

# A Discrete Analogue of the Half-Logistic Distribution

Alessandro Barbiero

*Department of Economics, Management and Quantitative Methods*  
*Università degli Studi di Milano*  
 Milan, Italy  
 alessandro.barbiero@unimi.it

Asmerilda Hitaj

*Department of Economics*  
*Università degli Studi dell'Insubria*  
 Varese, Italy  
 asmerilda.hitaj@uninsubria.it

**Abstract**—In lifetime modeling, the observed measurements are usually discrete in nature, because the values are measured to only a finite number of decimal places and cannot really constitute all points in a continuum. For example, the survival time of a cancer patient can be measured as the number of months he/she survives. Then, even if the lifetime (of a patient, a device, etc.) is intrinsically continuous, it is reasonable to consider its observations as coming from a discretized distribution generated from an underlying continuous model. In this work, a discrete random distribution, supported on the non-negative integers, is obtained from the continuous half-logistic distribution by using a well-established discretization technique, which preserves the functional form of the survival function. Its main statistical properties are explored, with a special focus on the shape of the probability mass function and the determination of the first two moments; we discuss and compare, both theoretically and empirically, two different methods for estimating its unique parameter. This discrete random distribution can be used for modeling data exhibiting excess of zeros and over-dispersion, which are features often met in the insurance and ecology fields: an example of application is illustrated. An extension of this discrete distribution is finally suggested, by considering the generalized half-logistic distribution, which introduces a second shape parameter allowing for greater flexibility.

**Index Terms**—count distribution, discretization, logistic distribution, survival function

## I. INTRODUCTION

The half-logistic distribution is a random distribution obtained by folding at zero the logistic distribution centered around the origin [1]. Thus, if  $W$  is a random variable (rv) that follows the logistic distribution, then the transformation  $Y = |W|$  follows the half-logistic distribution. Its probability density function (pdf) is

$$f_y(y) = \frac{2\theta e^{-\theta y}}{(1 + e^{-\theta y})^2}, \quad y \geq 0, \theta > 0; \quad (1)$$

its cumulative distribution function (cdf) is

$$F_y(y) = P(Y \leq y) = \frac{1 - e^{-\theta y}}{1 + e^{-\theta y}}, \quad y \geq 0,$$

and its survival function (sf) is

$$S_y(y) = P(Y \geq y) = \frac{2e^{-\theta y}}{1 + e^{-\theta y}}, \quad y \geq 0. \quad (2)$$

The expressions for the expected value and the variance are

$$\mathbb{E}(Y) = \log(4)/\theta, \quad \sigma_Y^2 = [\pi^2/3 - (\log 4)^2]/\theta^2.$$

Since the ratio between variance and expected value is  $\frac{1}{\theta} \frac{\pi^2/3 - (\log 4)^2}{\log 4}$ , we conclude that the half-logistic distribution is over-dispersed if  $\theta < \theta_0 = \frac{\pi^2/3 - (\log 4)^2}{\log 4} \approx 0.9868$ ; it is under-dispersed if  $\theta > \theta_0$ .

The half-logistic distribution, along with generalizations and modifications thereof (see for example [2], where a distribution is obtained by compounding half-logistic and Lomax distributions) can be used to model continuous phenomena, such as lifetimes, in several different fields as actuarial science, medical and biological sciences, and engineering. However, in many real-life situations where the phenomenon under study is modeled by a continuous distribution, it has been observed that the phenomenon is often recorded as an integer-valued one instead of real-valued, either because of its intrinsic nature or because of the limitation of measuring instruments; this fact justifies the introduction of a discrete version of the existing continuous distributions. With this background, the primary goal of this paper is to provide a discrete analog of the half-logistic distribution given in (1), which can be used for modeling discrete lifetimes or genuine count variables.

The paper is organized as follows. In Section II, we present the new discrete model, some of its properties and other results which are useful in simulation studies. Section III addresses the estimation of the parameter of the new model, by considering two estimation methods which are then compared through a Monte Carlo simulation study. Section IV illustrates an application to real data. In the final section, some conclusions are drawn and possible research perspectives are outlined.

## II. A DISCRETE HALF-LOGISTIC DISTRIBUTION: DEFINITION AND MAIN PROPERTIES

A discrete analogue can be obtained from a continuous distribution based on different methods, according to which feature of the “parent” distribution one wants to maintain (the pdf, the sf, the moments, and so on). A thorough review of such discretization methods is presented in [3]. The most popular and straightforward discretization technique is the one preserving the functional form of the sf. A discrete analogue

of a continuous distribution defined on the positive real half-line is obtained by setting its pmf equal to the difference  $S_y(x) - S_y(x+1)$ , where  $x = 0, 1, 2, \dots$ , and  $S_y(\cdot)$  is the sf of the parent continuous model. This discretization method has been applied in the literature many times for deriving a discrete counterpart of continuous random distributions: just to name a few, in [4] a discrete Weibull distribution is obtained; [5] discusses a discrete Burr and a discrete Pareto distribution; [6] derives a two-parameter discrete gamma distribution corresponding to the continuous two-parameter gamma distribution.

Applying this discretization to the half-logistic sf of Equation (2), we obtain the pmf:

$$p(x) = 2e^{-\theta x} / [1 + e^{-\theta x}] - 2e^{-\theta(x+1)} / [1 + e^{-\theta(x+1)}], \quad (3)$$

$$x = 0, 1, 2, \dots$$

For this discrete half-logistic distribution, and in general for any discrete analogue of a continuous rv supported on  $\mathbb{R}^+$ , the sf is thus the same as that of the continuous model at each non-negative integer value  $x$ :

$$S(x) = P(X \geq x) = \sum_{t \geq x} p(t) = \sum_{t \geq x} S_y(t) - S_y(t+1)$$

$$= S_y(x).$$

This distribution was actually introduced in [7], by using an alternative parametrization for the parent model.

By construction, since the probability  $p(x)$  of any non-negative integer value  $x$ , being equal to the difference  $S_y(x) - S_y(x+1)$ , corresponds to the integral of the pdf  $f_y(y)$  of the continuous distribution between  $x$  and  $x+1$ , and being  $f_y(y)$  in (1) a strictly decreasing function for  $y > 0$ , it descends that  $p(x)$  is strictly decreasing too and has a unique mode at 0. Figure 1 displays the pmf of the half-logistic distribution for four different values of  $\theta$ .

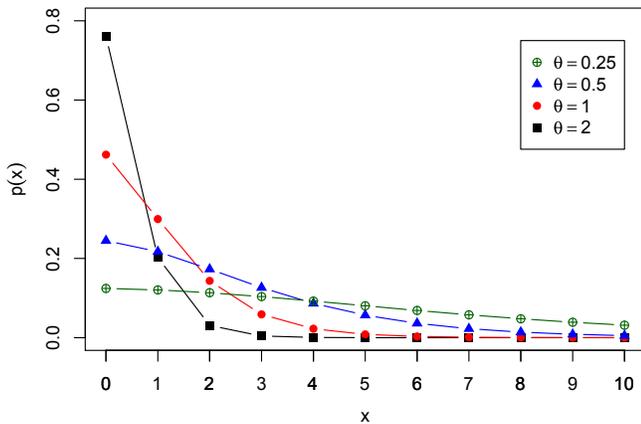


Fig. 1. Probability mass function (truncated at  $x = 10$ ) of the discrete half-logistic distribution for different values of  $\theta$ .

The cdf, for a positive integer value  $x$ , is given by

$$F(x) = P(X \leq x) = 1 - P(X > x) = 1 - S(x+1)$$

$$= \frac{1 - e^{-\theta(x+1)}}{1 + e^{-\theta(x+1)}}. \quad (4)$$

From (4), one can easily derive the expression of the quantile of order  $0 < u < 1$ , which is

$$x_u = \left\lceil \frac{1}{\theta} \log \frac{1+u}{1-u} - 1 \right\rceil. \quad (5)$$

In particular, the median is

$$x_{0.5} = \left\lceil \frac{1}{\theta} \log 3 - 1 \right\rceil.$$

The value 0 is then the median of this distribution for any  $\theta \geq \log 3$ .

A random number (non-negative integer) can be sampled from the proposed model through the usual inverse transformation method. Let  $u$  be a random number drawn from a uniform distribution on the unit interval. For fixed  $\theta$ , a random number following the discrete half-logistic distribution can be obtained by computing the right-hand side of Equation (5).

The expectation of the discrete half-logistic distribution is given by

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} x(1 - F(x)) = \sum_{x=1}^{\infty} \frac{2}{1 + e^{\theta x}},$$

that cannot be written in a closed form. The following relationship also holds

$$\sum_{x=1}^{\infty} \frac{2}{1 + e^{\theta x}} < \int_0^{\infty} \frac{2}{1 + e^{\theta x}} dx < \sum_{x=0}^{\infty} \frac{2}{1 + e^{\theta x}}$$

(see, e.g., [8] for similar result for the discrete Weibull distribution). In other terms, the mean  $\mathbb{E}(X)$  of the discrete half-logistic rv is such that

$$\mathbb{E}(X) < \mathbb{E}(Y) < \mathbb{E}(X) + 1,$$

or, better,

$$\mathbb{E}(Y) - 1 < \mathbb{E}(X) < \mathbb{E}(Y).$$

As for the variance,

$$\sigma_X^2 = \sum_{x=1}^{\infty} (2x-1)S(x) - [\mathbb{E}(X)]^2 = \sum_{x=1}^{\infty} \frac{4x-2}{1+e^{\theta x}} - [\mathbb{E}(X)]^2,$$

which again cannot be written in a closed form. Table I reports the values of  $\mathbb{E}(X)$  and  $\sigma_X^2$  for some values of  $\theta$ . By numerical inspection (one can still base on Table I), we can state that the discrete half-logistic distribution is over-dispersed for any value of  $\theta$ ; in particular, the coefficient of over-dispersion, given by the ratio  $\sigma_X^2 / \mathbb{E}(X)$ , is a decreasing function of  $\theta$ , tending asymptotically to 1. This contrasts with what occurring to the continuous parent distribution, which we have seen is under-dispersed or over-dispersed according to whether the  $\theta$  parameter is greater or smaller than a certain threshold.

TABLE I  
EXPECTED VALUE AND VARIANCE OF THE HALF-DISCRETE DISTRIBUTION  
FOR SOME VALUES OF THE PARAMETER  $\theta$

$\theta$	$\mathbb{E}(X)$	$\sigma_X^2$
0.1	13.37	136.77
0.2	6.44	34.17
0.5	2.29	5.44
1.0	0.93	1.33
2.0	0.28	0.30
4.0	0.04	0.04

### III. ESTIMATION

Parameter estimation based on an i.i.d. sample  $\mathbf{x} = (x_1, \dots, x_n)$ , which is assumed to come from the discrete half-logistic distribution, can be carried out resorting to different methods. Here we describe two possible options: the method of proportion and the maximum likelihood method.

#### A. Method of proportion

Since from (3) the probability of  $X$  being 0 is

$$p(0) = 1 - 2/(1 + e^\theta) = \frac{e^\theta - 1}{e^\theta + 1},$$

if we define  $\hat{p}_0$  as the relative frequency of zeros in the sample, we can find an estimate for  $\theta$  by equating  $p(0)$  to  $\hat{p}_0$  and solving with respect to the unknown parameter  $\theta$ :

$$\hat{\theta}_P = \log\left(\frac{1 + \hat{p}_0}{1 - \hat{p}_0}\right).$$

$\hat{p}_0$  must be neither 0 nor 1, in order to have a finite and positive value for  $\hat{\theta}_P$ . This method, named the method of proportion (see, e.g., [8]), is particularly suited for discrete random variables, is very intuitive, simple to apply and provides a closed-form expression for the estimate. Its main drawback is that it exploits only part of the information contained in the sample, i.e., the proportion of zeros. Therefore, it is not expected to be very efficient, especially when such proportion is small. In our case, however, we know that  $p(0) > p(k)$  for any  $k > 0$  and on average this happens also in terms of sample proportions; so using the value 0 for deriving the unknown parameter  $\theta$  is on average the best choice.

#### B. Maximum likelihood method

The log-likelihood function for the discrete half-logistic distribution can be written as:

$$\begin{aligned} \ell(\theta; \mathbf{x}) &= \log \prod_{i=1}^n p(x_i; \theta) = \sum_{i=1}^n \log p(x_i; \theta) \\ &= n \log 2 + n \log(1 - e^{-\theta}) - \theta \sum_{i=1}^n x_i + \\ &\quad - \sum_{i=1}^n \{\log[1 + e^{-\theta x_i}] + \log[1 + e^{-\theta(x_i+1)}]\}. \end{aligned} \quad (6)$$

By maximizing it with respect to the natural parameter space of  $\theta$ ,  $\mathbb{R}^+$ , one obtains the maximum likelihood estimate  $\hat{\theta}_{ML}$ :

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} \ell(\theta; \mathbf{x}).$$

The normal equation

$$\begin{aligned} \frac{d\ell(\theta; \mathbf{x})}{d\theta} &= \frac{ne^{-\theta}}{1 - e^{-\theta}} - \sum_{i=1}^n x_i + \sum_{i=1}^n \frac{x_i e^{-\theta x_i}}{1 + e^{-\theta x_i}} \\ &\quad + \sum_{i=1}^n \frac{(x_i + 1)e^{-\theta(x_i+1)}}{1 + e^{-\theta(x_i+1)}} = 0 \end{aligned} \quad (7)$$

does not have in general a closed-form solution for  $\theta$ ; hence,  $\hat{\theta}_{ML}$  has to be recovered numerically, by computing the root of (7) or directly maximizing (6).

#### C. An empirical comparison

In order to compare the statistical performance of the two estimation methods, we set up a preliminary simulation study. For several combinations of the parameter  $\theta$  and sample size  $n$ , we simulated a large number  $N = 10,000$  of samples, and for each single sample drawn  $\mathbf{x}_j$ , we computed both  $\hat{\theta}_{ML}^{(j)}$  and  $\hat{\theta}_P^{(j)}$ ,  $j = 1, 2, \dots, N$ . As performance indicators for the two estimators of  $\theta$ , we considered the Monte Carlo Bias,  $\text{Bias} = (1/N) \sum_{j=1}^N \hat{\theta}_j - \theta$ , and the Monte Carlo Root Mean Square Error,  $\text{RMSE} = (1/N) \sum_{j=1}^N (\hat{\theta}_j - \theta)^2$ , which can be regarded as approximations of the bias  $\mathbb{E}(\hat{\theta}) - \theta$  and of the root mean square error  $\mathbb{E}[(\hat{\theta} - \theta)^2]$ , respectively, which cannot be computed analytically due to the complex or not analytic form of either estimators. This Monte Carlo simulation study was carried out by using our own functions implemented in the the R programming environment [9].

In Table II we report the results obtained by setting  $\theta = 1$  and  $n = 20, 50$ , and  $100$ . In Figure 2, we use the boxplots to show the entire empirical distributions of the two estimators. It can be noted that for both estimators the bias is positive but very small, in particular for  $\hat{\theta}_P$ ; it decreases with  $n$ . The RMSE of  $\hat{\theta}_{ML}$  is smaller than that of the  $\hat{\theta}_P$  (drop of about 30% for the three scenarios considered), confirming that at least for the settings here considered the maximum likelihood estimator is more efficient than the estimator obtained through the method of proportion. Furthermore, as expected, the RMSEs of both estimators decrease as  $n$  increases. A more complete comparison can be attained by considering other values of  $\theta$  and  $n$ .

TABLE II  
MONTE CARLO BIAS AND ROOT MEAN SQUARE ERROR FOR THE TWO  
ESTIMATORS OF  $\theta$  FOR THE DISCRETE HALF-LOGISTIC DISTRIBUTION,  
WHEN  $n \in \{20, 50, 100\}$  AND  $\theta = 1$

sample size	estimator	Bias	RMSE
20	$\hat{\theta}_P$	0.0164	0.2973
	$\hat{\theta}_{ML}$	0.0409	0.2128
50	$\hat{\theta}_P$	0.0060	0.1813
	$\hat{\theta}_{ML}$	0.0146	0.1244
100	$\hat{\theta}_P$	0.0037	0.1255
	$\hat{\theta}_{ML}$	0.0070	0.0865

### IV. APPLICATION TO REAL DATA

In this section, we present a data set to examine the fitting of the proposed model. The data presented in the first two

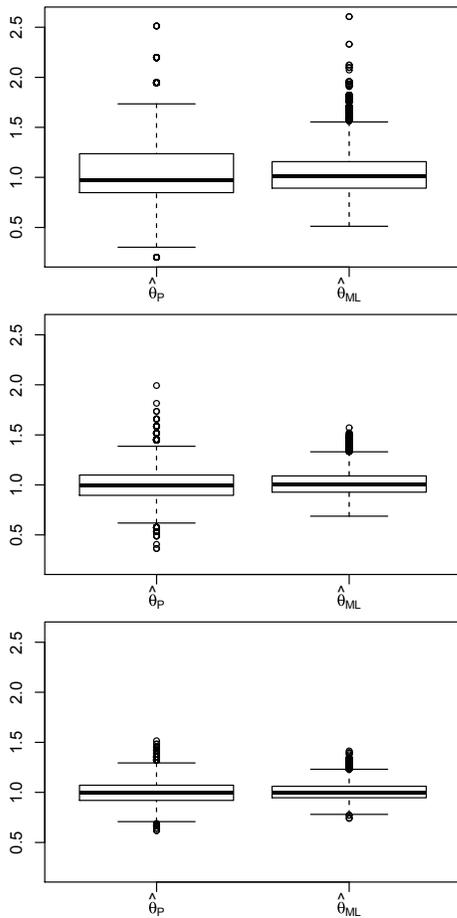


Fig. 2. Monte Carlo distribution of  $\hat{\theta}_P$  and  $\hat{\theta}_{ML}$  for  $\theta = 1$  and for  $n = 20$  (top graph),  $n = 50$  (center), and  $n = 100$  (bottom).

columns of Table III (counts of the number of European red mites on apple leaves which appears in [10], among others) are overdispersed since the sample variance is greater than the mean.

The estimate of  $\theta$  obtained from the maximum likelihood method is  $\hat{\theta} = 0.8454$ , with standard error 0.0596; the method of proportion leads to  $\hat{\theta}_P = 1.012$ . The third column of Table III displays the expected frequencies for the discrete half-logistic model fitted through the maximum likelihood method. The  $\chi^2$  test statistic, defined as  $\sum_{i=1}^K (O_i - E_i)^2 / E_i$ , where  $O_i$  and  $E_i$  are the observed and expected frequencies of the  $i$ -category,  $i = 1, \dots, K$ , is computed after pooling the last four categories (in order to have all the  $E_i$  greater than 5). It results 7.838 and the corresponding  $p$ -value is 0.0495, which gives a barely acceptable fit.

## V. DISCUSSION

We discussed the main properties and the inferential issues of a new discrete half-logistic distribution, which has a decreasing pmf with mode at zero for any value of the parameter  $\theta$ ; this feature makes it useful for fitting discrete lifetimes or count data with a large excess of zeros, often occurring in

TABLE III  
COUNTS OF THE NUMBER OF EUROPEAN RED MITES ON APPLE LEAVES

Number of mites per leaf	observed frequency	expected frequency
0	70	59.88
1	38	43.42
2	17	24.69
3	10	12.15
4	9	5.55
5	3	2.45
6	2	1.06
$\geq 7$	1	0.81

the insurance field (when modeling the number of claims) or in ecology. However, this feature of the pmf may represent a limitation for other applications. To overcome this, a possible generalization can be obtained starting from the generalized half-logistic distribution mentioned in [11], which introduces an additional shape parameter  $\alpha > 0$  and whose pdf is

$$f_{yG}(y) = \frac{\alpha \theta (2e^{-\theta y})^\alpha}{(1 + e^{-\theta y})^{\alpha+1}}$$

and sf

$$S_{yG}(y) = \left( \frac{2e^{-\theta y}}{1 + e^{-\theta y}} \right)^\alpha;$$

by setting  $\alpha = 1$ , the one-parameter half-logistic distribution of Equation (1) is obtained. The pmf of the generalized discrete half-logistic can be then defined as  $S_{yG}(x) - S_{yG}(x+1)$ . In this way, for any value of  $\theta$ , it is possible to calibrate the parameter  $\alpha$  in such a way that the pmf is no longer decreasing with a unique mode at zero. Future research will address the properties of this generalized distribution.

## REFERENCES

- [1] N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous univariate distributions, 2nd ed., vol.2. New York: John Wiley & Sons, Ltd, 1995.
- [2] M. Anwar, J. Zahoor, "The half-logistic Lomax distribution for lifetime modeling," J. Prob. Stat., Vol. 2018, Article ID 3152807, 2018.
- [3] S. Chakraborty, "Generating discrete analogues of continuous probability distributions-A survey of methods and constructions," J. Stat. Distrib. Appl., vol. 2(6) <https://doi.org/10.1186/s40488-015-0028-6>, 2015.
- [4] T. Nakagawa, S. Osaki, "The discrete Weibull distribution," IEEE Trans. Reliability, vol. 24(5), pp. 300-301, 1975.
- [5] H. Krishna, P.S. Pundir, "Discrete Burr and discrete Pareto distributions," Stat Methodology, vol. 6(2), pp. 177-188, 2009.
- [6] S. Chakraborty, D. Chakravarty, "Discrete gamma distributions: Properties and parameter estimations," Commun. Stat. Theory Methods, vol. 41(18), pp. 3301-3324, 2012.
- [7] S. Nadarajah, "A statistical analysis of Iraq body counts," Qual. Quant., vol. 49(1), pp. 21-37, 2015.
- [8] M.A. Khan, A. Khalique, A.M. Abouammoh, "On estimating parameters in a discrete Weibull distribution," IEEE Trans. Reliability, vol. 38(3), pp. 348-350, 1989.
- [9] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- [10] P.C. Consul, G.C. Jain, "A generalization of the Poisson distribution," Technometrics, vol. 15(4), pp. 791-799, 1973.
- [11] Y. Liu, Y. Shi, X. Bai, P. Zhan, "Reliability estimation of a NM-cold-standby redundancy system in a multicomponent stress-strength model with generalized half-logistic distribution," Phys. A, vol. 490, pp. 231-249, 2018.