

AN APPLICATION OF THE POISSON-WEIBULL MODEL FOR CENSURED DATA

Valdemiro Piedade VIGAS¹

Fábio PRATAVIERA²

Giovana Oliveira SILVA³

■ **ABSTRACT:** In this paper, we proposed the Poisson-Weibull distribution for the modeling of survival data. The motivation to study this model since, in addition to generalizing the Weibull distribution, which is widely used in several areas of knowledge among them the Survival and Reliability analysis, it presents great flexibility in the forms of the hazard function. The Poisson-Weibull distribution was created in a composition of discrete and continuous distributions where there is no information about which factor was responsible for the component failure, only the minimum lifetime value among all risks is observed. The maximum likelihood approach was used to estimate the parameters of the model. Also was conducted a simulation study to examine the mean, the bias, and the root of the mean square error of the maximum likelihood estimates of the proposed model according to the censoring percentages and sample sizes. The model selection criteria were also applied, in addition to graphic techniques such as TTT-Plot and Kaplan-Meier. Application to the real data set was used to illustrate the usefulness of the distribution.

■ **KEYWORDS:** Censored data; distributions composition; model selection; survival analysis.

¹Universidade Federal de Mato Grosso do Sul - UFMS, Instituto de Matemática, CEP: 79070-900, Campo Grande, MS, Brasil. E-mail: *valdemiro.vigas@ufms.br*

²Universidade de São Paulo - USP, Escola Superior de Agricultura Luíz de Queirós, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: *fabioprataviera828@gmail.com*

³Universidade Federal da Bahia - UFBA, Instituto de Matemática e Estatística, CEP: 40170-110, Salvador, BA, Brasil. E-mail: *giovana@ufba.br*

1 Introduction

In several applications of statistics, the response variable consists of the time until the occurrence interest event, and this time is generally called failure time or survival time. In studies of failure time, is common the presence censored observations and which consists of incomplete observation of the time of interest. Data with censored observations are present in several areas, such as public health, actuarial science, biomedical studies, demography, reliability, among others. For example, in animal science, according to Cardoso *et al.* (2009), several characteristics of importance economic show censored data, that is, that are not fully observed for all animals at the time of genetic evaluation. Characteristics such as longevity, prolificacy, and total female productivity are examples of censored data, because many animals are still reproducing at the time of evaluation and, therefore, only the lower limit of their phenotypic value is known. Records of these characteristics are considered as censoring and excluding these records from analysis or manipulating them as uncensored would yield biased results (GUEDES, 2014).

Some statistical techniques used in animal science, such as linear or generalized linear mixed model, may not consider the censored observations during the modeling, which can affect in the inference process. In survival analysis data modeling, the most of procedures lead to consideration of the non-informative censoring, i.e, censoring time distribution do not carry any information about the failure time distribution to parameters estimation (DOS SANTOS JUNIOR, 2012). The non-informative censoring condition was considered in this work.

Several probability distributions can be used to model the survival time distribution, but despite the existence in the literature of several of these distributions, the demand for new distributions is justified by the fact that the usual models such as the exponential, Weibull (WEIBULL, 1958), log-logistic (VERHULST, 1938), and Gamma (THOM, 1947) distributions, which are an of the bigger distributions used in survival and reliability area, often do not fit well with the actual data set under analysis. In this context, new distributions arise from distributions most used in the literature, as the Weibull distribution, because this one does not provide a reasonable fit for modeling of data with non-monotone hazard functions such as the bathtub shaped and the unimodal hazard functions which are common in reliability and biological studies (BARRETO-SOUZA *et al.*, 2008).

Various methods have been studied in the literature to generate new distributions. Marshall and Olkin (1997) presented a method for adding a parameter with application to the exponential and Weibull distributions. Another form to create new distributions is through the composition of discrete with continuous distributions. This procedure to create new probability distributions starts from the fact that it is not always possible to observe the exact value of the survival time, but the minimum or maximum value of this times. This occurs, for example, when the interest is to observe the lifetime of a system in series or in parallel where the observed time depends on the duration of a component set.

In recent years, several authors considered this approach to create new distributions, among them we can cite Adamidis and Loukas (1998) that proposed the exponential geometric distribution (EG) which was created through the duration of a series system. Specifically, it is considered that both the number and the identification of the components that caused the failure are not observable, but only the value of the minimum lifetime among them, where the number of components follows a geometric distribution and the time of duration of each component follows an exponential distribution. Following the same idea as EG distribution, Kus (2007) introduced the exponential-Poisson distribution (EP) inserted also in a serial system. Chahkandi and Ganjali (2009) introduced exponential power series (EPS), which contains the distributions cited EG, EP, and logarithmic exponential as special cases. Barreto-Souza *et al.* (2008) propose the Weibull-geometric distribution (WG) that generalizes the EG and Weibull distributions. Louzada *et al.* (2011) introduced the complementary exponential geometric distribution (CEG) that generalizes the EG distribution where the motivation is the duration of a system in parallel on what only the maximum lifetime for all risks is known. Yamachi *et al.* (2013) presented the exponentiated complementary exponential geometric (ECEG) distribution which is a generalization of the CEG distribution. Louzada *et al.* (2012) and Louzada *et al.* (2014) propose the CEG and ECEG distributions in long-term scenarios, respectively. Recently, Louzada *et al.* (2018) introduced the beta exponentiated Weibull geometric distribution, which contains some distributions as particular cases.

Thus, through the considerations mentioned we used in this paper the Poisson-Weibull distribution (PW) defined by Bereta *et al.* (2011), presenting as an original contribution the modeling of data with censored observations via a simulation study to some levels of censoring besides the application of these data. The choice of PW model, in addition to generalizing the Weibull distribution, which is one of the most used distributions in the area of survival and recent years is applied to data in the

unit interval (MAZUCHELI *et al.*, 2020), presents great flexibility in the forms of the hazard function. This paper is organized as follows. In Section 2, we present the PW distribution. The inferential procedure based on the maximum likelihood approach is introduced in Section 3. In Section 4 we shown the selection criteria. A simulation study to assess the performance of the PW model parameters estimates is presented in Section 5. In Section 6, the data set is analyzed, and the final remarks appear in Section 7.

2 The Poisson-Weibull distribution (PW)

The PW distribution proposed by Bereta *et al.* (2011) was created in the scenario in which it is considered that both the number of components and the identification of the component that caused the failure are not observable. On the other hand, the value of the minimum lifetime among them is registered. In this case, the number of components follows a Poisson distribution and the time of duration of each component follows a Weibull distribution. However, the PW distribution can be used in any other situation that fits well with the data according to the criteria used. The density function is defined by

$$f(t) = \frac{\alpha \exp \{ \alpha \exp [- (\beta t)^\gamma] - (\beta t)^\gamma \} \beta^\gamma t^{\gamma-1} \gamma}{\exp(\alpha) - 1}, \quad (1)$$

where $t > 0$, $\beta > 0$ is scale parameter, $\alpha > 0$ and $\gamma > 0$ are shape parameters.

The quantile, survival and hazard functions corresponding to (1) are given by, respectively,

$$t = Q(u) = \frac{[\log(\alpha) - \log \{ \log \{ \exp(\alpha) (1 - u) \} + u \}]^{1/\gamma}}{\beta}, \quad (2)$$

$$S(t) = \frac{\exp \{ \alpha \exp [- (\beta t)^\gamma] \} - 1}{\exp(\alpha) - 1}, \quad (3)$$

and

$$h(t) = \frac{\alpha \exp \{ \alpha \exp [- (\beta t)^\gamma] - (\beta t)^\gamma \} \beta^\gamma t^{\gamma-1} \gamma}{\exp \{ \alpha \exp [- (\beta t)^\gamma] \} - 1}. \quad (4)$$

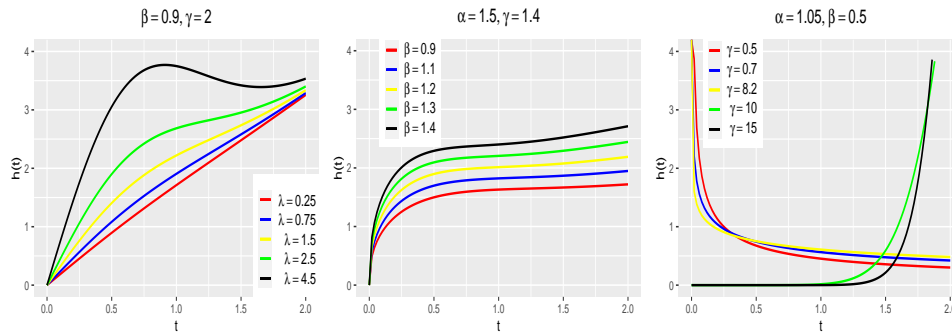


Figure 1 - Plots of the hazard function for PW distribution.

Figure 1 illustrates some of the possible shapes of the hazard function (4) for selected parameter values. Note in this figure, that the hazard function is quite flexible and can accommodate various forms, such as increasing, decreasing and unimodal. Applications of the PW distribution in survival studies for data without censored observations were investigated by Bereta *et al.* (2011).

3 Maximum likelihood estimation

Several methods can be used to estimate the probabilistic models parameters, and the most common is the maximum likelihood. One of the characteristics of this method is which allows the inclusion of censoring in its estimation process, which is not always possible with other estimation methods, for example, the least-squares method.

Maximum likelihood for PW distribution

Given a random sample of the size n composed by $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$ where $t_i = \min(T_i, C_i)$, T represents the survival time and C consists of the censoring time, which is independent of T . For each $i = 1, \dots, n$, the variable T_i has PW distribution with vector of parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^T$ and δ_i a random variable of censoring indicator, where the values 0 or 1 represent the censored or observed data, respectively. Consider the density and survival functions given by (1) and (3) respectively, the log-likelihood function $\ell(\boldsymbol{\theta})$ for the parameter vector $\boldsymbol{\theta}$, where the survival and censure times are independent and censure is non-informative, can be expressed as

$$\begin{aligned} \ell(\boldsymbol{\theta}) \propto & \sum_{i=1}^n \delta_i \log(\alpha \gamma \beta^\gamma t_i^{\gamma-1}) + \sum_{i=1}^n \delta_i [\alpha \exp\{-(\beta t_i)^\gamma\} - (\beta t_i)^\gamma] - \sum_{i=1}^n \delta_i \log(\exp(\alpha) - 1) + \\ & + \sum_{i=1}^n (1 - \delta_i) \log\{\exp\{\alpha \exp\{-(\beta t_i)^\gamma\} - 1\} - \sum_{i=1}^n (1 - \delta_i) \log(\exp(\alpha) - 1)\}. \end{aligned} \quad (5)$$

Maximum likelihood estimates (MLEs) for parameter vector $\boldsymbol{\theta}$ can be obtained by maximizing (5) from the resolution of the system of equations $U(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$. The components of the score vector $U(\boldsymbol{\theta})$ of the PW distribution are given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = & \frac{\sum_{i=1}^n \delta_i + \sum_{i=1}^n \delta_i \exp\{-(\beta t_i)^\gamma\}}{\alpha} - \frac{\sum_{i=1}^n \delta_i \exp(\alpha)}{\exp(\alpha) - 1} + \\ & + \sum_{i=1}^n (1 - \delta_i) \frac{(\exp\{\alpha \exp\{-(\beta t_i)^\gamma\}\}) \exp\{-(\beta t_i)^\gamma\}}{(\exp\{\alpha \exp\{-(\beta t_i)^\gamma\} - 1\}} + \frac{\sum_{i=1}^n (1 - \delta_i)}{\exp(\alpha) - 1}. \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} = & \sum_{i=1}^n \delta_i \frac{\gamma}{\beta} + \sum_{i=1}^n \delta_i \left(\frac{-\alpha (\beta t_i)^\gamma \gamma \exp\{-(\beta t_i)^\gamma\}}{\beta} - \frac{(\beta t_i)^\gamma \gamma}{\beta} \right) \\ & - \sum_{i=1}^n (1 - \delta_i) \frac{\alpha (\beta t_i)^\gamma \gamma (\exp\{\alpha \exp\{-(\beta t_i)^\gamma\})}{\beta (\exp\{\alpha \exp\{-(\beta t_i)^\gamma\} - 1)}. \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} = & \sum_{i=1}^n \delta_i \frac{\alpha \beta^\gamma t_i^{\gamma-1} + \alpha \gamma \beta^\gamma \log(\beta) t_i^{\gamma-1} + \alpha \beta^\gamma t_i^{\gamma-1} \log(t_i)}{\alpha \gamma \beta^\gamma t_i^{\gamma-1}} + \\ & + \sum_{i=1}^n \delta_i (-\alpha (\beta t_i)^\gamma \log(\beta t_i) \exp\{-(\beta t_i)^\gamma\} - (\beta t_i)^\gamma \log(\beta t_i)) \\ & - \frac{\sum_{i=1}^n (1 - \delta_i) \alpha (\beta t_i)^\gamma \log(\beta t_i) \exp\{\alpha \exp\{-(\beta t_i)^\gamma\} - (\beta t_i)^\gamma\}}{(\exp\{\alpha \exp\{-(\beta t_i)^\gamma\} - 1)}. \end{aligned}$$

Since there is no closed analytical way to find these estimators, we can use numerical methods for solving the system of equations. Thus, estimates of these parameters were obtained by numerical methods, using an iterative process. We used the command **optim** in software R through of the method BFGS.

4 Selection criteria

For the selection of the model that best fits the data, were used the criterion's: Global deviance (GD); AIC (Akaike's information criterion), and BIC (Bayesian information criterion). The GD , AIC , and BIC measures are defined by

$$GD = -2\ell(\hat{\theta}); \quad AIC = GD + 2p; \quad BIC = GD + 2\log(n),$$

where $\ell(\hat{\theta})$ is the maximized log-likelihood function, p is the number of parameters of the model and n is the sample size. Besides these criteria, hypothesis test, such as the likelihood ratio test (LR), can be taken into account due the WP distribution has other distributions such as particular cases. For test nested distributions, we compute the maximum values of the restricted (H_0) and unrestricted (H_1) log-likelihoods to construct the test statistic. Under H_0 and some regularity conditions, the distribution of the statistical likelihood ratio converges to a χ^2 distribution with degrees of freedom equal to the difference between the numbers of parameters of the unrestricted and restricted models.

For the comparison of the models in the boundary of the parameter space, for example, $H_0 : \alpha \rightarrow 0$ and $H_1 : \alpha > 0$, the distribution of the statistical test ω_n is a mixture with a weight (0.5 and 0.5) of distribution χ^2 with one degree of freedom, with a discrete distribution and with mass concentrated in the value 0, this is, $P(\omega_n \leq w) = \frac{1}{2} + \frac{1}{2}P(\chi_1^2 \leq w)$. Large positive values of ω_n give favorable evidence to the unrestricted model. For example, for a significance level of 5%, H_0 is rejected if $\omega_n > 2.7055$. More details in Maller *et al.* (2011).

5 Simulation study

To examine the performance of MLE's for the parameters of the PW distribution, a simulation study was made for different values of n and censored observations in each sample. In this study, were generating 1000 random samples simulated with the support of the software R. The values of the PW variable were generated from the inverse transformation method, where the parameter values were fixed in $\alpha = 1.0$, $\beta = 2.0$, $\gamma = 3.0$ and the censoring times C were sampled from the Uniform distribution in the interval $(0, \tau)$, where the τ assisted in the control of the censoring observations. Besides, the censoring percentages were considered equal to 0% ($\tau = 50.0$), 10% ($\tau = 3.6$) and 20% ($\tau = 2.0$) and the survival time observed of the variable in the simulation was considered through $t_i^* = \min(t_i, c_i)$. The process of this simulation is as follows:

1. Generate $u_i \sim U(0, 1)$;
2. Determine $t_i = F^{-1}(u_i) = \frac{(\log(\alpha) - \log(\log(\exp(\alpha)(1-u_i)+u_i)))^{1/\gamma}}{\beta}$;
3. Generate variable of censor c_i ; $C \sim U(0; \tau)$;
4. Find $t_i^* = \min(t_i, c_i)$;
5. If $t_i < c_i$ then $\delta_i = 1$, otherwise, $\delta_i = 0$, to $i = 1, \dots, n$.

For each combination of n , $B = 1000$ samples were generated and obtained the maximum likelihood estimates of the PW distribution by **optim** via the method **BFGS**.

The bias and the square root of the mean-squared error (RMSE) of the maximum likelihood estimates were also calculated. The formulas of bias and RMSE are defined as

$$\text{Bias} = \left[\left(\frac{1}{B} \sum_{j=1}^B \hat{\theta}_j \right) - \theta \right] \text{ and } \text{RMSE} = \sqrt{\frac{1}{B} \sum_{j=1}^B (\hat{\theta}_j - \theta)^2},$$

where $\hat{\theta}_j$ represents the maximum likelihood estimative of the parameter vector θ for the j -th replication, $j = 1, \dots, B$.

From the simulation results, shown in Table 1, it was observed that the estimates of the parameters of the PW distribution were close to the true value of the parameters and the RMSEs decay toward zero when the sample size n increases, as expected. In addition, the RMSE increased as the censoring percentages increased.

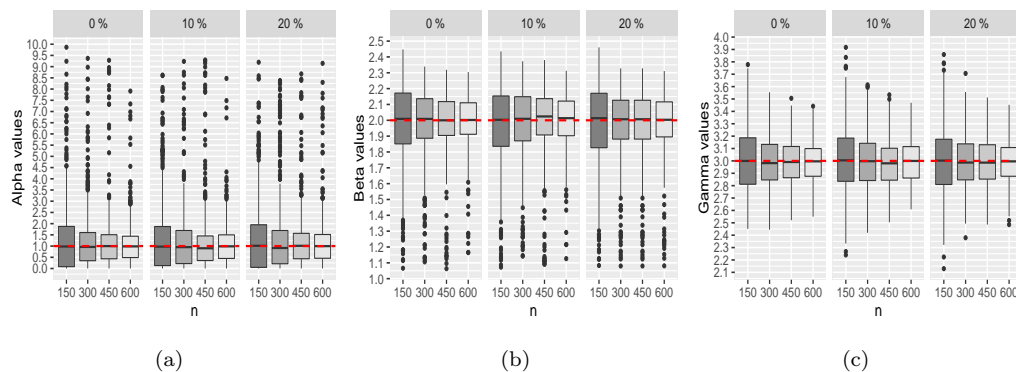


Figure 2 - Plots of simulation for PW distribution. (a) For α . (b) For β . (c) For γ .

Table 1 - Mean, Bias and RMSE of the estimates of the parameters of WP distribution with $\alpha = 1.0$, $\beta = 2$ and $\gamma=3.0$ for the percentage of censored (0%, 10% and 20%)

n	θ	0%			10%			20%		
		Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
150	α	1.2369	0.2369	1.3861	1.3042	0.3042	1.4415	1.2918	0.2918	1.4539
	β	1.9886	-0.0113	0.2287	1.9723	-0.0276	0.2358	1.9805	-0.0194	0.2427
	γ	3.0101	0.0101	0.2540	3.0089	0.0089	0.2535	2.9991	-0.0008	0.2660
300	α	1.1522	0.1522	1.1767	1.1627	0.1627	1.2231	1.2471	0.2471	1.4758
	β	1.9921	-0.0078	0.1934	1.9884	-0.0115	0.1996	1.9829	-0.0170	0.2297
	γ	2.9900	-0.0099	0.1976	3.0002	0.0002	0.2096	2.9951	-0.0048	0.2031
450	α	1.1166	0.1166	1.1048	1.0867	0.0867	1.1824	1.1662	0.1662	1.1567
	β	1.9953	-0.0046	0.1757	2.0047	0.0047	0.1865	1.9864	-0.0135	0.1911
	γ	2.9886	-0.0113	0.1767	2.9776	-0.0223	0.1812	2.9877	-0.0122	0.1885
600	α	1.0532	0.0532	0.8579	1.0475	0.0475	0.8525	1.1174	0.1174	1.0411
	β	2.0008	0.0008	0.1521	2.0054	0.0054	0.1552	1.9934	-0.0065	0.1764
	γ	2.9899	-0.0100	0.1575	2.9878	-0.0121	0.1676	2.9894	-0.0105	0.1623

As the results of the maximum likelihood estimates using the BFGS algorithm were satisfactory, we did not verify other iterative methods to estimate parameters of the PW distribution, such as the EM algorithm.

6 Application

The purpose of this section was to verify the applicability of the PW distribution on a dataset extracted from the literature where there are censored observations. The data set was extracted from the book of the Lee and Wang (2003, p.231) and refers to the survival time of 137 bladder cancer patients in which the interest is to study the time until their remission, where 7% of the patients are censored. Bladder cancer is one of the most common cancers of the urinary tract and one of the most incident in the world. In addition to smoking, white and elderly man present the most probability to develop this disease¹.

Initially, to obtain more information on the survival time of the bladder cancer patients, it was made an analysis of these times without considering the censored observations. In this case, see results in Table 2. Note that the median time of remission of the patients was approximate of 6 months, which indicates that

¹<https://www.inca.gov.br/tipos-de-cancer/cancer-de-bexiga>

approximately 50% of patients had the survival time larger than 6 months and its mean survival time was 9 months. It can also be observed that 25% of patients had a lifetime approximately less than 3 months or greater than 11 months. Furthermore, the lifetime of the patients was between 0.0800 and 79.0500 months, which implies a greater variability of the lifetime.

Table 2 - Summary of survival time of the patients for the bladder cancer data

Minimum	1° Quartile	Median	Mean	3° Quartile	Maximum	Standard deviation
0.0800	3.3600	6.2500	9.3430	11.7900	79.0500	10.3421

Table 3 gives a descriptive summary of these data based on Kaplan-Meier method. This table presents the values of the median and its respective 95% confidence interval. The percentage of censored observations for survival times is 7%.

Table 3 - Descriptive statistics by Kaplan-Meier for bladder cancer data

n	Event	Censored	Median	95% CI	
				2.50%	97.50%
137	128	9	6.93	5.34	7.93

In addition, we present in Figure 3 an empirical analysis of the behavior of the hazard function of the bladder cancer data. The graphical method based on the total time test (TTT), called of TTT-Plot (AARSET, 1987), was used to verify the hazard function on observed times of the cancer patients. This method is advantageous when there is information on the hazard function of the studied variable. According to AARSET (1987), the empirical version of the TTT plot is defined as $G(r/n) = [(\sum_{i=1}^r Y_{i:n}) - (n-r)Y_{r:n}]/(\sum_{i=1}^r Y_{i:n})$, where $r = 1, \dots, n$ and $Y_{i:n}$ represent the order statistics of the sample. AARSET (1987) showed that the hazard function is constant if the TTT plot is presented graphically as a straight diagonal, the hazard function is increasing (or decreasing) if the TTT plot is concave (or convex). The hazard function is U-shaped if the TTT plot is convex and then concave, otherwise, the hazard function is unimodal. The TTT-Plot related to bladder cancer data in Figure 3 (b) reveals a unimodal shape since the plot shows a first concave and then convex curvature. So, we can try using the PW distribution for the modeling of data because this distribution assumes some forms of the hazard function, among them, the unimodal.

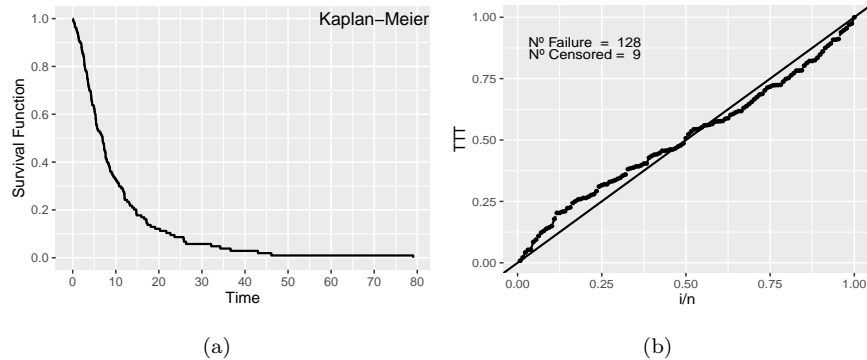


Figure 3 - Plots for bladder cancer data: (a) Kaplan-Meier; (b) TTT-Plot.

Table 4 shows the maximum likelihood estimates (MLE's) with selection criteria (GD , AIC and BIC) and standard errors for the parameters of PW distribution with their particular cases: exponential-Poisson ($\gamma = 1$); Weibull ($\alpha \rightarrow 0$); exponential ($\gamma = 1$ and $\alpha = 1$). It can be observed that the PW distribution had the lowest value of the GD and AIC in relation to the other distributions, indicating that this model is more appropriate to the data.

Table 4 - Selection criteria and MLEs of the model parameters: exponential (E), exponential-Poisson (EP), Weibull (W) and Poisson-Weibull (PW) for bladder cancer data

θ	E		EP		W		PW	
	Estimates	E.P.	Estimates	E.P.	Estimates	E.P.	Estimates	E.P.
α	-	-	9.934	4.235	-	-	3.939	1.749
β	0.100	0.008	0.011	0.005	0.098	0.008	0.038	0.014
γ	-	-	-	-	1.053	0.068	1.262	0.085
GD	845.4		844.8		844.8		837.7	
AIC	847.4		848.8		848.8		843.7	
BIC	850.3		854.6		854.6		852.5	

As the PW model is reduced in sub-models, the likelihood ratio (LR) test was used to verify if any sub-models fit better than the PW distribution. The values of LR test are presented in Table 5. Analyzing this Table, it can be observed that the PW distribution was favorable in all three hypothesis tests, thus this distribution was considered to be the most appropriate to the data. These results are corroborated by the plot in Figure 4, which shows the fitted survival function of PW distribution closed to the empirical Kaplan-Meier (Figure 4a) and fitted hazard

function (Figure 4b).

Table 5 - LR tests for the bladder cancer data

Models	Hypotheses	Statistic ω_n	p -value
PW vs EP	$H_0 : \gamma = 1$ vs $H_1 : H_0$ is false	7.6710	0.0056
PW vs Weibull	$H_0 : \alpha \rightarrow 0$ vs $H_1 : H_0$ is false	7.0406	0.0065
PW vs exponential	$H_0 : \gamma = \alpha = 1$ vs $H_1 : H_0$ is false	7.6710	0.0215

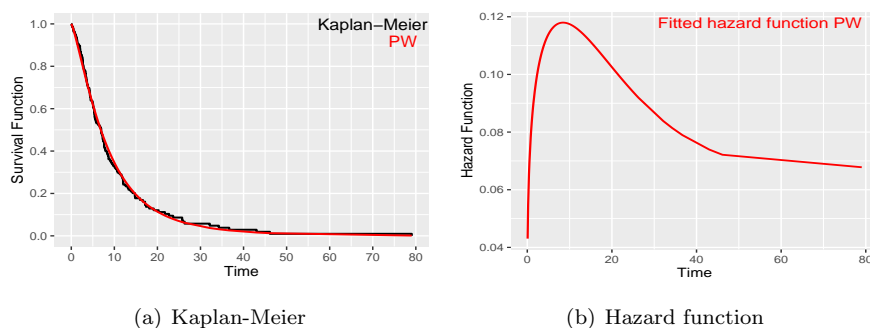


Figure 4 - Kaplan-Meier curve with the survival and hazard functions estimated of the PW distribution for the bladder cancer data.

From the considerations mentioned, the PW model was considered as the most appropriate for the bladder cancer data. Thus, the final model has the survival function given by

$$S(t) = \frac{\exp \left\{ 3.939 \exp \left[- (0.038t)^{1.262} \right] \right\} - 1}{\exp (3.939) - 1}, \quad (6)$$

where $t > 0$ represents the lifetime of the patients with bladder cancer. According to equation (6), it is possible to get some information for the bladder cancer data:

- The probability that a patient with bladder cancer will not have a remission of the disease after 4 months is given by

$$S(4) = \frac{\exp \left\{ 3.939 \exp \left[- (0.038(4))^{1.262} \right] \right\} - 1}{\exp (3.939) - 1} = 0.6939$$

This value indicates that the chance of a patient present evidence of the disease after 4 months is approximately 70%.

- The probability that a patient with bladder cancer will not have a remission of the disease after 12 months is given by

$$S(12) = \frac{\exp\left\{3.939 \exp\left[-(0.038(12))^{1.262}\right]\right\} - 1}{\exp(3.939) - 1} = 0.2735$$

This value indicates that the chance of a patient present evidence of the disease after 12 months is approximately 27%.

- The median time , which can be estimated via the quantile function, is given by equation (2). Thus, the median time value:

$$t = Q(0.50) = \frac{[\log(3.939) - \log\{\log\{\exp(3.939)(1 - 0.50)\} + 0.5\}]^{1/1.262}}{0.038} = 6.8599$$

7 Final remarks

In this paper, the Poisson-Weibull distribution with the presence of censored observations was used to analyze data. This distribution was created in a composition of discrete and continuous distributions, where there is no information about which factor was responsible for the component failure, only the minimum lifetime value among all risks is observed. Different simulation studies were adopted to study the means, the biases, and the root of mean squared error of the ML estimates of the proposed model for different values of n and censored observations where it was verified good results. Finally, an application of the PW distribution was presented as an alternative for the fit the bladder cancer data. Through graphical analysis of the TTT-Plot and Kaplan-Meier, the observed values of the GD , AIC , and BIC criteria, and the likelihood ratio test, it can be noted that the PW model fitted well to the data. Thus, it is expected that this model is useful for fitting other datasets that present some forms of the hazard function, among them, the unimodal.

Acknowledgments

We would like thanks CNPq Scholarship - Brazil (166774/2020-0) and the reviewers and editors for their comments and suggestions.

VIGAS, V. P.; PRATAVIERA, F.; SILVA, G. O. Uma aplicação do modelo Poisson-Weibull para dados censurados. *Rev. Bras. Biom.*, Lavras, v.39, n.4, p.505-521, 2021.

■ **RESUMO:** Neste artigo, propomos a distribuição Poisson-Weibull para a modelagem dos dados de sobrevivência. A motivação para estudar este modelo é que, além de generalizar a distribuição Weibull, que é amplamente utilizada em diversas áreas do conhecimento entre elas a análise de sobrevivência e confiabilidade, apresenta grande flexibilidade nas formas da função de risco. A distribuição Poisson-Weibull foi criada em uma composição de distribuições discreta e contínua onde não há informação sobre qual fator foi responsável pela falha do componente, apenas o valor mínimo de vida entre todos os riscos é observado. A abordagem de máxima verossimilhança foi usada para estimar os parâmetros do modelo. Também foi realizado um estudo de simulação para examinar a média, o viés e a raiz do erro quadrático médio das estimativas de máxima verossimilhança modelo proposto de acordo com os percentuais de censura e tamanhos da amostra. Critérios de seleção de modelos também foram aplicados, além de técnicas gráficas como TTT-Plot e Kaplan-Meier. Uma aplicação ao conjunto de dados foi usada para ilustrar a utilidade da distribuição.

■ **PALAVRAS-CHAVE:** Análise de sobrevivência; composição de distribuições; dados censurados; seleção de modelos.

References

AARSET, M. V. How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, v.36, n.1, p.106-108, 1987.

ADAMIDIS, K.; LOUKAS, S. A lifetime distribution with decreasing failure rate. *Statistics Probability Letters*, v.39, n.1, p.35-42, 1998.

BARRETO-SOUZA, W.; MORAIS, A. L. D.; CORDEIRO, G. M. The Weibull-Geometric Distribution. *Journal of Statistical Computation and Simulation*, v.81, n.5, p.645-657, 2011.

BERETA, E. M. P.; LOUZADA, F.; FRANCO, M. The Poisson-Weibull distribution. *Advances and Applications in Statistics*, v.22, n.2, p.117-118, 2011.

CANCHO, V. G.; LOUZADA-NETO, F.; BARRIGA, G. D. C. The Poisson-exponential lifetime distribution. *Computational Statistics and Data Analysis*, v.55, n.1, p.677-686, 2011.

CARDOSO, F. F.; ROSA, G. D. M.; TEMPELMAN, R. J.; TEMPELMAN, ROBERT; TORRES JUNIOR, R. D. A. Modelos hierárquicos bayesianos para estimação robusta e análise de dados censurados em melhoramento animal. *Revista Brasileira de Zootecnia*, v.38, p.72-80, 2009.

CHAHKANDI, M; GANJALI, M. On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, v.53, 12, p.4433-4440, 2009.

DOS SANTOS JUNIOR, P. C. *Análise de sobrevivência na presença de censura informativa*, Ph.D. thesis, Universidade Federal de Minas Gerais, 2012.

GUEDES, D. G. P. *Avaliação genética da eficiência reprodutiva em vacas pardas-suíças por meio da análise de sobrevivência*, 2014. 81p. Dissertação (Mestrado em Produção Animal). Universidade Federal do Rio Grande do Norte, Macaíba-RN, 2014.

KUS, C. A new lifetime distribution. *Computation Statistical Data Analysis*, v.51, n.9, p.4497-4509, 2007.

LEE, E.T.; WANG, J.W. *Statistical methods for survival data analysis*. 3.ed. Wiley, New Jersey, 2003. 534p.

LOUZADA, F.; ROMAN, M.; CANCHO, V. The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart. *Computational Statistics and Data Analysis*, v.55, n.8, p.2516-2524, 2011.

LOUZADA, F; CANCHO, V. G; ROMAN, M; LEITE, J. G. A new long-term lifetime distribution induced by a latent complementary risk framework. *Journal of Applied Statistics*, v.39, n.10, p.2209-2222, 2012.

LOUZADA, F; YAMACHI C. Y; MARCHI V. A; FRANCO M. A. P. The long-term exponentiated complementary exponential geometric distribution under a latent complementary causes framework, *TEMA (São Carlos)*, v.15, n.1, p.19-35, 2014.

LOUZADA, F; ELBATAL, I; GRANZOTTO D. C. T. The Beta Exponentiated Weibull Geometric Distribution: Modeling, structural properties, estimation and an application to a cervical intraepithelial neoplasia dataset, *Revista Brasileira de Biometria*, v.36, n.4, p.942-967, 2018.

MALLER, R.; ZHOU, X. Testing for the Presence of Immune or Cured Individuals in Censored Survival Data. *Biometrics*, v.51, p.1197-1205, 1995.

MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the Exponential and Weibull families. *Biometrika*, v.84, n.3, p.641-652, 1997.

MAZUCHELI, J.; MENEZES, A. F. B.; FERNANDES, L. B.; DE OLIVEIRA, R. P.; GHITANY, M. E. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *Journal of Applied Statistics*, v.47, n.6, p.954-974, 2020.

THOM, H. C. S. A Note on the Gamma distribution, *Statistical Laboratory, Iowa State College*, 1947 (Manuscript)

VERHULST P. F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, v.10, p.113-121, 1838.

WEIBULL, W. A statistical distribution function of wide applicability. *AMSE Journal of Applied Mechanics* , v.18, n.3, p.292-297, 1951.

YAMACHI, C. Y.; ROMAN, M.; LOUZADA, F.; FRANCO, M. A. P.; CANCHO, V. G. The Exponentiated Complementary Exponential Geometric Distribution. *Journal of Modern Mathematics Frontier*, v.2, n.2, 2013.

Received on 25.09.2020.

Approved after revised on 03.05.2021.

Supplementary material

```
#####  
##### PW Functions #####  
#####  
# Density function  
dPW <- function(x,alpha,beta,gamma){  
  pdf <- ((alpha*gamma*(beta^gamma))/(exp(alpha)-1))  
    *(x^(gamma-1))*exp(-(beta*x)^gamma)*exp(alpha*(exp(-(beta*x)^gamma)))  
  }  
# Survival function  
sPW <- function(x,alpha,beta,gamma){  
  s <- (exp(alpha*exp(-(x*beta)^gamma))-1)/(exp(alpha)-1)  
}  
# Hazard function  
hPW <- function(x,alpha,beta,gamma){  
  h <- dPW(x,alpha,beta,gamma)/sPW(x,alpha,beta,gamma)  
}  
#####  
##### Log-likelihood Function #####  
#####  
log.L <- function(parameter){  
  alpha <- parameter[1]  
  beta <- parameter[2]  
  gamma <- parameter[3]  
  dens.pW <- ((alpha*gamma*(beta^gamma))/(exp(alpha)-1))  
    *(time^(gamma-1))*exp(-(beta*time)^gamma)*  
    exp(alpha*(exp(-(beta*time)^gamma)))  
  surv.pW <- (exp(alpha*exp(-(time*beta)^gamma))-1)/(exp(alpha)-1)  
  l.vero <- -sum(delta*log(dens.pW)+(1-delta)*(log(surv.pW)))  
  return(l.vero)  
}  
#####  
##### Maximization by optim #####  
#####  
results <- optim(c(3.93962349, 0.03874076, 1.26282809),  
log.L,method = "BFGS",hessian=TRUE)  
results
```