dimensional sequence $\{\mathbf{B}_j = (B_j^{(1)}, \ldots, B_j^{(k)})\}$ of random vectors with marginal Bernoulli distributions such that for each $i = 1, \ldots, k$ the sequence $\{B_j^{(i)}\}$ of univariate Bernoulli trials have independent components and defines the distribution of N_i via $N_i = \inf\{n : B_n^{(i)} = 1\}$. While this idea has appeared already in connection with multivariate geometric distribution (see, e.g., Marshall and Olkin, 1985), it does not seem to have been explored much for other cases. Even in the geometric case, one can obtain a multitude of new multivariate distributions with given (geometric) margins by *dropping the assumption* that the underlying multivariate Bernoulli variables $\mathbf{B}_j$ are IID, as typically assumed in this construction (Marshall and Olkin, 1985), but still maintaining the independence of the components within each univariate sequence $\{B_j^{(i)}\}$. In comparison with the popular copula-based methodologies for constructing multivariate models with given marginal distributions, our approach is computationally simpler and allows for an intuitive interpretation of the model.

While we do not aim at an exhaustive account of the proposed methodology, particularly in the multivariate case, our goal is to present important properties of this construction in the univariate case (Section 2) along with illustrative examples (Section 3). Finally, we provide a brief introduction to the multivariate case (Section 4). We hope that the ideas and results presented below will be useful to researchers in statistics and those in other disciplines working with modeling data.

2. Main results

In this section we present our main result showing that every discrete distribution on $\mathbb{N}_0 = \{0, 1, \ldots\}$ has an interpretation as a waiting time for the first success in a sequence of independent Bernoulli trials. The waiting time is understood as the number of failures before the first success, although this can also be re-formulated in terms of the total number of trials. We also provide a probabilistic interpretation of this result and go over several important properties, followed by illustrative examples. We start with the main result.

Proposition 2.1. *Let N be a discrete random variable on $\mathbb{N}_0$ and let $\{B_n, n \in \mathbb{N}_0\}$ be a sequence of independent Bernoulli trials with the probabilities of success given by*

$$q_n = \mathbb{P}(B_n = 1) = \frac{\mathbb{P}(N = n)}{\mathbb{P}(N \ge n)} = \mathbb{P}(N = n \mid N \ge n), \quad n \in \mathbb{N}_0, \tag{5}$$

whenever $\mathbb{P}(N \ge n) > 0$ and with $q_n = 1$ when $\mathbb{P}(N \ge n) = 0$. Then,

$$N \stackrel{d}{=} \inf\{n \in \mathbb{N}_0 : B_n = 1\}. \tag{6}$$

Proof. For each $n \in \mathbb{N}_0$, set $p_n = \mathbb{P}(N = n)$ and let $\{B_n, n \in \mathbb{N}_0\}$ be a sequence of independent Bernoulli trials such that $\mathbb{P}(B_n = 1) = q_n$ with the $\{q_n\}$ defined above. Further, let $M = \inf\{n \in \mathbb{N}_0 : B_n = 1\}$. We show below that $M \stackrel{d}{=} N$, proving the result. We accomplish this by showing that for each $n \in \mathbb{N}_0$ we have $\mathbb{P}(N \ge n) = \mathbb{P}(M \ge n)$. Note that for $n = 0$ we have $\mathbb{P}(N \ge 0) = \mathbb{P}(M \ge 0) = 1$. Next, assume that the distribution of N is unbounded, so that $\mathbb{P}(N \ge n) > 0$ for all $n \in \mathbb{N}$. Then, for each $n \in \mathbb{N}$, we have

$$\mathbb{P}(M \ge n) = \mathbb{P}(B_0 = 0, \ldots, B_{n-1} = 0) = \prod_{j=0}^{n-1} (1 - q_j), \tag{7}$$

where the $\{q_j\}$ are given by (5). Since $1 - q_j = \mathbb{P}(N \ge j + 1)/\mathbb{P}(N \ge j)$, we have

$$\prod_{j=0}^{n-1} (1 - q_j) = \frac{\mathbb{P}(N \ge 1)}{\mathbb{P}(N \ge 0)} \cdot \frac{\mathbb{P}(N \ge 2)}{\mathbb{P}(N \ge 1)} \cdots \frac{\mathbb{P}(N \ge n)}{\mathbb{P}(N \ge n - 1)} = \frac{\mathbb{P}(N \ge n)}{\mathbb{P}(N \ge 0)} = \mathbb{P}(N \ge n) \tag{8}$$

for all $n \in \mathbb{N}$. Thus, $\mathbb{P}(N \ge n) = \mathbb{P}(M \ge n)$ for each $n \in \mathbb{N}_0$, as desired. Finally, assume that the support of the distribution of N is bounded, so that $p_k > 0$ for some $k \in N_0$ while $p_j = 0$ for $j \ge k + 1$. In this case, the $\{q_n\}$ are given by (5) for all $n \le k$ and $q_j = 1$ for all $j \ge k$. Moreover, for all $n \le k$ Eqs. (7)–(8) still hold while for $n \ge k + 1$ we have $\mathbb{P}(M \ge n) = \mathbb{P}(N \ge n) = 0$. Thus, we conclude that $\mathbb{P}(N \ge n) = \mathbb{P}(M \ge n)$ for each $n \in \mathbb{N}_0$. This completes the proof. □

Remark 2.2. Let us note that according to (6), the probabilities $p_n = \mathbb{P}(N = n)$ are given by

$$p_n = q_n \prod_{j=0}^{n-1}(1 - q_j), \quad n \in \mathbb{N}_0, \tag{9}$$

with the product above understood to be 1 for $n = 0$. Moreover, the quantity q_n in (5) is actually the hazard rate function of the distribution of N, which uniquely determines this distribution (Shaked et al., 1995). The sequence $\{q_n\}$ is also determined uniquely by the sequence $\{p_n\}$ under the convention that $q_n = 1$ when $\mathbb{P}(N \geq n) = 0$. While Eq. (9) appeared in Shaked et al. (1995), there was no interpretation as a waiting time distribution mentioned in their work.

Remark 2.3. It should be pointed out that not every sequence of independent Bernoulli trials leads to a proper probability distribution of N with probabilities $p_n = \mathbb{P}(N = n)$ defined via (9). If for some $k \in N_0$ we have $\mathbb{P}(B_k = 1) = 1$ then the resulting variable N is bounded above by k. However, if all the Bernoulli probabilities $\mathbb{P}(B_j = 1) = q_j$ are less than 1, then, as shown in Salvia and Bollinger (1982), the formula in (9) defines a legitimate set of probabilities on N_0 if and only if

$$\sum_{i=0}^{\infty} q_i = \infty. \tag{10}$$

This will always be true for monotonically non-decreasing sequences $\{q_n\}$, but in general may fail, particularly when $\{q_n\}$ converges to zero with a high rate. For example, while the harmonic sequence q_n given by the right-hand-side in (1) satisfies (10), the sequence $q_n = \alpha/(n+1)^2$ with $\alpha \in (0, 1)$ does not. Thus, when $\{q_n\}$ is taken to be the set of probabilities $\{\tilde{p}_n\}$ of an unbounded distribution on $\mathbb{N}_0$, then it does not yield a proper distribution of N in (6). However, if in this setting we set $q_n = 1 - \tilde{p}_n$, then the distribution of N will be proper. In this case, N represents the size of the string of the form $(0, 1, \ldots, k)$ that may come up (right away) when observing IID values of a random variable $\tilde{N}$ with probabilities given by $\{\tilde{p}_n\}$. We refer to Chapters VIII and XIII of Feller (1968) for classical examples of waiting times of various patterns of successes and failures connected with Bernoulli sequences.

Remark 2.4. The result in Proposition 2.1 has the following probabilistic interpretation. Consider a sequence of independent random variables $\{X_n, n \in \mathbb{N}_0\}$, where $X_n \stackrel{d}{=} N|N \geq n$ whenever $\mathbb{P}(N \geq n) > 0$ and $X_n = n$ otherwise. Further, define a sequence $\{B_n, n \in \mathbb{N}_0\}$ of independent Bernoulli random variables via $B_n = \mathbf{I}(X_n = n)$. Since the probabilities of success of the $\{B_n\}$ are given by (5), it follows that the variable N admits the representation (6) based on this particular $\{B_n\}$. In other words, N is the waiting time (in terms of the number of failures) for the event $X_n = n$.

Remark 2.5. Note that if $\mathbb{P}(N = i) = 0$ for some $i \in \mathbb{N}_0$ then $q_i = \mathbb{P}(B_i = 1) = 0$ as well, so that $\mathbb{P}(N \neq i) = 1$. Further, if the support of the distribution of N is bounded, so that $\mathbb{P}(n_1 \leq N \leq n_2) = 1$ for some $0 \leq n_1 \leq n_2 < \infty$, then we have $q_i = 0$ for $i < n_1$ (so that $\mathbb{P}(N = i) = 0$) and $q_{n_2} = 1$ (so that $\mathbb{P}(N \leq n_2) = 1$). In this situation we also set $q_i = 1$ for all $i > n_2$.

2.1. Further properties

We start with the following result concerning monotonic stochastic behavior of the distribution of N under monotonic transformations of the associated sequence of Bernoulli probabilities $\{q_n\}$. In the result below the ordinary stochastic order $X \leq_{\rm st} Y$ of two random variables X and Y is equivalent to the condition that $F_Y(x) \leq F_X(x)$ for all $x \in \mathbb{R}$ where $F_X(\cdot)$ and $F_Y(\cdot)$ are the CDFs of X and Y, respectively (see, e.g., Belzunce et al., 2016).

Proposition 2.6. *Let $i \in \{1, 2\}$. If N_i is a discrete random variable on $\mathbb{N}_0$ with an associated sequence $\{q_j^{(i)}\}$ of Bernoulli probabilities, where for each $j \in \mathbb{N}_0$ we have $q_j^{(1)} \geq q_j^{(2)}$, then $N_1 \leq_{st} N_2$.*

Proof. The result is a consequence of (8). □

We now proceed with several results related to common transformations, where in each case we provide an interpretation in terms of the underlying sequence of Bernoulli variables and/or their probabilities. For univariate transformations, we assume that N is a random variable with support in $\mathbf{N}_0$, probabilities $\{p_j\}$, and an associated Bernoulli sequence $\{B_j\}$ where $\mathbb{P}(B_j = 1) = q_j$, $j \in \mathbb{N}_0$. The transformed variable, denoted by $\tilde{N}$, has probabilities denoted by $\{\tilde{p}_j\}$ and an associated Bernoulli sequence $\{\tilde{B}_j\}$ where $\mathbb{P}(\tilde{B}_j = 1) = \tilde{q}_j$.

2.1.1. Truncation below

Let $\tilde{N} = N|N \geq k$ for some $k \in \mathbb{N}$, so that $\tilde{p}_j = \mathbb{P}(\tilde{N} = j) = 0$ for $j = 0, \ldots, k-1$ and

$$\tilde{p}_j = \mathbb{P}(N = j|N \geq k) = \frac{p_j}{\sum_{i \geq k} p_i} \quad \text{for } j \geq k. \tag{11}$$

Simple algebra shows that $\tilde{q}_j = \mathbb{P}(\tilde{B}_j = 1) = 0$ for $j = 0, \ldots, k-1$ and

$$\tilde{q}_j = \frac{\tilde{p}_j}{\sum_{m \geq j} \tilde{p}_m} = \frac{p_j / \sum_{i \geq k} p_i}{\sum_{m \geq j} p_m / \sum_{i \geq k} p_i} = \frac{p_j}{\sum_{m \geq j} p_m} = q_j \quad \text{for } j \geq k. \tag{12}$$

For a random variable that is truncated from below, the sequence of Bernoulli trials associated with it has the same probabilities of success as those associated with the original random variable for indexes at or above the truncation level, and these probabilities are equal to zero for indexes below the truncation level. Thus, in view of Proposition 2.6, we have $N \le_{st} \tilde{N}$.

2.1.2. Truncation above

Similarly, if $\tilde{N} = N | N \le k$ for some $k \in \mathbb{N}_0$ then $\tilde{p}_j = \mathbb{P}(\tilde{N} = j) = 0$ for $j = k+1, k+2, \ldots$ and

$$\tilde{p}_j = \mathbb{P}(N = j | N \le k) = \frac{p_j}{\sum_{i \le k} p_i} \quad \text{for } j \le k. \tag{13}$$

Simple algebra shows that in this case we have

$$\tilde{q}_j = \frac{\tilde{p}_j}{\sum_{m=j}^{k} \tilde{p}_m} = \frac{p_j / \sum_{i \le k} p_i}{\sum_{m=j}^{k} p_m / \sum_{i \le k} p_i} = \frac{p_j}{\sum_{m=j}^{k} p_m} \quad \text{for } j \le k. \tag{14}$$

Since the Bernoulli probabilities of the original random variable N are given by $q_j = p_j / \sum_{m=j}^{\infty} p_m$, we see that $q_j \le \tilde{q}_j$ for $j \le k$ so that by (8) we have $\mathbb{P}(\tilde{N} \ge n) \le \mathbb{P}(N \ge n)$ for $n \le k$. On the other hand, for $n > k$ we have $\mathbb{P}(\tilde{N} \ge n) = 0$, so the above relation between the two tail probabilities is also true. This leads to the conclusion that $\tilde{N} \le_{st} N$, as expected. The conclusion also follows from Proposition 2.6 with the convention that $\tilde{q}_j = 1$ for all $j > k$.

2.1.3. Translation

Let $\tilde{N} = N + k$ for some $k \in \mathbb{N}$. Then,

$$\tilde{p}_j = P(\tilde{N} = j) = \begin{cases} p_{j-k} & \text{for } j \ge k \\ 0 & \text{for } j < k. \end{cases}$$

It follows that $\tilde{q}_j = 0$ for $j < k$ and $\tilde{q}_j = q_{j-k}$ for $j \ge k$. Thus, for a random variable that is shifted up by $k \in \mathbb{N}$, the sequence of Bernoulli trials associated with it has the probabilities of success of the original random variable shifted according to the translation parameter k for indexes at or above k and zeros for indexes lower than k.

2.1.4. Zero-inflation

Let $\tilde{N}$ be a zero-inflated N, which is a mixture of a point mass at zero taken with probability $\omega \in (0,1)$ and the distribution of N taken with probability $1 - \omega$, so that

$$\tilde{p}_j = \mathbb{P}(\tilde{N} = j) = \begin{cases} \omega + (1-\omega) p_0 & \text{for } j = 0 \\ (1-\omega) p_j & \text{for } j > 0. \end{cases}$$

It follows that $\tilde{q}_0 = \omega + (1-\omega) q_0$ and

$$\tilde{q}_j = \frac{(1-\omega) p_j}{\sum_{i \ge j} (1-\omega) p_i} = q_j \quad \text{for } j > 0.$$

The sequence of Bernoulli trials for zero-inflated variable has the same probabilities of success as the original variable except for the first element of the sequence.

2.1.5. Minima and mixtures

Consider a random vector $(N_1, \ldots, N_k)$ of mutually independent discrete random variables on $\mathbb{N}_0$ with probabilities given by $p_n^{(i)} = \mathbb{P}(N_i = n)$, $n \in \mathbb{N}_0$. Further, let $\{B_j^{(i)}, j \in \mathbb{N}_0\}$ be the Bernoulli sequence associated with N_i, where the sequences $\{B_j^{(i)}\}$ are mutually independent across $i = 1, \ldots, k$ and $\mathbb{P}(B_j^{(i)} = 1) = q_j^{(i)}$, $j \in \mathbb{N}_0$. The first result below provides the Bernoulli sequence associated with the random variable

$$\tilde{N} = \min\{N_1, \ldots, N_k\}. \tag{15}$$

Proposition 2.7. *In the above setting, the Bernoulli sequence associated with $\tilde{N}$ in (15) is given by*

$$\tilde{B}_j = \max\{B_j^{(1)}, \ldots, B_j^{(k)}\}, \quad j \in \mathbb{N}_0, \tag{16}$$

so that

$$\tilde{q}_j = \mathbb{P}(\tilde{B}_j = 1) = 1 - \prod_{i=1}^{k} (1 - q_j^{(i)}), \quad j \in \mathbb{N}_0. \tag{17}$$

Proof. The proof is straightforward. □

We now consider a random variable $\tilde{N}$ with distribution which is a mixture of the distributions of N_i, so that

$$\tilde{p}_n = \mathbb{P}(\tilde{N} = n) = \sum_{j=1}^{k} \omega_j p_n^{(j)}, \quad n \in \mathbb{N}_0, \tag{18}$$

where the $\{\omega_j\}$ are non-negative weights that sum to 1. It turns out that the Bernoulli probabilities $\tilde{q}_j$ associated with the variable $\tilde{N}$ also arise as a mixture, as shown in the result below.

Proposition 2.8. *In the above setting, the Bernoulli sequence $\{\tilde{B}_j, j \in \mathbb{N}_0\}$ associated with $\tilde{N}$ in (18) has probabilities of the form:*

$$\tilde{q}_j = \mathbb{P}(\tilde{B}_j = 1) = \sum_{i=1}^{k} \omega_{ij} q_j^{(i)}, \quad j \in \mathbb{N}_0, \tag{19}$$

where

$$\omega_{ij} = \frac{\omega_i \mathbb{P}(N_i \geq j)}{\sum_{l=1}^{k} \omega_l \mathbb{P}(N_l \geq j)}, \quad i = 1, \dots, k, j \in \mathbb{N}_0. \tag{20}$$

Proof. This is a straightforward consequence of the definition of $\tilde{q}_n$ and the fact that $\mathbb{P}(\tilde{N} \geq j) = \sum_{i=1}^{k} \omega_i \mathbb{P}(N_i \geq j)$, $j \in \mathbb{N}_0$. □

Remark 2.9. We note that in the special case where $k = 2$ and N_1 is a point mass at zero, the result in Proposition 2.8 reduces to that connected with zero-inflation, discussed in Section 2.1.4.

3. Illustrative examples

As shown in Proposition 2.1, all the classical discrete distributions on $\mathbb{N}_0$, including the Poisson, the binomial, the negative binomial, and the logarithmic distributions, arise as a waiting time for the first success of a particular Bernoulli sequence. The probabilities of success of their underlying Bernoulli sequences are provided by the hazard rates of these distributions, which are all well-known and studied in the literature. This provides an additional interpretation of these distributions, which may be useful in certain applications. Some of these distributions appear in the examples below, which are focused on the lesser known and new distributions, illustrating the main results of this paper.

3.1. Discretization of continuous random variables

There is a large body of literature devoted to various discretization schemes that generate discrete distributions on $\mathbb{N}$ or $\mathbb{N}_0$ from their continuous counterparts. In one particularly popular scheme, a discrete variable N is generated by taking the integer part of X (see, e.g., Chakraborty, 2015), so that $N = n$ whenever $n \leq X < n + 1$, leading to $p_n = \mathbb{P}(N = n) = S(n) - S(n + 1)$, where $S(\cdot)$ is the SF of X. Moreover, in this scheme N and X share the same tail probabilities: $\mathbb{P}(N \geq n) = \mathbb{P}(X \geq n)$ for integer values of n. If the underlying X has distribution supported on $\mathbb{R}_+$, then its discrete counterpart N will have a discrete distribution on $\mathbb{N}_0$. In this case the success probabilities q_n of the Bernoulli sequence connected with N have a simple representation in terms of the SF of X, shown in the result below.

Proposition 3.1. *Let X have a continuous distribution on $\mathbb{R}_+$ given by the SF $S(\cdot)$, and let N be defined as the integer part of X. Then N admits the representation given by (6), where the success probabilities $\{q_n\}$ of the Bernoulli sequence $\{B_n\}$ are given by*

$$q_n = \mathbb{P}(B_n = 1) = 1 - \frac{S(n+1)}{S(n)}, \quad n \in \mathbb{N}_0. \tag{21}$$

Proof. The proof is straightforward. □

3.1.1. Discrete Weibull distribution

If the random variable X in Proposition 3.1 has a Weibull distribution, given by the SF $S(x) = \exp(-\lambda x^{\beta})$, $x \in \mathbb{R}_+$, where $\lambda > 0$ is a scale parameter and $\beta > 0$ is a shape parameter, then upon discretization we obtain a discrete Weibull distribution (see, e.g., Nakagawa and Osaki, 1975). According to Propositions 2.1 and 3.1, the resulting discrete Weibull random variable N admits the waiting time representation (6), where the success probabilities $\{q_n\}$ of the Bernoulli sequence $\{B_n\}$ are given by

$$q_n = \mathbb{P}(B_n = 1) = 1 - e^{-\lambda[(n+1)^{\beta} - n^{\beta}]}, \quad n \in \mathbb{N}_0. \tag{22}$$

These probabilities are increasing for $\beta > 1$ and decreasing for $0 < \beta < 1$, while for $\beta = 1$ they are constant. In the latter case, the Weibull distribution is exponential and its discretization is a geometric distribution with parameter $p = 1 - e^{-\lambda}$, denoted by $GEO(p)$ and given by the PDF

$$p_n = \mathbb{P}(N = n) = p(1 - p)^n, \quad n \in \mathbb{N}_0. \tag{23}$$

3.1.2. Discrete Pareto distribution

If the random variable X in Proposition 3.1 has Pareto Type II (Lomax) distribution, given by the SF $S(x) = (1+\sigma\alpha x)^{-1/\alpha}$, $x \in \mathbb{R}_+$, where $\sigma > 0$ is a scale parameter and $\alpha > 0$ is a tail parameter, then upon discretization we obtain a discrete Pareto distribution (see, e.g., Buddana and Kozubowski, 2014). According to Propositions 2.1 and 3.1, the resulting random variable N admits the waiting time representation (6), where the success probabilities $\{q_n\}$ of the Bernoulli sequence $\{B_n\}$ are given by

$$q_n = \mathbb{P}(B_n = 1) = 1 - \left(\frac{1+\sigma\alpha n}{1+\sigma\alpha(n+1)}\right)^{1/\alpha}, \quad n \in \mathbb{N}_0. \tag{24}$$

These probabilities are monotonically decreasing to zero for any $\alpha, \sigma > 0$, and the discrete Pareto distribution has a power-law tail. At the boundary value of $\alpha = 0$ the probabilities q_n in (24) are constant, given by $p = 1 - e^{-\sigma}$, and we recover the geometric distribution (23). Further, in the special case where $\alpha = 1$ and $\sigma = 1$, the probabilities q_n in (24) turn into $q_n = 1/(2+n)$. If in this case the distribution of N is shifted up by $k = 1$, then according to the results of Section 2.1.3 we obtain the harmonic probabilities (1) with $w_1 = w_2 = 1$. The corresponding random variable on $\mathbb{N}$, with probabilities given by the right-hand-side in (3), provides the waiting time for the first record connected with a sequence of IID continuous random variables, as discussed in the introduction.

3.2. Geometric distribution

Let N have a geometric distribution with the PDF (23), so that $q_n = p$ for all $n \in \mathbb{N}_0$. If $\tilde{N} \stackrel{d}{=} N|N \geq k$ for some $k \in \mathbb{N}_0$, then according to the results of Section 2.1.1 we have $\tilde{q}_j = p$ for $j \geq k$ and $\tilde{q}_j = 0$ for $j < k$. In turn, by the results of Section 2.1.3, the distribution of $\tilde{N}$ is also a translation of the distribution of N up by the same k, that is $\tilde{N} \stackrel{d}{=} N + k$. Thus we have recovered the well-known *memoryless property* of the geometric distribution: If $N \sim GEO(p)$ then $N \stackrel{d}{=} N - k|N \geq k$ for all $k \in \mathbb{N}_0$. On the other hand, if $\tilde{N} \stackrel{d}{=} N|N \leq k$ for some $k \in \mathbb{N}$, then, by the results of Section 2.1.2, the $\{q_n\}$ sequence of $\tilde{N}$ is given by

$$\tilde{q}_n = \frac{p}{1-(1-p)^{k+1-n}}, \quad n \leq k, \tag{25}$$

which is monotonically increasing to the largest value of $q_k = 1$. By convention, we also set $\tilde{q}_n = 1$ for all $n > k$.

Consider now a random variable $\tilde{N}$ defined via $\tilde{N} = \min\{n : \tilde{B}_n = 1\}$, where $\tilde{q}_n = \mathbb{P}(\tilde{B}_n = 1) = 1 - p_n$, $n \in \mathbb{N}_0$, and the $\{p_n\}$ are the geometric probabilities in (23). It is easy to see that the PDF of $\tilde{N}$ is given by

$$\tilde{p}_n = \mathbb{P}(\tilde{N} = n) = p^n(1-p)^{n(n-1)/2} - p^{n+1}(1-p)^{n(n+1)/2}, \quad n \in \mathbb{N}_0, \tag{26}$$

while the tail probability of $\tilde{N}$ is of the form $\mathbb{P}(\tilde{N} \geq n) = p^n(1-p)^{n(n-1)/2}$, $n \in \mathbb{N}_0$. The latter is the product of $\mathbb{P}(N_1 \geq n) = p^n$ and $\mathbb{P}(N_2 \geq n) = (1-p)^{n(n-1)/2}$, where $N_1 \sim GEO(1-p)$ and N_2 has the PDF given by $p_n^{(2)} = \mathbb{P}(N_2 = n) = (1-p)^{n(n-1)/2} - (1-p)^{n(n+1)/2}$, $n \in \mathbb{N}$. This shows that $\tilde{N} \stackrel{d}{=} \min\{N_1, N_2\}$, where N_1 and N_2 are independent. Since the q-sequences of N_1 and N_2 are given by $q_n^{(1)} = 1 - p$ and $q_n^{(2)} = 1 - (1-p)^n$, respectively, we have $\tilde{q}_n = 1 - (1-q^{(1)})(1-q^{(2)})$, so our conclusion is also consistent with Proposition 2.7. The distribution of N_2 is obtained as follows. By the results of Section 2.1.3, we have $N_2 \stackrel{d}{=} M + 1$, where M is a discrete distribution on $\mathbb{N}_0$ with the q-sequence given by $q_n^M = 1 - (1-p)^{n+1} = 1 - S(n+1)/S(n)$, where $S(x) = (1-p)^{x(x+1)/2}$ $(x > 0)$ is the SF of a continuous random variable X. By Proposition 3.1, M is the integer part of this X. Finally, note that since $g(x) = x(x+1)/2$ is a bijection on $[0, \infty)$, it follows that $X \stackrel{d}{=} h(-E/\log(1-p))$, where E is standard exponential and the function $h(\cdot)$ is the inverse of $g(\cdot)$.

3.3. Sibuya distribution

When $w_1 = \alpha \in (0,1)$ and $w_2 = 1 - \alpha$, then the harmonic probabilities (1) turn into α/n $(n \in \mathbb{N})$, and the distribution of N defined via (2) is known as the Sibuya distribution (see, e.g., Kozubowski and Podgórski, 2018). This N provides the *number of trials* before the first success in a sequence of independent Bernoulli trials with the above probability of success. According to the results of Section 2.1.3, this distribution is a translation by $k = 1$ of a Sibuya distribution on N_0, which represents the *number of failures* before the first success, and for which we have

$$q_n = \mathbb{P}(B_n = 1) = \frac{\alpha}{n+1}, \quad n \in \mathbb{N}_0. \tag{27}$$

According to (9), the probabilities of the random variable N defined via (2) for which the $\{q_n\}$ in Proposition 2.1 are given by (27), are as follows:

$$p_n = \mathbb{P}(N = n) = (1-\alpha)(1-\alpha/2)\dots(1-\alpha/n)\frac{\alpha}{n+1} = \binom{\alpha}{n+1}(-1)^n, \quad n \in \mathbb{N}_0. \tag{28}$$

With the sequence $\{q_n\}$ slowly decreasing to zero, the distribution (28) is heavy tailed, and the tail probability of N has power-law asymptotics (see, e.g., Corollary 2.6 in Kozubowski and Podgórski, 2018):

$$\mathbb{P}(N \geq n) \sim \frac{1}{\Gamma(1-\alpha)}\frac{1}{n^\alpha} \quad \text{as } n \to \infty. \tag{29}$$

Consequently, the moments of order α and above of this distribution do not exist. On the other hand, a related distribution for which the Bernoulli probabilities are given by $\tilde{q}_n = 1 - q_n$ and the corresponding "waiting-time" random variable $\tilde{N}$ has probabilities of the form[1]

$$\tilde{p}_n = \mathbb{P}(\tilde{N} = n) = \frac{\alpha}{1}\frac{\alpha}{2}\cdots\frac{\alpha}{n}\left(1 - \frac{\alpha}{n+1}\right) = \frac{\alpha^n}{n!} - \frac{\alpha^{n+1}}{(n+1)!}, \quad n \in \mathbb{N}_0, \tag{30}$$

has a very light tail, since here we have

$$\mathbb{P}(\tilde{N} \geq n) = \frac{\alpha^n}{n!}, \quad n \in \mathbb{N}_0. \tag{31}$$

Since the tail probability in (31) is the product of the tail probabilities of N_1 and N_2, where $\mathbb{P}(N_1 \geq n) = \alpha^n$ and $\mathbb{P}(N_2 \geq n) = 1/n!$, it follows that $\tilde{N} \stackrel{d}{=} \min\{N_1, N_2\}$, where N_1 and N_2 are independent. This is in agreement with Proposition 2.7, since $\tilde{q}_n = 1 - (1 - q_n^{(1)})(1 - q_n^{(2)})$, $n \in \mathbb{N}_0$, where $q_n^{(1)} = 1 - \alpha$ are the success probabilities of the Bernoulli sequence corresponding to the (geometric) variable $N_1 \sim GEO(1-\alpha)$ and $q_n^{(2)} = 1 - 1/(n+1)$ are the success probabilities of the Bernoulli sequence corresponding to N_2. This is a special case of the result below, which can be proven along the same lines.

Proposition 3.2. *For each $\alpha, \beta \in [0,1]$, let X_α have the distribution given by the PDF (30) and let N_β have a geometric $GEO(1-\beta)$ distribution. Then,* $\min\{N_\beta, X_\alpha\} \stackrel{d}{=} X_{\alpha\beta}$.

4. Remarks on multivariate extensions

The results presented above lead to a new method for constructing multivariate discrete distributions of $\mathbf{N} = (N_1, \dots, N_k)$ *with any given univariate marginals* supported on $\mathbb{N}_0$ and with a nontrivial dependence structures. Indeed, one may be able to build such distributions starting with a k-dimensional sequence $\{\mathbf{B}_j = (B_j^{(1)}, \dots, B_j^{(k)})\}$ of random vectors with univariate marginal Bernoulli distributions such that for each $i = 1, \dots, k$ the sequence $\{B_j^{(i)}\}$ of univariate Bernoulli trials have independent components and $N_i = \inf\{n : B_n^{(i)} = 1\}$.

Below we sketch this idea in the bivariate case $k = 2$. Since we are interested in bivariate distributions with fixed marginals, it is convenient to describe the Bernoulli random vector $\mathbf{B} = (B^{(1)}, B^{(2)})$ from the perspective of its marginal distributions, using marginal probabilities of success $q^{(i)} = \mathbb{P}(B_i = 1) \in [0,1]$, $i = 1, 2$. The joint probabilities of $\mathbf{B}$ can be conveniently described via a 2×2 table (see Table 1). With the marginals being fixed, the distribution is driven by the parameter α, with range given by

$$D(q^{(1)}, q^{(2)}) = \left\{\alpha : 0 \vee \frac{q^{(1)} + q^{(2)} - 1}{q^{(1)} \wedge q^{(2)}} \leq \alpha \leq 1\right\}, \tag{32}$$

where $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. We denote this distribution by $BB(\alpha, q^{(1)}, q^{(2)})$. We note that the independent case arises when $\alpha = q^{(1)} \vee q^{(2)}$, which always falls within the boundaries in (32), while the upper and lower boundaries for α correspond to extreme positive and negative linear dependence, respectively.

Table 1
A general bivariate Bernoulli distribution $BB(\alpha, q^{(1)}, q^{(2)})$.

$B^{(1)} \backslash B^{(2)}$	1	0	
1	$\alpha(q^{(1)} \wedge q^{(2)})$	$q^{(1)} - \alpha(q^{(1)} \wedge q^{(2)})$	$q^{(1)}$
0	$q^{(2)} - \alpha(q^{(1)} \wedge q^{(2)})$	$1 - q^{(1)} - q^{(2)} + \alpha(q^{(1)} \wedge q^{(2)})$	$1 - q^{(1)}$
	$q^{(2)}$	$1 - q^{(2)}$	

Suppose we wish to construct a bivariate distribution of $\mathbf{N} = (N_1, N_2)$ on $\mathbb{N}_0^2$ with the N_i having particular marginal distributions defined through the set of probabilities $\{p_j^{(i)}\}$ such that $p_j^{(i)} = \mathbb{P}(N_i = j)$, $j \in \mathbb{N}_0$. These $\{p_j^{(i)}\}$ then "connect" with the success probabilities $\{q_j^{(i)}\}$ of the underlying Bernoulli sequences $\{B_j^{(i)}\}$ via Proposition 2.1, so that $N_i = \inf\{j : B_j^{(i)} = 1\}$. Thus, given a particular bivariate sequence $\{\mathbf{B}_j = (B_j^{(1)}, B_j^{(2)}), j \in \mathbb{N}_0\}$ such that $\mathbf{B}_j \sim BB(\alpha_j, q_j^{(1)}, q_j^{(2)})$, $j \in \mathbb{N}_0$, and where for each $i = 1, 2$ the Bernoulli variables $\{B_j^{(i)}\}$ are mutually independent across $j \in \mathbb{N}_0$, the variables N_1 and N_2 defined via $N_i = \inf\{j : B_j^{(i)} = 1\}$ have the required marginal distributions. There are multiple ways of constructing such multivariate Bernoulli distributions (cf. Lovison, 2006), with the simplest scenario being the case where the bivariate vectors $\mathbf{B}_j$ are mutually independent. This was the underlying assumption in this construction where it was used earlier to define bivariate geometric distribution (see, e.g., Marshall and Olkin, 1985 and references therein). However, this work (and other similar works) focused on the geometric model assume that the $\mathbf{B}_j$ are actually IID, with the same parameter α across different j, so that $\mathbf{B}_j \sim BB(\alpha, q^{(1)}, q^{(2)})$, where the $q^{(1)}$ $q^{(2)}$ are the (constant) sequences corresponding to the marginal geometric variables $N_i \sim GEO(q^{(i)})$, $i = 1, 2$. However, the assumption of constant α in the BB distribution can be relaxed with the $\mathbf{B}_j$ still being mutually independent but not identically distributed. To illustrate this, we re-visit the geometric case, where $N_i \sim GEO(q^{(i)})$, $i = 1, 2$ and where $\mathbf{B}_j \sim BB(\alpha_j, q^{(1)}, q^{(2)})$ are mutually independent across $j \in \mathbb{N}_0$,

[1] This distribution was briefly mentioned in Salvia and Bollinger (1982), see their example b on p. 458. However, the expression for the relevant probability given therein was not stated correctly.

where each α_j satisfies the boundary condition (32) but otherwise can vary arbitrarily with j. It is straightforward to derive the joint PDF of (N_1, N_2), where $N_i = \inf\{j : B_j^{(i)} = 1\}$, leading to

$$p_{ij} = \mathbb{P}(N_1 = i, N_2 = j) = \begin{cases} q^{(2)}(1-q^{(2)})^{j-i-1}(q^{(1)} - \alpha_i q)\prod_{k=0}^{i-1}(1 - q^{(1)} - q^{(2)} + \alpha_k q) & i < j, \\ \alpha_i q \prod_{k=0}^{i-1}(1 - q^{(1)} - q^{(2)} + \alpha_k q) & i = j, \\ q^{(1)}(1-q^{(1)})^{i-j-1}(q^{(2)} - \alpha_j q)\prod_{k=0}^{j-1}(1 - q^{(1)} - q^{(2)} + \alpha_k q) & i > j, \end{cases}$$

where $q = q^{(1)} \wedge q^{(2)}$. It can be shown that this is a well defined distribution on $\mathbb{N}_0^2$. Moreover, this family of bivariate distributions with geometric $GEO(q^{(i)})$ marginals of N_i is more general than that studied in the literature (Marshall and Olkin, 1985), where it was assumed that all $\{\alpha_i\}$ are equal.

Remark 4.1. There is a multitude of different multivariate geometric distributions in the literature, some of which are related to the construction via Bernoulli trials discussed above. Nair et al. (2018) presented six different constructions. Other methods were developed in Jayakumar and Mundassery (2007), Omey and Minkova (2014), Roy (1993), Phatak and Sreehari (1981), and Kimpton et al. (2022), among others. It is worth mentioning that the model studied by Omey and Minkova (2014), which was defined via a *single sequence* of independent trials *with three different outcomes*, is actually equivalent to the above construction via IID bivariate Bernoulli trials where $q^{(1)} + q^{(2)} < 1$ and $\alpha = 0$ (corresponding to the extreme negative correlation). Another popular model studied by Phatak and Sreehari (1981) and others (see, e.g., Jayakumar and Mundassery, 2007) can also be related to the above scheme, although the N_1 and N_2 are defined by different means. Namely, assuming that we observe a sequence of IID bivariate Bernoulli trials $\mathbf{B}_j \sim BB(0, q^{(1)}, q^{(2)})$ where $q^{(1)} + q^{(2)} < 1$ (so that the event $(1,1)$ can never be observed), the variables N_1 and N_2 represent the numbers of events $(1,0)$ and $(0,1)$, respectively, before the first event $(0,0)$ is observed. It should be noted that while the marginal distributions of these N_1 and N_2 are indeed geometric, the parameters are no longer given by $q^{(1)}$ and $q^{(2)}$.

Data availability

No data was used for the research described in the article.

Acknowledgments

All authors thank the reviewer for her/his thoughtful comments and suggestions that led to improvement of this manuscript. This paper was drafted during D.M. research visit to UNR, and revised during A.P. and T.K. visit to The Centre de Recerca Matemàtica (CRM). A.P. and T.K. thank CRM for their support. D.M. is grateful for the support received from the *Fundació Ferran Sunyer i Balaguer* in the form of a travel grant, facilitating her research stay at the University of Nevada Reno (UNR). This work was also supported by the Ministerio de Ciencia e Innovación (Spain) [research grant PID2022-137414OB-I00]. T.K.'s travel was partially supported by the UNR VPRI travel grant. A.P.'s travel was partially supported by the Department of Mathematics and Statistics, UNR.

References

Belzunce, F., Martinez-Riquelme, C., Mulero, J., 2016. An Introduction To Stochastic Orders. Academic Press, Amsterdam.
Buddana, A., Kozubowski, T.J., 2014. Discrete Pareto distributions. Econ. Qual. Control 29 (2), 143–156.
Chakraborty, S., 2015. Generating discrete analogs of continuous probability distributions. J. Stat. Distrib. Appl. 2 (Article 6), 1–30.
Chen, S.X., Liu, J.S., 1997. Statistical applications of the Poisson-Binomial and conditional Bernoulli distributions. Statist. Sinica 7 (4), 875–892.
Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3, 87–112.
Feller, W., 1968. An Introduction to Probability Theory and its Applications, vol. 1, John Wiley & Sons.
Gnedin, A., Pitman, J., 2007. Poisson representation of a Ewens fragmentation process. Combin. Probab. Comput. 16 (6), 819–827.
Hoshino, N., 2001. Applying Pitman's sampling formula to microdata disclosure risk assessment. J. Official Statist. 17 (4), 499–520.
Huillet, T., Möhle, M., 2023. On Bernoulli trials with unequal harmonic success probabilities. Metrika http://dx.doi.org/10.1007/s00184-023-00913-5.
Jayakumar, K., Mundassery, D.A., 2007. On bivariate geometric distribution. Statistica 67 (4), 389–404.
Kimpton, L., Challenor, P., Wynn, H., 2022. Modelling correlated Bernoulli data part I: Theory and run lengths. http://dx.doi.org/10.48550/arXiv.2211.16921.
Kozubowski, T.J., Podgórski, K., 2018. A generalized Sibuya distribution. Ann. Inst. Statist. Math. 70 (4), 855–887.
Lovison, G., 2006. A matrix-valued Bernoulli distribution. J. Multivariate Anal. 97, 1573–1585.
Marshall, A.W., Olkin, I., 1985. A family of bivariate distributions generated by the bivariate Bernoulli distribution. J. Amer. Statist. Assoc. 80 (390), 332–338.
Nair, U., Sankaran, P.G., Balakrishnan, N., 2018. Reliability Modelling and Analysis in Discrete Time. Academic Press.
Najnudel, J., Pitman, J., 2020. Feller coupling of cycles of permutations and Poisson spacings in inhomogeneous Bernoulli trials. Electron. Commun. Probab. 25, 73, 11pp.
Nakagawa, T., Osaki, S., 1975. The discrete Weibull distribution. IEEE Trans. Reliab. 24 (5), 300–301.
Omey, E., Minkova, L.D., 2014. Bivariate geometric distributions. C. R. Acad. Bulgare Sci. 67 (9), 1201–1210.
Phatak, A.G., Sreehari, M., 1981. Some characterizations of bivariate geometric distribution. J. Indian Statist. Assoc. 19, 141–146.
Pitman, J., Yor, M., 1997. The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. Ann. Probab. 25 (2), 855–900.
Rényi, A., 1976. On outstanding values of a sequence of observations. In: Selected Papers of A. Rényi. Vol. 3, Akadémiai Kiadó, Budapest, pp. 50–65.
Roy, D., 1993. Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. J. Multivariate Anal. 46 (2), 362–373.
Salvia, A.A., Bollinger, R.C., 1982. On discrete hazard functions. IEEE Trans. Reliab. R-31 (5), 458–459.
Shaked, M., Shanthikumar, J.G., Valdez-Torres, J.B., 1995. Discrete hazard rate functions. Comput. Oper. Res. 22 (4), 391–402.
Sibuya, M., 2014. Prediction in Ewens-Pitman sampling formula and random samples from number partitions. Ann. Inst. Statist. Math. 66 (5), 833–864.