

Bayesian estimation of a power law process with incomplete data

HU Junming^{1,2,3}, HUANG Hongzhong^{1,3,*}, and LI Yanfeng^{1,3}

1. School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. School of Transportation and Automotive Engineering, Xihua University, Chengdu 610039, China; 3. Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: Due to the simplicity and flexibility of the power law process, it is widely used to model the failures of repairable systems. Although statistical inference on the parameters of the power law process has been well developed, numerous studies largely depend on complete failure data. A few methods on incomplete data are reported to process such data, but they are limited to their specific cases, especially to that where missing data occur at the early stage of the failures. No framework to handle generic scenarios is available. To overcome this problem, from the point of view of order statistics, the statistical inference of the power law process with incomplete data is established in this paper. The theoretical derivation is carried out and the case studies demonstrate and verify the proposed method. Order statistics offer an alternative to the statistical inference of the power law process with incomplete data as they can reformulate current studies on the left censored failure data and interval censored data in a unified framework. The results show that the proposed method has more flexibility and more applicability.

Keywords: incomplete data, power law process, Bayesian inference, order statistics, repairable system.

DOI: [10.23919/JSEE.2021.000021](https://doi.org/10.23919/JSEE.2021.000021)

1. Introduction

Reliability is a crucial characteristic for products because their higher reliability always leads to less loss and greater competitiveness. Many efforts have been made to meet expected specifications of product reliability in practice [1]. For example, the reliability growth management incorporates corrective measures to fix the defects that have surfaced and the reliability is enhanced gradually by the implementation of the test–analysis–fix–test cycle for product prototypes [2]. In addition to the reliability growth, sound maintenance policies and procedures also help to

ensure that product reliability is maintained at the benchmark level after it comes into the operation. For both reliability growth and maintenance, failure data collection and analysis are the primary concerns [3]. Over the past decades, multiple approaches like Bayesian network [4–6], renewal process [7], trend-renewal process [8,9], Markov process [10–12], nonhomogeneous Poisson process [13–15], and modulated power law process [16], have been proposed to help analysts to make informed decisions.

Although there are different models, the power law process (PLP) is commonly used for failure data analysis due to its flexibility and ease of use. After the Army Materiel Systems Analysis Activity initially uses it to plan, track, and project a reliability growth program [17], the International Electrotechnical Commission recommends it for reliability growth modeling [18]. Moreover, the PLP is not restricted to dealing with data produced by the improvement program and is also suitable for handling data from other scenarios such as minimal repairs [19]. Although the procedures that deal with parameters estimation, confidence intervals, prediction intervals, and the goodness-of-fit test of the PLP have been well established [20], the PLP still attracts much attention. Multiple investigations—the PLP model selection taking into account system heterogeneities [21], Bayesian inference of the PLP under different maintenance policies [22], statistical inference of the PLP considering competing risk [23], the generalized confidence interval of the scale parameter [24], conjugate prior for Bayesian inference of the PLP [25], and parametric bootstrap confidence interval method [26]—have been conducted.

Most of the literature use the PLP to perform the data analysis, which assumes that the failure times are fully known, that is to say, the successive failure times of the repairable systems are monitored and accurately recorded. However, failure data may not be detected or obtained

Manuscript received June 20, 2020.

*Corresponding author.

This work was supported by the National Natural Science Foundation of China (51775090).

due to various uncontrollable reasons. For example, human error, lack of experience, and malfunction of sensors may trigger any unobserved or unrecorded failure. Sometimes, the data may be found to be unfaithful or even incorrect. In practice, it often occurs that the exact failure times are not recorded and that only the number of failures occurred within a given period is known. Incomplete data can reduce the representativeness of the data sample and result in a distorted inference in turn [27]. The interest in the incomplete data problem is therefore growing. Guo et al. [28] proposed a three-parameter Weibull failure rate for the analysis of incomplete wind turbine failure data, where a new time parameter was introduced into the Weibull rate function to represent the past running time.

Statistical inferences of incomplete data have also been discussed for the PLP. Yu et al. [29] carried out a statistical inference and prediction analysis of the PLP with left censored data by using the classical approach in which the missing data occurred at the early stage of the research. This work laid the groundwork for further studies. Based on the obtained results, Tian et al. [30] established Bayesian estimation and prediction of the PLP with left censored data; Chumnaul and Sepehrifar [31] derived a generalized confidence interval for the scale parameter of the PLP with left censored failure data; Na and Chang [32] conducted a trend analysis for tank maintenance where the left censored data are from multiple machines and no information on the number of missing data is available. Besides, some researchers have performed reliability studies that are subject to interval censored data. For example, Taghipour and Banjevic [33] conducted a trend analysis of the PLP with interval censored data. Peng et al. [34] suggested a framework to evaluate the dynamic reliability of repairable systems with interval censored failure data in which the imputation method was used to fill out missing data.

The above-mentioned methods deal with incomplete data based on their respective assumptions. They are limited to their specific cases, either left censored data or interval censored data, and no unified framework is available for handling all situations. In addition, the suggested methods are complex and lack flexibility since the successive failure times of the PLP are mutually dependent. Considering that order statistics can help manipulate the dependent variables into independent variables, it can provide an alternative viewpoint for dealing with missing data and simplify the related mathematical derivations. This paper proposes a generic approach for making statistical inferences of the PLP with incomplete data.

The rest of the paper is organized as follows. Section 2 introduces the preliminaries on order statistics and the

PLP. After the likelihood function of the observed data is derived from the perspective of order statistics, Bayesian procedures of estimating unknown parameters of the PLP are discussed in Section 3. Section 4 provides case studies to explain and validate the proposed process, where both simulated and actual failure data are used. Finally, Section 5 summarizes the conclusions of this study.

2. Preliminaries

In this section, we briefly review some basic properties that are necessary for later discussion and mathematical derivation.

2.1 Order statistics

Given any random variable X_1, X_2, \dots, X_n , the i th order statistic, $X_{i:n}$ ($i = 1, 2, \dots, n$), is defined by sorting the corresponding realizations of these random variables in an ascending order as $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$.

When the random variable X_1, X_2, \dots, X_n is sampled from the same population, its realization x_1, x_2, \dots, x_n is independent and identically distributed. Let $f(x)$ and $F(x)$ be the associated probability density function (PDF) and cumulative distribution function (CDF). The joint PDF of

x_1, x_2, \dots, x_n is equal to $\prod_{i=1}^n f(x_i)$. Further, there are $n!$ permutations of $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ to generate the same sample, i.e., each of these permutations has an equal chance of being the same order statistics. Thus, there is an $n!$ to one map from the observations to the order statistics. The joint PDF of order statistics $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ is

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \quad (1)$$

where $n!$ is the factorial of n .

For any two order statistics $X_{i:n}$ and $X_{j:n}$ ($1 \leq i < j \leq n$), the joint PDF of $X_{i:n}$ and $X_{j:n}$ is given without proof as

$$\frac{f_{i,j}(x_i, x_j) = n! f(x_i) f(x_j) \cdot [F(x_i)]^{i-1} [F(x_j) - F(x_i)]^{j-i-1} [1 - F(x_j)]^{n-j}}{(i-1)!(j-i-1)!(n-j)!} \quad (2)$$

More details about (2) can be referred to [35]. Three identities, which are related to the derivation of (2) and also useful for further theoretical development, are described as follows:

$$\begin{aligned} \int_{-\infty}^{x_i} \dots \int_{-\infty}^{x_2} f(x_1) \dots f(x_{i-1}) dx_1 \dots dx_{i-1} &= \\ \int_{-\infty}^{x_i} f(x_{i-1}) dx_{i-1} \dots \int_{-\infty}^{x_3} f(x_2) dx_2 \int_{-\infty}^{x_2} f(x_1) dx_1 &= \\ \int_{-\infty}^{x_i} f(x_{i-1}) dx_{i-1} \dots \int_{-\infty}^{x_3} f(x_2) dx_2 F(x_2) &= \end{aligned}$$

$$\int_{-\infty}^{x_i} f(x_{i-1}) dx_{i-1} \cdots \int_{-\infty}^{x_4} f(x_3) dx_3 \frac{[F(x_3)]^2}{2} = \int_{-\infty}^{x_i} f(x_{i-1}) dx_{i-1} \cdots \int_{-\infty}^{x_5} f(x_4) dx_4 \frac{[F(x_4)]^3}{2 \times 3} = \frac{[F(x_i)]^{i-1}}{(i-1)!}. \quad (3)$$

By integration of order statistics over their domain, (3) can be easily derived. Similarly, there exist

$$\int_{x_i}^{x_j} \cdots \int_{x_i}^{x_{i+2}} f(x_{i+1}) \cdots f(x_{j-1}) dx_{i+1} \cdots dx_{j-1} = \frac{[F(x_j) - F(x_i)]^{j-i-1}}{(j-i-1)!}, \quad (4)$$

and

$$\int_{x_j}^{\infty} \cdots \int_{x_j}^{x_{j+2}} f(x_{j+1}) \cdots f(x_n) dx_{j+1} \cdots dx_n = \frac{[1 - F(x_j)]^{n-j}}{(n-j)!}. \quad (5)$$

2.2 Power law process

The PLP is a nonhomogeneous Poisson process with a time-dependent failure intensity that is proportional to a power of time t . Let α and β denote the scale parameter and shape parameter separately. The failure intensity of the PLP, $\nu(t)$, is

$$\nu(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}, \quad \alpha > 0; \beta > 0. \quad (6)$$

Because the failure intensity of the PLP has the same form as the failure rate of Weibull distribution, the PLP is also known as the Weibull process [36]. If $\beta > 1$, the system is deteriorating. If $\beta < 1$, the system reliability is improving. If $\beta = 1$, the system has a constant failure intensity and the PLP reduces to a homogeneous Poisson process.

The mean value function $m(t)$ of the PLP is

$$m(t) = E[N(t)] = \int_0^t \nu(x) dx = \left(\frac{t}{\alpha}\right)^\beta \quad (7)$$

where $N(t)$ denotes the number of failures observed up to time t . Furthermore, according to the property of the non-homogeneous Poisson process [37], the probability of $N(t) = n$ is

$$P\{N(t) = n\} = \frac{[m(t)]^n e^{-m(t)}}{n!}. \quad (8)$$

Let $\mathbf{T} = (T_1, T_2, \dots, T_n)$ denote the consecutive failure times of a PLP where T_i is the i th failure time, t_c is the time at which the data collection ceases. In practice, there are two types of truncated data: failure truncated and time truncated. If the data collection process is ceased at the arrival time of the n th failure, the data collected in this

way are failure truncated. If the data collection process is terminated at a prefixed time, the collected data are called time truncated data. Thus, for failure truncated case, t_c is equal to t_n , and for the time truncated case, t_c is predetermined.

Given the truncated time t_c and the observed values $T_1 = t_1, T_2 = t_2, \dots, T_n = t_n$, the joint PDF is

$$f(t_1, t_2, \dots, t_n, t_c) = \left[\prod_{i=1}^n \nu(t_i) \right] e^{-m(t)} = \left[\prod_{i=1}^n \nu(t_i) \right] e^{-\int_0^{t_c} \nu(t) dt} = \frac{\beta^n}{\alpha^{n\beta}} e^{-\left(\frac{t_c}{\alpha}\right)^\beta} \prod_{i=1}^n t_i^{\beta-1}. \quad (9)$$

The conditional density function of $\mathbf{T} = (T_1, T_2, \dots, T_n)$ given $N(t_c) = n$ is

$$f(t_1, t_2, \dots, t_n, t_c | N(t_c) = n) = \frac{f(t_1, t_2, \dots, t_n, t_c)}{P[N(t_c) = n]} = n! \prod_{i=1}^n \frac{\nu(t_i)}{m(t_c)}. \quad (10)$$

Equation (10) indicates that on the condition of $N(t_c) = n$, the successive failure times $\mathbf{T} = (T_1, T_2, \dots, T_n)$ are distributed as order statistics from the distribution with PDF as

$$f(y) = \frac{\nu(y)}{m(t_c)} = \frac{\beta}{t_c} \left(\frac{y}{t_c}\right)^{\beta-1}, \quad 0 \leq y \leq t_c, \quad (11)$$

and the corresponding CDF is

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{m(y)}{m(t_c)} = \left(\frac{y}{t_c}\right)^\beta, & 0 \leq y \leq t_c. \\ 1, & y > t_c \end{cases} \quad (12)$$

Equation (12) shows that for the time truncated case, the successive failure data are that of n order statistics; while for the failure truncated case, the successive failure data are that of $n-1$ order statistics.

3. Analysis of incomplete data

To deal with the incomplete data, a unified framework combining the expectation maximization algorithm and Bayesian inference is proposed in this section.

3.1 Incomplete data

Incomplete failure data are common for repairable systems in practical situations. Different missing locations lead to different incompleteness forms. In this paper, the failure data are classified into two types: complete data and incomplete data. All failures are gathered and recorded for complete data. On the other hand, failure data in which certain elements are missing are classified into incomplete data.

Three types of failure data, complete data, left censored data, and interval censored data are shown in Fig. 1, where the symbol \times stands for a failure and t_i is the i th

failure time. If the failure time is unrecorded, a subscript $*$ is used for indication. For both complete data and incomplete data, t_c is the truncated time.

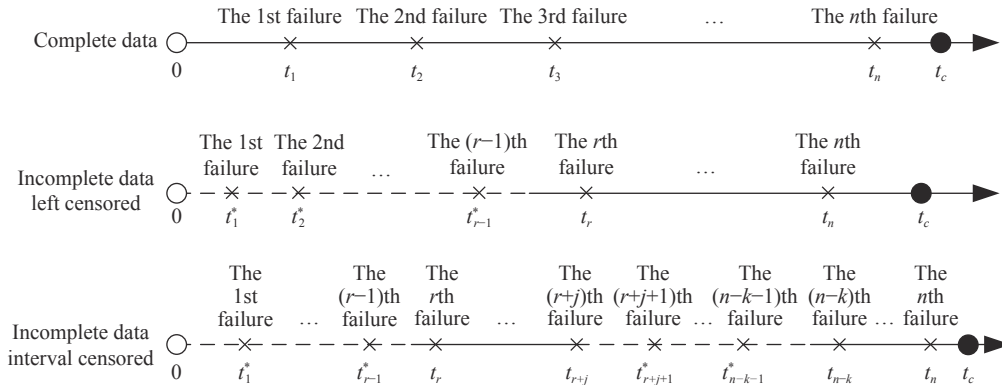


Fig. 1 Schematics of failure data

When the missing location is at the early stage, we term the data collected as left censored data. For example, the first $r-1$ ($1 \leq r < n$) failures are assumed missing in Fig. 1, so the missing data are denoted by $\mathbf{Y}_m = \{t_i^*\}_{i=1}^{r-1}$ and the observed data are $\mathbf{Y}_o = \{t_1, \dots, t_n\}$. If the missing occurs at certain periods, the data collected are regarded as interval censored data. For interval censored data in Fig. 1, the failures occurring at the interval $(0, t_r]$ and $[t_{r+j}, t_{n-k}]$ are assumed missing. Thus, the missing data are denoted by $\mathbf{Y}_m = \{t_1^*, \dots, t_{r-1}^*, t_{r+j+1}^*, \dots, t_{n-k-1}^*\}$ and the observed data are denoted as $\mathbf{Y}_o = \{t_r, \dots, t_{r+j}, t_{n-k}, \dots, t_n\}$.

3.2 Likelihood function

Dempster et al. [38] first presented a general approach to maximize the likelihood estimation of parameters with incomplete data. This approach assumes that there is a many-one mapping from the complete data to the incomplete data. Let the observed data and the missing data be denoted by \mathbf{Y}_o and \mathbf{Y}_m respectively. The likelihood of complete data $f(\mathbf{Y}_o, \mathbf{Y}_m | \alpha, \beta)$ is related to the likelihood of incomplete data $f(\mathbf{Y}_o | \alpha, \beta)$ by

$$f(\mathbf{Y}_o | \alpha, \beta) = \int f(\mathbf{Y}_o, \mathbf{Y}_m | \alpha, \beta) d\mathbf{Y}_m. \quad (13)$$

3.2.1 Likelihood of left censored data

For left censored data shown in Fig. 1, $\mathbf{Y}_m = \{t_1^*, t_2^*, \dots, t_{r-1}^*\}$ and $\mathbf{Y}_o = \{t_r, \dots, t_c\}$. The likelihood of \mathbf{Y}_o can be derived by integrating (13) with respect to $\{t_i^*\}_{i=1}^{r-1}$, i.e.,

$$f(\mathbf{Y}_o | \alpha, \beta) = f(t_r, \dots, t_c | \alpha, \beta) = \int \dots \int f(\mathbf{Y}_m, \mathbf{Y}_o | \alpha, \beta) dt_1^* \dots dt_{r-1}^* = P\{N(t_c) = n\} \cdot \int \dots \int f(\mathbf{Y}_m, \mathbf{Y}_o | N(t_c) = n, \alpha, \beta) dt_1^* \dots dt_{r-1}^*. \quad (14)$$

As proofed in Section 2, the failure times of PLP are order statistics conditioning on $N(t_c) = n$ and its associa-

ted PDF and CDF are given in (11) and (12), respectively. Taking advantage of (2) and (3), and substituting (8) and (10) into (14), we can easily integrate (14) with respect to every t_i^* and obtain the likelihood of \mathbf{Y}_o as

$$f(\mathbf{Y}_o | \alpha, \beta) = \frac{\beta^{n+1-r} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{\alpha^{n\beta} (r-1)!} t_r^{(r-1)\beta} \prod_{i=r}^n t_i^{\beta-1}. \quad (15)$$

Equation (16) provides a mathematical proof of (15).

$$\begin{aligned} f(\mathbf{Y}_o | \alpha, \beta) &= P\{N(t_c) = n\} \cdot \int \dots \int f(\mathbf{Y}_m, \mathbf{Y}_o | N(t_c) = n, \alpha, \beta) dt_1^* \dots dt_{r-1}^* = \\ &= \frac{[m(t_c)]^n e^{-m(t_c)}}{n!} \cdot n! \prod_{i=r}^n \frac{v(t_i)}{m(t_c)} \int \dots \int \frac{v(t_j)}{m(t_c)} dt_1^* \dots dt_{r-1}^* = \\ &= \frac{[m(t_c)]^n e^{-m(t_c)}}{n!} \cdot n! \prod_{i=r}^n \frac{v(t_i)}{m(t_c)} \cdot \frac{[F(t_r)]^{r-1}}{(r-1)!} = \\ &= \frac{[m(t_c)]^n e^{-m(t_c)}}{n!} \cdot \frac{n!}{(r-1)!} \prod_{i=r}^n \frac{\beta t_i^{\beta-1}}{\alpha^\beta} \cdot \left[\left(\frac{t_r}{t_c}\right)^\beta\right]^{r-1} = \\ &= \frac{\left(\frac{t_c}{\alpha}\right)^\beta e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{(r-1)!} \cdot \prod_{i=r}^n \frac{\beta t_i^{\beta-1}}{t_c^\beta} \cdot \left(\frac{t_r}{t_c}\right)^{\beta(r-1)} = \\ &= \frac{\beta^n \alpha^{-\beta n} e^{-\left(\frac{t_c}{\alpha}\right)^\beta} \beta^{n-r+1} t_c^{-\beta(n-r+1)} \left(\prod_{i=r}^n t_i^{\beta-1}\right) t_c^{-\beta(r-1)} t_r^{\beta(r-1)}}{(r-1)!} = \\ &= \frac{\beta^{n-r+1} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{\alpha^{n\beta} (r-1)!} t_r^{\beta(r-1)} \prod_{i=r}^n t_i^{\beta-1}. \quad (16) \end{aligned}$$

3.2.2 Likelihood of interval censored data

Similarly, for the interval-censored case where $\mathbf{Y}_m = \{t_1^*, \dots, t_{r-1}^*, t_{r+j+1}^*, \dots, t_{n-k-1}^*\}$ and $\mathbf{Y}_o = \{t_r, \dots, t_{r+j}, t_{n-k}, \dots, t_n\}$, we can obtain the likelihood of \mathbf{Y}_o as

$$f(\mathbf{Y}_o|\alpha,\beta) = \frac{e^{-\left(\frac{t_c}{\alpha}\right)^\beta} \beta^{j+k+2} \left(t_{n-k}^\beta - t_{r+j}^\beta\right)^{n-k-r-j-1} t_r^{\beta(r-1)}}{\alpha^{\beta n} (n-k-r-j-1)! (r-1)!} \cdot \prod_{i=r}^{r+j} t_i^{\beta-1} \cdot \prod_{i=n-k}^n t_i^{\beta-1}. \quad (17)$$

The mathematical proof of (17) can be referred to (18).

$$\begin{aligned} f(\mathbf{Y}_o|\alpha,\beta) &= f(t_r, \dots, t_{r+j}, t_{n-k}, \dots, t_n, t_c|\alpha,\beta) = \\ &= \frac{\{m(t_c)\}^n e^{-m(t_c)}}{n!} \cdot n! \prod_{i=r}^{r+j} \frac{\nu(t_i)}{m(t_c)} \cdot \frac{\{F(t_r)\}^{r-1}}{(r-1)!} \cdot \\ &= \frac{\{F(t_{n-k}) - F(t_{r+j})\}^{n-k-r-j-1}}{(n-k-r-j-1)!} \cdot \prod_{i=n-k}^n \frac{\nu(t_i)}{m(t_c)} = \\ &= \frac{\left(\frac{t_c}{\alpha}\right)^{\beta n} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{n!} \cdot n! \prod_{i=r}^{r+j} \frac{\beta t_i^{\beta-1}}{\alpha^\beta} \cdot \frac{\left\{\left(\frac{t_r}{t_c}\right)^\beta\right\}^{r-1}}{(r-1)!} \cdot \\ &= \frac{\left\{\left(\frac{t_{n-k}}{t_c}\right)^\beta - \left(\frac{t_{r+j}}{t_c}\right)^\beta\right\}^{n-k-r-j-1}}{(n-k-r-j-1)!} \cdot \prod_{i=n-k}^n \frac{\beta t_i^{\beta-1}}{\alpha^\beta} = \\ &= \frac{t_c^{\beta n} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{\alpha^{\beta n}} \cdot \prod_{i=r}^{r+j} \frac{\beta t_i^{\beta-1}}{t_c^\beta} \cdot \frac{\left(\frac{t_r}{t_c}\right)^{\beta(r-1)}}{(r-1)!} \cdot \\ &= \frac{\left\{\left(\frac{t_{n-k}}{t_c}\right)^\beta - \left(\frac{t_{r+j}}{t_c}\right)^\beta\right\}^{n-k-r-j-1}}{(n-k-r-j-1)!} \cdot \prod_{i=n-k}^n \frac{\beta t_i^{\beta-1}}{t_c^\beta} = \\ &= \frac{t_c^{\beta n} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{\alpha^{\beta n}} \cdot \frac{\beta^{j+1} \prod_{i=r}^{r+j} t_i^{\beta-1}}{t_c^{\beta(j+1)}} \cdot \frac{t_r^{\beta(r-1)}}{t_c^{\beta(1-r)} (r-1)!} \cdot \\ &= \frac{\left\{\left(\frac{t_{n-k}}{t_c}\right)^\beta - \left(\frac{t_{r+j}}{t_c}\right)^\beta\right\}^{n-k-r-j-1}}{(n-k-r-j-1)!} \cdot \frac{\beta^{k+1} \prod_{i=n-k}^n t_i^{\beta-1}}{t_c^{\beta(k+1)}} = \\ &= \frac{t_c^{\beta n} e^{-\left(\frac{t_c}{\alpha}\right)^\beta}}{\alpha^{\beta n}} \cdot \frac{\beta^{j+1} \prod_{i=r}^{r+j} t_i^{\beta-1}}{t_c^{\beta(j+1)}} \cdot \frac{t_r^{\beta(r-1)}}{t_c^{\beta(r-1)} (r-1)!} \cdot \\ &= \frac{t_c^{-\beta(n-k-r-j-1)} \left(t_{n-k}^\beta - t_{r+j}^\beta\right)^{n-k-r-j-1}}{(n-k-r-j-1)!} \cdot \frac{\beta^{k+1} \prod_{i=n-k}^n t_i^{\beta-1}}{t_c^{\beta(k+1)}} = \\ &= \frac{e^{-\left(\frac{t_c}{\alpha}\right)^\beta} \beta^{j+k+2} \left(t_{n-k}^\beta - t_{r+j}^\beta\right)^{n-k-r-j-1} t_r^{\beta(r-1)}}{\alpha^{\beta n} (n-k-r-j-1)! (r-1)!} \cdot \prod_{i=r}^{r+j} t_i^{\beta-1} \cdot \prod_{i=n-k}^n t_i^{\beta-1} \end{aligned} \quad (18)$$

3.3 Bayesian inference

The Bayesian inference provides a useful approach to estimate unknown parameters in which the prior information and likelihood information are fused and more reasonable information on the unknown parameters can be yielded from the posterior. Let \mathbf{S} stand for the observed data and \mathbf{P} be the unknown parameters of interest. Under the Bayesian framework, the parameters \mathbf{P} are assumed to be random variables. Assuming the associated prior PDF of \mathbf{P} is $f(\mathbf{P})$ and the likelihood function of \mathbf{S} conditional on \mathbf{P} is $L(\mathbf{S}|\mathbf{P})$, the posterior PDF of \mathbf{P} , $f(\mathbf{P}|\mathbf{S})$, is determined by

$$f(\mathbf{P}|\mathbf{S}) = \frac{f(\mathbf{P})L(\mathbf{S}|\mathbf{P})}{\int_{-\infty}^{+\infty} f(\mathbf{P})L(\mathbf{S}|\mathbf{P})d\mathbf{P}}. \quad (19)$$

The denominator of (19) is a normalizing constant. By this mechanism, the prior information $f(\mathbf{P})$ is updated to the posterior $f(\mathbf{P}|\mathbf{S})$.

For the PLP, the scale parameter α and shape parameter β are the parameters of interest, i.e., $\mathbf{P} = (\alpha, \beta)$. If α and β are independent mutually, $f(\mathbf{P})$ is equal to $f(\alpha)f(\beta)$. The likelihood function of the observed failure data is determined by the failure data type. For complete failure data, $L(\mathbf{S}|\mathbf{P})$ has the form of (9). For the left-censored data, $L(\mathbf{S}|\mathbf{P})$ is defined by (15), and for the interval censored data, $L(\mathbf{S}|\mathbf{P})$ is given in (17).

4. Case study

In this section, we utilize two examples to illustrate and validate the proposed method. In the simulation case, all the data are generated in the same simulation. The Bayesian inference for the complete data, left censored data and interval censored data are conducted, and the inference results are compared. In the engineering case, real failure data from a reliability growth testing of a prototype vehicle are analyzed.

4.1 Simulation case

According to (10), the successive failure data of the PLP can be viewed as ordered statistics with the PDF and CDF as (11) and (12). This property provides a mechanism to simulate the PLP. For a time truncated case, simulation procedures for failure data are listed in Table 1.

Table 1 PLP simulation procedures

Step	Content
1	Set the parameters α , β and the truncated testing time t_c
2	Generate the failure number n from the Poisson distribution with mean parameter $\lambda = (t_c/\alpha)^\beta$
3	Generate n random numbers u_i from the standard uniform distribution
4	Generate the failure times \mathbf{T} , where t_i is equal to $t_i = t_c \cdot (r_i)^{\frac{1}{\beta}}$
5	Sort t_i in an ascending order

A total of 27 failure times are simulated based on the procedures listed in Table 1, where the parameters $\alpha = 50$, $\beta = 0.6$, and $t_c = 10^4$. The successive failure times simulated are 4.4*, 29.8*, 104.4*, 217.8*, 656.1, 657.3, 925.9, 1 624.8, 1 717.1, 2 043.0, 2 509.4*, 2 524.4*, 2 664.8, 3 470.6, 3 673.4, 3 776.1, 4 599.9, 4 931.4, 5 431.5, 5 437.2, 5 609.5, 6 532.4, 8 202.6, 8 398.9, 9 154.4, 9 380.7, 9 969.6. Both the left censored and interval censored data are considered in this study. For the left censored data, the exact failure times for the first four failures are assumed unknown and are indicated by *. That is, $n = 27$ and $r = 5$ for the left censored case. For the interval censored case, additional failure times are assumed missing besides the first four failures times. All the missing data are denoted by *. That is, $n = 27$, $r = 5$, $j = 5$, and $k = 14$ for the interval censored case.

Under the Bayesian framework, the prior knowledge, such as statistical data, expert opinions, historical data, or experiences from similar systems, can help to create a reasonable prior distribution. Prior elicitation, however, requires significant effort. Therefore, for the unknown parameters α and β in this study, noninformative prior distributions characterized by uniform distributions are adopted.

To verify and compare the efficiency of the proposed approach, all the cases use the same noninformative prior distributions where α and β are considered to be mutually independent. Using the Markov chain Monte Carlo (MCMC) sampling, the point estimation and 95% credible intervals are easily obtained from the posterior distribution. The results of the Bayesian inference are listed in Table 2.

Table 2 Inference results of different data types

Data type	Mean estimation		Credible interval	
	α	β	α	β
Complete data	40.61	0.591 5	[21.17, 59.05]	[0.505 1, 0.675 9]
Left censored	42.14	0.600 7	[21.55, 59.24]	[0.514 9, 0.685 2]
Interval censored	41.91	0.599 8	[21.57, 59.20]	[0.511 7, 0.683 0]

As shown in Table 2, the point estimates of α and β are similar to each other, and all the point estimates of β are very close to the actual value. The estimate thus captures the characterization of the system. In addition, all the credible intervals of α and β are valid. There is no significant difference between circumstances. It is verified that the proposed method is effective.

4.2 Engineering case

Two engineering cases from an aircraft generator and a prototype vehicle are used to illustrate the proposed methodology. For the first example, Chumnaul and Sepehrifar [31] gave the maximum likelihood estimates (MLE), classical confidence intervals (CCI), and generalized confidence intervals (GCI) derived from generalized pivotal quantity. The same data are used here for the comparative analysis. For the prototype vehicle data, we

focus on determining whether the system is improving.

4.2.1 Aircraft generator data

The testing of an aircraft generator is ceased at the 13th failure. The observed failure data are 55*, 166*, 205*, 341, 488, 567, 731, 1 308, 2 050, 2 453, 3 115, 4 017, and 4 596. For comparative purposes, a mathematical transformation $\gamma = \alpha^{-\beta}$ is performed for consistency, so the failure intensity reduces to $\nu(t) = \gamma\beta t^{\beta-1}$ and the parameters of interest in this case are γ and β .

We assume $r = 1, 2, 3, 4$, where $r - 1$ is the number of missing failures at the early stage of testing. Similar to the left censored data of Subsection 4.1, the noninformative prior and the MCMC algorithm are used. The point estimates and credible intervals obtained are tabulated in Table 3, where the MLE, CCI and GCI provided by Chumnaul and Sepehrifar [31] are also listed.

Table 3 Parameter interval estimates for complete ($r = 1$) data and incomplete ($r \geq 2$) data

r	$\hat{\gamma}$		$\hat{\beta}$		95% credible interval of γ		95% confidence interval of γ		95% confidence interval of γ	
	Mean	MLE	Mean	MLE	MCMC	Width	CCI	Width	GCI	Width
1	0.131 5	0.107 2	0.551 2	0.569 0	[0.060 6, 0.246 9]	0.186 3	[0.039 4, 2.104 5]	2.065 1	[0.008 1, 1.424 4]	1.416 4
2	0.146 6	0.123 8	0.536 0	0.551 9	[0.073 8, 0.257 2]	0.183 4	[0.046 3, 2.681 1]	2.634 8	[0.010 3, 1.614 0]	1.603 7
3	0.150 6	0.167 6	0.530 3	0.515 9	[0.075 4, 0.262 3]	0.186 9	[0.059 0, 5.265 2]	5.206 2	[0.013 7, 2.074 1]	2.060 5
4	0.151 3	0.183 5	0.529 3	0.505 2	[0.074 6, 0.265 1]	0.190 5	[0.065 6, 7.211 6]	7.146 0	[0.015 1, 2.343 7]	2.328 6

Table 3 shows that $\hat{\beta}$ and $\hat{\gamma}$ decrease slightly as the value of r increases. The estimates of β obtained from the proposed method has no significant difference from the MLE, while there is a discrepancy between the estimates of γ . The reason may be that γ depends on β , and $\hat{\beta}$ can affect $\hat{\gamma}$.

When the performances of interval estimates are compared, the interval widths obtained from GCI and CCI increase as r increases, and CCI has a bigger width than GCI. However, all the widths of credible intervals are closed to each other and are less than CCI and GCI. It indicates that the proposed method can yields precise results.

4.2.2 Prototype vehicle data

Reliability growth is an effective way to improve the system reliability. The failure data analysis can help identify whether the fixes are effective and the reliability is improving. A reliability growth testing for a prototype vehicle has been conducted where some potential failure modes are surfaced and corrective actions are incorporated. However, the exact failure data of the initial testing stage are not documented, and only the failure number is reported. The exact failure times recorded are 673.6, 697.4, 796.1, 892.7, 906.3, 967.2, 1 066, 1 108. What is more, the exact failure times for the first 16 failures are not recorded, and the testing is terminated at 1 200 k.

To assess the reliability of the system after this stage of testing, the PLP is used to fit the data. Obviously, the data is the left censored type. Using the above methods, the marginal posterior distribution of β is obtained as shown in Fig. 2. The posterior mean of β is 0.65, with a 95% credible interval of (0.37, 1.14). The probability that β is less than one is quite large. It implies that the failure rate decreases over time. Therefore, the system performance is improved and the fixes implemented are effective.

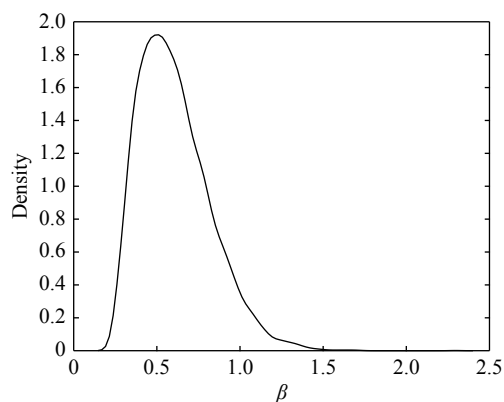


Fig. 2 Marginal posterior density of β

Similarly, it can be derived that the posterior mean of α is 15.77, with a 95% credible interval of (0.15,78.71).

Taking advantage of the inference results of α and β , the system reliability can be tracked and the present or potential level of reliability can be measured.

5. Conclusions

This paper develops a method for statistical inference of the PLP with incomplete data from the viewpoint of order statistics. It extends the application scenarios of the PLP because it can not only handle the left censored failure data, but also can be used for interval censored failure data. The mathematical derivation is comprehensive, and the case studies demonstrate and validate the efficiency of this approach.

Further research may focus on developing frequentist approaches for the interval censored case since frequentist approaches provide alternatives to Bayesian inference. For the parameters of interest, closed expressions on maximization likelihood estimators, confidence intervals, and prediction intervals can also be obtained from the point of view of order statistics. In addition, since the order statistics can simplify the successive failure times that are mutually dependent to be independent identically distributed, other methods such as the bootstrap approach and the Monte Carlo expectation maximization algorithm are also worth being studied for the incomplete data analysis. Comparative studies between different approaches are also required.

It should be noted that the incomplete data analyzed in this paper are limited to the situation where the exact time for failure is missing but the number of failures is known. Further research should pay attention to the situation where both the failure number and failure times are missed.

Acknowledgment

The authors extend sincere gratitude to the Technology Institute of Armored Force for the data provided.

References

- [1] JIN T D. Reliability engineering and services. Hoboken, New Jersey: Wiley, 2019.
- [2] HU J M, HUANG H Z, LI Y F. Reliability growth planning based on information gap decision theory. Mechanical Systems and Signal Processing, 2019, 133: 106274. DOI: 10.1016/j.ymssp.2019. 106274.
- [3] LINDQVIST B H. On the statistical modeling and analysis of repairable systems. Statistical Science, 2006, 21(4): 532–551.
- [4] MI J, LI Y F, PENG W, et al. Reliability analysis of complex multi-state system with common cause failure based on evidential networks. Reliability Engineering & System Safety, 2018, 174: 71–81.
- [5] LI Y F, HUANG H Z, MI J, et al. Reliability analysis of multi-state systems with common cause failures based on Bayesian network and fuzzy probability. Annals of Opera-

- tions Research, 2019. DOI: 10.1007/s10479-019-03247-6.
- [6] LI Y F, LIU Y, HUANG T, et al. Reliability assessment for systems suffering common cause failure based on Bayesian networks and proportional hazards model. *Quality and Reliability Engineering International*, 2020, 36(7): 2509–2520.
 - [7] SAKURAHARA T, O'SHEA N, CHENG W C, et al. Integrating renewal process modeling with probabilistic physics of failure: application to loss of coolant accident (LOCA) frequency estimations in nuclear power plants. *Reliability Engineering & System Safety*, 2019, 190: 106479. DOI: 10.1016/j.res.2019.04.032.
 - [8] GAMIZ M L, LINDQVIST B H. Nonparametric estimation in trend-renewal processes. *Reliability Engineering & System Safety*, 2016, 145: 38–46.
 - [9] YANG Q Y, HONG Y L, ZHANG N L, et al. A copula based trend renewal process model for analysis of repairable systems with multitype failures. *IEEE Trans. on Reliability*, 2017, 66(3): 509–602.
 - [10] Wu B, Cui L R, Fang C. Reliability analysis of semi-Markov systems with restriction on transition times. *Reliability Engineering & System Safety*, 2019, 190: 106516. DOI: 10.1016/j.res.2019.106516.
 - [11] Wu Bei, Cui L R. Reliability evaluation of Markov renewal shock models with multiple failure mechanisms. *Reliability Engineering & System Safety*, 2020, 202: 107051. DOI: 10.1016/j.res.2020.107051.
 - [12] Dhulipala S L N, Flint M M. Series of semi-Markov processes to model infrastructure resilience under multihazards. *Reliability Engineering & System Safety*, 2020, 193: 106659. DOI: 10.1016/j.res.2019.106659.
 - [13] WANG J Y, WU Z B, SHU Y J, et al. An optimized method for software reliability model based on nonhomogeneous Poisson process. *Applied Mathematical Modelling*, 2016, 40(13/14): 6324–6339.
 - [14] KAMRANFAR H, ETMINAN J, CHAHKANDI M. Statistical inference for a repairable system subject to shocks: classical vs. Bayesian. *Journal of Statistical Computation and Simulation*, 2019, 90(1): 112–137.
 - [15] DAS S, KUNDU D, DEWANJI A. Software reliability modeling based on NHPP for error occurrence in each fault with periodic debugging schedule. *Communications in Statistics—Theory and Methods*, 2020. DOI: 10.1080/03610926.2020.1828462.
 - [16] TANG Y C, FU J Y, LIU W C, et al. Bayesian analysis of repairable systems with modulated power law process. *Applied Mathematical Modelling*, 2017, 44: 357–373.
 - [17] Department of Defense. MIL-HDBK-89C. Reliability growth management. 2011.
 - [18] International Electronical Commission. IEC 61164. Reliability growth—statistical test and estimation methods edition 2.0. 2004.
 - [19] ASFAW Z G, LINDQVIST B H. Extending minimal repair models for repairable systems: a comparison of dynamic and heterogeneous extensions of a nonhomogeneous poisson process. *Reliability Engineering & System Safety*, 2015, 140: 53–58.
 - [20] International Electronical Commission. IEC 61710. Power law model—goodness-of-fit tests and estimation methods. 2013.
 - [21] OLIVEIRA DE M D, COLOSIMO E A, GILARDONI G L. Power law selection model for repairable systems. *Communications in Statistics—Theory and Methods*, 2013, 42(4): 570–578.
 - [22] OLIVEIRA DE M D, COLOSIMO E A, GILARDONI G L. Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference*, 2012, 142(5): 1151–1160.
 - [23] SOMBOONSAVATDEE A, SEN A. Statistical inference for power law process with competing risks. *Technometrics*, 2015, 57(1): 112–122.
 - [24] WANG B, XIE M, ZHOU J. Generalized confidence interval for the scale parameter of the power law process. *Communications in Statistics—Theory and Methods*, 2013, 42(5): 898–906.
 - [25] WANG Y P, LU Z Z. Bayesian inference and prediction analysis of the power law process based on a natural conjugate prior. *Journal of Statistical Computation and Simulation*, 2015, 85(5): 881–898.
 - [26] WANG Y, LU Z Z, XIAO S N. Parametric bootstrap confidence interval method for the power law process with applications to multiple repairable systems. *IEEE Access*, 2018, 6: 49157–49169.
 - [27] SI W J, YANG Q Y, MONPLAISIR L, et al. Reliability analysis of repairable systems with incomplete failure time data. *IEEE Trans. on Reliability*, 2018, 67(3): 1043–1059.
 - [28] GUO H, WATSON S, TAVNER P, et al. Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation. *Reliability Engineering & System Safety*, 2009, 94(6): 1057–1063.
 - [29] YU J W, TIAN G L, TANG M L. Statistical inference and prediction for the Weibull process with incomplete observations. *Computational Statistics & Data Analysis*, 2008, 52(3): 1587–1603.
 - [30] TIAN G L, TANG M L, YU J W. Bayesian estimation and prediction for the power law process with left-truncated data. *Journal of Data Science*, 2011, 9: 445–470.
 - [31] CHUMNAUL J, SEPEHRIFAR M. Generalized confidence interval for the scale parameter of the power law process with incomplete failure data. *Computational Statistics & Data Analysis*, 2018, 128: 17–33.
 - [32] NA I Y, CHANG W J. Multi-system reliability trend analysis model using incomplete data with application to tank maintenance. *Quality and Reliability Engineering International*, 2017, 33(8): 2385–2395.
 - [33] TAGHIPOUR S, BANJEVIC D. Trend analysis of the power law process using expectation-maximization algorithm for data censored by inspection intervals. *Reliability Engineering & System Safety*, 2011, 96(10): 1340–1348.
 - [34] PENG Y, WANG Y, ZI Y Y, et al. Dynamic reliability assessment and prediction for repairable systems with interval-censored data. *Reliability Engineering & System Safety*, 2017, 159: 301–309.
 - [35] AHSANULLAH M, NEVZOROV V B, SHAKIL M. An introduction to order statistics. Paris: Atlantis Press, 2013.
 - [36] HORST R. The Weibull distribution: a handbook. New York: CRC Press, 2008.
 - [37] CHA J H, FINKELSTEIN M. Point processes for reliability analysis: shocks and repairable systems. Cham: Springer, 2018.
 - [38] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977, 39(1): 1–38.

Biographies



HU Junming was born in 1987. He is currently a lecturer at Xihua University and also a Ph.D. candidate at the University of Electronic Science and Technology of China. He received his M.S. degree in weapon system and utilization engineering from Academy of Engineering Physics in 2012. His research interests include system reliability analysis, reliability growth, maintenance modeling, and uncertainty quantification.

E-mail: hujunming@std.uestc.edu.cn



HUANG Hongzhong was born in 1963. He received his Ph.D. degree in reliability engineering from Shanghai Jiao Tong University in 1999. He is a full professor at the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China. He has held visiting appointments at several universities in the USA, Canada, and several Asian countries. He is

an ISEAM fellow, a technical committee member of ESRA, a co-editor-in-chief of the International Journal of Reliability and Applications, and editorial board members of several international journals. His research interests include reliability engineering, optimization design, fuzzy sets theory, and product development.

E-mail: hzhuang@uestc.edu.cn



LI Yanfeng was born in 1981. He received his Ph.D. degree in mechatronics engineering from the University of Electronic Science and Technology of China in 2013. He is an associate professor in the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China. His research interests include reliability modeling and analysis of complex systems, dynamic fault tree analysis, Bayesian network modeling, and probabilistic inference.

E-mail: yanfengli@uestc.edu.cn