Proceedings of International Mathematical Sciences
ISSN: 2717-6355, URL: https://dergipark.org.tr/tr/pub/pims
Volume 5 Issue 2 (2023), Pages 76-80.
Doi: https://doi.org/10.47086/pims.1346708

.

AN ALTERNATIVE DISCRETE ANALOGUE OF THE HALF-LOGISTIC DISTRIBUTION

ALESSANDRO BARBIERO*, ASMERILDA HITAJ**
*DEPARTMENT OF ECONOMICS, MANAGEMENT AND QUANTITATIVE METHODS, UNIVERSITÀ DEGLI STUDI DI MILANO, MILAN, ITALY. ORCID NUMBER: 0000-0002-5072-5662
**DEPARTMENT OF ECONOMICS, UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA, VARESE, ITALY. ORCID NUMBER: 0000-0002-1921-8435

Abstract. A discrete version of the continuous half-logistic distribution is introduced, which is based on the minimization of the Cramér distance between the corresponding continuous and step-wise cumulative distribution functions. The expression of the probability mass function is derived in analytic form and some properties of the distribution are discussed, as well as sample estimation. A comparison is also made with a discrete version already proposed in the literature, which is based on a different rationale. An application to real data is finally presented.

1. Introduction

The half-logistic distribution is a random distribution supported on $\mathbb{R}^+$ obtained by folding the logistic distribution about the origin [1]. Thus, if Y is a random variable (rv) following the logistic distribution with parameter $\theta > 0$, with cumulative distribution function (cdf) $F_Y(y) = \frac{1}{1+e^{-\theta y}}$ and probability density function (pdf) $f_Y(y) = \frac{\theta e^{-\theta y}}{(1+e^{-\theta y})^2}$, the rv $X = |Y|$ follows the the half-logistic distribution with the same parameter θ; its pdf is

$$f(x) = \frac{2\theta e^{-\theta x}}{(1+e^{-\theta x})^2}, \quad x \in \mathbb{R}^+, \theta \in \mathbb{R}^+; \tag{1.1}$$

its cdf is

$$F(x) = \frac{2}{1+e^{-\theta x}} - 1 = \frac{2e^{\theta x}}{e^{\theta x}+1} - 1 = \frac{e^{\theta x}-1}{e^{\theta x}+1} = 1 - \frac{2}{1+e^{\theta x}}, \quad x \in \mathbb{R}^+. \tag{1.2}$$

The expectation is $\mu = \log 4/\theta$. [2] introduced a discrete analogue of the half-logistic distribution, defined through (1.1) or (1.2), by imposing the matching of

2020 *Mathematics Subject Classification.* Primary: 62E15, 62F10, 62N05.
Key words and phrases. Discretization; Distance minimization; Survival function.

Submitted on 21.08.2023, Accepted on 30.08.2023.
Communicated by Müjgan TEZ and Kadri Ulaş AKAY.

the survival function (sf) $P(X \geq x)$ at each integer value of the support, i.e., defining the probability mass function (pmf) as $p(x) = F(x+1) - F(x)$. The pmf of the discrete analogue of the half-logistic distribution has thus the following expression:

$$p_i = p(i) = 2\left[1 + e^{-\theta(i+1)}\right]^{-1} - 2\left[1 + e^{-\theta i}\right]^{-1}, i = 0, 1, 2, \dots \tag{1.3}$$

In this paper, we introduce and discuss an alternative discrete version of the continuous half-logistic distribution by following a different approach, based on the minimization of a discrepancy measure between the continuous cdf of the parent distribution and the step-wise cdf of the discrete counterpart [3]. The distance chosen is the Cramér distance, defined as

$$d(F, G) = \int_{\mathbb{R}} |F(x) - G(x)|^2 \mathrm{d}x, \tag{1.4}$$

where F and G are the continuous and step-wise cdf of the continuous random distribution and of its discrete version, respectively. The paper is structured as follows: In the next section, we provide the general solution to the problem stated above and then derive the "optimal" discrete counterpart of the half-logistic distribution, by providing the analytic expression of its pmf and some properties. The third section is devoted to sample estimation and discusses the maximum likelihood method, the method of moment and the method of proportion. The fourth and final section presents an application to a real dataset, on which the proposed discrete distribution is fitted.

2. Definition of an alternative discrete version of the half-logistic distribution

If G is a stepwise cdf, supported on the non-negative integers $i \in \{0, 1, 2, \dots\}$, which can be seen as a discrete version of a continuous cdf F, supported on the positive half-line, letting $Q_i = G(i)$, the Cramér distance (1.4) can be rewritten as

$$d(F, G) = \sum_{i=0}^{\infty} |F(x) - Q_i|^2.$$

By minimizing the function above with respect to the Q_i's, we obtain the "optimal" values as $Q_i = \int_i^{i+1} F(x)\mathrm{d}x$ [3]. The optimal discrete analogue of the half-logistic distribution has then cumulative probabilities given by

$$Q_i = \int_i^{i+1} \left(1 - \frac{2}{1 + e^{\theta x}}\right) \mathrm{d}x = 1 - 2 + \left[\frac{2\log(1 + e^{\theta x})}{\theta}\right]_i^{i+1} = \frac{2}{\theta} \log \frac{1 + e^{\theta(i+1)}}{1 + e^{\theta i}} - 1,$$

for $i = 0, 1, 2, \dots$, so that the probabilities are

$$\begin{cases} p_0 = Q_0 = \dfrac{2}{\theta} \log \dfrac{1 + e^{\theta}}{2} - 1 \\ p_i = Q_i - Q_{i-1} = \dfrac{2}{\theta} \log \dfrac{(1 + e^{\theta(i+1)})(1 + e^{\theta(i-1)})}{(1 + e^{\theta i})^2}, \quad i = 1, 2, \dots. \end{cases} \tag{2.1}$$

It can be proved that $p_0 < p_1$ if θ is smaller than $\theta^* = 2.12255$. Conversely, for any $\theta > \theta^*$, $p_0 > p_1$, whereas if $\theta = \theta^*$, it follows that $p_0 = p_1$. It can be also proved that $p_i > p_{i+1}$ for any $i \geq 1$. As a direct consequence of the two results above, we have that the proposed alternative discrete counterpart of the half-logistic distribution is unimodal with mode equal to 1 if $\theta < \theta^*$, with mode

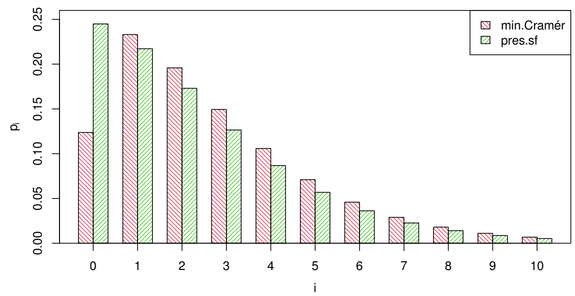

Figure 1. Pmf of the proposed discrete counterpart, based on Eq. (2.1), and of the discrete counterpart proposed by [2], Eq. (1.3), based on the preservation of the sf; $\theta = 1/2$.

equal to 0 if $\theta > \theta^*$; it is bimodal with modes at 0 and 1 if $\theta = \theta^*$. This is a very relevant difference with respect to the model of Eq. (1.3), which is unimodal with mode at 0. Figure 1 displays, for the integers 0 to 10, the probabilities for the two models when $\theta = 1/2$. It can be easily shown that the expectation of the alternative discrete half-logistic coincides with that of the parent continuous distribution. This is a general property holding for the discrete counterparts of positive rvs obtained by minimizing the Cramér distance (1.4). In fact, denoting the continuous rv and its optimal counterpart by X and $\tilde{X}$, respectively, and recalling an alternative formulation of the expected value for non-negative rvs, one shows that

$$\mathbb{E}(\tilde{X}) = \sum_{i=0}^{\infty}(1 - Q_i) = \sum_{i=0}^{\infty}\left(1 - \int_i^{i+1} F(x)\mathrm{d}x\right) = \int_0^{\infty}(1 - F(x))\mathrm{d}x = \mathbb{E}(X).$$

Determining if and when the proposed discretization is "better", according to some appropriate criterion, than the usual one, based on the matching of the sf, can the the object of further study.

3. Parameter estimation

Given an iid sample $(x_1, x_2, \ldots, x_n)$ which we assume to come from the alternative discrete half-logistic distribution (2.1), the unknown parameter θ can be estimated by resorting to one of the following methods.

3.1. **Maximum likelihood method.** The maximum likelihood estimate $\hat{\theta}_{ML}$ of θ is the value maximizing the log-likelihood function $\ell(\theta; x_1, \ldots, x_n) = \sum_{i=1}^n \log p_{x_i}(\theta)$. Due to the complicated expression of the pmf, it is not possible to derive a closed-form expression of $\hat{\theta}_{ML}$, but any standard optimization routine can be used in order to obtain it numerically.

3.2. **Method of moment.** By equating this expectation of the proposed model to the sample mean $\bar{x} = \sum_{i=1}^n x_i/n$, one derives the moment estimate as $\hat{\theta}_M = \log 4/\bar{x}$.

3.3. **Method of proportion.** This method, suitable for discrete distributions, consists in considering an assigned support value and determining the value of the parameter for which the probability of that value equals the corresponding relative sample frequency. One can consider matching the probability of 0, available from (2.1), and the corresponding relative sample frequency of zeros, $\hat{p}_0$. After

TABLE 1. Distribution of number of outbreaks of strikes, from [4]

count	observed frequency	theoretical frequency
0	46	51.39
1	76	69.69
2	24	25.61
3	9	7.01
($\geq$)4	1	2.30
total	156	156

simple algebraic steps, one obtains the following equation in $\omega = e^\theta$, $2\omega^{(1+\hat{p}_0)/2} - \omega - 1 = 0$, which yields a unique root $\hat{\omega}_P$ and the corresponding estimate $\hat{\theta}_P = \log \hat{\omega}_P$.

4. A REAL DATA EXAMPLE

We consider the dataset presented in [4] and reported in Table 1. Fitting the data through the alternative half-logistic might be plausible, since the mode is 1. The MLE of θ is 1.4238 and using this estimate we reconstruct the theoretical frequencies, which are displayed in the last column of Table 1. Pooling the last two counts (3 and 4), we calculate the usual chi-square statistic, $X^2 = \sum_{i=0}^{3}(n_i - n_i^*)^2/\hat{n}_i$, where n_i and n_i^* are the observed and theoretical frequencies of the count i; its value is 1.2896 and the approximate p-value of the chi-square test is 0.5247, thus indicating a more than satisfactory fit of the model. The maximum value of the log-likelihood function is -188.104; the AIC value is 378.208. All these results, if compared to those of the statistical models analyzed in [5], highlight that the alternative discrete half-logistic distribution has a superior goodness-of-fit.

Acknowledgments. This work was presented at the 7th International Conference of Mathematical Sciences (ICMS 2023) [6].

REFERENCES

[1] N. Balakrishnan, Order statistics from the half logistic distribution, J. Stat. Comput. Simul. 20(4)(1985) 287–309.
[2] A. Barbiero and A. Hitaj, A discrete analogue of the half-logistic distribution, in 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 64–67.
[3] A. Barbiero and A. Hitaj, 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), 2021, pp. 338–341.
[4] M. S. Ridout and P. Besbeas, An empirical model for underdispersed count data, Stat. Model. 4(1) (2004) 77–89.
[5] S. Chakraborty and R. D. Gupta, Exponentiated geometric distribution: another generalization of geometric distribution, Comm. Statist. Simulation Comput. 44(6) (2015) 1143–1157.
[6] A. Barbiero, A. Hitaj, An alternative discrete analogue of the half-logistic distribution, in Abstract Book of ICMS 2023, p.84.

ALESSANDRO BARBIERO,
VIA CONSERVATORIO 7, 20122 MILAN, ITALY. PHONE: (+39)0250321533 ORCID NUMBER: 0000-0002-5072-5662
Email address: alessandro.barbiero@unimi.it

Asmerilda Hitaj,
via Monte Generoso 71, Varese, Italy. Phone: (+39)0332395546 Orcid number: 0000-0002-1921-8435
Email address: asmerilda.hitaj@uninsubria.it