Research Article

Parameter Estimation for Type III Discrete Weibull Distribution: A Comparative Study

Alessandro Barbiero

Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Via Conservatorio, 7-20122 Milan, Italy

Correspondence should be addressed to Alessandro Barbiero; alessandro.barbiero@unimi.it

The type III discrete Weibull distribution can be used in reliability analysis for modeling failure data such as the number of shocks, cycles, or runs a component or a structure can overcome before failing. This paper describes three methods for estimating its parameters: two customary techniques and a technique particularly suitable for discrete distributions, which, in contrast to the two other techniques, provides analytical estimates, whose derivation is detailed here. The techniques' peculiarities and practical limits are outlined. A Monte Carlo simulation study has been performed to assess the statistical performance of these methods for different parameter combinations and sample sizes and then give some indication for their mindful use. Two applications of real data are provided with the aim of showing how the type III discrete Weibull distribution can fit real data, even better than other popular discrete models, and how the inferential procedures work. A software implementation of the model is also provided.

1. Introduction

Most reliability studies assume that time is continuous, and continuous probability distributions such as exponential, gamma, Weibull, normal, and lognormal are commonly used to model the lifetime of a component or a structure. These distributions and the methods for estimating their parameters are well known. In many practical situations, however, lifetime is not measured with calendar time: for example, when a machine works in cycles or on demands and the number of cycles or demands before failure is observed; or when the regular operation of a system is monitored once per period, and the number of time periods successfully completed is observed. Moreover, reliability data are often grouped into classes or truncated according to some censoring criterion. In all these cases, lifetime is modeled as a discrete random variable (r.v.). Indeed, not much work has been done in discrete reliability. Generally, most reliability concepts for continuous lifetimes have been adapted to the discrete case; in particular, discrete analogues of continuous distributions have been introduced [1]. In this context, geometric and negative binomial distributions are the corresponding discrete alternatives for the exponential and gamma

distributions, respectively. Yet, discrete lifetime distributions can be defined without any reference to a continuous counterpart. Bracquemond and Gaudoin [2] provided an exhaustive survey on discrete lifetime distributions.

The (two-parameter) continuous Weibull distribution is one of the most widely used stochastic distributions for modeling the life of a component; it is very flexible since, for different chosen shape parameters, it models either increasing or decreasing failure rates [3]. As a discrete alternative to the continuous Weibull distribution, three main forms have been proposed. The first one was introduced in Nakagawa and Osaki [4] and is referred to as type I discrete Weibull; it mimics the cumulative distribution function of the continuous Weibull distribution. The second one (type II) was proposed and studied in Stein and Dattero [5]; it mimics the hazard rate of its continuous counterpart. The third, which is referred to as type III and this paper considers, was introduced in Padgett and Spurrier [6]: their approach does not start from the continuous Weibull distribution but tries to generalize the notions of hazard rate and mean residual life to the discrete case [7]. From a different perspective and with a different objective, Roy and Dasgupta [8] proposed a discretization method for continuous random variables for

computing reliability in complex stress-strength models, with a specific application to the Weibull r.v.; however, the support of the discrete r.v. thus provided is not the set of nonnegative integers.

Type III discrete Weibull (henceforth simply discrete Weibull) r.v. is not similar in functional form to any of the functions describing a continuous Weibull distribution. It is defined by the following cumulative distribution:

$$F\left(x;c,\beta\right) = P\left(X \le x;c,\beta\right) = 1 - e^{-c\sum_{j=1}^{x+1} j^\beta} \tag{1}$$

$$x = 0, 1, \ldots$$

with $c > 0$ and $\beta \ge -1$ (and not $\beta \in \mathbb{R}$, as stated in Padgett and Spurrier [6]) or equivalently by the following probability mass function:

$$P\left(X = 0;c,\beta\right) = p\left(0;c,\beta\right) = 1 - e^{-c}$$

$$P\left(X = x;c,\beta\right) = p\left(x;c,\beta\right) = e^{-c\sum_{j=1}^{x} j^\beta}\left[1 - e^{-c(x+1)^\beta}\right] \tag{2}$$

$$x = 1, 2, \ldots$$

which can be more compactly rewritten as follows:

$$P\left(X = x;c,\beta\right) = p\left(x;c,\beta\right) = e^{-c\sum_{j=1}^{x} j^\beta}\left[1 - e^{-c(x+1)^\beta}\right]$$

$$x = 0, 1, 2, \ldots, \tag{3}$$

letting $\sum_{j=1}^{x} j^\beta = 0$ if $x = 0$. The corresponding survival function is

$$S\left(x\right) = P\left(X \ge x;c,\beta\right) = e^{-c\sum_{j=1}^{x} j^\beta} \quad x = 0, 1, \ldots, \tag{4}$$

and the hazard rate (or failure rate) function is

$$h\left(x\right) = \frac{p\left(x\right)}{S\left(x\right)} = 1 - e^{-c(x+1)^\beta} \quad x = 0, 1, \ldots. \tag{5}$$

The complexity of the expression of the probability mass function (3) has somehow hindered the use and diffusion of this discrete model. A plot of the probability mass function of the discrete Weibull for three combinations of its parameters is given in Figure 1. Note that for $\beta = 0$, the discrete Weibull reduces to a geometric r.v. with parameter $1 - \exp(-c)$, characterized then by a constant failure rate; for $\beta > 0$, the distribution has an increasing failure rate, whereas for $\beta < 0$, it has a decreasing failure rate. The first and second moments cannot in general be expressed in a closed form but can only be expressed as an infinite series

$$\mathbb{E}\left(X\right) = \sum_{x=0}^{+\infty} x P\left(X = x;c,\beta\right) = \sum_{x=1}^{+\infty} e^{-c\sum_{j=1}^{x} j^\beta}, \tag{6}$$

$$\mathbb{E}\left(X^2\right) = \sum_{x=0}^{+\infty} x^2 P\left(X = x;c,\beta\right)$$

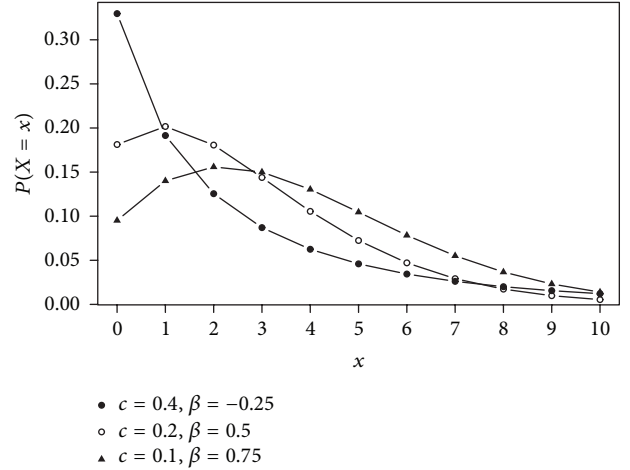$$= \sum_{x=1}^{+\infty} 2x e^{-c\sum_{j=1}^{x} j^\beta} - \mathbb{E}\left(X\right). \tag{7}$$

FIGURE 1: Probability mass function of type III discrete Weibull distribution for three combinations of parameters.

The first moment (6) is finite for $\beta > -1$ or for $\beta = -1$ and $c > 1$; the second moment (7) is finite for $\beta > -1$ or for $\beta = -1$ and $c > 2$ (see the appendix).

This study first describes and discusses procedures for estimating parameters c and β of the discrete Weibull r.v., one of which is an original adaptation of a technique used for some discrete distributions, emphasizing their practical limits (Section 2). An extensive Monte Carlo study, implemented through an adhoc package developed under the R environment, assesses and compares the performance of these estimators, in terms of bias and variability, for different combinations of the parameters and sample sizes (Section 3). The estimation procedures are applied to two datasets taken from the literature (Section 4), and, finally, the summarizing remarks conclude the paper (Section 5).

2. Parameters Point Estimation

Focusing on the point estimation of the parameters of the discrete Weibull r.v., based on an observed simple random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of size n, three techniques are now described: the method of proportion, which is strictly related to the specific features of the distribution function of the discrete Weibull r.v.; the maximum likelihood method; and the method of moments. Below, we assume that both parameters are unknown. Special attention will be devoted to detecting samples leading to implausible estimates or to no estimate at all with either technique.

Method of Proportion. This method was originally introduced in Khan et al. [9] for type I discrete Weibull and was extended by Jazi et al. [10] to discrete inverse Weibull. It relies upon the estimation of a probability (or two or more probabilities, according to the number of parameters involved) using the corresponding sample proportion(s). For the present model, we have $P(X = 0) = 1 - \exp(-c)$, which can be estimated through the proportion of 0's in the sample, y/n, where $y = \sum_{i=1}^{n} I_{x_i=0}$ denotes the number of 0's in the sample,

I_A the indicator function, which equals 1 if A is true and 0 if A is false. Thus, an estimate of c is

$$\widehat{c}_P = -\log\left(1 - \frac{y}{n}\right). \tag{8}$$

Similarly, the probability $P(X = 1) = \exp(-c)[1 - \exp(-c2^{\beta})]$ can be estimated by the proportion of 1's in the sample, z/n, with $z = \sum_{i=1}^{n} I_{x_i=1}$. Consequently, also substituting for c its estimate (8) (see the appendix), β is estimated by

$$\widehat{\beta}_P = \frac{\log\left\{\log\left(1 - y/n - z/n\right) / \log\left(1 - y/n\right) - 1\right\}}{\log 2}. \tag{9}$$

Though the method of proportion easily provides an analytical expression for both estimates, an apparent weakness is that this method does not exploit all the information contained in the sample, since the estimates involve only the rates of 0's and 1's. Moreover, it fails to provide feasible estimates for β if there are no 0's in the sample. In this case, $\widehat{c}_P = 0$, which is a boundary value for c, and $\widehat{\beta}_P$ in (9) cannot be computed. This is particularly common for small samples, and for small values of c. If in the sample there are no 1's ($z = 0$), then by formula (9), the estimate $\widehat{\beta}_P$ is not available again. The method fails in computing an estimate for β even if the sample contains only 0's and 1's ($y + z = n$).

The method of proportion can also lead to "implausible" estimates of β, that is, estimates that do not belong to its parameter space: $\widehat{\beta}_P < -1$. This happens when

$$\frac{\log\left(1 - y/n - z/n\right)}{\log\left(1 - y/n\right)} - 1 < \frac{1}{2}, \tag{10}$$

which is equivalent to say

$$1 - \frac{y}{n} - \frac{z}{n} > \left(1 - \frac{y}{n}\right)^{3/2}, \tag{11}$$

or in terms of z

$$\frac{z}{n} < \left(1 - \frac{y}{n}\right) - \left(1 - \frac{y}{n}\right)^{3/2}. \tag{12}$$

The probability of an "implausible" estimate could be theoretically derived from the last inequality, remembering that y and z are the first two marginal distributions of a trinomial r.v. with parameters (n, $p_0 = P(X = 0)$, $p_1 = P(X = 1)$, $p_2 = 1 - p_0 - p_1$). Figure 2 shows the combinations of y/n and z/n leading to such implausible estimates as a subset of the triangular region ($0 \leq y/n \leq 1$; $0 \leq z/n \leq 1 - y/n$). For example, in the sample $\mathbf{x} = (0, 0, 0, 1, 2, 2, 3, 5, 6, 8)$, with $y/n = 0.3$ and $z/n = 0.1$, the method of proportion provides, through (8) and (9), the estimates $\widehat{c}_P = 0.357$ and $\widehat{\beta}_P = -1.210$.

Maximum Likelihood Method. Having defined the log-likelihood function as $l(c, \beta; \mathbf{x}) = \log \prod_{i=1}^{n} P(X = x_i; c, \beta)$, we obtain

$$l(c, \beta; \mathbf{x}) = -c\sum_{i=1}^{n}\sum_{j=1}^{x_i} j^{\beta} + \sum_{i=1}^{n} \log\left[1 - e^{-c(x_i+1)^{\beta}}\right]. \tag{13}$$
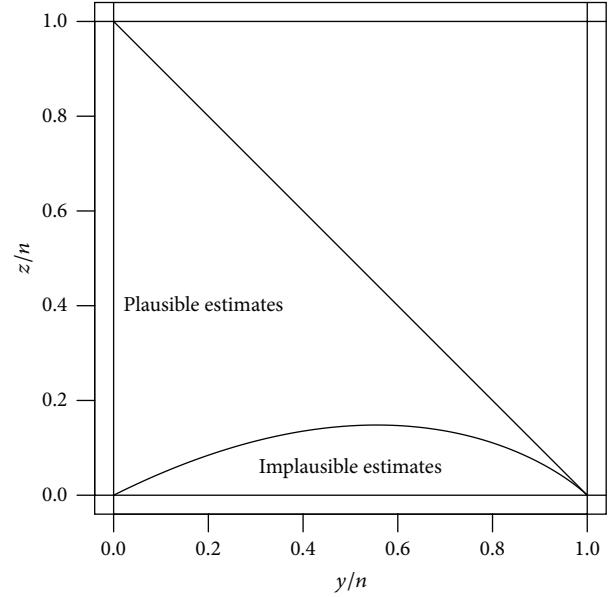

Figure 2: "Plausible" and "implausible" regions for the estimate of β through the method of proportion.

The maximum likelihood estimates of c and β ($\widehat{c}_{\mathrm{ML}}, \widehat{\beta}_{\mathrm{ML}}$) are defined as the values that maximize the log-likelihood function

$$\left(\widehat{c}_{\mathrm{ML}}, \widehat{\beta}_{\mathrm{ML}}\right) = \arg\max l\left(c, \beta; \mathbf{x}\right). \tag{14}$$

The two first derivatives of $l(c, \beta; \mathbf{x})$ are quite easily computed,

$$\begin{aligned}
\frac{\partial l\left(c, \beta; \mathbf{x}\right)}{\partial c} &= \sum_{i=1}^{n}\left[\frac{(x_i + 1)^{\beta} e^{-c(x_i+1)^{\beta}}}{1 - e^{-c(x_i+1)^{\beta}}} - \sum_{j=1}^{x_i} j^{\beta}\right] \\
&= \sum_{i=1}^{n} \frac{e^{-c(x_i+1)^{\beta}} \sum_{j=1}^{x_i+1} j^{\beta} - \sum_{j=1}^{x_i} j^{\beta}}{1 - e^{-c(x_i+1)^{\beta}}}, \\
\frac{\partial l\left(c, \beta; \mathbf{x}\right)}{\partial \beta} & \\
&= c\sum_{i=1}^{n} \frac{e^{-c(x_i+1)^{\beta}} \sum_{j=1}^{x_i+1} j^{\beta} \log j - \sum_{j=1}^{x_i} j^{\beta} \log j}{1 - e^{-c(x_i+1)^{\beta}}},
\end{aligned} \tag{15}$$

but, as already noted in Padgett and Spurrier [6], the solution to the maximization of $l(c, \beta; \mathbf{x})$ (with the constraints that c and β belong to in their natural parameter spaces) cannot be derived in a closed form, by equating the expressions in (15) to zero, but can be obtained only numerically, for example, using one of the functions `nlm`, `optim`, and `Rsolnp` in the R programming environment, which allows the user to solve nonlinear constrained or unconstrained minimization/maximization problems. Even this method cannot be successfully applied to every possible sample; in particular, the method fails in providing a solution if $y + z = n$ (i.e., if the sample contains only 0's and 1's). In this case, in fact, it can be easily proved that the first-order β derivative of

the log-likelihood is never null; the log-likelihood does not have an absolute maximum in the parameter space. Let us consider, for example, the following sample: $\mathbf{x} = (0, 1)$. The corresponding likelihood function is given by

$$L(\mathbf{x}; c, \beta) = \left(1 - e^{-c}\right) e^{-c} \left(1 - e^{-c2^{\beta}}\right) \qquad (16)$$

and the log-likelihood

$$l(c; \beta) = -c + \log\left(1 - e^{-c}\right) + \log\left(1 - e^{-c2^{\beta}}\right). \qquad (17)$$

The first-order derivatives are

$$\frac{\partial l(c, \beta; x_i)}{\partial c} = -1 + \frac{e^{-c}}{1 - e^{-c}} + \frac{2^{\beta} e^{-c2^{\beta}}}{1 - e^{-c2^{\beta}}}, \qquad (18a)$$

$$\frac{\partial l(c, \beta; x_i)}{\partial \beta} = \frac{e^{-c2^{\beta}} c 2^{\beta} \log 2}{1 - e^{-c2^{\beta}}}. \qquad (18b)$$

The derivative (18b) is never null but tends to zero for $\beta \to +\infty$; when $\beta \to +\infty$ the third addend in the derivative (18a) goes to zero, and solving the equation $-1 + \exp(-c)/[1 - \exp(-c)] = 0$ one can obtain $\widehat{c}^{*}_{\mathrm{ML}} = \log 2$. Note that this is the same estimate the method of proportion supplies.

Method of Moments. Through this method, the parameter estimates are obtained solving, in terms of c and β, the equations $\mathbb{E}(X) = \mu_1 = m_1$ and $\mathbb{E}(X^2) = \mu_2 = m_2$, where m_1 and m_2 are the first and second sample moments: $m_1 = \sum_{i=1}^{n} x_i/n$, $m_2 = \sum_{i=1}^{n} x_i^2/n$. Since they cannot be solved analytically, as suggested in Khan et al. [9], one can numerically minimize, with respect to c and β, the following quadratic loss function:

$$\mathscr{L}(c, \beta) = \left(m_1 - \mu_1\right)^2 + \left(m_2 - \mu_2\right)^2. \qquad (19)$$

The task can be carried out, for example, again using the functions `nlm`, `optim`, or `solnp` in the R environment [11]. The solution is the couple $(\widehat{c}_M, \widehat{\beta}_M)$, which should correspond to the value $\mathscr{L}(\widehat{c}_M, \widehat{\beta}_M) = 0$. Note that the minimization should be subject to the natural constraints $c > 0$ and $\beta > -1$. Without these constraints, the minimization algorithm may get stuck in implausible intermediate-step solutions. Indeed, the `optim` function under the R environment seems to provide excellent results in terms of convergence to the optimal solution, even without setting the constraints on c and β. As to the launch values for c and β, required by any minimization function, one can set $\beta = \beta_0 = 0$ and $c = c_0 = \log[(1 + \overline{x})/\overline{x}]$. This pair of values is always computable and feasible (unless the sample contains all 0's) and ensures at the first iteration that $m_1 = \mathbb{E}(X)$: in fact, recalling (6), $\mathbb{E}(X; c_0, \beta_0) = \sum_{i=1}^{+\infty} \exp(-c_0 \sum_{j=1}^{i} 1) = \sum_{i=1}^{+\infty} \exp(-ic_0) = \exp(-c_0)/[1 - \exp(-c_0)] = \overline{x}$.

When the sample contains only 0's and 1's, the method of moments is not applicable. To see this, first consider the probability mass function of the discrete Weibull r.v., and note that by letting β tend to $+\infty$ in (3), it degenerates into a r.v. that takes only two values: 0 with probability $1 - \exp(-c)$ and

1 with probability $\exp(-c)$. It is then clear that if the sample is made up of a fraction r of 0's and a fraction $(1 - r)$ of 1's, the equality of both first and second moments computed on the sample and on the original r.v. holds for $1 - \exp(-c) = r$; that is, $c = -\log(1 - r)$ and $\beta \to +\infty$. In this case, the loss function $\mathscr{L}(c, \beta)$ does not admit an absolute minimum, it only admits an inferior limit.

For some samples, the numerical minimization procedure can require a huge computation time, much larger than that required by the maximum likelihood method. This is due to the iterated calculation of the first and second moments, which is itself numerical and particularly time consuming for the negative values of β (in this case, in fact, the convergence of the series in (6) and (7) is slower).

Given the complexity of the estimators derived through the methods listed in this section (and only the method of proportion provides an analytical expression for them), not as much can be analytically derived about their statistical properties for finite sample size, that is, bias and variability. When the method of proportion can be applied, it provides a consistent estimator for c; $\widehat{\beta}_P$ is consistent as well, but nothing can be said about their unbiasedness (see the analogous discussion in Khan et al. [9] for type I discrete Weibull). For large samples, the general properties of the estimators derived from the maximum likelihood method and the method of moments can be recalled. In the next section, an extensive simulation study is presented, which was performed to investigate the performance of the estimation methods and outline some practical advice for their use.

3. Simulation Study

The estimators presented in the previous section were investigated through an extensive Monte Carlo study; they were compared in terms of bias (B), defined as $B(\widehat{\theta}_E) = \mathbb{E}(\widehat{\theta}_E) - \theta$, and root mean square error (RMSE), defined as $\mathrm{RMSE}(\widehat{\theta}_E) = \sqrt{\mathbb{E}[(\widehat{\theta}_E - \theta)^2]}$, where θ denotes one of the two parameters (c or β) and $\widehat{\theta}_E$ denotes one of the three corresponding estimators, according to the method indicated by the subscript E (method of proportion, P; maximum likelihood method, ML and method of moments, M).

3.1. Simulation Design. In this study, several parameter combinations and sample sizes ($n = 20, 50, 100$) were considered. The values (c, β) were chosen in order to explore a large spectrum of the discrete Weibull distribution, in particular to comprise increasing, constant, and decreasing failure rates. At the same time, the parameters were set in order to keep the discrete nature of the distribution reasonable: values entailing a nonnegligible probability for a large number of integer values were deliberately excluded (in this case, a continuous r.v. should be preferred to model failure data); parameter values ensuring a nonnegligible probability for just the first integers were avoided as well (one assumes the component that is monitored is likely to last for more than 1 or 2 cycles). Table 1 shows the combinations of values of c and β examined in the simulation study along with

TABLE 1: Parameter combinations and corresponding expected value, standard deviation, and 0.99-quantile for the simulation study.

c	β	$\mathbb{E}(X)$	SD(X)	$F^{-1}(0.99)$
0.1	1	3.01	2.06	9
0.1	0.75	3.64	2.67	11
0.1	0.5	4.61	3.69	16
0.1	0.25	6.26	5.61	25
0.2	0.75	2.18	1.78	7
0.2	0.5	2.62	2.30	10
0.2	0.25	3.30	3.19	14
0.2	0	4.52	4.99	23
0.4	0.5	1.38	1.42	6
0.4	0.25	1.63	1.80	8
0.4	0	2.03	2.48	11
0.4	−0.25	2.76	3.95	18
0.6	0.25	1.03	1.28	5
0.6	0	1.22	1.64	7
0.6	−0.25	1.53	2.35	11
0.6	−0.5	2.19	4.21	20
0.8	0	0.82	1.22	5
0.8	−0.25	0.98	1.62	7
0.8	−0.5	1.29	2.56	12
0.9	−0.5	1.04	2.10	10

the corresponding expected value, standard deviation, and 99% quantile. A note about the computation of the expected value and standard deviation is due: they are calculated numerically—see formulas (6) and (7)—considering the first n_{max} integers, with $n_{max} = 2F^{-1}(1 - \epsilon)$ and ϵ as small as possible (here, $\epsilon = 0.0001$, which actually proved to be a satisfactory practical choice for the considered scenarios).

3.2. Software Implementation. The simulation study was based on 5,000 Monte Carlo replications and carried out under the R environment [11]. In particular, the necessary code was developed to implement the probability mass function, the cumulative distribution function and its inverse, and the random generation for the discrete Weibull model; to compute the first and second moments; and to to realize the algorithms corresponding to each estimation method. This code, structured as an R package, `DiscreteWeibull`, is freely available in the CRAN repository [12].

3.3. Simulation Results. Table 2 shows the bias and root mean square error for both estimators derived from each of the three methods, under each combination of the two parameters and for each sample size. The results took into account the applicability limits of the three methods, computing the Monte Carlo averages over the feasible samples only. For the smallest sample size ($n = 20$) and, more rarely, for $n = 50$, the method of proportion met a certain percentage of infeasible samples under several scenarios, the highest (about 24%) with $n = 20$, $c = 0.1$, and $\beta = 0.25$. The maximum likelihood method and the method of moments encountered only a few

infeasible samples under some scenarios, whose rate was in any case always smaller than 1%.

Note that strictly speaking, to compare the behavior of the estimators under different scenarios, one should use their relative bias and relative root mean square error as performance indexes, that is, the bias and the root mean square error divided by the MC expected value of the estimator. This way of proceeding would provide a more correct "double" reading of the results, assessing the performance of a single estimator moving through different scenarios and comparing the performance of the three estimators under a fixed scenario. Our choice of using "absolute" instead of "relative" indexes was partially dictated by the necessity of including negative and null values for β, which would represent difficulty when computing the relative indexes (the denominator could take very small values, dramatically amplifying the corresponding absolute value of the index and making the reading arduous).

Trying to synthesize the results, let us start with the largest sample size ($n = 100$) and the estimators of parameter c. It is quite evident that all three estimators show a modest bias, regardless of the values of c and β. The maximum likelihood method looks the best in this sense, while the method of moments shows the largest bias in absolute value for high values of c and negative values of β, and it is often negatively biased. As to the root mean square error, the method of moments and the maximum likelihood method present very similar values under each scenario; the method of proportion shows a larger value, especially for small values of c. The magnitude of RMSE seems to be much more affected by the value of c rather than by β for all the estimators.

As to the estimators of β, on average, the method of proportion provides the least biased estimator, unless c is too small, while the maximum likelihood method and, to a larger extent, the method of moments are significantly biased. While M and ML estimators always overestimate the true value of the parameter, the P estimator seems to underestimate it for lower values. In absolute value, the bias of the ML estimator increases as c increases for a fixed β; it increases as β increases for a fixed c. The bias of the M estimator increases as c increases for a fixed β; it increases as β decreases for a fixed c. In terms of variability, the M and ML estimators perform in a similar way for each parameter combination, while the P estimator is much more variable and presents very different values moving through the scenarios: its RMSE is from 2 to 6 times larger than the corresponding value of its competitors. The worst scenarios for the estimator, in this sense, are those connected with $c = 0.1$, and this is quite reasonable. In fact, in this case, the probability $P(X = 0)$ equals 0.095, which corresponds to the expected fraction of zeros in the sample, and this small value leads to a modest performance of the P method, which uses only proportions of 0's and 1's of the samples and discards the other information.

Decreasing the sample size to 50 and 20, the bias in absolute value and the root mean square error of the estimators for each scenario obviously increase. However, the behaviors and trends exposed for $n = 100$ hold still. The bias of the estimators of β derived by the method of moments

TABLE 2: Simulation results: bias (B) and root mean square error (RMSE) for the estimators of c and β.

c	P		ML		M		β	P		ML		M	
	B	RMSE	B	RMSE	B	RMSE		B	RMSE	B	RMSE	B	RMSE
							$n = 100$						
0.1	0.001	0.032	0.000	0.024	0.001	0.025	1	0.033	0.636	0.028	0.181	0.020	0.183
0.1	0.001	0.032	0.000	0.024	0.001	0.025	0.75	0.025	0.658	0.025	0.162	0.017	0.165
0.1	0.001	0.032	0.000	0.023	0.001	0.024	0.50	0.014	0.683	0.022	0.143	0.017	0.145
0.1	0.001	0.032	0.000	0.022	0.000	0.023	0.25	−0.001	0.718	0.019	0.124	0.019	0.127
0.2	0.001	0.048	0.000	0.040	0.000	0.040	0.75	0.009	0.475	0.029	0.179	0.029	0.177
0.2	0.001	0.048	0.000	0.039	−0.001	0.039	0.5	0.003	0.494	0.025	0.158	0.027	0.158
0.2	0.001	0.048	0.000	0.038	−0.002	0.038	0.25	−0.003	0.512	0.022	0.140	0.029	0.143
0.2	0.001	0.048	0.000	0.037	−0.003	0.038	0	−0.008	0.537	0.019	0.120	0.031	0.128
0.4	0.003	0.072	0.000	0.065	−0.001	0.065	0.5	0.000	0.377	0.030	0.187	0.033	0.186
0.4	0.003	0.072	0.000	0.064	−0.002	0.064	0.25	−0.004	0.395	0.027	0.167	0.035	0.168
0.4	0.003	0.072	0.000	0.063	−0.005	0.063	0	−0.010	0.412	0.025	0.147	0.038	0.152
0.4	0.003	0.072	−0.001	0.062	−0.010	0.064	−0.25	−0.017	0.430	0.023	0.127	0.047	0.139
0.6	0.004	0.093	0.000	0.087	−0.002	0.086	0.25	0.004	0.359	0.036	0.195	0.043	0.195
0.6	0.004	0.093	0.000	0.086	−0.005	0.086	0	−0.001	0.371	0.032	0.175	0.046	0.178
0.6	0.004	0.093	−0.001	0.085	−0.011	0.085	−0.25	−0.007	0.386	0.030	0.154	0.055	0.163
0.6	0.004	0.093	−0.001	0.083	−0.029	0.088	−0.5	−0.013	0.406	0.028	0.131	0.077	0.152
0.8	0.006	0.111	0.001	0.106	−0.005	0.105	0	−0.001	0.363	0.040	0.204	0.056	0.204
0.8	0.006	0.111	0.001	0.105	−0.011	0.105	−0.25	−0.005	0.379	0.038	0.182	0.064	0.188
0.8	0.006	0.111	0.000	0.104	−0.027	0.106	−0.5	−0.009	0.397	0.035	0.159	0.085	0.175
0.9	0.008	0.123	0.001	0.116	−0.027	0.116	−0.5	−0.012	0.398	0.040	0.176	0.091	0.188
							$n = 50$						
0.1	0.001	0.046	0.000	0.034	0.001	0.034	1	0.061	0.943	0.059	0.269	0.046	0.261
0.1	0.001	0.046	−0.001	0.033	0.001	0.033	0.75	0.040	0.978	0.053	0.240	0.045	0.236
0.1	0.001	0.046	−0.001	0.032	0.000	0.032	0.50	0.021	1.012	0.047	0.211	0.042	0.208
0.1	0.001	0.046	−0.001	0.031	−0.001	0.031	0.25	0.005	1.047	0.040	0.182	0.043	0.181
0.2	0.002	0.067	−0.001	0.055	−0.002	0.055	0.75	0.013	0.703	0.059	0.262	0.060	0.258
0.2	0.002	0.067	−0.001	0.054	−0.002	0.054	0.5	0.003	0.732	0.052	0.232	0.057	0.231
0.2	0.002	0.067	−0.002	0.053	−0.004	0.053	0.25	−0.014	0.761	0.047	0.205	0.059	0.209
0.2	0.002	0.067	−0.002	0.052	−0.007	0.052	0	−0.026	0.801	0.041	0.176	0.063	0.186
0.4	0.005	0.100	−0.002	0.091	−0.004	0.090	0.5	0.005	0.537	0.065	0.278	0.071	0.274
0.4	0.005	0.100	−0.002	0.089	0.072	0.088	0.25	−0.004	0.562	0.058	0.247	0.072	0.247
0.4	0.005	0.100	−0.002	0.088	−0.011	0.086	0	−0.014	0.588	0.053	0.218	0.078	0.223
0.4	0.005	0.100	−0.003	0.086	−0.020	0.086	−0.25	−0.027	0.626	0.048	0.189	0.091	0.203
0.6	0.007	0.131	−0.002	0.123	−0.006	0.120	0.25	0.005	0.518	0.074	0.291	0.089	0.287
0.6	0.007	0.131	−0.002	0.122	−0.011	0.119	0	−0.005	0.542	0.068	0.262	0.094	0.263
0.6	0.007	0.131	−0.003	0.120	−0.022	0.117	−0.25	−0.016	0.564	0.063	0.231	0.107	0.240
0.6	0.007	0.131	−0.004	0.117	−0.047	0.119	−0.5	−0.025	0.594	0.058	0.199	0.136	0.225
0.8	0.012	0.158	0.000	0.151	−0.010	0.147	0	−0.009	0.534	0.085	0.309	0.114	0.306
0.8	0.012	0.158	−0.001	0.149	−0.021	0.145	−0.25	−0.018	0.556	0.079	0.278	0.127	0.282
0.8	0.012	0.158	−0.002	0.147	−0.045	0.145	−0.5	−0.029	0.586	0.075	0.245	0.154	0.264
0.9	0.014	0.175	−0.002	0.165	−0.045	0.160	−0.5	−0.029	0.590	0.086	0.273	0.167	0.288
							$n = 20$						
0.1	0.004	0.075	0.000	0.055	0.002	0.054	1	−0.156	1.154	0.145	0.472	0.122	0.443
0.1	0.004	0.075	0.000	0.054	0.001	0.052	0.75	−0.125	1.139	0.129	0.421	0.111	0.388
0.1	0.004	0.075	0.000	0.053	0.000	0.050	0.50	−0.082	1.128	0.113	0.368	0.105	0.336
0.1	0.004	0.075	0.000	0.051	−0.003	0.047	0.25	−0.013	1.103	0.098	0.317	0.105	0.292

TABLE 2: Continued.

c	P		ML		M		β	P		ML		M	
	B	RMSE	B	RMSE	B	RMSE		B	RMSE	B	RMSE	B	RMSE
0.2	0.008	0.110	−0.001	0.090	−0.002	0.087	0.75	0.011	1.126	0.141	0.455	0.143	0.441
0.2	0.008	0.110	−0.001	0.088	−0.004	0.084	0.5	0.006	1.153	0.124	0.402	0.136	0.389
0.2	0.008	0.110	−0.001	0.086	−0.008	0.081	0.25	0.005	1.165	0.110	0.354	0.135	0.346
0.2	0.008	0.110	−0.002	0.084	−0.014	0.077	0	0.027	1.163	0.097	0.304	0.141	0.306
0.4	0.013	0.163	−0.002	0.147	−0.006	0.142	0.5	0.004	0.912	0.156	0.476	0.170	0.462
0.4	0.013	0.163	−0.002	0.145	−0.012	0.138	0.25	−0.018	0.948	0.140	0.426	0.169	0.416
0.4	0.013	0.163	−0.003	0.142	−0.021	0.133	0	−0.040	0.979	0.129	0.378	0.177	0.375
0.4	0.013	0.163	−0.004	0.139	−0.035	0.128	−0.25	−0.038	0.992	0.115	0.328	0.192	0.340
0.6	0.023	0.216	0.001	0.199	−0.010	0.189	0.25	−0.017	0.883	0.171	0.499	0.202	0.482
0.6	0.023	0.214	−0.001	0.196	−0.020	0.184	0	−0.043	0.922	0.163	0.458	0.214	0.447
0.6	0.023	0.216	−0.002	0.193	−0.037	0.179	−0.25	−0.056	0.949	0.149	0.405	0.228	0.407
0.6	0.023	0.216	−0.004	0.190	−0.072	0.179	−0.5	−0.063	0.972	0.138	0.354	0.263	0.383
0.8	0.031	0.265	0.002	0.247	−0.019	0.233	0	−0.025	0.885	0.199	0.534	0.252	0.517
0.8	0.031	0.265	0.001	0.244	−0.036	0.227	−0.25	−0.035	0.905	0.191	0.494	0.274	0.487
0.8	0.031	0.265	−0.002	0.241	−0.069	0.225	−0.5	−0.052	0.935	0.182	0.443	0.306	0.459
0.9	0.038	0.292	0.000	0.267	−0.068	0.250	−0.5	−0.029	0.914	0.205	0.492	0.331	0.502

P: method of proportion, M: method of moments, and ML: maximum likelihood method.

and the maximum likelihood method becomes substantial especially for large values of c and negative values of β; in these cases, the method of proportion is less biased, but its RMSE is still much larger than those of its competitors. For high values of c (viz., equal to or larger than 0.8) and $n = 20$, the bias in absolute value of the estimators of c for the method of proportion and the method of moments tends to become much more substantial than that for the method of maximum likelihood.

Looking at the overall values of RMSE for the estimators of c and β obtained through the three methods, it is also evident that much more uncertainty is associated with the point estimation of the second parameter.

Figure 3 shows the MC distributions of the estimators of c and β for $n = 50$ and three combinations of parameters. From the analysis, it emerges that the positive bias of $\widehat{\beta}_{\mathrm{ML}}$ and $\widehat{\beta}_M$ is due to the presence of a certain number of samples providing estimates much larger than the true value of β, while the MC medians of both estimators are very close to it. These boxplots also emphasize the presence of implausible values for the estimates of β yielded by the method of proportion ($\widehat{\beta}_P < -1$).

Trying to synthesize all the results presented here—which are not, however, exhaustive—the method of proportion, despite its straightforward analytical derivation of the estimators, performs worse overall (especially in terms of variability) than the method of moments and the maximum likelihood method, and it is competitive only for some specific scenarios, namely, for medium/high values of c, where it can better exploit the information contained in the sample.

4. Applications to Real Data

The methods that have been illustrated and empirically investigated in the previous sections are applied to two

TABLE 3: Empirical frequency distributions for the first example dataset.

cat.	0	1	2	3	4	5	6	7	8	9	10	11
f_i	20	10	11	10	2	3	3	0	0	1	1	1

TABLE 4: Estimates of the parameters of the discrete Weibull for the first example dataset.

		Point estimators			
$\widehat{c}_P$	$\widehat{\beta}_P$	$\widehat{c}_{\mathrm{ML}}$	$\widehat{\beta}_{\mathrm{ML}}$	$\widehat{c}_M$	$\widehat{\beta}_M$
0.389	−0.518	0.361	0.068	0.356	0.083

datasets. The first one, called J1 [13], contains the number of failures of software observed over 62 weeks, and its frequency distribution is reported in Table 3. Assuming that the statistical distribution underlying the data is a type III discrete Weibull, it is possible to compute the estimates yielded by the three estimators described in Section 3, which are reported in Table 4. Note that the method of moments and the maximum likelihood method provide very similar estimates for c and β; the method of proportion supplies an estimate of β that is quite different from the other two methods (moreover, it is negative), while the estimate of c, $\widehat{c}_P$ is quite close to $\widehat{c}_M$ and $\widehat{c}_{\mathrm{ML}}$. If we want to test the goodness of fit of the discrete Weibull model for the data, we can use the chi-square statistic $T = \sum_{i=1}^{k}(f_i - np_i)^2/(np_i)$, where f_i denotes the category frequencies and p_i denotes the probability of an observation falling into the ith category under the study model, such that $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} f_i = n$. If the model holds, T is asymptotically distributed as a chi-square with $k - 1 - p$ degrees of freedom, where p is the number of parameters to be estimated (in this case, $p = 2$). Before computing the empirical value of T, under each model we have to group
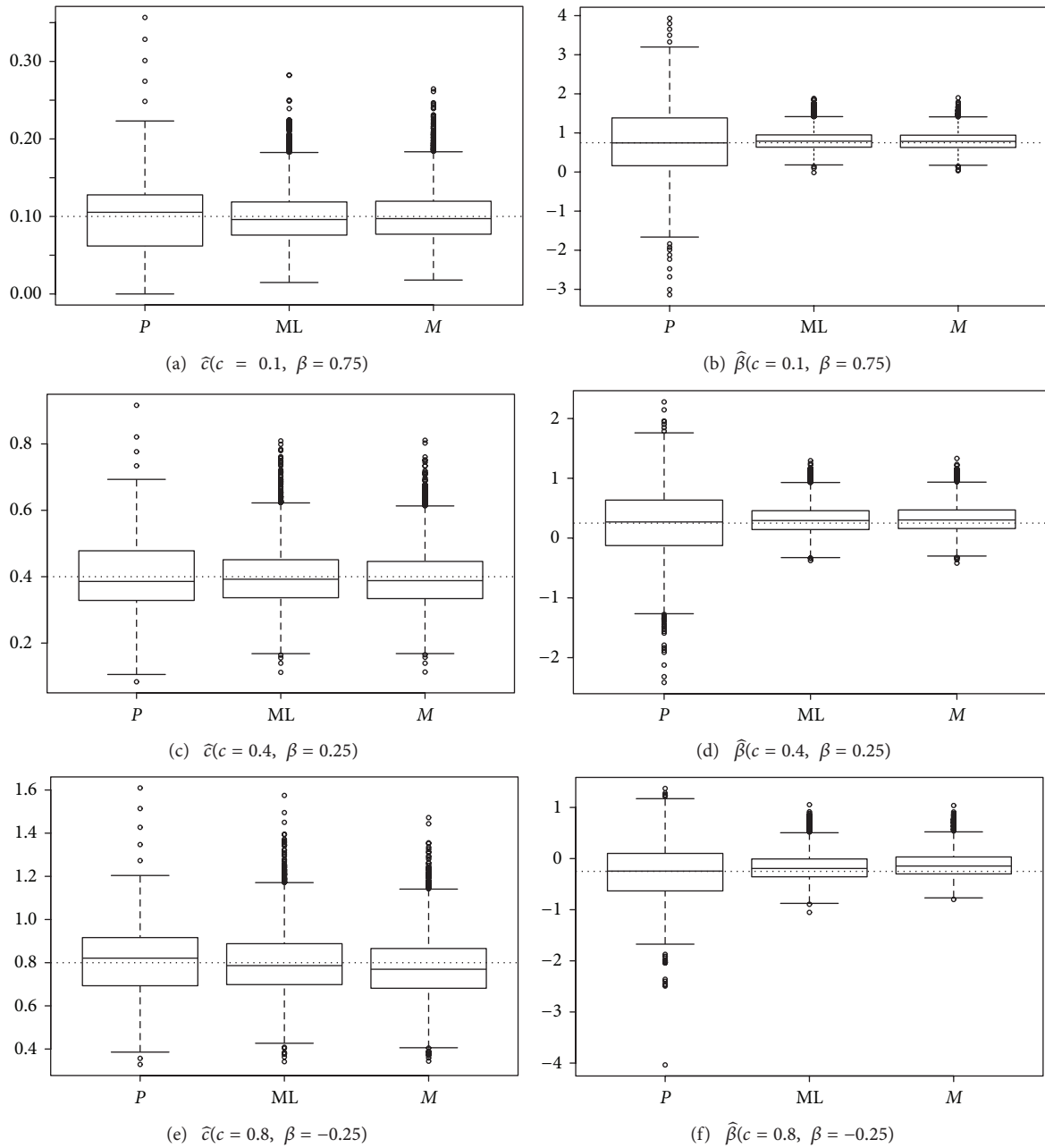
(a) $\widehat{c}(c = 0.1, \beta = 0.75)$

(b) $\widehat{\beta}(c = 0.1, \beta = 0.75)$

(c) $\widehat{c}(c = 0.4, \beta = 0.25)$

(d) $\widehat{\beta}(c = 0.4, \beta = 0.25)$

(e) $\widehat{c}(c = 0.8, \beta = -0.25)$

(f) $\widehat{\beta}(c = 0.8, \beta = -0.25)$

FIGURE 3: Boxplot of Monte Carlo distributions of $\widehat{c}$ and $\widehat{\beta}$ for three parameter combinations ($n = 50$ and 5,000 MC runs). The dashed horizontal lines indicate the true value of the parameters c and β.

the categories in such a way that all the expected frequencies np_i are not smaller than 5. Table 5 reports these groupings. The goodness-of-fit chi-square test accepts the type III Weibull model with the ML and M parameter estimates (statistic's value of 3.8843 and P value of 0.2742 for ML, 3.8514 and 0.2780 for M), while it refuses the model with the P parameter estimates (statistic's value of 10.2323 and P-value of 0.0060) at a 5% nominal level of significance. Again, there is a notable difference between the results for the method of proportion and the maximum likelihood method and the method of moments.

The second example considers the data regarding accidents to 647 women working on H. E. Shells during five weeks [14], whose frequency distribution is reported in Table 6. Even if these are not strictly failure data, in the meaning explained in Section 1, nevertheless they are count data, and the discrete Weibull r.v. can be used to model them. The parameter estimates yielded by the three methods are reported in Table 7. The estimates of c are again very close to each other; all the estimates of β are negative and quite small in absolute value. Note that all the pairs $(\widehat{c}, \widehat{\beta})$ fall outside the scenarios explored in the simulation study; nevertheless,

TABLE 5: Empirical and theoretical frequency distributions for the first example dataset.

cat.	0	1	2	3-4	≥5	
Empirical	20	10	11	12	9	
P	20.0	10.0	6.3	7.7	17.9	
cat.	0	1	2	3	4-5	≥6
Empirical	20	10	11	10	5	6
ML	18.8	13.6	9.5	6.6	7.5	6.0
M	18.6	13.6	9.6	6.6	7.6	5.9

TABLE 6: Empirical frequency distribution for the second example dataset.

cat.	0	1	2	3	4	5
f_i	447	132	42	21	3	2

TABLE 7: Estimates of the parameters of the discrete Weibull for the second example dataset.

Point estimators					
$\hat{c}_P$	$\hat{\beta}_P$	$\hat{c}_M$	$\hat{\beta}_M$	$\hat{c}_{ML}$	$\hat{\beta}_{ML}$
1.174	−0.122	1.153	−0.0169	1.162	−0.0474

the empirical distribution of the data (many 0's and 1's) is favorable to the method of proportion, and the large sample size ($n = 647$) should ensure that all three methods are reliable. Testing the goodness-of-fit of the type III discrete Weibull model with the ML parameters, which requires grouping the last two categories (then $k = 5$), the value of the chi-square statistic with $k − 1 − p = 2$ degrees of freedom is 3.9595, and the corresponding P-value is 0.1381. Note that this model fits the data better than the natural negative binomial model considered by Greenwood and Yule [14] and the generalized Poisson model proposed by Consul and Jain [15].

5. Conclusion

This paper examined three estimators for the parameters of the type III discrete Weibull random variable, which represents an alternative distribution to the geometric and negative binomial for modeling discrete reliability data, and can ensure increasing and decreasing failure rates. One of the methods presented (method of proportion) has recently been introduced in discrete models, and here it is newly phrased; it provides a closed form for the estimates of both parameters. The other two methods (maximum likelihood method and method of moments) are standard approaches for estimating parameters, but due to the complex expression of the probability mass function, they provide the estimates as a numerical solution to a minimization/maximization problem. It follows that not as much can be said about the statistical properties of the estimators for a finite sample size; then, an extensive Monte Carlo simulation study was carried out to assess their behavior. Far from giving a definitive solution to the problem, the study highlighted that the method of proportion, when applicable, can provide reliable estimates

even for small sample sizes only under specific parameter configurations, whereas under other configurations it may provide poor results, especially in terms of the accuracy of the estimator of the second parameter. The other two standard methods can be usefully adopted under most scenarios; caution is necessary since some parameter combinations may lead to nonnegligible bias of the corresponding estimators. The paper stressed the potential practical applicability limits of each method, also in terms of computational burden, and illustrated their use through two examples with real data.

Appendix

A. Existence of First- and Second-Order Moments

Let us denote with a_n, $n \geq 1$, the general element of the series in (6), $a_n = \exp[-c\sum_{j=1}^n j^\beta]$. $\mathbb{E}(X)$ exists finite if and only if the series is convergent. If $\beta \geq 0$, the series is convergent according to the ratio criterion, since the limit $l_1 = \lim_{n\to\infty} a_{n+1}/a_n = \lim_{n\to\infty} \exp[-c(n+1)^\beta]$ is 0 if $\beta > 0$ and is $\exp(-c) < 1$ if $\beta = 0$.

If $-1 < \beta < 0$, note that $\sum_{j=1}^n j^\beta > \int_1^n j^\beta dj = n^{\beta+1}/(\beta+1) - 1/(\beta+1)$, and then

$$e^{-c\sum_{j=1}^n j^\beta} < e^{-c[n^{1+\beta}/(1+\beta)-1/(1+\beta)]}$$
$$= e^{c/(1+\beta)}e^{-cn^{1+\beta}/(1+\beta)}. \tag{A.1}$$

The series with the minoring term is convergent (e.g., using the comparison criterion with the harmonic converging series) and thus the original series in (6), too.

If $\beta = -1$, note that $\sum_{j=1}^n 1/j > \int_1^n 1/j \, dj = \log n$ and then

$$e^{-c\sum_{j=1}^n 1/j} < e^{-c\log n} = n^{-c}. \tag{A.2}$$

The series with the minoring term is convergent for $c > 1$ and thus the original series in (6) is too. Moreover, since $\int_1^n 1/j \, dj = \log n > \sum_{j=2}^n 1/j$ for $n \geq 2$

$$e^{-c\sum_{j=1}^n 1/j} = e^{-c(1+\sum_{j=2}^n 1/j)}$$
$$= e^{-c}e^{-c\sum_{j=2}^n 1/j} > e^{-c}e^{-c\log n} = e^{-c}n^{-c}. \tag{A.3}$$

The series with the minoring term is divergent for $c \leq 1$ and thus the original series in (6) is too.

If $\beta < -1$, then $\lim_{n\to\infty} a_n$ is greater than zero, and the series is divergent.

Let us now denote with b_n, $n \geq 1$, the general element of the series in (7), $b_n = 2n\exp[-c\sum_{j=1}^n j^\beta]$; then $\mathbb{E}(X^2) = \sum_{n=1}^\infty b_n - \mathbb{E}(X)$. If $\beta \geq 0$, the series is convergent according to the ratio criterion, since the limit $l_2 = \lim_{n\to\infty} b_{n+1}/b_n = \lim_{n\to\infty}(n+1)/n \cdot \exp[-c(n+1)^\beta]$ is 0 if $\beta > 0$ and is $\exp(-c) < 1$ if $\beta = 0$; thus $\mathbb{E}(X^2)$ is finite.

If $-1 < \beta < 0$,

$$2ne^{-c\sum_{j=1}^n j^\beta} < 2ne^{-c[n^{1+\beta}/(1+\beta)-1/(1+\beta)]}$$
$$= 2e^{c/(1+\beta)}ne^{-cn^{1+\beta}/(1+\beta)}. \tag{A.4}$$

The series with the minoring term is convergent (e.g., using the comparison criterion with the harmonic converging series) and thus the original series in (7) is too. Then $\mathbb{E}(X^2)$ is finite.

If $\beta = -1$,

$$2ne^{-c\sum_{j=1}^{n} 1/j} < 2ne^{-c\log n} = 2n^{-c+1}. \tag{A.5}$$

The series with the minoring term is convergent for $c > 2$ and thus the original series in (7) is too. Then $\mathbb{E}(X^2)$ is finite. Moreover, for $n \geq 2$,

$$
\begin{aligned}
2ne^{-c\sum_{j=1}^{n} 1/j} &= 2ne^{-c(1+\sum_{j=2}^{n} 1/j)} = 2e^{-c}ne^{-c\sum_{j=2}^{n} 1/j} \\
&> 2e^{-c}ne^{-c\log n} = 2e^{-c}n^{-c+1}.
\end{aligned}
\tag{A.6}
$$

The series with the minoring term is divergent for $c \leq 2$, and thus the original series in (7) is too. $\mathbb{E}(X^2)$ is not finite.

For $\beta < -1$, $\mathbb{E}(X^2)$ is not finite, since $\mathbb{E}(X)$ is not finite and $\mathbb{E}(X^2) \geq \mathbb{E}(X)$.

B. Derivation of $\widehat{\beta}_P$

Equating the probability $P(X = 1) = \exp(-c)[1 - \exp(-c2^{\beta})]$ to the sample proportion z/n, we get

$$\exp\left(-c2^{\beta}\right) = 1 - \frac{z}{n}\exp\left(c\right), \tag{B.1}$$

and after taking the natural logarithm of both sides of the equation twice, the second time after changing signs, we obtain

$$\beta = \frac{\log\left\{-\log\left[1 - (z/n)\exp\left(c\right)\right]/c\right\}}{\log 2}. \tag{B.2}$$

Substituting c with its estimate (8) in (B.2) we get

$$\beta = \frac{\log\left\{\log\left[1 - (z/n)/\left(1 - y/n\right)\right]/\log\left(1 - y/n\right)\right\}}{\log 2}, \tag{B.3}$$

and further simplifying it

$$\beta = \frac{\log\left[\log\left(1 - y/n - z/n\right)/\log\left(1 - y/n\right) - 1\right]}{\log 2}, \tag{B.4}$$

which represents the estimate $\widehat{\beta}_P$.

Acknowledgments

References

[1] A. Khalique, "On discrete failure-time distributions," *Reliability Engineering and System Safety*, vol. 25, no. 2, pp. 99–107, 1989.

[2] C. Bracquemond and O. Gaudoin, "A survey on discrete lifetime distributions," *International Journal of Reliability, Quality and Safety Engineering*, vol. 10, no. 1, pp. 69–98, 2003.

[3] D. N. Prabhakar Murthy, M. Xie, and R. Jiang, *Weibull Models*, John Wiley & Sons, Hoboken, NJ, USA, 2004.

[4] T. Nakagawa and S. Osaki, "The discrete Weibull distribution," *IEEE Transactions on Reliability*, vol. 24, no. 5, pp. 300–301, 1975.

[5] W. E. Stein and R. Dattero, "A new discrete Weibull distribution," *IEEE Transactions on Reliability*, vol. 33, no. 2, pp. 196–197, 1984.

[6] W. J. Padgett and J. D. Spurrier, "Discrete failure models," *IEEE Transactions on Reliability*, vol. 34, no. 3, pp. 253–256, 1985.

[7] H. Rinne, *The Weibull Distribution—A Handbook*, CRC Press, Boca Raton, Fla, USA, 2009.

[8] D. Roy and T. Dasgupta, "Evaluation of reliability of complex systems by means of a discretizing approach Weibull set-up," *International Journal of Quality and Reliability Management*, vol. 19, no. 6, pp. 792–801, 2002.

[9] M. S. A. Khan, A. Khalique, and A. M. Abouammoh, "On estimating parameters in a discrete Weibull distribution," *IEEE Transactions on Reliability*, vol. 38, no. 3, pp. 348–350, 1989.

[10] M. A. Jazi, C.-D. Lai, and M. H. Alamatsaz, "A discrete inverse Weibull distribution and estimation of its parameters," *Statistical Methodology*, vol. 7, no. 2, pp. 121–132, 2010.

[11] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011.

[12] A. Barbiero, "DiscreteWeibull: Discrete Weibull distribution," R package version 1.0, http://CRAN.R-project.org/package=DiscreteWeibull, 2013.

[13] A. P. Nikora and M. R. Lyu, "Software reliability measurement experience," in *Handbook of Software Reliability Engineering*, M. R. Lyu, Ed., McGraw-Hill, New York, NY, USA, 1996.

[14] M. Greenwood and G. U. Yule, "An enquiry into the nature of frequency distributions representative of multiple happenings," *Journal of the Royal Statistical Society A*, vol. 83, pp. 255–279, 1920.

[15] P. C. Consul and G. C. Jain, "A generalization of the Poisson distribution," *Technometrics*, vol. 15, pp. 791–799, 1973.