Journal of Statistical Theory and Practice (2025) 19:39
https://doi.org/10.1007/s42519-025-00447-1

ORIGINAL ARTICLE

A new discrete exponential distribution: properties and applications

Alessandro Barbiero[1] · **Asmerilda Hitaj**[2]

Accepted: 6 April 2025

Abstract

In this work, we propose a novel discrete counterpart to the continuous exponential random variable. It is defined on $N_0 = \{0, 1, 2, \dots\}$ and is constructed to have a step-wise cumulative distribution function that minimizes the Cramér distance to the continuous cumulative distribution function of the exponential random variable. We show that its distribution is a particular case of the zero-modified geometric distribution. The probability mass function is analyzed in detail, and the characteristic function is derived, from which the moments of the distribution can be readily obtained. The failure rate function, the zero-modification index, Shannon's entropy, and the stress-strength reliability parameter are also derived and discussed. Parameter estimation is examined, by considering the maximum likelihood method, the method of moments, and the least-squares method. A two-parameter generalization is also introduced and investigated. A real data analysis is provided, where the proposed distribution is fitted to a data set and compared to a well-known counting distribution. Finally, an application of the proposed discrete model is presented, focusing on the determination of the distribution of a compound sum of i.i.d. continuous random variables, with a specific application to the insurance field.

Keywords Count distribution · Cramér distance · Discretization · Exponential distribution · Panjer's recursive formula

A. Barbiero and A. Hitaj have contributed equally to this work. The authors are members of GNAMPA-INdAM.

✉ Alessandro Barbiero
alessandro.barbiero@unimi.it

Asmerilda Hitaj
asmerilda.hitaj@uninsubria.it

1 Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan, MI, Italy

2 Department of Economics, Università degli Studi dell'Insubria, via Monte Generoso 71, 21100 Varese, VA, Italy

Published online: 27 May 2025

1 Introduction

Discretization of continuous variables has received significant attention in different fields, such as machine learning, reliability theory, and actuarial science (see, among others, [1]). Many statistical models and techniques exist that potentially involve continuous random variables (rvs), but whose algorithms cannot be applied unless the continuous features are first discretized. More generally, the discretization of continuous (random) quantities can help the researcher find an approximate solution to the problem at hand if it cannot be solved analytically. In such circumstances, thanks to the greatly increasing availability of computational power, one may also turn to Monte Carlo simulation. However, as highlighted by [2, 3], this tool is not exempt from flaws, as it inevitably involves a sampling error. Some simple statistical applications where the discretization approach is useful are discussed in [4], namely, the determination of the distribution of the sum or of the product of two independent rvs not belonging to the same parametric family. Moving to a real-life application, discretization is an important step in the numerical evaluation of the distribution of the total claims amount of non-life portfolios in the so-called collective risk model, since well-known results on recurrence relations for probabilities only apply to discrete distributions, whereas single claim sizes are typically modeled by continuous rvs [5]. In such cases, discretization represents a viable alternative to the inversion theorem or the convolution integral. Another important application is scenario generation in stochastic programming, where one deals with a decision problem that is subject to uncertainty, typically modeled by a continuous probability distribution. For general continuous distributions, the stochastic optimization model is not solvable in general; many solution methods rely on approximating the underlying continuous probability distribution by a probability measure based on a finite number of scenarios, also termed realizations or atoms [6, 7]. Therefore, a crucial task for the researcher is to find an appropriate discrete distribution generated from the underlying continuous model, which possibly preserves one or more important features of it (see, for example, [8]).

Let us consider an absolutely continuous rv Y with cumulative distribution function (cdf) $F(y) = \mathbb{P}(Y \le y)$ and probability density function (pdf) $f(y)$. Let us denote with X a discrete counterpart to Y. If the support of Y is the whole real line $\mathbb{R}$, then it is reasonable to assume that the support $\mathcal{S}_X$ of X is $\mathbb{Z}$; if the support of Y is the positive real line $\mathbb{R}^+$, then one can assume that $\mathcal{S}_X = \mathbb{N} = \{1, 2, \dots\}$ or $\mathcal{S}_X = \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Several proposals are available in the literature for constructing a discrete analog to a continuous probability distribution. The most popular method is based on the preservation of the survival function (sf) or, complementarily, of the cdf, at each integer value. Focusing on the sf, if Y is a continuous positive rv, the probability mass function (pmf) of the discrete analog is set equal to

$$p_i = F(i+1) - F(i), \tag{1}$$

for $i \in \mathbb{N}_0$; if Y is real-valued, the definition of p_i is still the same, with $i \in \mathbb{Z}$. This technique has the remarkable advantage that the resulting pmf always has a closed-form expression, if this occurs for $F(y)$, and that $\mathbb{P}(X \ge i) = \mathbb{P}(Y \ge i)$ (thus the value of the sf is preserved at each integer value); in fact,

$$\mathbb{P}(X \geq i) = \sum_{j \geq i} p_j = \sum_{j \geq i} [F(j+1) - F(j)] = 1 - F(i) = P(Y \geq i).$$

Examples of discrete analogs obtained using this procedure are the discrete Weibull [9], the discrete Burr and Pareto [10], and the discrete normal distribution [11].

Another technique consists of resembling the pdf of Y directly by setting the pmf of X equal, at each integer value, to the corresponding value of f, properly scaled by a normalizing constant:

$$p_i = f(i) / \sum_{j \in \mathcal{S}_X} f(j) \tag{2}$$

(providing that $f(i)$ is finite for all i). This technique, although intuitive, is not able, in general, to produce a closed-form expression for the p_i, since the infinite series sum, representing the normalizing constant, cannot always be calculated analytically, or it can be expressed just in terms of some special functions. A case where such a technique yields an analytic expression for the probabilities is the Laplace or double exponential distribution [12]. Only when one considers the exponential distribution, the two techniques (i.e., preservation of the sf and resembling of the pdf) provide the same discrete analog, which proves to be the geometric distribution. In fact, the resulting pmf of the discrete analogue of an exponential rv with parameter λ obtained by applying (1) is

$$p_i = F(i+1) - F(i) = e^{-\lambda i} \left(1 - e^{-\lambda}\right), i \in \mathbb{N}_0, \tag{3}$$

which defines a geometric distribution, supported over $\mathbb{N}_0$, with parameter $\pi = 1 - e^{-\lambda}$. By applying the second technique summarized by (2), we obtain

$$p_i = \frac{\lambda e^{-\lambda i}}{\sum_{j=0}^{\infty} \lambda e^{-\lambda j}} = \frac{\lambda e^{-\lambda i}}{\lambda / (1 - e^{-\lambda})} = e^{-\lambda i} \left(1 - e^{-\lambda}\right),$$

which coincides with (3). In this work, we present an alternative discrete exponential distribution, derived by minimizing the Cramér distance to its parent continuous distribution. A preliminary version of this methodology was outlined in [13] and subsequently discussed in [14].

The layout of the paper is the following: In the next section, we derive the discrete analogue and present its main properties, especially related to reliability concepts. Section 3 examines parameter estimation. Section 4 proposes and discusses a two-parameter generalization of the discrete exponential distribution and updates the estimation procedures. Section 5 fits the discrete exponential distribution to a real dataset. Section 6 applies the proposed distribution for providing an approximate solution to a well-known problem in the insurance field. The last section contains some final remarks.

2 Derivation of a new discrete exponential distribution

2.1 Minimization of Cramér distance between continuous and step-wise cdfs

If one has to compare two rvs, one of which is absolutely continuous and the other discrete, it can be convenient to use a discrepancy measure based on the two cdfs, which are continuous and piecewise constant, respectively. Letting Y be the continuous rv with cdf $F(y)$ and pdf $f(y)$, and X a discrete rv with cdf $G(x)$, a simple class of discrepancy measures can be set up as follows:

$$d(F, G) = \int_{\mathbb{R}} |F(x) - G(x)|^r w(x)\mathrm{d}x, \tag{4}$$

where $w(x)$ is a positive weight function and $r > 0$ a positive parameter. When $r = 2$, (4) originates several well-known statistical distances according to different choices for $w(x)$ (see, for example, [15]). In particular, $w(x) = 1$ leads to the Cramér distance.

Let us assume that the continuous rv Y is supported over $\mathbb{R}^+$ and its discrete analog X is supported over $\mathbb{N}_0$. Then, an optimal discrete counterpart to F can be defined as the cdf G, within the set of all the cdfs supported over $\mathbb{N}_0$, minimizing the selected distance [13]. Now, a step-wise cdf G is fully identified by the values $Q_i = G(i)$, $i \in \mathbb{N}_0$. Therefore, the distance (4) is actually an infinite sum of contributions; focusing on the Cramér distance:

$$d_C(F, G) = \sum_{i=0}^{\infty} \int_i^{i+1} |F(x) - Q_i|^2 \mathrm{d}x. \tag{5}$$

The discrete counterpart minimizing (5) is then obtained by computing and equating to zero its first-order derivative with respect to each Q_i, then solving for the optimal Q_i, and finally checking for the second-order condition. The first-order condition on the Q_i is $\int_i^{i+1}(F(x) - Q_i)\mathrm{d}x = 0$, which implies

$$Q_i^* = \int_i^{i+1} F(x)\mathrm{d}x, \quad i = 0, 1, 2, \ldots. \tag{6}$$

These Q_i^* actually represent an global minimum point; in fact, by computing the second-order derivative of (5) we obtain $\partial^2 d_C(F, G)/\partial Q_i^2 = 2 > 0$ for all $i = 0, 1, 2, \ldots$, $\partial^2 d_C(F, G)/\partial Q_i \partial Q_j = 0$ for all $i \neq j$. Clearly, if the primitive function of F can be written in a closed form, so can the Q_i^* in (6) and the corresponding pmf. It can be proved that $F(i) < Q_i^* < F(i + 1)$ for any non-negative integer i [13].

The optimal probability p_0^* equals Q_0^*, whereas the other probabilities are obtained by difference as $p_i^* = Q_i^* - Q_{i-1}^*$, $i = 1, 2, \ldots$.

We note that since for a continuous positive rv Y with a finite first moment we can write $\mathbb{E}(Y) = \int_0^{+\infty}(1 - F(y))\mathrm{d}y$, and similarly, for a non-negative count rv X with cdf G and finite first moment, the following identity holds: $\mathbb{E}(X) = \sum_{i=0}^{\infty}(1 - G(i))$.

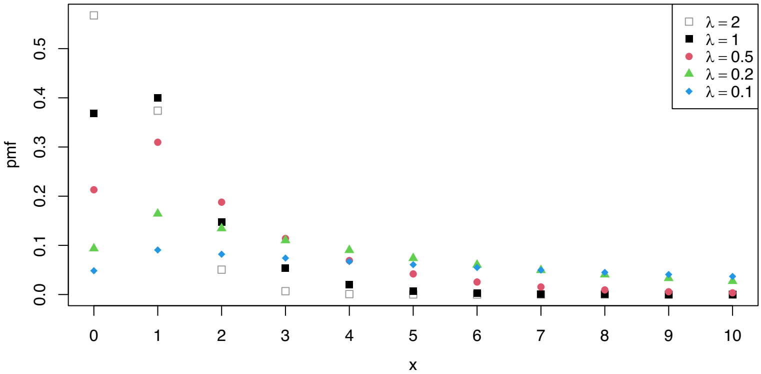


Fig. 1 Graph of the pmf (8) for $x = 0, 1, \ldots, 10$ and various values of λ

Then, for the discrete analog minimizing the Cramér distance, recalling (6), we derive that

$$\mathbb{E}(X)=\sum_{i=0}^{\infty}\left[1-\int_i^{i+1}F(y)\mathrm{d}y\right]=\sum_{i=0}^{\infty}\int_i^{i+1}(1-F(y))\mathrm{d}y=\int_0^{+\infty}(1-F(y))\mathrm{d}y=\mathbb{E}(Y);$$

the proposed discrete counterpart preserves the first raw moment of the parent rv.

2.2 A new discrete exponential distribution

The one-parameter exponential distribution is a very well-known continuous probability distribution; many popular stochastic models for lifetimes can be viewed as its extensions, which exploit the nice mathematical properties of the exponential function [16]. The cdf is $F(y) = 1 - e^{-\lambda y}$ for $y > 0$, where $\lambda > 0$ is the rate parameter; the pdf is $f(y) = \lambda e^{-\lambda y}$ for $y > 0$. The expectation is $1/\lambda$ and the variance is $1/\lambda^2$. For this rv, the first-order condition (6) becomes

$$Q_i^* = \int_i^{i+1}(1 - e^{-\lambda x})\mathrm{d}x = 1 - \frac{1}{\lambda}\left[e^{-\lambda i} - e^{-\lambda(i+1)}\right] = 1 - \frac{e^{-\lambda i}}{\lambda}(1 - e^{-\lambda}). \quad (7)$$

The probabilities of the optimal discrete counterpart can then be derived as

$$\begin{cases} p_0 &= 1 - \dfrac{1 - e^{-\lambda}}{\lambda} = \dfrac{\lambda - 1 + e^{-\lambda}}{\lambda} \\ p_i &= \dfrac{(1 - e^{-\lambda})^2 e^{-\lambda(i-1)}}{\lambda}, i = 1, 2, \ldots \end{cases} \quad (8)$$

Figure 1 displays the pmf (8) for different values of λ (0.1, 0.2, 0.5, 1, and 2).

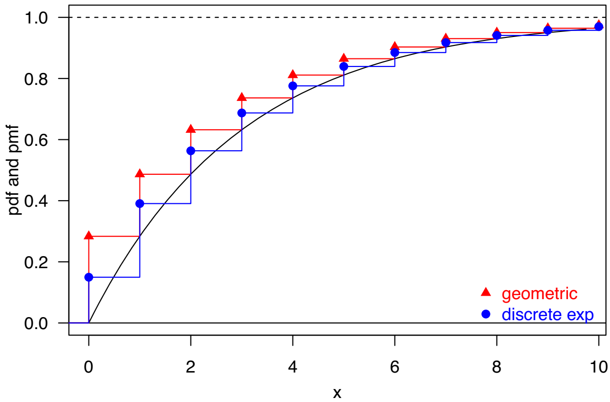


Fig. 2 Graph of the cdf of an exponential rv with parameter $\lambda = 1/3$ and graphs of the step-wise cdfs of a geometric rv with parameter $\pi = 1 - e^{-\lambda} \approx 0.28347$ and of a discrete exponential rv with parameter λ

After examining the expression of its pmf, one can note that the discrete exponential rv with parameter λ can be regarded as a finite mixture between a degenerate rv at 0, with weight $1 - (1 - e^{-\lambda})/\lambda$, and a geometric rv sitting on $\mathbb{N}$ with success parameter $1 - e^{-\lambda}$ and weight $(1 - e^{-\lambda})/\lambda$. It turns out that (8) is a particular case of the two-parameter modified geometric distribution, mentioned for example in [17], equation (2.11), with the parameter π equal to $(1 - e^{-\lambda})/\lambda$; or, equivalently, (8) is a particular case of a zero-modified geometric distribution [18, p.74] with $\alpha = \dfrac{\lambda - 1 + e^{-\lambda}}{\lambda}$. The pmf (8) can also be rewritten, following the reparametrization $\pi = 1 - e^{-\lambda}$, to which corresponds $\lambda = -\log(1 - \pi)$, as

$$\begin{cases} p_0 &= 1 + \dfrac{\pi}{\log(1-\pi)} \\ p_i &= -\dfrac{\pi^2(1-\pi)^{i-1}}{\log(1-\pi)} \end{cases} \tag{9}$$

Figure 2 displays the cdf of an exponential rv with rate parameter $\lambda = 1/3$; superimposed, the step-wise cdfs of a geometric distribution with parameter $\pi = 1 - e^{-\lambda}$ (the "standard" discrete analogue of the exponential obtained by applying (1)) and of a discrete exponential distribution with the same parameter λ. We notice that the step-wise cdf of the geometric is always greater than the cdf of the exponential rv (the former left-attains the latter at non-negative integers), whereas the step-wise cdf of the discrete exponential "interpolates" the continuous cdf, as we showed algebraically in the previous section.

Table 1 summarizes the main properties of the proposed discrete exponential distribution. Details about the algebraic derivations, as well as graphs of the relevant functions, are available in the Appendix A.

Table 1 Main properties of the discrete exponential distribution with pmf given by Eq. (8). See Appendix A for details

Property	Expression
Mode	0 if $\lambda > \lambda^*$; 1 if $\lambda < \lambda^*$; 0 and 1 if $\lambda = \lambda^*$ ($\lambda^* \approx 1.151387$)
Ch. function	$\phi_X(t) = \dfrac{\lambda - 1 + e^{-\lambda}}{\lambda} + \frac{(1-e^{-\lambda})^2 e^{\lambda}}{\lambda} \frac{e^{it-\lambda}}{1-e^{it-\lambda}}, i = \sqrt{-1}, t \in \mathbb{R}$
Expected value	$1/\lambda$
Variance	$\frac{e^{-\lambda}(\lambda+1)+\lambda-1}{\lambda^2(1-e^{-\lambda})}$
Quantile function	$\left\lceil -\frac{1}{\lambda} \log \frac{\lambda(1-u)}{1-e^{-\lambda}} \right\rceil, 0 < u < 1$
Failure rate function	$r_0 = p_0 = (\lambda - 1 + e^{-\lambda})/\lambda, r_i = 1 - e^{-\lambda}$ for $i = 1, 2, \dots$
ZMI	$1 + \lambda \log \frac{\lambda-1+e^{-\lambda}}{\lambda}$
Stress-strength reliability	$R = P(X < Y) = \frac{\lambda_1-1+e^{-\lambda_1}}{\lambda_1} + \frac{(1-e^{-\lambda_1})^2(1-e^{-\lambda_2})}{\lambda_1\lambda_2} \frac{1}{1-e^{-(\lambda_1+\lambda_2)}}$
Shannon entropy	$e^{-\lambda} + \log\lambda - \frac{\lambda-1+e^{-\lambda}}{\lambda} \log(\lambda - 1 + e^{-\lambda}) - \frac{2(1-e^{-\lambda})}{\lambda} \log(1 - e^{-\lambda})$
Right-tail deviation	$\frac{\sqrt{\lambda(1-e^{-\lambda})}}{1-e^{-\lambda/2}} - 1$

3 Parameter estimation

As for any parametric family of (discrete) distributions, there are several methods that can be employed for estimating the unknown parameter λ of the discrete exponential distribution, based on an iid sample $(x_1, x_2, \dots, x_n)$.

3.1 Method of moment

Since $\mathbb{E}(X) = 1/\lambda$, one can equate the sample mean $\bar{x}$ to $\mathbb{E}(X)$ and obtain the method of moment's estimate as $\hat{\lambda}_M = 1/\bar{x}$.

3.2 Maximum likelihood method

Recalling the expression of the pmf in (8) and denoting by n_i the sample frequency of the value i and by $x_{(n)}$ the largest observed value, the log-likelihood function takes on the following expression:

$$
\begin{aligned}
\ell(\lambda) &= n_0 \ln\left[\frac{\lambda - 1 + e^{-\lambda}}{\lambda}\right] + \sum_{i=1}^{x_{(n)}} n_i \left[-\ln\lambda + 2\ln(1 - e^{-\lambda}) - \lambda(i-1)\right] \\
&= n_0 \ln\left[\frac{\lambda - 1 + e^{-\lambda}}{\lambda}\right] + (n - n_0)\left[-\ln\lambda + 2\ln(1 - e^{-\lambda}) + \lambda\right] - n\bar{x}\lambda
\end{aligned}
$$

and the first-order derivative is:

$$\begin{aligned}\ell'(\lambda) &= n_0 \frac{1-e^{-\lambda}}{\lambda-1+e^{-\lambda}} - \frac{n_0}{\lambda} + (n-n_0)\left[-\frac{1}{\lambda} + \frac{2e^{-\lambda}}{1-e^{-\lambda}} + 1\right] - n\bar{x} \\ &= n_0 \frac{1-e^{-\lambda}}{\lambda-1+e^{-\lambda}} - \frac{n}{\lambda} + 2(n-n_0)\frac{e^{-\lambda}}{1-e^{-\lambda}} + n - n_0 - n\bar{x}.\end{aligned}$$

One can find the MLE of λ numerically by either directly maximizing $\ell(\lambda)$ or solving the equation $\ell'(\lambda) = 0$. Since $\ell'(\lambda)$ is continuous and $\lim_{\lambda\to 0^+} \ell'(\lambda) = +\infty$ and $\lim_{\lambda\to+\infty} \ell'(\lambda) = n(1-\bar{x}) - n_0 < 0$, if $n_0 + n_1 < n$, or $\lim_\lambda(\lambda) = 0^-$, if $n_0 + n_1 = n$, there exists (at least) one positive value of λ where $\ell'(\lambda)$ is zero (provided that $n_0 \neq n$). Now, the second order derivative of the log-likelihood function is

$$\ell''(\lambda) = n_0 \frac{\lambda e^{-\lambda} + e^{-\lambda} - 1}{(\lambda - 1 + e^{-\lambda})^2} + \frac{n}{\lambda^2} - 2(n-n_0)\frac{e^{-\lambda}}{(1-e^{-\lambda})^2}.$$

It can be proven that $\lim_{\lambda\to 0^+} \ell''(\lambda) = -\infty$ whereas $\lim_{\lambda\to+\infty} \ell''(\lambda) = 0^+$, since for sufficiently large λ, the term n/λ^2 in the expression of ℓ'' tends to dominate the two other terms, which are both negative. Then, since ℓ'' is first negative and then positive, we deduce that ℓ' is first decreasing and then increasing, and for what we said before, we conclude that the log-likelihood function it admits a unique global maximum, and then a unique MLE of λ, $\hat{\lambda}_{ML}$, exists.

The Fisher Information for the iid sample can be then computed as $I_n(\lambda) = -\mathbb{E}(\ell''(\lambda))$. Since $\mathbb{E}(n_0) = np_0$, we can write

$$\begin{aligned}I_n(\lambda) &= -n\frac{\lambda - 1 + e^{-\lambda}}{\lambda}\frac{\lambda e^{-\lambda} + e^{-\lambda} - 1}{(\lambda - 1 + e^{-\lambda})^2} - \frac{n}{\lambda^2} + 2n\frac{1-e^{-\lambda}}{\lambda}\frac{e^{-\lambda}}{(1-e^{-\lambda})^2} \\ &= n\frac{e^{-\lambda}(\lambda^2 - 3\lambda - 2) + e^{-2\lambda}(\lambda^2 + 3\lambda + 1) + 1}{\lambda^2(\lambda - 1 + e^{-\lambda})(1-e^{-\lambda})}\end{aligned}$$

It is quite easy to check that $I_n(\lambda) \approx n/\lambda^2$ if $\lambda \ll 1$. For any unbiased estimator $\hat{\lambda}$ of λ, by applying the Cramér-Rao inequality we have that $\mathrm{Var}(\hat{\lambda}) \geq 1/I_n(\lambda)$. The MLE of λ is asymptotically normal: for n sufficiently large, $\hat{\lambda}_{ML} \overset{\text{approx}}{\sim} \mathcal{N}(\lambda_0, 1/I_n(\lambda_0))$, where $\lambda_0 > 0$ is the true value of the parameter λ.

3.3 Least-squares method

For the proposed model, the survival function $S(i) = P(X \geq i)$ can be written as $S(i) = e^{-\lambda(i-1)}(1-e^{-\lambda})/\lambda$, for $i \geq 1$, or

$$\ln S(i) = -\lambda i + \lambda + \log(1-e^{-\lambda}) - \log\lambda$$

By letting $y = \ln S(i)$, $x = i$, $b = -\lambda$ and $a = \lambda + \log(1-e^{-\lambda}) - \log\lambda$, the above equation can be read as a simple linear equation, $y = a + bx$. The unknown parameter

λ can be then estimated based on the observed sample $(x_1, x_2, \ldots, x_n)$, considering only the observations where $x_j > 0$, by substituting the unknown values $S(i)$ with the sample version $\hat{S}(i) = \sum_{j=1}^{n} \mathbb{1}(x_j \geq i)/n$, and then computing $\hat{b}$, the ordinary least-squares estimate of the coefficient b for the model above. Then the estimate of λ is simply $\hat{\lambda}_{LS} = -\hat{b}$.

Another estimation method, the method of proportion, is described in Appendix B.

4 Generalization

The one-parameter model we introduced is quite easy to handle, yet it does not allow a great level of flexibility. To widen its scope, one can think of extending it; to do this, at least two ways are available. The first one consists of considering the generalized exponential distribution, as defined in [19], and then applying the minimization criterion resulting in (7) to it. The second one consists in directly generalizing the discrete model in (8) by adding one parameter.

The first way is problematic; in fact, the cdf of the generalized exponential distribution is $F(x) = (1 - e^{-\lambda x})^{\alpha}$, with $\alpha > 0$ being the additional shape parameter. If we want to apply the proposed criterion, then we obtain the optimal Q_i^* as $Q_i^* = \int_i^{i+1} (1 - e^{-\lambda x})^{\alpha} \mathrm{d}x$, where the integral cannot be written down in a simple, closed form.

If one pursues the second strategy, a generalized discrete exponential distribution with parameters λ and $\alpha > 0$ can be defined via its cdf, whose expression is

$$F(i; \lambda, \alpha) = \left[1 - \frac{e^{-\lambda i}}{\lambda}(1 - e^{-\lambda})\right]^{\alpha}, i = 0, 1, 2, \ldots; \tag{10}$$

the probabilities are obtained as usual as $p_0 = F(0; \lambda, \alpha)$ and $p_i = F(i; \lambda, \alpha) - F(i-1; \lambda, \alpha)$, $i = 1, 2, \ldots$. Needless to say that despite the closed-form expression of the pmf, the extended model is less manageable than the one-parameter model - this is the cost one pays for adding flexibility. What can be still handled with relative simplicity are the quantile function and parameter estimation, which we will describe in the following.

Since $F(x; \lambda, \alpha_1) > F(x; \lambda, \alpha_2)$ for all x if $\alpha_1 < \alpha_2$, the effect of the additional parameter α is clear: increasing α, with λ fixed, pushes the random distribution towards larger values, as can be seen by considering Figure 3 where for each λ in $\{0.2, 0.5, 1\}$ and for each α in $\{0.5, 0.75, 1.25, 2\}$ the pmf is displayed (limited to the first non-negative integers).

Checking (10), a result is readily obtained. Assume that $\{X_i\}$, $i = 1, 2, \ldots, n$, are independent rvs following the same discrete exponential distribution with parameter λ; then, the cdf of $X_{n:n} = \max\{X_1, \ldots, X_n\}$ is

$$P(X_{n:n} \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \left[1 - \frac{e^{-\lambda x}}{\lambda}(1 - e^{-\lambda})\right]^{n},$$

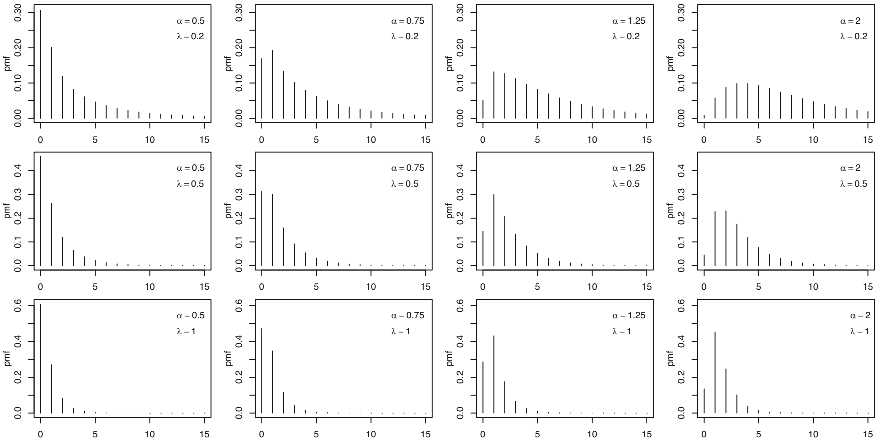


Fig. 3 Graphs of the pmf of the generalized discrete exponential distribution for different value combinations of its two parameters λ and α. Note that the scale for the y-axis is the same for the graphs on the same row, but varies across the rows

which corresponds to the cdf of the generalized discrete exponential distribution with parameters λ and n.

From (10), it is also easy to derive the quantile of level γ:

$$x_\gamma(\lambda, \alpha) = \left\lceil -\frac{1}{\lambda} \log \frac{\lambda(1-\gamma^{1/\alpha})}{1-e^{-\lambda}} \right\rceil.$$

Expectation and variance cannot be determined as easily as for the one-parameter discrete exponential distribution. However, one can use the practical formulas reported in [20] and write

$$\mathbb{E}(X; \lambda, \alpha) = \sum_{i=0}^{\infty} (1 - F(i; \lambda, \alpha)) \tag{11a}$$

$$\mathbb{E}(X^2; \lambda, \alpha) = \sum_{i=0}^{\infty} (2i+1)\,(1 - F(i; \lambda, \alpha))\,. \tag{11b}$$

The infinite series sum in (11a) and (11b) can be truncated to some appropriate upper bound; one can choose a priori a high-order quantile, or determine it a posteriori as the limit beyond which the sum converges reliably and stably. If λ is small, the infinite series sums above can be approximated by the following integrals, where the cumulative probabilities are substituted by $(1-e^{-\lambda x})^\alpha$, the cdf of the generalized exponential rv with parameters λ and α:

$$\mathbb{E}(X; \lambda, \alpha) \approx \int_0^\infty \left[1 - (1-e^{-\lambda x})^\alpha\right] \mathrm{d}x$$

$$\mathbb{E}(X^2; \lambda, \alpha) \approx \int_0^\infty \left[2x(1 - (1-e^{-\lambda x})^\alpha)\right] \mathrm{d}x$$

These integrals can be numerically computed using, for example, the `integrate` function in the `R` programming environment [21].

4.1 Estimation

The method of moments is no longer straightforward since there are no closed-form expressions for the first two moments. Hence, a numerical root-search or a numerical minimization must be implemented for finding the couple $(\hat{\lambda}_M, \hat{\alpha}_M)$ ensuring the matching of theoretical and sample moments. One can consider the loss-type function

$$\mathcal{L}(\lambda, \alpha; x_1, \ldots, x_n) = [\mathbb{E}(X; \lambda, \alpha) - \bar{x}]^2 + [\mathbb{E}(X^2; \lambda, \alpha) - \hat{\mu}_2]^2,$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $\hat{\mu}_2 = \sum_{i=1}^{n} x_i^2/n$, and find its global minimum point $(\hat{\lambda}_M, \hat{\alpha}_M)$, where the function is expected to equal zero, numerically.

The maximum likelihood method increases in analytic complexity; however, it is quite easy to derive the maximizer $(\hat{\lambda}, \hat{\alpha})$ of the log-likelihood function numerically, for example using the `mle2` function in the `bmle` package of `R`. Along with the point estimates, `mle2` is able to calculate a $(1-\alpha) \cdot 100\%$ confidence interval for λ and α separately, based on inverting a spline fit to the likelihood profile (the default). Appendix B describes the method of proportion for the generalized model.

5 Data analysis

We consider the data set summarized in Table 5.3 in [22], which reports the observed frequencies of home injuries of 122 experienced men during the period 1937-1947. For these data, the mean is about 0.984, the variance 1.557, the skewness 1.347, and the kurtosis 4.120. We first fit the one-parameter exponential distribution to the data, followed by the generalized discrete exponential distribution. The fitted frequencies obtained by plugging in the MLEs of λ (and α) are displayed in Table 2, where summary goodness-of-fit indexes are also displayed along with the MLEs. It is evident that the one-parameter model does not provide an adequate fit for the data: simply comparing the observed and theoretical frequencies for values 0 and 1 reveals notable discrepancies. When moving to the generalized model, we observe a significantly improved fit: the theoretical and observed frequencies differ only slightly. This qualitative assessment is confirmed by numerical goodness-of-fit measures. The AIC for the generalized discrete exponential distribution is 339.35; the chi-squared test statistic (after aggregating the last two counts) is 0.42, and the corresponding p-value is 0.809, so quite close to 1. We also fit another very common discrete distribution, the negative binomial (see the last column of Table 2). It turns out to provide a slightly better, yet comparable, fit than the generalized discrete exponential distribution in terms of AIC, which equals 338.36, whereas the chi-squared statistic suggests a slightly better fit for the proposed discrete distribution.

Figure 4 displays in the same graph the empirical cdf and the estimated cdf for the data set, fitted by the generalized discrete exponential distribution.

Table 2 Home injuries of 122 experienced men during the period 1937-1947 (from Table 5.3 in [22])

x_i	n_i	D.Exp	D.Gen.Exp.	Neg.Bin
0	58	43.78	57.42	57.54
1	34	48.45	36.19	33.78
2	14	18.44	14.79	16.73
3	8	7.02	6.91	7.77
4	6	2.67	3.36	3.49
(≥)5	2	1.64	3.32	2.70
total	122	122	122	122
$\hat{\lambda}$		0.966(0.092)	0.689(0.114)	-
$\hat{\alpha}$		-	0.588(0.103)	-
$\hat{n}$		-	-	1.454(0.561)
$\hat{p}$		-	-	0.597(0.097)
$\ell_{\max}$		-172.64	-167.67	-167.18
AIC		347.29	339.35	338.36
X^2 (p-value)		10.80(0.013)	0.42(0.809)	0.46(0.794)

The third column displays the theoretical frequencies for the discrete exponential distribution (D.Exp); the fourth column for the generalized discrete exponential distribution (D.Gen.Exp); the last column for the negative binomial distribution (Neg.Bin). The MLEs estimates of the parameter(s) of the discrete distributions are also reported along with their standard errors and some goodness-of-fit measures

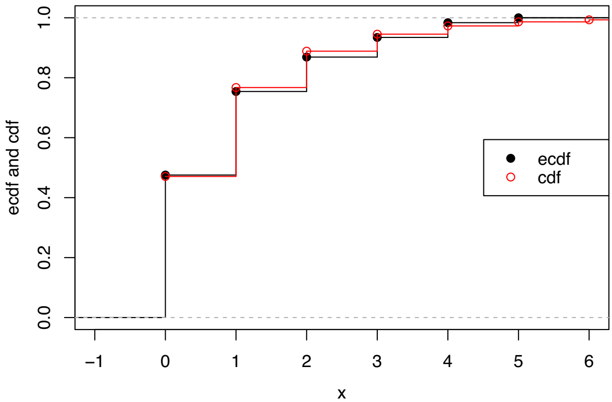


Fig. 4 Empirical and estimated cdfs for the home injures data set fitted by the generalized discrete exponential distribution

6 Application

In situations where it is much easier to handle a discrete rather than a continuous rv, one can consider replacing the latter with a discrete approximation or counterpart thereof. In the insurance sector, the collective risk model is used to construct aggregate losses from a claim count distribution and a claim size distribution. Consider the aggregate loss $S_N = \sum_{i=1}^{N} X_i$, where the X_i are iid rvs modeling the N claim amounts occurring over a fixed time window, with N being a count rv independent from the X_i. The exact distribution of S_N, which is a compound sum, is generally difficult to derive analytically, since it involves integrals that can be solved only numerically. One can resort to Monte Carlo simulation, but simulating individual claims can be a lengthy process when the expected number of claims is large. Alternatively, one may resort to approximation via the normal or three-parameter Gamma distribution, etc., but the approximation may be not satisfactory if the claim size distribution is heavy-tailed. Alternatively, if N belongs to the $(a, b, 0)$ class of discrete distributions (which comprises the Poisson, the binomial, and the negative binomial), one can think of discretizing the X_i (which are typically continuous) into, say, $\tilde{X}_i$, and calculate the (approximate) distribution of S_N, say $\tilde{S}_N$, by resorting to Panjer recursive formula:

$$g_x = P(\tilde{S}_N = x) = \frac{1}{1 - ap_0} \sum_{k=1}^{x} \left(a + \frac{bk}{x}\right) p_k g_{x-k}, x = 1, 2, \ldots, \tag{12}$$

where a and b are the parameters of the $(a, b, 0)$ class and p_k is the pmf of the discrete $\tilde{X}$ [23]. We recall that the pmf of a distribution belonging to the $(a, b, 0)$ class satisfies

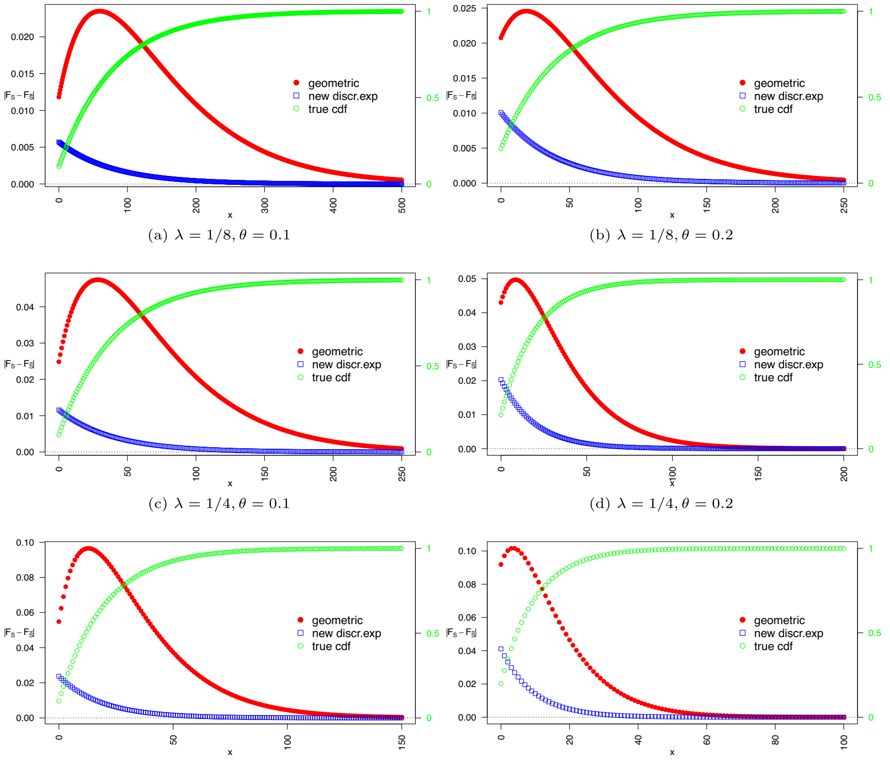
$$f_k = \left(a + \frac{b}{k}\right) f_{k-1}, k = 1, 2, \ldots$$

The starting value for the evaluation of the approximate distribution of S_N through (12) is g_0, whose value is equal to $g_0 = f_0 + \sum_{j=1}^{\infty} f_j p_0^j$.

In our application, we examine the combination of $X_i \sim \text{Exp}(\lambda)$ with $N \sim \text{Geom}(\theta)$. In this case, the distribution of S_N can be determined exactly: S_N is a mixed-type rv, with a discrete mass at 0 and a still exponentially distributed continuous component:

$$P(S_N \leq x) = \begin{cases} 0 & x < 0 \\ \theta + (1-\theta)(1 - e^{-\theta\lambda x}) & x \geq 0. \end{cases}$$

Therefore, for this particular case, there is no need of finding an approximation since the exact distribution is readily available in a closed-form. Nevertheless, we performed discretization followed by application of Panjer's formula in order to evaluate its performance. We considered the two discretizations of an exponential rv X discussed in this work and determined the pmf and the cdf of $\tilde{S}_N$, obtained as $\tilde{S}_N = \sum_{i=1}^{N} \tilde{X}_i$, using (12), with $a = 1 - \theta$ and $b = 0$ (the class parameters corresponding to the geometric distribution). We used $\lambda \in \{1/8, 1/4, 1/2\}$ and $\theta \in \{0.1, 0.2\}$. Figure 5, where the absolute error between the true and the approximated cdfs are compared for all the six artificial scenarios, show that the proposed discretization (the discrete exponential distribution) performs better than the existent one (the geometric). In fact, for each scenario, the former, plugged in Panjer recursive formula, leads to absolute errors uniformly smaller than the geometric, i.e., for any x, the absolute error induced by the new discrete exponential distribution is dominated by the absolute error induced by the geometric. In particular, the absolute error for the new discrete exponential



(a) $\lambda = 1/8, \theta = 0.1$ (b) $\lambda = 1/8, \theta = 0.2$

(c) $\lambda = 1/4, \theta = 0.1$ (d) $\lambda = 1/4, \theta = 0.2$

(e) $\lambda = 1/2, \theta = 0.1$ (f) $\lambda = 1/2, \theta = 0.2$

Fig. 5 Comparison of the approximation provided by the geometric distribution and the proposed discrete exponential distribution in the assessment of the true cdf of a geometric compound sum with exponentially-distributed individual claims. We reported the absolute errors (left y-axis) and the values of the true cdf (right y-axis). Note that the scale of the x-axis and y-axis is not the same across the panels

is strictly monotonically decreasing to zero and tends to zero much faster than the geometric as x goes to ∞. This confirms that the discrete exponential distribution introduced in this work is a valid substitute or proxy of the continuous exponential distribution.

We note that the approximation improves as λ decreases for a fixed θ, which is expected: for small values of λ, the discrete analogues more closely resemble the parent exponential random variable. Additionally, decreasing the parameter θ while keeping λ fixed in the geometric distribution of N also improves the approximation (i.e., reduces absolute errors), particularly for the one resulting from the proposed discretization.

7 Final remarks

We derived a discrete exponential distribution supported on the set of non-negative integers, obtained by minimizing the Cramér distance from the continuous exponential distribution. The random distribution has a simple closed-form expression for the probabilities and other relevant functions, and its pseudo-random simulation is straightforward. A two-parameter generalization is also explored, which offers greater flexibility for fitting real data. The distance-based derivation applied to the exponential model can also be extended to other continuous random distributions, allowing for the construction of discrete analogs that may serve as approximations of the original models.

Appendix A Properties of the discrete exponential distribution

A.1 Mode(s)

From (8), it is easy to see that $p_i > p_{i+1}$ for all $i = 1, 2, \ldots$. Moreover, it can be numerically proven that $p_0 > p_1$ if and only if $\lambda > \lambda^* \approx 1.151387$, whereas $p_0 < p_1$ if and only if $\lambda < \lambda^*$. Therefore, for $\lambda > \lambda^*$ the distribution has a unique mode at 0 and has a decreasing pmf; for $\lambda < \lambda^*$ the distribution has a unique mode at 1; and if $\lambda = \lambda^*$ the distribution has modes at 0 and 1.

A.2 Characteristic function and moments

The expression of the characteristic function for the proposed discrete exponential distribution can be obtained by applying its definition for discrete rvs, $\phi_X(t) = \sum_{j=0}^{\infty} p_j e^{itj}$, where i is now the imaginary unit: $i = \sqrt{-1}$. Then, after some steps, we derive:

$$\begin{aligned}\phi_X(t) &= p_0 + \sum_{j=1}^{\infty} \frac{(1-e^{-\lambda})^2 e^{-\lambda(j-1)}}{\lambda} e^{itj} = p_0 + \frac{(1-e^{-\lambda})^2 e^{\lambda}}{\lambda} \sum_{j=1}^{\infty} e^{(it-\lambda)j} \\ &= \frac{\lambda - 1 + e^{-\lambda}}{\lambda} + \frac{(1-e^{-\lambda})^2 e^{\lambda}}{\lambda} \frac{e^{it-\lambda}}{1-e^{it-\lambda}}.\end{aligned}$$

If we compute the first-order derivative of $\phi_X(t)$ with respect to t, we obtain

$$\phi'_X(t) = \frac{i e^{it-\lambda}}{(1-e^{it-\lambda})^2} \frac{(1-e^{-\lambda})^2 e^{\lambda}}{\lambda},$$

which, evaluated at zero, provides

$$\phi'_X(0) = \frac{i}{\lambda} = i\mathbb{E}(X);$$

therefore, the expectation is $1/\lambda$: this result confirms what we mentioned in Section 2.1. The second-order derivative of the cf is

$$\phi''_X(t) = -\frac{e^{\lambda}(1-e^{-\lambda})^2}{\lambda} \frac{e^{it-\lambda}(1+e^{it-\lambda})}{(1-e^{it-\lambda})^3},$$

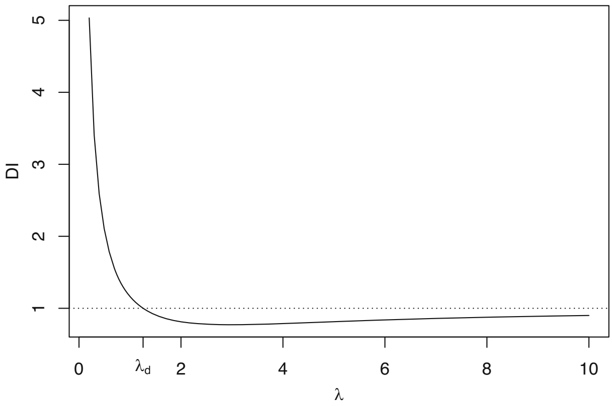


Fig. 6 Graph of the DI for the discrete exponential distribution as a function of λ

from which the second moment can be easily obtained:

$$\mathbb{E}(X^2) = \phi''_X(0)/i^2 = \frac{1+e^{-\lambda}}{\lambda(1-e^{-\lambda})};$$

the variance is therefore $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{e^{-\lambda}(\lambda+1)+\lambda-1}{\lambda^2(1-e^{-\lambda})}$, and the dispersion index $\mathrm{DI} = \frac{e^{-\lambda}(\lambda+1)+\lambda-1}{\lambda(1-e^{-\lambda})}$. The plot of DI is reported in Figure 6. For $\lambda < \lambda_d = 1.256438$, DI is larger than 1 (and as λ goes to 0, it tends monotonically to ∞) thus indicating overdispersion; for $\lambda > \lambda_d$, DI is between 0 and 1, thus indicating underdispersion, and tends asymptotically to 1 as λ goes to ∞. This is a notable difference with respect to the geometric distribution, which is overdispersed ($\mathrm{DI} = 1/\pi$) for any value of its parameter $\pi \in (0, 1)$.

The third-order derivative of the cf is

$$\phi'''_X(t) = -\frac{e^{\lambda}(1-e^{-\lambda})^2}{\lambda}\frac{ie^{it-\lambda}(4e^{it-\lambda}+e^{2(it-\lambda)}+1)}{(1-e^{it-\lambda})^4},$$

therefore the third moment is equal to

$$\mathbb{E}(X^3) = \phi'''(0)/i^3 = \frac{4e^{-\lambda}+e^{-2\lambda}+1}{\lambda(1-e^{-\lambda})^2}.$$

The fourth-order derivative of the cf is

$$\phi^{(4)}_X(t) = \frac{e^{\lambda}(1-e^{-\lambda})^2}{\lambda}\frac{e^{it-\lambda}(11e^{it-\lambda}+11e^{2(it-\lambda)}+e^{3(it-\lambda)}+1)}{(1-e^{it-\lambda})^5},$$

and then the fourth moment is

$$\mathbb{E}(X^4) = \frac{11e^{-\lambda}+11e^{-2\lambda}+e^{-3\lambda}+1}{\lambda(1-e^{-\lambda})^3}.$$

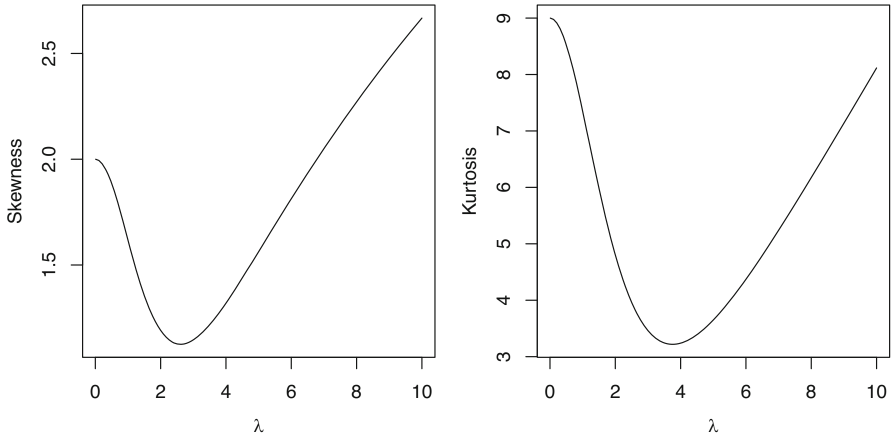


Fig. 7 Graph of skewness and kurtosis of the discrete exponential distribution as functions of λ

From the results above, it is easy, though algebraically cumbersome, to find the expressions of the usual measures of skewness and kurtosis. The graphs of skewness and kurtosis as functions of λ are displayed in Figure 7. From them, one can see that for any value of the parameter λ, the discrete model is positively skewed and leptokurtic. We notice that when λ tends to zero, the values of skewness and kurtosis converge to the corresponding values of the exponential distribution (2 and 9, respectively).

A.3 Quantile function

From the espression of the cdf (7), it is easy to derive that the quantile of level u, $0 < u < 1$, is given by

$$x_u(\lambda) = \left\lceil -\frac{1}{\lambda} \log \frac{\lambda(1-u)}{1-e^{-\lambda}} \right\rceil,$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Based on this expression, it is then straightforward to generate a random number from the proposed distribution with assigned parameter λ. It is sufficient to generate a random number from a standard uniform distribution, say u, and compute the quantile $x_u(\lambda)$.

A.4 Failure rate function

For a count rv X, a naïve failure rate function can be defined as

$$r_i = p_i / P(X \geq i) = p_i / (1 - F(i-1)).$$

For the proposed distribution, since for $i \geq 1$, $1 - F(i-1) = (1 - e^{-\lambda})^2 e^{-\lambda(i-1)}/\lambda$, it can be shown that $r_0 = p_0 = (\lambda - 1 + e^{-\lambda})/\lambda$ and r_i is constant and equal to

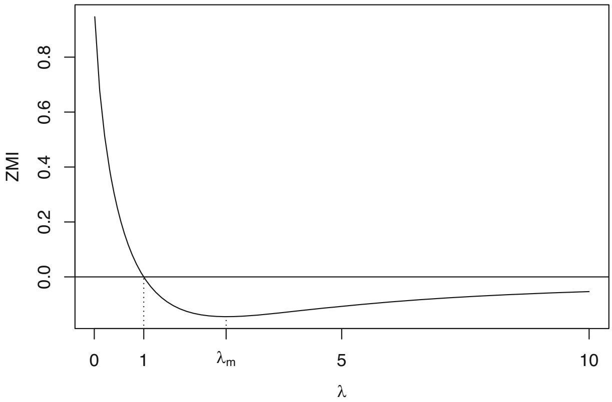


Fig. 8 Graph of the ZMI for the discrete exponential distribution

$1 - e^{-\lambda} > r_0$ for $i \geq 1$. So the property of constant failure rate holding for the exponential (and also for the geometric) distribution is somewhat kept or, better, is kept for all the positive values of the support. This was expected, as we observed in Section 2.2 that this rv can be viewed as a mixture of a degenerate rv at 0 and a geometric rv on $\mathbb{N}$.

A.5 ZMI

The Zero Modification Index (ZMI) for a discrete random distribution is defined as $\text{ZMI} = 1 + \log p_0/\mathbb{E}(X)$. According to whether $\text{ZMI} > 0$ or $\text{ZMI} < 0$, we say that the distribution is zero-inflated or zero-deflated. For the proposed distribution, its expression is equal to

$$\text{ZMI} = 1 + \lambda \log \frac{\lambda - 1 + e^{-\lambda}}{\lambda}.$$

ZMI is a decreasing-increasing function with λ, satisfying $\lim_{\lambda \to 0^+} \text{ZMI} = 1$, taking on the zero value at $\lambda = 1$, and asymptotically tending to 0 as λ goes to ∞. Its minimum value is approximately -0.1444245, which is attained at $\lambda_m \approx 2.664476$. The graph of ZMI is displayed in Figure 8.

We remark that the expression for the ZMI of the geometric distribution with parameter $\pi = 1 - e^{-\lambda}$ is $1 + \frac{\pi}{1-\pi} \log \pi$ or, equivalently, $1 + \frac{1-e^{-\lambda}}{e^{-\lambda}} \log(1 - e^{-\lambda})$, which is strictly decreasing with π (or λ) and takes values in the unit interval.

A.6 Stress-strength reliability

In a stress-strength model, we consider a mechanical or electrical component that exhibits an intrinsic strength Y and is subjected to an external stress X. The component works as long as X does not exceed Y. X and Y are assumed to be rvs and the probability that the component works is given by the reliability parameter $R = P(X \leq Y)$. Typically, X and Y are assumed to be independent and are often modeled by continuous rvs, although discrete distributions can sometimes be employed (see [24] for geometric, [25] for Poisson, [26] for negative binomial stress and strength, [27] for geometric stress and Poisson strength). If we assume that X and Y are independent discrete exponential rvs with parameter λ_1 and λ_2, respectively, then R takes on the following expression

$$\begin{aligned}
R &= \sum_{x=0}^{\infty}\sum_{y=x}^{\infty} p_1(x;\lambda_1)\cdot p_2(y;\lambda_2) = \sum_{x=0}^{\infty} p_1(x;\lambda_1)\sum_{y=x}^{\infty} p_2(y;\lambda_2)\\
&= \sum_{x=0}^{\infty} p_1(x;\lambda_1)[1-F_2(x-1;\lambda_2)]\\
&= \frac{\lambda_1-1+e^{-\lambda_1}}{\lambda_1} + \sum_{x=1}^{\infty}\frac{(1-e^{-\lambda_1})^2 e^{-\lambda_1(x-1)}}{\lambda_1}\frac{e^{-\lambda_2(x-1)}}{\lambda_2}(1-e^{-\lambda_2})\\
&= \frac{\lambda_1-1+e^{-\lambda_1}}{\lambda_1} + \frac{(1-e^{-\lambda_1})^2(1-e^{-\lambda_2})}{\lambda_1\lambda_2}\frac{1}{1-e^{-(\lambda_1+\lambda_2)}}.
\end{aligned}$$

If X and Y are identically distributed with a common parameter $\lambda = \lambda_1 = \lambda_2$, then

$$R = \frac{\lambda-1+e^{-\lambda}}{\lambda} + \frac{(1-e^{-\lambda})^3}{\lambda^2(1-e^{-2\lambda})} = 1 - \frac{1-e^{-\lambda}}{\lambda} + \frac{(1-e^{-\lambda})^2}{\lambda^2}\cdot\frac{1}{1+e^{-\lambda}},$$

which is clearly greater than 0.5 for any $\lambda > 0$.

A.7 Shannon entropy

Shannon entropy is a measure of uncertainty of a rv; for a discrete rv X with pmf p_x, it is defined as $H(X) = -\sum p_x \log p_x$ [28] and for the proposed distribution we can then write $H(X) = -p_0 \log p_0 - \sum_{i=1}^{\infty} p_i \log p_i$, and then, by substituting the expression in (8) into the formula above,

$$\begin{aligned}
H(X;\lambda) &= -\frac{\lambda-1+e^{-\lambda}}{\lambda}\log\frac{\lambda-1+e^{-\lambda}}{\lambda}\\
&\quad - \frac{(1-e^{-\lambda})^2}{\lambda}\sum_{x=1}^{\infty} e^{-\lambda(i-1)}\left[-\lambda(i-1) + \log\frac{(1-e^{-\lambda})^2}{\lambda}\right]\\
&= -\frac{\lambda-1+e^{-\lambda}}{\lambda}\log\frac{\lambda-1+e^{-\lambda}}{\lambda}\\
&\quad - \frac{(1-e^{-\lambda})^2}{\lambda}\left[-\frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})^2} + \frac{1}{1-e^{-\lambda}}\log\frac{(1-e^{-\lambda})^2}{\lambda}\right]\\
&= -\frac{\lambda-1+e^{-\lambda}}{\lambda}\log\frac{\lambda-1+e^{-\lambda}}{\lambda} + e^{-\lambda} - \frac{1-e^{-\lambda}}{\lambda}\log\frac{(1-e^{-\lambda})^2}{\lambda}\\
&= e^{-\lambda} + \log\lambda - \frac{\lambda-1+e^{-\lambda}}{\lambda}\log(\lambda-1+e^{-\lambda}) - \frac{2(1-e^{-\lambda})}{\lambda}\log(1-e^{-\lambda})
\end{aligned}$$

By graphing Shannon entropy as a function of the parameter λ, one notices that it is strictly decreasing with λ, and that $\lim_{\lambda\to 0^+} H(X;\lambda) = +\infty$ and $\lim_{\lambda\to+\infty} H(X;\lambda) = 0$.

We recall that for a geometric rv with parameter $\pi = 1 - e^{-\lambda}$, Shannon entropy takes on the expression

$$H(X;\pi) = \frac{-(1-\pi)\log(1-\pi) - \pi\log\pi}{\pi}.$$

A.8 Right-tail deviation

To capture the tail behaviour of a non-negative rv X with survival function $S(x) = P(X \geq x)$, [29] proposed a tail index, called right-tail deviation, defined as

$$T(X) = \frac{\sum_{i=1}^{\infty}\sqrt{S(i)}}{\mathbb{E}(X)} - 1.$$

For the discrete exponential rv, its expression turns out to be

$$T(X;\lambda) = \frac{\sum_{i=1}^{\infty}\sqrt{e^{-\lambda(i-1)}(1-e^{-\lambda})/\lambda}}{1/\lambda} - 1 = \frac{\sqrt{\lambda(1-e^{-\lambda})}}{1-e^{-\lambda/2}} - 1$$

$T(X;\lambda)$ is strictly increasing with λ; $\lim_{\lambda\to 0^+} T(X;\lambda) = 1$ and $\lim_{\lambda\to+\infty} T(X;\lambda) = +\infty$.

Appendix B Parameter estimation

B.1 Method of proportion

For the discrete exponential distribution with pmf (8), one can think of equating the probability of a given value of the support (e.g., 0 or 1, which are the only two candidates for being the mode of the theoretical distribution) to the corresponding sample proportion, and solve the resulting equation with respect to the unknown parameter λ. The solution (if it exists and is unique) is indicated as $\hat{\lambda}_P$.

If we consider the value 0, then, after denoting the proportion of zeros in the sample as $\hat{p}_0$, we obtain the following equation

$$p_0(\lambda) = \frac{\lambda - 1 + e^{-\lambda}}{\lambda} = \hat{p}_0.$$

$p_0(\lambda)$ is a continuous and strictly increasing function with λ, satisfying $\lim_{\lambda\to 0^+} p_0(\lambda) = 0$ and $\lim_{\lambda\to+\infty} p_0(\lambda) = 1$, so the method of proportion always returns a feasible estimate $\hat{\lambda}$, providing that $0 < \hat{p}_0 < 1$. To see that $p_0(\lambda)$ is strictly increasing, we just need to compute its first-order derivative, which is equal to

$$p_0'(\lambda) = \frac{1-(1+\lambda)e^{-\lambda}}{\lambda^2}$$

and is strictly positive for $\lambda > 0$.

If we consider the value 1 and its probability, $p_1(\lambda) = (1 - e^{-\lambda})^2/\lambda$, as a function of λ, one can easily prove that it is continuous, with $\lim_{\lambda \to 0^+} p_1(\lambda) = 0$, $\lim_{\lambda \to +\infty} p_1(\lambda) = 0$, and that it has an absolute maximum at $\lambda_{1,M} \approx 1.256432$, whose value is $p_{1,\max} \approx 0.4072644$. In fact, the first-order derivative of $p_1(\lambda)$ is

$$p_1'(\lambda) = -\frac{e^{-2\lambda}(e^{\lambda} - 1)(-2\lambda + e^{\lambda} - 1)}{\lambda^2}$$

which is null only if $\lambda = \lambda_{1,M}$. Therefore, denoting with $\hat{p}_1$ the sample proportion of ones, the method of proportion does not return a feasible estimate if $\hat{p}_1 > p_{1,\max}$; it returns two different estimates if $0 < \hat{p}_1 < p_{1,\max}$; it returns $\lambda_{1,M}$ if $\hat{p}_1 = p_{1,\max}$.

The method of proportion can still be applied to the generalized discrete exponential model (10). However, since the parameters to be estimated are now two, one needs to consider two values and the corresponding probabilities and sample proportions. If we consider the values 0 and 1, since

$$\begin{cases} Q_0 = p_0 \qquad amp; = \left[1 - \frac{1}{\lambda}(1 - e^{-\lambda})\right]^{\alpha} \\ Q_1 = p_0 + p_1 \; amp; = \left[1 - \frac{e^{-\lambda}}{\lambda}(1 - e^{-\lambda})\right]^{\alpha}, \end{cases} \tag{13}$$

then, by taking the logarithm of both sides of the first equation and solving for α, we obtain

$$\alpha = \log p_0 \Big/ \log\left[1 - \frac{1}{\lambda}(1 - e^{-\lambda})\right].$$

By dividing the second equation by the first in (13), we derive

$$\left(\frac{p_1 + p_0}{p_0}\right)^{1/\alpha} = \frac{\lambda - e^{-\lambda}(1 - e^{-\lambda})}{\lambda - (1 - e^{-\lambda})},$$

and taking the logarithm of both sides and plugging-in the expression for α derived at the previous step we obtain

$$\log\left[1 - \frac{1}{\lambda}(1 - e^{-\lambda})\right]\frac{\log(p_0 + p_1) - \log p_0}{\log p_0} = \log\frac{\lambda - e^{-\lambda}(1 - e^{-\lambda})}{\lambda - (1 - e^{-\lambda})},$$

from which

$$\lambda^c\left[\lambda - e^{-\lambda}(1 - e^{-\lambda})\right] = [\lambda - (1 - e^{-\lambda})]^{c+1},$$

with $c = [\log(p_0 + p_1) - \log p_0]/\log p_0$. Then, in order to find an estimate for λ, one has to solve the equation

$$g(\lambda) = \lambda^{\hat{c}}[\lambda - e^{-\lambda}(1 - e^{-\lambda})] - [\lambda - (1 - e^{-\lambda})]^{\hat{c}+1} = 0$$

where $\hat{c} = [\log(\hat{p}_0 + \hat{p}_1) - \log\hat{p}_0]/\log\hat{p}_0$. Let us denote the root of the equation above as $\hat{\lambda}_P$; it exists and is unique providing that $\hat{p}_0 > 0$, $\hat{p}_1 > 0$, and $\hat{p}_0 + \hat{p}_1 < 1 -$

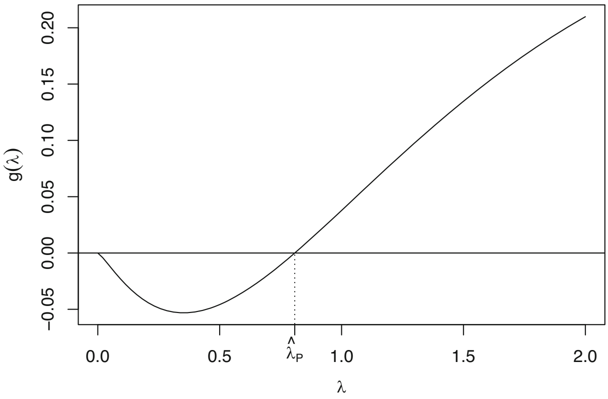


Fig. 9 Example of graph of the function $g(\lambda)$ employed in the estimation through the method of proportion

in this case, $-1 < \hat{c} < 0$. The estimate of α is finally obtained as $\hat{\alpha}_P = \log \hat{p}_0 / \log[1-(1-e^{-\hat{\lambda}_P})/\hat{\lambda}_P]$.

Figure 9 displays the graph of $g(\lambda)$ and the estimate $\hat{\lambda}_P$ for a sample whose proportion of zeros and ones are 0.15 and 0.4, respectively. It turns out that $\hat{\lambda}_P \approx 0.8078636$ and $\hat{\alpha}_P \approx 1.637782$.

Acknowledgements We thank one anonymous referee for their comments, which improved the clarity and quality of the manuscript.

Author Contributions The authors contributed equally to this work.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. The first author acknowledges support by the PRIN2022 project "The effects of climate change in the evaluation of financial instruments" financed by the "Ministero dell'Università e della Ricerca" with grant number 20225PC98R, CUP Code: G53D23001960006.

Data Availability Not applicable

Code Availability Relevant R code is available on GitHub: https://github.com/alessandro-barbiero/discrete.exp.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Consent for publication The manuscript has been read and approved by all named authors.

Ethics approval and consent to participate Not applicable

Materials availability Supplementary material is available in Appendix A and Appendix B.

References

1. Garcia S, Luengo J, Sáez J, Lopez V, Herrera F (2013) A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. IEEE Transactions on Knowledge and Data Engineering 25(4):734–750
2. Bickel J, Lake L, Lehman J (2011) Discretization, simulation, and Swanson's (inaccurate) mean. SPE Economics & Management 3(03):128–140
3. Bickel J (2012) Discretization, simulation, and the value of information. SPE Economics & Management 4(04):198–203
4. Luceño A (1999) Discrete approximations to continuous univariate distributions-an alternative to simulation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61(2):345–352
5. Panjer H (1981) Recursive evaluation of a family of compound distributions. ASTIN Bulletin: The Journal of the IAA 12(1):22–26
6. Heitsch H, Römisch W (2011) Scenario tree generation for multi-stage stochastic programs. In: Bertocchi M, Consigli G, Dempster MAH (eds) Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies. Springer, NY, pp 313–341
7. Pflug G, Pichler A (2011) Approximations for probability distributions and stochastic optimization problems. In: Bertocchi M, Consigli G, Dempster MAH (eds) Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies. Springer, NY, pp 343–387
8. Lai C-D (2013) Issues concerning constructions of discrete lifetime models. Quality Technology & Quantitative Management 10(2):251–262
9. Nakagawa T, Osaki S (1975) The discrete Weibull distribution. IEEE Transactions on Reliability 24(5):300–301
10. Krishna H, Pundir P (2009) Discrete Burr and discrete Pareto distributions. Statistical Methodology 6(2):177–188
11. Roy D (2003) The discrete normal distribution. Communications in Statistics - Theory and Methods 32(10):1871–1883
12. Inusah S, Kozubowski T (2006) A discrete analogue of the Laplace distribution. Journal of Statistical Planning and Inference 136:1090–1102
13. Barbiero A, Hitaj A (2021) A new method for building a discrete analogue to a continuous random variable based on minimization of a distance between distribution functions. In: 2021 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 338–341. IEEE
14. Barbiero A, Hitaj A (2024) Minimizing distance between distribution functions: discrete counterparts to continuous random variables with applications in non-life insurance and stochastic reliability. Statistics 58(5):1140–1168
15. Székely G, Rizzo M (2013) Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference 143(8):1249–1272
16. Pal N, Jin C, Lim W (2006) Handbook of Exponential and Related Distributions for Engineers and Scientists. CRC Press, Boca Raton, FL
17. Minkova L (2012) Distributions in insurance risk models. Doctor of Science Thesis, FMI, Sofia University, Bulgaria
18. Dickson D (2016) Insurance Risk and Ruin. Cambridge University Press, Cambridge, UK
19. Gupta R, Kundu D (1999) Generalized exponential distributions. Australian & New Zealand Journal of Statistics 41(2):173–188
20. Chakraborti S, Jardim F, Epprecht E (2019) Higher-order moments using the survival function: The alternative expectation formula. The American Statistician 73(2):191–194
21. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. https://www.R-project.org/ (2024)

22. Consul P (1989) Generalized Poisson Distributions: Properties and Applications. Marcel Dekker Inc., New York
23. Klugman SA, Panjer HH, Willmot GE (2012) Loss Models: from Data to Decisions, vol 715. John Wiley & Sons, Hoboken, NJ
24. Maiti S (1995) Estimation of $P(X \leq Y)$ in the geometric case. Journal of the Indian Statistical Association 33(2):87–91
25. Barbiero A (2013) Inference on reliability of stress-strength models for Poisson data. Journal of Quality and Reliability Engineering **2013** Article ID 530530
26. Sathe Y, Dixit U (2001) Estimation of $P[X \leq Y]$ in the negative binomial distribution. Journal of Statistical Planning and Inference 93(1–2):83–92
27. Obradovic M, Jovanovic M, Milosevic B, Jevremovic V (2015) Estimation of $P\{X \leq Y\}$ for geometric-Poisson model. Hacettepe Journal of Mathematics and Statistics 44(4):949–964
28. Shannon C (1951) Prediction and entropy of printed English. Bell System Technical Journal 30(1):50–64
29. Wang S (1998) An actuarial index of the right-tail risk. North American Actuarial Journal 2(2):88–101