

Statistical Inference for the Entropy of the Transmuted Weibull Distribution under Adaptive Progressive Type-II Censored Samples

Shixiao Xiao

Chengyi College, Jimei University

Xinyu Wu

Chengyi College, Jimei University

Yanqiu Zeng

yqzeng1982@jmu.edu.cn

Chengyi College, Jimei University

Xue Hu

Nanchang Jiaotong Institute

Chengxi Hong

Chengyi College, Jimei University

Article

Keywords: Transmuted Weibull distribution, Shannon entropy, adaptive progressive Type-II hybrid censoring, Bayesian inference, Markov chain Monte Carlo

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10563041/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Statistical Inference for the Entropy of the Transmuted Weibull Distribution under Adaptive Progressive Type-II Censored Samples

Abstract

Introduction

Preliminaries and Theoretical Background

Transmuted Weibull Distribution

Definition of Shannon Entropy and Its Derivation for the TWD

Maximum Likelihood Estimation

Construction of the Joint Likelihood Function

Parameter Estimation Methods

Existence and Uniqueness of the Maximum Likelihood Estimator

Asymptotic Confidence Intervals

Bootstrap Confidence Interval

Bayesian Inference

Prior and Posterior Distributions

MCMC Algorithm Implementation

Bayesian Estimators of Shannon Entropy

Highest Posterior Density Credible Interval for Shannon Entropy

Monte Carlo Simulation Study

Numerical Computation Methods

Simulation Design

Empirical Analysis

Conclusion

References

Acknowledgements

Funding

Author contributions

Competing interests

Table 9. Design Table

Statistical Inference for the Entropy of the Transmuted Weibull Distribution under Adaptive Progressive Type-II Censored Samples

Shixiao Xiao^{1†}, Xinyu Wu^{2†}, Yanqiu Zeng^{3*}, Xue Hu⁴, Chengxi Hong⁵

¹ Affiliation 1; Chengyi College, Jimei University, Xiamen 361021, China

² Affiliation 2; Chengyi College, Jimei University, Xiamen 361021, China

³ Affiliation 3; Chengyi College, Jimei University, Xiamen 361021, China

⁴ Affiliation 4; Nanchang Jiaotong Institute, Nanchang, 330100, China

⁵ Affiliation 5; Chengyi College, Jimei University, Xiamen 361021, China

*Correspondence: yqzeng1982@jmu.edu.cn.

†These authors contributed equally to this work.

Author one (xiaoshixiao@jmu.edu.cn).

Author two(202441084005@jmu.edu.cn).

Author three(yqzeng1982@jmu.edu.cn).

Author four(09116@ncjti.edu.cn).

Author five(hongcx0929@jmu.edu.cn).

Abstract

This paper investigates statistical inference for the Shannon entropy of the Transmuted Weibull distribution under adaptive progressive Type-II hybrid censoring. First, the sign error in the existing analytical expression of the Shannon entropy is corrected, and maximum likelihood estimators of the model parameters and the entropy are obtained. Asymptotic and Bootstrap confidence intervals are then constructed. Under Gamma and uniform prior distributions, Bayesian estimators under the squared error and LINEX loss functions, together with highest posterior density credible intervals, are obtained using a hybrid Gibbs–Metropolis–Hastings Markov chain Monte Carlo algorithm. Monte Carlo simulations show that the Bayesian approach generally outperforms the maximum likelihood method in terms of bias, mean squared error, and interval estimation, particularly for small samples and heavy censoring. A real bladder cancer remission-time dataset further demonstrates the effectiveness and practical applicability of the proposed method.

Keywords: Transmuted Weibull distribution; Shannon entropy; adaptive progressive Type-II hybrid censoring; Bayesian inference; Markov chain Monte Carlo.

Introduction

In reliability engineering and survival analysis, many commonly used reliability measures are fundamentally based on the lifetime distribution of products. Therefore, the choice of an appropriate lifetime distribution not only affects the accuracy of characterizing failure behaviors but also has a direct impact on the precision of reliability assessment and subsequent statistical inference. Owing to its flexible shape parameter, the Weibull distribution has been widely adopted in these fields. However, when dealing with data exhibiting non-monotonic hazard rates or pronounced skewness, the classical Weibull distribution often fails to adequately capture the underlying distributional characteristics,

thereby limiting its modeling capability [1, 2]. To overcome these limitations, Aryal and Tsokos [3] introduced the Transmuted Weibull Distribution (TWD) based on the Weibull distribution by employing the quadratic rank transmutation map (QRTM) and incorporating an additional transmutation parameter λ . This construction preserves the desirable properties of the Weibull distribution while substantially enhancing its flexibility and adaptability in modeling complex lifetime data. Since its introduction, research on the TWD has progressed from investigations of its distributional properties to statistical inference under increasingly sophisticated censoring schemes. Ahmad and Ahmad et al. [4] systematically studied the structural properties of the TWD and derived several important statistical characteristics, including the mean, harmonic mean, standard deviation, moment generating function, skewness, and kurtosis, thereby establishing a theoretical foundation for subsequent studies. Pobočíková et al. [5] further evaluated the practical performance of the TWD through comprehensive goodness-of-fit analyses and demonstrated that the TWD provides a better fit than the classical Weibull distribution for complex lifetime data, thereby confirming the practical advantages of introducing the transmutation parameter. Building upon these studies, Yousaf et al. [6] investigated Bayesian estimation of the TWD parameters under both complete samples and Type-I censored samples using the squared error, precautionary, and quadratic loss functions, and compared the performance of the resulting estimators in terms of posterior risk. Subsequently, Yousaf et al. [7] extended the TWD to a two-component mixture model and derived the Bayesian estimators and credible intervals for the mixing proportion and component parameters under Type-I censoring, demonstrating that the TWD is also well suited for modeling heterogeneous lifetime populations. Additional studies on the TWD can be found in Refs. [8-10].

In practical lifetime experiments, complete lifetime observations are often unavailable because of limitations on experimental duration and testing costs. Consequently, censored samples have become an inevitable and widely encountered form of data in reliability and survival analysis [11]. Conventional Type-I censoring allows the total test duration to be predetermined, but the resulting number of observed failures may be insufficient, leading to a loss of statistical efficiency. In contrast, Type-II censoring guarantees a predetermined number of observed failures but provides no control over the total testing time, which may substantially increase experimental costs. Therefore, these two censoring schemes reduce testing costs by controlling either the experimental duration or the number of observed failures. In addition to these conventional censoring schemes, progressive censoring, first introduced by Cohen and Clifford [12], allows a predetermined number of surviving units to be randomly withdrawn during the experiment. By removing test units before the termination of the experiment, progressive censoring can further reduce both testing time and experimental cost. To combine the advantages of different censoring mechanisms, several hybrid censoring schemes have subsequently been developed. Balakrishnan and Aggarwala [13] proposed the progressive Type-II censoring scheme, in which a specified number of surviving units are randomly removed at each observed failure time. This design ensures that the predetermined number of failures is observed while making more efficient use of experimental resources. However, when highly reliable products are tested, the total duration of a progressive Type-II censoring experiment may still be excessively long. To address this issue, Kundu and Joarder [14] combined progressive Type-II censoring with Type-I censoring and proposed the progressive hybrid censoring scheme, in which a preassigned termination time T is introduced to limit the experimental duration. Nevertheless, this scheme may yield too few or even no observed failures,

thereby reducing the effectiveness of statistical inference. To overcome this drawback, Cho et al. [15] developed the generalized progressive hybrid censoring scheme, which guarantees a minimum number of observed failures by specifying a predetermined failure number, thereby improving the reliability of statistical inference. Building upon this framework, Ng et al. [16] further proposed the Adaptive Progressive Type-II Hybrid Censoring (APHC-T2) scheme. Under this scheme, the removal strategy can be adaptively adjusted according to the observed failure process during the experiment. Consequently, the experimental duration can be effectively controlled while ensuring that a sufficient number of failure observations are collected, thereby achieving a desirable balance between experimental efficiency and the accuracy of statistical inference.

The APHC-T2 scheme can be regarded as a combination of the Type-I censoring and progressive Type-II censoring schemes. It introduces a preassigned termination time T while allowing the actual experimental duration to exceed T when necessary. If the predetermined number of failures, m , is observed before the termination time T , the experiment is terminated immediately. Otherwise, if fewer than m failures have occurred by time T , the censoring scheme is instantaneously modified to accelerate the completion of the experiment. The detailed implementation procedure is described as follows:

Suppose that the initial sample consists of n units. Let m ($m < n$) denote the predetermined number of failures to be observed, T the preassigned termination time, and $R = (R_1, R_2, \dots, R_m)$ the predetermined censoring scheme. When the i th failure is observed, its lifetime is recorded as $X_{i:m:n}$, and simultaneously, R_i surviving units are randomly withdrawn from the remaining test units, where $i = 1, 2, \dots, m$. Clearly, the initial sample size and the censoring scheme satisfy $R_1 + R_2 + \dots + R_m = n - m$. Depending on the relationship between the occurrence time of the m th failure, $X_{m:m:n}$, and the preassigned termination time T , the APHC-T2 scheme gives rise to the following two cases:

If the m th failure occurs before the preassigned termination time T , that is, $X_{m:m:n} < T$, the experiment is terminated immediately after recording $X_{m:m:n}$. In this case, the censoring scheme remains identical to the predetermined censoring scheme.

If the m th failure occurs after the preassigned termination time T , that is, $X_{m:m:n} > T$, the predetermined censoring scheme must be modified. Suppose that only r ($r < m$) failures have been observed before time T , that is, $X_{r:m:n} < T < X_{r+1:m:n}$. To accelerate the completion of the experiment, no surviving units are withdrawn after the lifetime of the $r+1$ th failed unit has been recorded; that is, $R_{r+1} = R_{r+2} = \dots = R_{m-1} = 0$. When the lifetime of the m th failed unit is recorded, all remaining surviving units are withdrawn simultaneously. Consequently, the predetermined censoring scheme is modified to $R = (R_1, R_2, \dots, R_r, 0, \dots, 0, R_m)$, where $R_m = n - m - (R_1 + R_2 + \dots + R_r)$.

Owing to its adaptive mechanism, the APHC-T2 scheme not only ensures that a sufficient number of failure observations are collected but also effectively controls the experimental duration within an acceptable range. Consequently, it has attracted considerable attention and has been widely applied to statistical inference for various lifetime distributions. Lv et al. [17] investigated parameter estimation for the Gompertz distribution based on APHC-T2 samples and derived the MLEs and Bayesian estimators

under several loss functions. In addition, they constructed confidence intervals and credible intervals using the Bootstrap and MCMC methods, respectively. Elshahhat and Nassar [18] performed Bayesian survival analysis for the Hjorth distribution under APHC-T2 and obtained Bayesian estimates of the model parameters and reliability function within the MCMC framework. Mustafa and Ahmed [19] systematically studied the MLEs and Bayesian estimators of the model parameters, reliability function, and coefficient of variation for the Power Lomax distribution under APHC-T2, and further constructed asymptotic confidence intervals. Hussam and ALMetwally [20] investigated step-stress accelerated life testing for the Perks distribution under APHC-T2. Both classical and Bayesian approaches were employed to estimate the model parameters and reliability function, and asymptotic as well as Bootstrap confidence intervals were subsequently developed.

Shannon [21] inspired by the concept of entropy in thermodynamic systems, introduced information entropy as a quantitative measure of information content. Subsequently, Cover [22] extended Shannon's concept by defining the Shannon entropy of continuous random variables. This extension has greatly advanced the development of information theory and provided a new perspective for quantifying the amount of information or uncertainty associated with random variables. In reliability engineering, Shannon entropy has been widely used to characterize the degradation of the remaining lifetime of a system [23]. In medical survival analysis, it has been employed to quantify the uncertainty associated with human mortality [24]. Further applications of Shannon entropy can be found in Refs [25-27].

Since the entropy of a random variable depends on its underlying lifetime distribution, the statistical inference of entropy under different lifetime distributions has attracted considerable attention. To date, substantial research has been devoted to entropy estimation based on censored samples. Gong and Yin [28] derived the MLEs and Bayesian estimators of the Shannon and Rényi entropies for the generalized inverse exponential distribution under progressive Type-II censoring and compared the performance of the resulting estimators under various loss functions. Hassan [29] investigated the estimation of several entropy measures for the Weibull distribution under both Bayesian and classical frameworks and comprehensively evaluated the performance of different estimators under the squared error, LINEX, and generalized entropy loss functions. Alotaibi et al. [30] further studied the statistical inference of the Shannon, Rényi, and Tsallis entropies for the inverse Weibull distribution based on unified hybrid censored samples. Ren et al. [31] investigated the MLE of the Shannon entropy for the inverse exponential Rayleigh distribution under progressive Type-II censoring and obtained the corresponding Bayesian estimator using the Lindley approximation and MCMC methods. Additional studies on the statistical inference of entropy can be found in Refs [32-35].

The aforementioned studies have established an important foundation for the statistical inference of Shannon entropy under censored samples. However, research on the TWD remains limited. Ahmad and Ahmad et al. [4] derived the fundamental statistical characteristics and structural properties of the TWD; however, their work was confined to theoretical investigations and did not address parameter estimation under any censoring scheme. Yousaf et al. [6] investigated Bayesian estimation of the TWD parameters based on complete samples and Type-I censored samples. Nevertheless, since Type-I censoring controls only the experimental duration without guaranteeing a sufficient number of observed failures, the accuracy of statistical inference may be compromised when highly reliable products are tested. Moreover, the statistical inference of Shannon entropy was not considered. Khan et

al. [36] derived an analytical expression for the Shannon entropy of the TWD while investigating its distributional properties. However, the derived expression contains an overall sign error, resulting in every term having the opposite sign of the correct expression. Yousaf et al. [7] further investigated parameter estimation for a two-component TWD mixture model under Type-I censoring but likewise did not consider the statistical inference of Shannon entropy. Motivated by these observations, this paper first corrects the sign error in the analytical expression for the Shannon entropy of the TWD reported by Khan et al. [36]. On this basis, the MLE and Bayesian estimation of the Shannon entropy are investigated under the APHC-T2 scheme. The remainder of this paper is organized as follows:

The remainder of this paper is organized as follows. Section 2 presents the fundamental properties of the TWD and derives the analytical expression for its Shannon entropy. Section 3 establishes the likelihood equations for the distribution parameters based on APHC-T2 samples and proves the existence and uniqueness of the MLEs. By virtue of the invariance property of the MLEs, the MLE of the Shannon entropy is subsequently obtained. Considering the limitations of asymptotic inference for small sample sizes, ACI and Bootstrap confidence intervals are further constructed for the Shannon entropy. Section 4 assumes Gamma priors and a uniform prior for the model parameters, and derives the Bayesian estimators of the Shannon entropy under the squared error loss (SEL) and LINEX loss (LL) functions, respectively. A hybrid Gibbs sampler is then implemented to perform the MCMC algorithm, from which the Bayesian estimates and HPD credible intervals of the Shannon entropy are obtained. Section 5 evaluates and compares the performance of the proposed estimation methods through Monte Carlo simulations. Section 6 illustrates the proposed methods using a real medical survival dataset. Finally, Section 7 concludes the paper and discusses directions for future research.

Preliminaries and Theoretical Background

Transmuted Weibull Distribution

A random variable X is said to follow a three-parameter Weibull distribution, denoted by $X \sim W(a, b, c)$, if its cumulative distribution function (CDF) and probability density function (PDF) are given by:

$$F_W(x; a, b, c) = 1 - \exp\left(-\left(\frac{x-c}{b}\right)^a\right), \quad x > 0 \quad (1)$$

$$f_W(x; a, b, c) = \frac{a}{b} \left(\frac{x-c}{b}\right)^{a-1} \exp\left(-\left(\frac{x-c}{b}\right)^a\right), \quad x > 0 \quad (2)$$

where $a > 0$, $b > 0$, and $c \in \mathbb{R}$ denote the shape, scale, and location parameters, respectively.

Setting $c = 0$, reduces Eqs.(1) and (2) to the CDF and PDF of the two-parameter Weibull distribution, given in Eqs. (3) and (4), respectively. Accordingly X is said to follow a two-parameter Weibull distribution, denoted by $X \sim W(a, b)$.

$$G_0(x; a, b) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right), x > 0 \quad (3)$$

$$g_0(x; a, b) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right), x > 0 \quad (4)$$

Shaw and Buckle [37] proposed the QRTM. Given a baseline distribution with CDF and PDF, the corresponding transmuted distribution is defined by:

$$F_T(x; \lambda) = (1 + \lambda)G(x) - \lambda G(x)^2 \quad (5)$$

$$f_T(x; \lambda) = g(x)[(1 + \lambda) - 2\lambda G(x)] \quad (6)$$

Where λ is the transmutation parameter, and $G(x)$ and $g(x)$ denote the CDF and PDF of the baseline distribution, respectively. According to the QRTM, any continuous distribution can be adopted as the baseline distribution. By substituting the CDF and PDF of the baseline distribution into Eqs.(5) and (6), the corresponding family of transmuted distributions can be generated.

By taking the two-parameter Weibull distribution as the baseline distribution, Aryal and Tsokos [3] introduced the TWD, whose CDF and PDF are given by:

$$F(x; a, b, \lambda) = \left[1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)\right] \left[1 + \lambda \exp\left(-\left(\frac{x}{b}\right)^a\right)\right], x > 0 \quad (7)$$

$$f(x; a, b, \lambda) = \frac{a}{b^a} x^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right) \left[1 - \lambda + 2\lambda \exp\left(-\left(\frac{x}{b}\right)^a\right)\right], x > 0 \quad (8)$$

Where $a > 0$, $b > 0$, and $|\lambda| \leq 1$, with λ denoting the transmutation parameter. A random variable X with the above CDF and PDF is said to follow the TWD, denoted by $X \sim TW(a, b, \lambda)$. Figures 1 and 2 illustrate the shapes of the CDF and PDF of the TWD for different parameter settings.

It is worth noting that the TWD is an extended model capable of providing greater flexibility for analyzing complex data and includes several widely used distributions as special cases. When $a = 1$, the TWD reduces to the transmuted exponential distribution, as discussed by Aryal and Tsokos [3]. When $\lambda = 0$, the TWD reduces to the standard Weibull distribution. When $a = \lambda = 1$, the resulting distribution becomes an exponential distribution with parameter $b/2$.

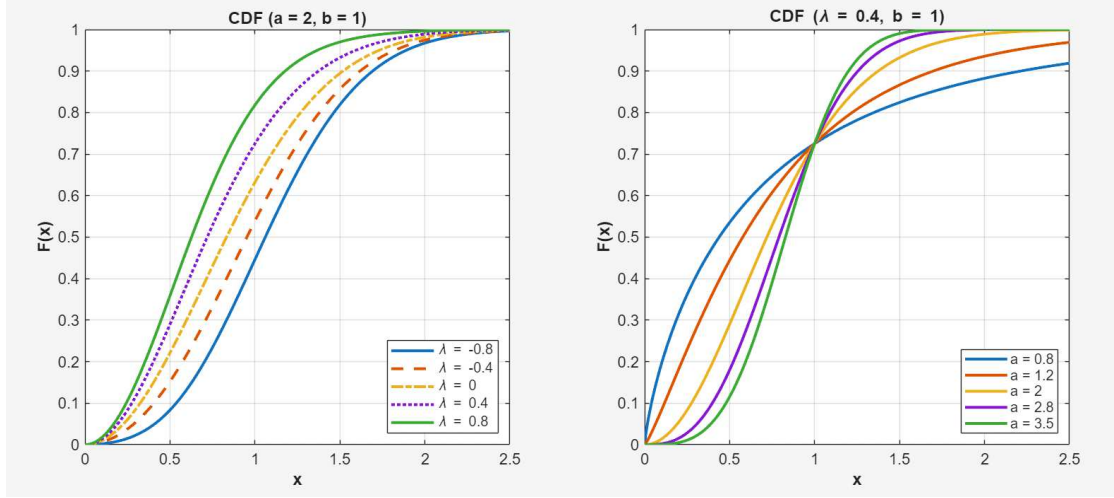


图 1 TWD 的 CDF

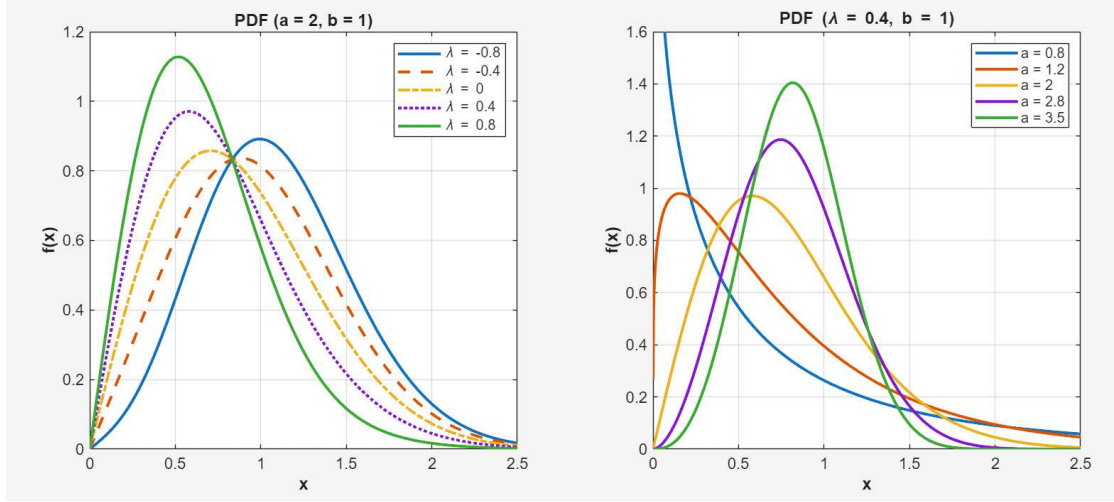


Figure 2. PDF of the TWD

Definition of Shannon Entropy and Its Derivation for the TWD

For a continuous random variable X with PDF $f(x)$, the Shannon entropy $H(f)$ is defined as:

$$H(f) = E(-\ln f(x)) \quad (9)$$

Theorem. Suppose that X follows the TWD, denoted by $X \sim TW(a, b, \lambda)$, with PDF $f(x; a, b, \lambda)$ given by Eq.(8). Let $\theta = (a, b, \lambda)$. Then, the Shannon entropy of the TWD is given by:

$$\begin{aligned} H(\theta) &= E[-\ln f(x; a, b, \lambda)] \\ &= -\ln(a) + \ln(b) - (a-1)E\left(\ln \frac{x}{b}\right) + E\left[\left(\frac{x}{b}\right)^a\right] - E\left[\ln\left(1 - \lambda + 2\lambda e^{-(x/b)^a}\right)\right] \end{aligned} \quad (10)$$

Proof. The expectation in Eq.(10) can be obtained through the following steps:

$$E\left[\ln\left(\frac{x}{b}\right)\right] = \Gamma\left(\frac{1}{a}\right)(1) - \lambda \ln(2)^{\frac{1}{a}}$$

$$E \left[\ln \left(1 - \lambda + 2\lambda e^{-\left(\frac{x}{b}\right)^a} \right) \right] = \ln(1 - \lambda) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{2\lambda}{1 - \lambda} \right)^k \int_0^{\infty} e^{-k\left(\frac{x}{b}\right)^a} f(x; a, b, \lambda) dx$$

The above integral can be evaluated as follows:

$$\begin{aligned} K &= \int_0^{\infty} \frac{a}{b} \left(\frac{x}{b} \right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} \left[1 - \lambda + 2\lambda e^{-\left(\frac{x}{b}\right)^a} \right] dx \\ K &= (1 - \lambda) \int_0^{\infty} \frac{a}{b} \left(\frac{x}{b} \right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} dx + 2\lambda \int_0^{\infty} \frac{a}{b} \left(\frac{x}{b} \right)^{a-1} e^{-2\left(\frac{x}{b}\right)^a} dx \\ K &= \frac{1 - \lambda}{k + 1} + \frac{2\lambda}{k + 2} \end{aligned}$$

The expectation in Eq.(10) can be obtained through integration as follows:

$$E \left[\ln \left(1 - \lambda + 2\lambda e^{-\left(\frac{x}{b}\right)^a} \right) \right] = \ln(1 - \lambda) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{2\lambda}{1 - \lambda} \right)^k \left(\frac{1 - \lambda}{k + 1} + \frac{2\lambda}{k + 2} \right)$$

Finally, we obtain:

$$\begin{aligned} H(\theta) &= -\ln(a) + \ln(b) - (a - 1) \left[\Gamma\left(\frac{1}{a}\right) (1 - \lambda \ln(2))^{\left(\frac{1}{a}\right)} \right] + \left(1 - \frac{\lambda}{2} \right) - \ln(1 - \lambda) - \\ &\quad \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{2\lambda}{1 - \lambda} \right)^k \left(\frac{1 - \lambda}{k + 1} + \frac{2\lambda}{k + 2} \right) \end{aligned} \quad (11)$$

Maximum Likelihood Estimation

Construction of the Joint Likelihood Function

Suppose that $X_{1:m:n}, X_{2:m:n}, \dots, X_{m:m:n}$ are the observed APHC-T2 samples from n test units, where each $X_{i:m:n}$ ($0 < i < m$) follows the $X \sim TW(a, b, \lambda)$. Under the APHC-T2 scheme, the likelihood function of the parameter vector a, b, λ is given by:

$$L(a, b, \lambda | x) = \left(\frac{a}{b^a} \right)^m \left[\prod_{i=1}^m x_i^{a-1} \right] e^{-\frac{1}{b^a} \sum_{i=1}^m (R_i + 1) x_i^a} \prod_{i=1}^m \left[1 - \lambda + 2\lambda e^{-\left(x_i/b\right)^a} \right] \quad (12)$$

Where $x = (x_{1:m:n}, x_{2:m:n}, \dots, x_{m:m:n})$ represents the observed failure times.

The corresponding log-likelihood function is given by:

$$\begin{aligned} \ln L(a, b, \lambda | x) &= m(\ln a - a \ln b) + (a - 1) \sum_{i=1}^m \ln x_i - \sum_{i=1}^m (1 + R_i) \left(\frac{x_i}{b} \right)^a \\ &\quad + \sum_{i=1}^m \ln \left[1 - \lambda + 2\lambda e^{-\left(\frac{x_i}{b}\right)^a} \right] + \sum_{i=1}^m R_i \ln \left[1 - \lambda + \lambda e^{-\left(\frac{x_i}{b}\right)^a} \right] \end{aligned} \quad (13)$$

The statistical inference of MLEs under the APHC-T2 scheme has been extensively investigated. Early studies mainly focused on the optimality properties of estimators. Balakrishnan and Mitra [38] derived the MLEs, uniformly minimum variance unbiased estimators, and best linear unbiased estimators for the two-parameter exponential distribution system, and established recursive relationships for the moments of variances and covariances. Subsequently, research has been extended to more complex model settings. For the Lomax distribution under competing risks, optimal censoring schemes were designed by maximizing the expected Fisher information, and the corresponding parameter estimation problem was further investigated [39]. Elsherpieny et al. [40] applied MLEs to model discrimination between the Weibull and Log-normal distributions. They constructed discrimination statistics based on the values of the log-likelihood functions and numerically evaluated the probability of correct model selection under small sample sizes.

Parameter Estimation Methods

Under the APHC-T2 scheme, the log-likelihood function of the TWD parameters is first constructed. Subsequently, the MLE equations for the shape parameter a , scale parameter b , and transmutation parameter λ are derived. The detailed procedures are given as follows.

For convenience of notation, some commonly used terms are simplified as follows:

$$\text{Let } A_i = \left(\frac{x_i}{b} \right)^a$$

$$\text{Let } E_i = e^{-A_i} = e^{-\left(\frac{x_i}{b} \right)^a}$$

$$\text{Let } Y_i = \ln \left(\frac{x_i}{b} \right)$$

Taking the partial derivative with respect to the shape parameter a :

$$\frac{\partial \ln L}{\partial a} = \frac{m}{a} + \sum_{i=1}^m Y_i \left[1 - A_i - R_i A_i - \frac{2\lambda E_i A_i}{1 - \lambda + 2\lambda E_i} - \frac{R_i \lambda E_i A_i}{1 - \lambda + \lambda E_i} \right] \quad (14)$$

Taking the partial derivative with respect to the scale parameter b :

$$\frac{\partial \ln L}{\partial b} = \frac{a}{b} \left(-m + \sum_{i=1}^m \left[A_i + R_i A_i + \frac{2\lambda E_i A_i}{1 - \lambda + 2\lambda E_i} + \frac{R_i \lambda E_i A_i}{1 - \lambda + \lambda E_i} \right] \right) \quad (15)$$

Taking the partial derivative with respect to the transmutation parameter λ :

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^m \frac{2E_i - 1}{1 - \lambda + 2\lambda E_i} + \sum_{i=1}^m \frac{R_i(E_i - 1)}{1 - \lambda + \lambda E_i} \quad (16)$$

The solution to the system of equations(17) provides the MLEs of the model parameters. Since this system generally has no closed-form solution, numerical optimization methods, such as the Newton–Raphson algorithm, are employed to solve the equations and obtain the parameter estimates $\hat{a}, \hat{b}, \hat{\lambda}$.

$$\begin{cases} \frac{\partial \ln L}{\partial a} = 0 \\ \frac{\partial \ln L}{\partial b} = 0 \\ \frac{\partial \ln L}{\partial \lambda} = 0 \end{cases} \quad (17)$$

The Newton–Raphson algorithm is employed to obtain the MLEs of the parameters a, b, λ .

Suppose that the MLE of the parameter vector θ is denoted by $\hat{\theta} = (\hat{a}, \hat{b}, \hat{\lambda})^T$. The iterative procedure is described as follows:

- (1) Take the logarithm of the likelihood function to obtain the log-likelihood function.
- (2) Calculate the first-order and second-order derivatives of the log-likelihood function.
- (3) Initialize the parameter values, where the sample mean or other empirical values are commonly used as the initial values.
- (4) Evaluate the first-order and second-order derivatives using the current parameter values.
- (5) Update the parameter values using the Newton–Raphson iterative formula:

$$\hat{\theta}_{new} = \hat{\theta}_{old} - Hess(\hat{\theta}_{old})^{-1} \nabla \ln L(\hat{\theta}_{old}) \quad (18)$$

(6) Repeat Steps (4) and (5) until the parameter estimates converge or the maximum number of iterations is reached. The final parameter values are regarded as the MLEs.

In the above procedure, $\nabla \ln L(\theta)$ denotes the vector of first-order partial derivatives of the log-likelihood function with respect to θ , namely, the score vector. Let $Hess(\theta)$ represent the Hessian matrix of the log-likelihood function with respect to θ , and $Hess(\theta)^{-1}$ denote the inverse of the Hessian matrix.

Based on the invariance property of MLEs, the estimates of the model parameters obtained during the iterative procedure can be directly substituted into Eq.(11) to obtain the MLE of the Shannon entropy. For notational convenience, it is denoted by $\hat{H}_{ML}$, and its expression is given as follows:

$$\begin{aligned} \hat{H}_{ML} = & -\ln(\hat{a}) + \ln(\hat{b}) - (\hat{a} - 1) \left[\Gamma\left(\frac{1}{\hat{a}}\right) (1) - \hat{\lambda} \ln(2) \left(\frac{1}{\hat{a}}\right) \right] \\ & + \left(1 - \frac{\hat{\lambda}}{2}\right) - \ln(1 - \hat{\lambda}) - \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{2\hat{\lambda}}{1-\hat{\lambda}}\right)^k \left(\frac{1-\hat{\lambda}}{k+1} + \frac{2\hat{\lambda}}{k+2}\right) \end{aligned} \quad (19)$$

Existence and Uniqueness of the Maximum Likelihood Estimator

Theorem. Let $x_1 < x_2 < \dots < x_m$ be APHC-T2 samples from $TW(a, b, \lambda)$, where $a > 0, b > 0, \lambda \in [-1, 1]$. Then, the MLEs of the parameter vector a, b, λ exist and are unique.

Proof. Denote the log-likelihood function by $\ln L(a, b, \lambda)$, By setting the partial derivatives of $\ln L(a, b, \lambda)$ with respect to the three parameters equal to zero, the likelihood equations(17) are obtained. The existence and uniqueness of the roots of each equation are then investigated separately for the corresponding parameters.

(1) Uniqueness with respect to the parameter a :

The function $\frac{\partial \ln L}{\partial a}$ is continuous with respect to $a \in (0, +\infty)$. We analyze its limiting behavior at the boundaries.

When $a \rightarrow 0^+$ the term involving $\frac{m}{a} \rightarrow +\infty$ dominates in the likelihood equation, while the remaining terms remain bounded under this limit. Therefore:

$$\lim_{a \rightarrow 0^+} \frac{\partial \ln L}{\partial a} = +\infty \quad (20)$$

When $a \rightarrow +\infty$, for the observations satisfying x_i , the term involving $A_i \rightarrow -\infty$ appears, and thus:

$$\lim_{a \rightarrow +\infty} \frac{\partial \ln L}{\partial a} = -\infty \quad (21)$$

According to the intermediate value theorem, the equation $\frac{\partial \ln L}{\partial a} = 0$ has at least one root in the interval $(0, +\infty)$. Furthermore, according to Eq. (14), $\frac{\partial \ln L}{\partial a}$ is strictly monotonically decreasing with respect to a . Hence, the equation has a unique solution. Therefore, the MLE of $\hat{a}$ exists and is unique.

(2) Uniqueness with respect to the parameter b :

The function $\frac{\partial \ln L}{\partial b}$ is continuous with respect to $b \in (0, +\infty)$. The limiting behavior at the boundaries is investigated as follows.

When $b \rightarrow 0^+$, according to Eq.(15), the term involving $A_i \rightarrow +\infty$, therefore, $E_i \rightarrow 0$ dominates in the likelihood equation. Since the growth rate of the latter term is much faster than that of the former term, we have:

$$\lim_{b \rightarrow 0^+} \frac{\partial \ln L}{\partial b} = +\infty \quad (22)$$

When $b \rightarrow +\infty$, so that $A_i \rightarrow 0$, hence, $E_i \rightarrow 1$, and $-\frac{ma}{b}$ is always negative, therefore:

$$\lim_{b \rightarrow +\infty} \frac{\partial \ln L}{\partial b} = -\infty \quad (23)$$

By the intermediate value theorem, the equation $\frac{\partial \ln L}{\partial b} = 0$ has at least one root in $(0, +\infty)$.

Moreover, according to Eq. (15), $\frac{\partial \ln L}{\partial b}$ is strictly decreasing with respect to b . Hence, the equation has

a unique solution. Therefore, the MLE of $\hat{b}$ exists and is unique.

(3) Uniqueness with respect to the parameter λ :

The function $\frac{\partial \ln L}{\partial \lambda}$ is continuous with respect to $\lambda \in [-1, 1]$. The signs at the endpoints of the parameter space are investigated.

When $\lambda \rightarrow -1^+$, the positive terms associated with $1 - \lambda + 2\lambda E_i$ in the likelihood equation dominate, and thus,

$$\left. \frac{\partial \ln L}{\partial \lambda} \right|_{\lambda=-1^+} > 0 \quad (24)$$

When $\lambda \rightarrow 1^-$, the signs of the above terms are reversed, and the negative terms dominate, yielding:

$$\left. \frac{\partial \ln L}{\partial \lambda} \right|_{\lambda=1^-} < 0 \quad (25)$$

By the intermediate value theorem, a solution for $\hat{\lambda}$ exists. According to Eq. (33) in Section 3.4, the numerators of all terms are squares of real numbers, and they are preceded by negative signs, while the denominators remain positive within the parameter space. Therefore, $\frac{\partial \ln L}{\partial \lambda}$ is strictly decreasing with respect to λ , implying that the equation has a unique root. Hence, the MLE of $\hat{\lambda}$ exists and is unique within $(-1, 1)$.

Combining the above results, the likelihood equations (17) have a unique solution. Therefore, the MLE of the parameter vector (a, b, λ) denoted by $(\hat{a}, \hat{b}, \hat{\lambda})$, exists and is unique.

Asymptotic Confidence Intervals

To quantify the uncertainty associated with the entropy estimation, the ACI is constructed based on the asymptotic theory. The ACI method obtains the gradient vector by taking the first-order partial derivatives of the entropy function with respect to the parameters. Combined with the asymptotic covariance matrix of the parameter estimators derived from the Fisher information matrix, the Delta method is employed to calculate the asymptotic variance of the estimator. Subsequently, the corresponding ACI is constructed under the framework of large-sample theory.

The Fisher information matrix $I(\theta)$ is defined as the matrix composed of the negative expected values of the second-order derivatives of the log-likelihood function, and its inverse matrix represents the asymptotic covariance matrix of the parameter estimators.

$$I(\theta) = -E \begin{bmatrix} \frac{\partial^2 \ln L}{\partial a^2} & \frac{\partial^2 \ln L}{\partial a \partial b} & \frac{\partial^2 \ln L}{\partial a \partial \lambda} \\ \frac{\partial^2 \ln L}{\partial b \partial a} & \frac{\partial^2 \ln L}{\partial b^2} & \frac{\partial^2 \ln L}{\partial b \partial \lambda} \\ \frac{\partial^2 \ln L}{\partial \lambda \partial a} & \frac{\partial^2 \ln L}{\partial \lambda \partial b} & \frac{\partial^2 \ln L}{\partial \lambda^2} \end{bmatrix} \quad (26)$$

In this context, the elements are as follows:

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{m}{a^2} - \sum_{i=1}^m A_i Y_i^2 - \sum_{i=1}^m \frac{4\lambda^2 E_i^2 A_i^2 Y_i^2}{(1-\lambda+2\lambda E_i)^2} \quad (27)$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial a \partial b} = & -\frac{m}{b} + \sum_{i=1}^m \frac{A_i}{b} + \sum_{i=1}^m \frac{a A_i}{b} Y_i - \sum_{i=1}^m \frac{2\lambda a A_i^2 E_i Y_i}{b(1-\lambda+2\lambda E_i)} \\ & + \sum_{i=1}^m \frac{2\lambda a A_i E_i Y_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{2\lambda A_i E_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{8\lambda^2 E_i^2 A_i^2 Y_i}{(1-\lambda+2\lambda E_i)^2} \end{aligned} \quad (28)$$

$$\frac{\partial^2 \ln L}{\partial a \partial \lambda} = -\sum_{i=1}^m \frac{2E_i A_i Y_i}{(1-\lambda+2\lambda E_i)^2} - \sum_{i=1}^m \frac{R_i E_i A_i Y_i}{(1-\lambda+\lambda E_i)^2} \quad (29)$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial b \partial a} = & -\frac{m}{b} + \sum_{i=1}^m \frac{A_i}{b} + \sum_{i=1}^m \frac{a A_i}{b} Y_i - \sum_{i=1}^m \frac{2\lambda a A_i^2 E_i Y_i}{b(1-\lambda+2\lambda E_i)} \\ & + \sum_{i=1}^m \frac{2\lambda a A_i E_i Y_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{2\lambda A_i E_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{4\lambda^2 E_i^2 A_i^2 Y_i}{b(1-\lambda+2\lambda E_i)^2} \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial b^2} = & \frac{ma}{b^2} - \sum_{i=1}^m \frac{a A_i}{b^2} - \sum_{i=1}^m \frac{a^2 A_i}{b^2} + \sum_{i=1}^m \frac{2\lambda a A_i^2 E_i}{b^2(1-\lambda+2\lambda E_i)} - \sum_{i=1}^m \frac{2\lambda a A_i E_i}{b^2(1-\lambda+2\lambda E_i)} \\ & - \sum_{i=1}^m \frac{4\lambda^2 a^2 A_i^2 E_i^2}{b^2(1-\lambda+2\lambda E_i)^2} - \sum_{i=1}^m \frac{R_i a A_i}{b^2} - \sum_{i=1}^m \frac{R_i a^2 A_i}{b^2} + \sum_{i=1}^m \frac{R_i \lambda a A_i^2 E_i}{b^2(1-\lambda+\lambda E_i)} \\ & - \sum_{i=1}^m \frac{R_i \lambda a A_i E_i}{b^2(1-\lambda+\lambda E_i)} - \sum_{i=1}^m \frac{R_i \lambda^2 a^2 A_i^2 E_i^2}{b^2(1-\lambda+\lambda E_i)^2} \end{aligned} \quad (31)$$

$$\frac{\partial^2 \ln L}{\partial b \partial \lambda} = -\sum_{i=1}^m \frac{R_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{A_i + 2\lambda A_i E_i}{b(1-\lambda+2\lambda E_i)^2} \quad (32)$$

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\sum_{i=1}^m \frac{(2E_i-1)^2}{(1-\lambda+2\lambda E_i)^2} - \sum_{i=1}^m \frac{R_i(E_i-1)^2}{(1-\lambda+\lambda E_i)^2} \quad (33)$$

$$\frac{\partial^2 \ln L}{\partial \lambda \partial a} = -\sum_{i=1}^m \frac{2E_i A_i Y_i}{(1-\lambda+2\lambda E_i)^2} - \sum_{i=1}^m \frac{R_i E_i A_i Y_i}{(1-\lambda+\lambda E_i)^2} \quad (34)$$

$$\frac{\partial^2 \ln L}{\partial \lambda \partial b} = \sum_{i=1}^m \frac{R_i}{b(1-\lambda+2\lambda E_i)} + \sum_{i=1}^m \frac{A_i(1-\lambda+2\lambda E_i)}{b(1-\lambda+2\lambda E_i)^2} + \sum_{i=1}^m \frac{2\lambda^2 E_i A_i}{b(1-\lambda+2\lambda E_i)^2} \quad (35)$$

Since the expectations involved in the Fisher information matrix are difficult to obtain analytically, the observed Fisher information matrix is usually employed instead. The Fisher information matrix can be approximated by:

$$I(\hat{\theta}) = - \begin{bmatrix} \frac{\partial^2 \ln L}{\partial a^2} & \frac{\partial^2 \ln L}{\partial a \partial b} & \frac{\partial^2 \ln L}{\partial a \partial \lambda} \\ \frac{\partial^2 \ln L}{\partial b \partial a} & \frac{\partial^2 \ln L}{\partial b^2} & \frac{\partial^2 \ln L}{\partial b \partial \lambda} \\ \frac{\partial^2 \ln L}{\partial \lambda \partial a} & \frac{\partial^2 \ln L}{\partial \lambda \partial b} & \frac{\partial^2 \ln L}{\partial \lambda^2} \end{bmatrix}_{\theta=\hat{\theta}}$$

In this case, the inverse Fisher information matrix $I^{-1}(\hat{\theta})$ evaluated at the MLEs can be directly used to approximate the inverse of the observed Fisher information matrix, that is:

$$I^{-1}(\hat{\theta}) = \begin{bmatrix} Var(\hat{a}) & Cov(\hat{a}, \hat{b}) & Cov(\hat{a}, \hat{\lambda}) \\ Cov(\hat{b}, \hat{a}) & Var(\hat{b}) & Cov(\hat{b}, \hat{\lambda}) \\ Cov(\hat{\lambda}, \hat{a}) & Cov(\hat{\lambda}, \hat{b}) & Var(\hat{\lambda}) \end{bmatrix}$$

Where, $Var(\cdot)$ denotes the variance of the estimator and $Cov(\cdot, \cdot)$ represents the covariance between two estimators.

To obtain the asymptotic covariance matrix of the entropy estimator, the partial derivatives of the entropy function $H(\theta)$ with respect to each parameter need to be calculated. The gradient vector $k(\theta)$ is defined as :

$$k(\theta) = \nabla H(\theta) = \left(\frac{\partial H(\theta)}{\partial a}, \frac{\partial H(\theta)}{\partial b}, \frac{\partial H(\theta)}{\partial \lambda} \right)^T$$

By substituting the MLEs into the above expression, we obtain $\hat{k} = k(\hat{\theta})$. According to the Delta method, the asymptotic variance estimator of the MLE of the Shannon entropy $\hat{H}_{ML}$ is given by:

$$Var(\hat{H}_{ML}) = \hat{k}^T I^{-1}(\hat{\theta}) \hat{k}$$

Based on the above derivation, the asymptotic variance estimator of the Shannon entropy can be obtained, and the standard form of the ACI can consequently be constructed. The ACI for the Shannon entropy $100(1-\gamma)\%$ is given by:

$$[\hat{H}_{ML} - Z_{1-\gamma/2} \sqrt{Var(\hat{H}_{ML})}, \hat{H}_{ML} + Z_{1-\gamma/2} \sqrt{Var(\hat{H}_{ML})}] \quad (36)$$

Where, $Z_{1-\gamma/2}$ denotes the $(1-\gamma/2)$ quantile of the standard normal distribution. This interval represents the range in $1-\gamma$ which the true value of the Shannon entropy may lie with a given confidence level.

Bootstrap Confidence Interval

Confidence intervals are important tools for quantifying the uncertainty of parameters in statistical inference. Due to the randomness of samples, point estimators are subject to fluctuations, whereas interval estimation can provide a more comprehensive description of the possible range of parameter values. In small-sample cases or under complex distributional assumptions, the ACI method based on asymptotic normality may fail to achieve satisfactory coverage probabilities. To overcome this limitation, Efron [41] proposed the Bootstrap method as an effective alternative approach. The Bootstrap method constructs Bootstrap samples by repeatedly resampling the original observations with replacement and approximates the sampling distribution of a statistic based on these resampled samples, thereby obtaining interval estimates. Compared with conventional methods, the Bootstrap method does not rely on specific distributional assumptions and exhibits better applicability for skewed or heavy-tailed data. It has been widely validated in parameter estimation and interval inference for various complex models. For APHC-T2 samples, the original censoring mechanism should be preserved during the resampling procedure to ensure that the generated samples faithfully reflect the underlying data structure. Bootstrap confidence intervals can be constructed using several approaches, including the Bootstrap-t method, percentile method, and bias-corrected and accelerated (BCa) method. In this study, the Bootstrap-t method is adopted to construct the confidence interval for the Shannon entropy. The detailed procedure is described as follows:

- (1) A Bootstrap sample $x_1^*, x_2^*, x_3^*, \dots, x_m^*$ is generated by randomly drawing m observations with replacement from the original APHC-T2 sample $x = (x_1, x_2, \dots, x_m)$ of size m .
- (2) The MLE of the Shannon entropy is calculated based on the generated Bootstrap sample and denoted by $\hat{H}^*$.
- (3) The above procedure is repeated N times, resulting in N Bootstrap estimates of the Shannon entropy. These estimates are arranged in ascending order and denoted by $\hat{H}_1^*, \hat{H}_2^*, \hat{H}_3^*, \dots, \hat{H}_N^*$.
- (4) For a given confidence level $1 - \gamma$, the Bootstrap confidence interval of the Shannon entropy is obtained as follows, where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling and floor functions, respectively.

$$\left[\hat{H}_{(\lceil N\gamma/2 \rceil)}^*, \hat{H}_{(\lfloor N(1-\gamma/2) \rfloor)}^* \right] \quad (37)$$

Bayesian Inference

In reliability analysis and information measurement studies, MLEs may suffer from substantial bias or even computational instability under small sample sizes or heavy censoring levels. Bayesian inference provides an effective complement to overcome these limitations. Bayesian statistics treats unknown parameters as random variables and combines prior distributions with sample likelihoods through Bayes' theorem to obtain the posterior distribution. Specifically, the prior distribution reflects the prior knowledge about the parameters before observing the data, the likelihood function describes the probability of the observed data given the parameters, and the posterior distribution integrates both

sources of information. Under this framework, the uncertainty of parameters can be fully characterized in a probabilistic manner, and credible intervals possess a direct probabilistic interpretation. Even when the posterior distribution does not have an analytical form, numerical methods such as MCMC can still provide effective statistical inference. In recent years, Bayesian methods have demonstrated considerable potential in reliability modeling and lifetime data analysis, and a number of studies have also been conducted under the APHC-T2 framework. Nassar et al. [42] investigated both MLEs and Bayesian estimators of the Weibull distribution parameters based on APHC-T2 samples. Under gamma priors, Bayesian estimators were derived under the SEL and LL functions, respectively. By combining Lindley approximation and MCMC methods, their simulation results verified the superiority of the Bayesian approach. Xu and Gui [43] investigated the entropy estimation problem of the inverse Weibull distribution under the same censoring scheme. Bayesian estimators were obtained using Lindley approximation under several loss functions, and the resulting MSEs were smaller than those of the MLEs. Chen [44] studied Bayesian inference for the Gompertz distribution under the APHC-T2 scheme. The estimators and credible intervals were obtained via MCMC under gamma priors and SEL and LL functions, showing remarkable advantages for small samples and high censoring proportions. Zhao [45] conducted Bayesian inference for the parameters and entropy of the generalized Bilal distribution. By combining Lindley approximation and MCMC methods, the simulation results demonstrated satisfactory performance in terms of coverage probabilities and estimation accuracy.

In this study, a Bayesian inference framework for the TWD is established under APHC-T2 samples. The posterior distributions of the model parameters and the posterior expectation of the Shannon entropy are derived. By considering the SEL and LL functions, Bayesian point estimators and HPD credible intervals are obtained through the MCMC method.

Prior and Posterior Distributions

The choice of prior distributions directly affects the quality of posterior inference. A reasonable prior distribution can effectively combine existing knowledge with sample information, particularly in cases involving small sample sizes or high censoring proportions. The degree of informativeness of a prior distribution usually requires a trade-off between incorporating expert knowledge and avoiding excessive subjective influence. Therefore, the prior distributions are selected according to the characteristics of each parameter in this study.

According to the properties of the parameters, gamma and uniform priors are adopted in this study. For the shape parameter a and the scale parameter b independent gamma prior distributions are assumed. The support of the gamma distribution is the positive real domain, which is consistent with the parameter space of the Weibull family. Moreover, its hyperparameters can be flexibly adjusted to represent different levels of prior information. The gamma prior has been shown to possess desirable posterior properties in Bayesian inference with censored lifetime data. For the transmutation parameter λ , a uniform prior distribution is assumed. The uniform prior assigns equal probabilities to all possible values within the parameter space and is regarded as a non-informative prior. Thus, the posterior inference is primarily driven by the observed data, avoiding the introduction of subjective preferences.

Specifically, the gamma distribution is assumed as the prior distribution of a , denoted by $p(a)$. Its PDF is given by:

$$p(a) = \frac{b_0^{a_0}}{\Gamma(a_0)} a^{a_0-1} e^{-b_0 a}, a > 0 \quad (38)$$

Where, $a_0 > 0$ and $b_0 > 0$ are the shape and rate hyperparameters, respectively.

Similarly, the gamma distribution is assumed as the prior distribution of b , denoted by $p(b)$. Its PDF is given by:

$$p(b) = \frac{d_0^{c_0}}{\Gamma(c_0)} b^{c_0-1} e^{-d_0 b}, b > 0 \quad (39)$$

Where, $c_0 > 0$ and $d_0 > 0$ are the shape and rate hyperparameters, respectively.

For the transmutation parameter λ , a uniform distribution $U[-1, 1]$ is assumed as the prior distribution. This choice implies that every possible value of λ within the interval $[-1, 1]$ has an equal probability.

$$p(\lambda) = \frac{1}{2}, -1 \leq \lambda \leq 1 \quad (40)$$

Therefore, the joint prior distribution is given by:

$$p(\theta) = \left(\frac{b_0^{a_0}}{\Gamma(a_0)} a^{a_0-1} e^{-b_0 a} \right) \cdot \left(\frac{d_0^{c_0}}{\Gamma(c_0)} b^{c_0-1} e^{-d_0 b} \right) \cdot \left(\frac{1}{2} \right) \quad (41)$$

Where, $a > 0, b > 0, -1 \leq \lambda \leq 1$.

Let $X_{1:m:n}, X_{2:m:n}, \dots, X_{m:m:n}$ denote the m ordered failure observations obtained from the APHC-T2 experiment $x_{1:m:n}, x_{2:m:n}, \dots, x_{m:m:n}$, where each observation follows the TWD with the PDF given in Eq. (6). Under the Bayesian framework, the joint posterior distribution of the parameters θ is obtained by combining the prior distributions and the likelihood function through Bayes' theorem. The joint posterior distribution of the parameters is given by:

$$p(\theta|x) = \frac{p(a) \cdot p(b) \cdot p(\lambda) \cdot L(\theta|x)}{\int_0^{+\infty} \int_0^{+\infty} \int_{-1}^1 p(a) \cdot p(b) \cdot p(\lambda) \cdot L(\theta|x) d\lambda da db} \quad (42)$$

Substituting the prior distributions (23), (24), (25) and the likelihood function (12) into the above equation yields:

$$p(\theta|x) \propto \prod_{i=1}^m \frac{a}{b^a} x_i^{a-1} e^{-\left(\frac{x_i}{b}\right)^a} \left[1 - \lambda + 2\lambda e^{-\left(\frac{x_i}{b}\right)^a} \right] \left[1 - e^{-\left(\frac{x_i}{b}\right)^a} \right]^{R_i} \left[1 + \lambda e^{-\left(\frac{x_i}{b}\right)^a} \right]^{R_i} \cdot a^{a_0-1} e^{-b_0 a} \cdot b^{c_0-1} e^{-d_0 b} \quad (43)$$

Within the Bayesian framework, the parameter estimators are jointly determined by the posterior distribution and the selected loss function. The loss function quantifies the discrepancy between the estimated value and the true value of a parameter, and different loss functions lead to different optimal

Bayesian estimators. To comprehensively investigate the effectiveness and robustness of Shannon entropy estimation under different decision-theoretic criteria, the SEL and LL functions are considered in this study for comparative analysis, providing more comprehensive and robust entropy estimates.

● Squared Error Loss Function

The SEL function is one of the most commonly used loss functions. It assigns a larger penalty to larger deviations, and its form is given as follows:

$$L(\hat{H}, H) = (\hat{H} - H)^2$$

Under the SEL function, the Bayesian estimator $\hat{H}_{SEL}$ of the Shannon entropy $H(\theta)$ is obtained as follows, where $E[H(\theta)|x]$ denotes the posterior mean of $H(\theta)$:

$$\hat{H}_{SEL} = E[H(\theta)|x] = \int_0^{+\infty} \int_0^{+\infty} \int_{-1}^1 H(\theta) p(\theta|x) d\lambda da db \quad (44)$$

● LINEX Loss Function

The LL function is an asymmetric loss function that can flexibly characterize the situation where the costs of overestimation and underestimation are different. Its form is given as follows:

$$L(\hat{H}, H) = e^{\mu(\hat{H}, H)} - \mu(\hat{H}, H) - 1, \mu \neq 0$$

Under the LL function, the Bayesian estimator $\hat{H}_{LL}$ of the Shannon entropy $H(\theta)$ is given by:

$$\hat{H}_{LL} = -\frac{1}{\mu} \ln E[e^{-\mu H(\theta)} | x] = -\frac{1}{\mu} \ln \left[\int_0^{+\infty} \int_0^{+\infty} \int_{-1}^1 e^{-\mu H(\theta)} p(\theta|x) d\lambda da db \right] \quad (45)$$

MCMC Algorithm Implementation

Due to the complex form of the likelihood function of the TWD, the gamma and uniform priors generally do not constitute conjugate priors, resulting in a highly complicated analytical form of the joint posterior distribution $p(\theta|x)$. Therefore, the marginal posterior distributions and posterior

expectations cannot be obtained directly through integration. To address this issue, the MCMC method is introduced for numerical inference in this study. The MCMC method constructs a Markov chain whose stationary distribution is exactly the target posterior distribution. After a sufficiently long iteration process, the generated sequence of samples approximately follows the posterior distribution, thereby transforming the complicated posterior integration problem into a numerical sampling problem. Gibbs sampling is one of the most commonly used algorithms for implementing MCMC. Its basic idea is to sequentially sample each parameter from its full conditional posterior distribution given the remaining parameters at each iteration. In this way, the sampling problem of a high-dimensional joint posterior distribution is decomposed into a series of sampling problems from lower-dimensional conditional distributions. However, the full conditional posterior distributions of all parameters in the TWD do not have standard analytical forms, and thus a pure Gibbs sampler cannot be directly implemented.

Therefore, a hybrid Gibbs sampling approach is adopted in this study, where the Metropolis–Hastings (M-H) step is embedded to sample parameters whose full conditional posterior distributions do not

follow standard forms. This approach combines the computational efficiency of Gibbs sampling with the flexibility of the M-H algorithm, making it effective for handling the non-conjugate posterior distributions of TWD parameters.

Based on the joint posterior distribution in Eq.(42) , the full conditional posterior distributions of the three parameters are given as follows:

$$p(a|b, \lambda, x) \propto a^{m+a_0-1} e^{-b_0 a} \cdot b^{-ma} \cdot \prod_{i=1}^m x_i^{a-1} \cdot \prod_{i=1}^m \left[e^{-\left(\frac{x_i}{b}\right)^a} \left(1 - \lambda + 2\lambda e^{-\left(\frac{x_i}{b}\right)^a}\right) \left(1 + \lambda e^{-\left(\frac{x_i}{b}\right)^a}\right)^{R_i} \right] \quad (46)$$

$$p(b|a, \lambda, x) \propto b^{c_0-1-ma} e^{-d_0 b} \cdot \prod_{i=1}^m \left[e^{-\left(\frac{x_i}{b}\right)^a} \left(1 - \lambda + 2\lambda e^{-\left(\frac{x_i}{b}\right)^a}\right) \left(1 + \lambda e^{-\left(\frac{x_i}{b}\right)^a}\right)^{R_i} \right] \quad (47)$$

$$p(\lambda|a, b, x) \propto \prod_{i=1}^m \left[\left(1 - \lambda + 2\lambda e^{-\left(\frac{x_i}{b}\right)^a}\right) \left(1 + \lambda e^{-\left(\frac{x_i}{b}\right)^a}\right)^{R_i} \right] \quad (48)$$

According to Eqs.(46)—(48) , the full conditional distributions of all parameters θ are in non-standard forms. Therefore, a hybrid Gibbs sampling algorithm is employed to obtain posterior samples of the parameters. For the parameters a and b , which are constrained to be positive, proposal distributions are constructed in the logarithmic space. Specifically, let $\tilde{a} = \ln a, \tilde{b} = \ln b$, which prevents the generated candidate values from falling into the negative domain and eliminates the need for truncation correction of the proposal distributions. For the parameter λ , a normal proposal distribution is directly adopted in the original parameter space, and the parameter constraint is ensured when calculating the acceptance probability. The detailed algorithm is described as follows.

Step 1: Set the initial values of the parameters $\theta^{(0)} = (a^{(0)}, b^{(0)}, \lambda^{(0)})$, where the MLEs are usually selected as the initial values. Set the total number of iterations M and the burn-in period M_0 .

Step 2: Suppose that the parameter values at the i th iteration are given by $\theta^{(i)} = (a^{(i)}, b^{(i)}, \lambda^{(i)})$.

The $i+1$ th iteration is performed as follows:

(2.1) Sampling $a^{(i+1)}$ using the MH algorithm. In the logarithmic space, a candidate value $\tilde{a}^* \sim N(\ln a^{(i)}, \sigma_a^2), N(\ln a^{(i)}, \sigma_a^2)$ is generated from the proposal distribution, Let $a^* = e^{\tilde{a}^*}$, where σ_a^2 is the (1,1) element of the inverse of the observed Fisher information matrix. Calculate the acceptance probability:

$$p_1 = \min \left\{ 1, \frac{p(a^*|b^{(i)}, \lambda^{(i)}, x)}{p(a^{(i)}|b^{(i)}, \lambda^{(i)}, x)} \cdot \frac{a^*}{a^{(i)}} \right\} \quad (49)$$

Where $\frac{a^*}{a^{(i)}}$ the Jacobian correction term introduced by the logarithmic transformation is included. A random number u_1 is generated from the uniform distribution $U(0,1)$, and is used to determine whether the candidate value is accepted.

$$a^{(i+1)} = \begin{cases} a^*, & u_1 \leq p_1 \\ a^{(i)}, & u_1 > p_1 \end{cases} \quad (50)$$

(2.2) Sampling $b^{(i+1)}$ using the MH algorithm. In the logarithmic space, a candidate value $\tilde{b}^* \sim N(\ln b^{(i)}, \sigma_b^2)$, $N(\ln b^{(i)}, \sigma_b^2)$ is generated from the proposal distribution, let $b^* = e^{\tilde{b}^*}$, where σ_b^2 is the (2,2) element of the inverse observed Fisher information matrix. The acceptance probability is calculated as:

$$p_2 = \min \left\{ 1, \frac{p(a^{(i+1)}, b^* | \lambda^{(i)}, x)}{p(a^{(i+1)}, b^{(i)} | \lambda^{(i)}, x)} \cdot \frac{b^*}{b^{(i)}} \right\} \quad (51)$$

A random number u_2 is generated from the uniform distribution $U(0,1)$, and is used to determine whether the candidate value is accepted.

$$b^{(i+1)} = \begin{cases} b^*, & u_2 \leq p_2 \\ b^{(i)}, & u_2 > p_2 \end{cases} \quad (52)$$

(2.3) Sampling $\lambda^{(i+1)}$ using the MH algorithm. The proposal distribution of λ is selected as a normal distribution $N(\lambda^{(i)}, \sigma_\lambda^2)$, where σ_λ^2 is the (3,3) element of the inverse observed Fisher information matrix.

(i) Generate a candidate value λ^* from the proposal distribution $N(\lambda^{(i)}, \sigma_\lambda^2)$. If $\lambda^* \notin (-1, 1)$, regenerate the candidate value. The acceptance probability is calculated as:

$$p_3 = \min \left\{ 1, \frac{p(a^{(i+1)}, b^{(i+1)}, \lambda^* | x)}{p(a^{(i+1)}, b^{(i+1)}, \lambda^{(i)} | x)} \right\} \quad (53)$$

(ii) Generate a random number u_3 from the uniform distribution $U(0,1)$, and is used to determine whether the candidate value is accepted.

$$\lambda^{(i+1)} = \begin{cases} \lambda^*, & u_3 \leq p_3 \\ \lambda^{(i)}, & u_3 > p_3 \end{cases} \quad (54)$$

Step 3: Substitute the current parameter $\theta^{(i+1)} = (a^{(i+1)}, b^{(i+1)}, \lambda^{(i+1)})$ into the Shannon entropy expression(11) to calculate the corresponding entropy value $H^{(i+1)}$. Set $i = i + 1$, return to Step 2, and repeat the procedure until M iterations are completed.

Step 4: Discard the first M_0 burn-in samples and retain the remaining $M - M_0$ effective samples for subsequent Bayesian estimation and interval estimation.

Bayesian Estimators of Shannon Entropy

After the convergence of the MCMC chain and the removal of the burn-in samples, the remaining $M - M_0$ effective posterior samples of the parameters, denoted by $(a^{(t)}, b^{(t)}, \lambda^{(t)}), t = M_0 + 1, \dots, M$, are substituted into Eq.(11) to obtain the corresponding samples of the Shannon entropy, denoted by $H^{(t)} = H(a^{(t)}, b^{(t)}, \lambda^{(t)})$.

Under the SEL function, the Bayesian estimator of the Shannon entropy is the posterior mean, which can be obtained by taking the arithmetic average of the effective entropy samples, given by:

$$\hat{H}_{\text{SEL}} = \frac{1}{M - M_0} \sum_{t=M_0+1}^M H^{(t)} \quad (55)$$

Under the LL function, the Bayesian estimator of the Shannon entropy $\hat{H}_{LL}$ is given by:

$$\hat{H}_{LL} = -\frac{1}{\mu} \ln \left(\frac{1}{M - M_0} \sum_{t=M_0+1}^M e^{-\mu H^{(t)}} \right) \quad (56)$$

Highest Posterior Density Credible Interval for Shannon Entropy

In Bayesian inference, the HPD credible interval is the shortest credible interval among all intervals with the same posterior probability coverage. It is more appropriate than the equal-tailed credible interval when the posterior distribution is asymmetric. Based on the MCMC samples obtained from the hybrid Gibbs sampling algorithm in Section 4.2, the HPD credible interval is constructed according to the following steps.

(1) Obtaining Shannon entropy samples. According to Steps 3 and 4 in Section 4.2, $N = M - M_0$ effective posterior samples of the Shannon entropy, denoted by $\{H^{(t)}\}_{M_0+1}^M$ have been obtained.

(2) Sorting the samples in ascending order. The above samples are arranged in ascending order and denoted by:

$$H_{(1)} \leq H_{(2)} \leq \dots \leq H_{(N)} \quad (57)$$

(3) Constructing the HPD credible interval. For a given significance level γ , the number of samples contained in the credible interval is calculated as:

$$k = \lfloor N \times (1 - \gamma) \rfloor \quad (58)$$

All possible starting points $j = 1, 2, \dots, N - k$ are considered, and the length of each candidate interval is calculated as:

$$L_j = H_{(j+k)} - H_{(j)} \quad (59)$$

The index L_j that minimizes j^* is selected, that is:

$$j^* = \arg \min_{1 \leq j \leq N-k} L_j \quad (60)$$

Therefore, the $100(1-\gamma)\%HPD$ HPD credible interval for the Shannon entropy is given by:

$$\left(H_{(j^*)}, H_{(j^*+k)} \right) \quad (61)$$

Monte Carlo Simulation Study

Monte Carlo simulation approximates the distribution of target statistics through a large number of random sampling procedures and is a commonly used numerical technique for solving complex inference problems where analytical solutions are difficult to obtain. In this study, several simulation experiments are designed under different combinations of censoring schemes and sample sizes. The statistical properties of the MLEs and Bayesian estimators are compared, and the effects of different loss functions on entropy estimation performance are systematically evaluated using criteria such as Bias and MSE. ◦

Numerical Computation Methods

In the Monte Carlo simulation study, APHC-T2 samples need to be generated. Specifically, the procedure for generating APHC-T2 censored data from $X \square TW(a_{real}, b_{real}, \lambda_{real})$ is described as follows, where $a_{real}, b_{real}, \lambda_{real}$ denotes the true parameter vector specified in the simulation study:

(1) Generate m random numbers from the uniform distribution $U(0,1)$ denoted by

$w_1, w_2, \dots, w_m$ ◦

(2) Set $v_i = w_i^{(i+R_m+R_{m-1}+\dots+R_{m-i+1})^{-1}}$, and then calculate $u_i = 1 - v_m v_{m-1} \dots v_{m-i+1}$, where $i = 1, 2, \dots, m$ ◦

(3) Determine the value of r . Specifically, find the index $X_{r:m:n} < T < X_{r+1:m:n}$ satisfying r , and remove $X_{r+2:m:n}, \dots, X_{m:m:n}$ observations.

(4) According to the censoring mechanism, adjust the removal scheme by setting

$R_i = 0 (i = r+1, r+2, \dots, m-1)$, and set $R_m = n - m - \sum_{i=1}^r R_i$, and obtain the final censoring scheme

$(R_1, \dots, R_r, 0, \dots, 0, R_m)$.

(5) Let $x_i = F^{-1}(u_i; a_{real}, b_{real}, \lambda_{real})$ then $(x_1, x_2, \dots, x_m)$ represents the APHC-T2 censored sample from $TW(a_{real}, b_{real}, \lambda_{real})$, with the corresponding censoring scheme $(R_1, R_2, \dots, R_r, 0, \dots, 0, R_m)$. Here, $F(\cdot)$ denotes the CDF of the TWD.

Simulation Design

Three APHC-T2 censoring schemes are considered in the simulation study, as presented in Table 1.

Table 1. Three different censoring schemes

Scheme	
A	$R_1 = R_2 = \dots = R_{m-1} = 0, R_m = n - m - \sum_{i=1}^{m-1} R_i$
B	$R_1 = R_2 = R_3 = \dots = R_{m-2} = 0, R_{m-1} = R_m = \frac{n-m}{2}$
C	$R_1 = R_2 = \dots = R_{m-2} = 0, R_m = 0, R_{m-1} = n - m - \sum_{i=1}^{m-1} R_i$

The parameter settings are specified as $a = 2, b = 1, \lambda = 0.5$. The hyperparameters of the gamma priors are set as $a_0 = 1, b_0 = 1, c_0 = 1, d_0 = 1$, and the true value of the Shannon entropy is 0.4759558028.

Table 2. Bias and MSE of the entropy estimates obtained by MLE and Bayesian methods under different censoring schemes

n	m	CS	MLE		SEL		LL(c=1)		LL(c=-1)	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
30	8	A	-0.4571	0.7415	0.0656	0.1648	0.0269	0.1448	0.2784	0.2323
		B	-0.4547	0.7340	0.0524	0.1630	-0.0103	0.1449	0.1356	0.1652
		C	-0.4126	0.5487	0.0893	0.1590	-0.0165	0.1209	0.1833	0.1699
	16	A	-0.4007	0.3787	0.0070	0.0619	-0.0163	0.0562	0.0761	0.0711
		B	-0.3886	0.3500	0.0174	0.0685	-0.0293	0.0570	0.0527	0.0619
		C	-0.3788	0.3487	0.0116	0.0703	-0.0083	0.0644	0.0579	0.0511
	24	A	-0.3094	0.2486	0.0039	0.0305	0.0052	0.0292	-0.0112	0.0230
		B	-0.3159	0.2506	0.0213	0.0311	-0.0028	0.0286	0.0434	0.0300
		C	-0.3205	0.2583	-0.0209	0.0276	-0.0003	0.0204	-0.0058	0.0219
	40	A	-0.4957	0.8322	0.1118	0.1890	-0.0106	0.1242	0.2434	0.2007
		B	-0.5014	0.8070	0.0951	0.2069	-0.0353	0.1944	0.2441	0.2213
		C	-0.4613	0.7855	0.0818	0.1698	-0.0513	0.1417	0.1750	0.1799
40	16	A	-0.3810	0.3848	0.0250	0.0629	-0.0674	0.0817	0.0498	0.0722
		B	-0.3901	0.3753	-0.0005	0.0851	-0.0385	0.0669	0.0858	0.0959
		C	-0.5079	0.8886	0.0072	0.0765	-0.0318	0.0733	0.0702	0.0665
	24	A	-0.3739	0.3135	0.0242	0.0398	-0.0328	0.0372	-0.0165	0.0371
		B	-0.3836	0.3115	-0.0181	0.0432	0.0105	0.0355	0.0232	0.0436
		C	-0.3692	0.2971	0.0175	0.0347	-0.0183	0.0335	0.0206	0.0359
	50	A	-0.5083	0.8871	0.1242	0.1669	-0.0158	0.1290	0.2893	0.2703
		B	-0.4607	0.6741	0.0898	0.1683	-0.0487	0.1596	0.3378	0.3161
		C	-0.4809	0.6448	0.0868	0.1426	-0.0293	0.1405	0.1897	0.2602
	16	A	-0.3464	0.3515	0.0458	0.0948	-0.0382	0.0751	0.0579	0.1107
		B	-0.3585	0.3548	0.0060	0.1025	0.0005	0.0784	0.1290	0.1126
		C	-0.3789	0.3172	0.0166	0.0754	-0.0331	0.0864	0.0630	0.0961
	24	A	-0.3719	0.3003	0.0056	0.0481	-0.0184	0.0464	-0.0175	0.0544

60	24	B	-0.3425	0.2809	0.0075	0.044	-0.0215	0.0493	0.0100	0.0509
		C	-0.3551	0.2857	0.0041	0.0391	-0.0489	0.0467	0.0368	0.0480
	8	A	-0.4729	0.8147	0.1377	0.1885	-0.0731	0.2046	0.2385	0.2657
		B	-0.5082	0.8755	0.0704	0.2101	0.0158	0.2019	0.2604	0.2832
		C	-0.4223	0.6287	0.0765	0.1410	-0.0808	0.1911	0.2696	0.2141
	16	A	-0.3400	0.3254	0.0142	0.0921	-0.0643	0.1027	0.1139	0.1182
		B	-0.3337	0.3375	0.0536	0.0727	-0.1005	0.1343	0.0922	0.1100
		C	-0.4037	0.3552	0.0674	0.1107	-0.0236	0.1182	0.0565	0.1064
	24	A	-0.3377	0.2787	0.0095	0.0541	-0.0360	0.0576	0.0076	0.0613
		B	-0.3308	0.2663	0.0277	0.0490	-0.0260	0.0462	0.0429	0.0519
		C	-0.3277	0.2673	0.0188	0.0508	-0.0247	0.0455	0.0103	0.0412
	8	A	-0.4987	0.9084	0.1140	0.2006	-0.0512	0.1609	0.1889	0.2133
		B	-0.5021	0.9311	0.1344	0.1672	-0.0968	0.1826	0.2939	0.2410
		C	-0.4304	0.6369	0.0519	0.1542	-0.0360	0.1983	0.2486	0.2511
	16	A	-0.3106	0.3320	0.0181	0.1215	-0.0146	0.1071	0.1118	0.1213
		B	-0.3657	0.3722	0.0238	0.1363	-0.0337	0.0838	0.0779	0.1222
		C	-0.4291	0.3752	0.0573	0.1095	-0.0131	0.0914	0.1081	0.1215
	24	A	-0.3264	0.2563	-0.0010	0.0678	-0.0359	0.0621	0.0147	0.0809
		B	-0.3249	0.2804	-0.0187	0.0577	-0.0674	0.0681	0.0856	0.0766
		C	-0.3277	0.2673	0.0277	0.0582	-0.0465	0.0668	0.0525	0.0736
80	8	A	-0.4369	0.7817	0.0116	0.2486	-0.0568	0.1666	0.1399	0.2717
		B	-0.4254	0.8504	0.0156	0.2288	-0.0604	0.2130	0.2929	0.3004
		C	-0.3827	0.5435	-0.0222	0.1588	-0.0980	0.1844	0.2509	0.2966
	16	A	-0.3615	0.4053	0.0459	0.1314	-0.1328	0.1753	0.1513	0.1365
		B	-0.3651	0.3828	0.03878	0.1285	-0.0508	0.1526	0.2125	0.1632
		C	-0.4797	0.4418	0.0186	0.1015	-0.0051	0.0955	0.1457	0.1277
	24	A	-0.3317	0.2864	0.0117	0.0821	-0.0331	0.0696	0.0744	0.0751
		B	-0.3035	0.2747	0.0236	0.0654	-0.0189	0.0656	0.0457	0.0755
		C	-0.4132	0.3117	0.0281	0.0735	-0.0367	0.0971	0.0303	0.0625

According to Table 2, the following conclusions can be drawn.

(1) By comparing the MLE and Bayesian estimators under the same censoring scheme, it can be observed that the Bayesian estimators generally outperform the MLE in terms of both the absolute value of Bias and MSE. The Bayesian estimator under the SEL function usually has the smallest absolute bias, while the Bayesian estimator under the LL function generally achieves smaller MSE values than the MLE. This indicates that Bayesian methods provide better estimation accuracy for the Shannon entropy estimation problem.

(2) From the perspective of MSE, as the sample size n increases, the MSE values of all three estimation methods show a clear decreasing trend, indicating that all estimators possess asymptotic consistency. In comparison, the Bayesian estimators generally yield smaller MSE values than the MLE under different sample sizes, demonstrating stronger stability and further validating the effectiveness of Bayesian methods in Shannon entropy estimation.

(3) For a fixed sample size n , the absolute values of Bias and MSE decrease with the increase in the effective number of observations m for all estimation methods. This indicates that more observed failure information contributes to improved estimation accuracy. When m is relatively small,

corresponding to a higher censoring proportion, the MLE exhibits larger Bias and MSE values, whereas the advantages of Bayesian estimators become more evident

(4) Regarding the comparison among different censoring schemes, Scheme C achieves relatively smaller MSE values in some cases, whereas Scheme A generally produces larger MSE values. Overall, the influence of different censoring schemes on Bayesian estimation results is weaker than that on the MLE results, indicating that Bayesian methods are less sensitive to censoring mechanisms and possess stronger robustness.

Table 3. Confidence intervals and coverage probabilities of the entropy estimates obtained by MLE and Bayesian methods under different censoring schemes.

n	m	CS	ACI	CP(%)	Bootstrap	CP(%)	HPD	CP(%)
			CI		CI		CI	
30	8	A	[-0.2406,1.4366]	88.00	[-1.1301,0.8684]	84.00	[-0.3633,1.4381]	90.00
		B	[-0.2823,1.5333]	84.00	[-1.1803,0.7495]	81.00	[-0.2620,1.5170]	96.00
		C	[-0.2390,1.4671]	86.00	[-1.0605,0.6679]	86.00	[-0.2316,1.4344]	98.00
	16	A	[0.0920,0.9214]	74.00	[-0.3213,0.7246]	86.00	[0.0444,0.9887]	98.00
		B	[0.1240,0.9749]	82.00	[-0.2968,0.7501]	86.00	[0.0641,0.8640]	94.00
		C	[0.0803,0.9789]	82.00	[-0.3310,0.6773]	89.00	[0.0931,0.9235]	98.00
	24	A	[0.2119,0.7558]	76.00	[-0.0113,0.6644]	81.00	[0.2141,0.7146]	93.00
		B	[0.2319,0.7823]	80.00	[-0.008,0.6698]	82.00	[0.2237,0.7439]	94.00
		C	[0.2215,0.7345]	72.00	[-0.0233,0.6446]	81.00	[0.2449,0.7802]	96.00
	8	A	[-0.2892,1.6785]	83.00	[-1.2960,0.8506]	89.00	[-0.3857,1.4963]	95.00
		B	[-0.2669,1.6751]	82.00	[-1.2515,0.8619]	89.00	[-0.3518,1.5935]	96.00
		C	[-0.2436,1.5821]	86.00	[-1.1867,0.6289]	84.00	[-0.2990,1.3765]	94.00
40	16	A	[0.0171,1.0903]	83.00	[-0.3701,0.8313]	86.00	[-0.0591,1.0232]	94.00
		B	[0.0362,1.1796]	79.00	[-0.4062,0.7854]	88.00	[-0.0126,1.0617]	96.00
		C	[0.0169,1.0808]	83.00	[-0.4224,0.6998]	89.00	[0.0310,0.9492]	92.00
	24	A	[0.1892,0.8203]	73.00	[-0.0989,0.6991]	84.00	[0.1631,0.8623]	95.00
		B	[0.1776,0.7776]	73.00	[-0.0943,0.7018]	83.00	[0.1852,0.8322]	94.00
		C	[0.1564,0.7815]	77.00	[-0.1036,0.6704]	82.00	[0.1501,0.7983]	94.00
	8	A	[-0.4139,1.5946]	89.00	[-1.3607,0.9335]	86.00	[-0.4643,1.6613]	93.00
		B	[-0.3197,1.7975]	84.00	[-1.4190,0.7990]	87.00	[-0.4039,1.6590]	98.00
		C	[-0.2550,1.6240]	85.00	[-1.2811,0.5854]	85.00	[-0.3992,1.4992]	96.00
	16	A	[-0.0405,1.1131]	86.00	[-0.4815,0.8230]	87.00	[-0.0724,1.1672]	96.00
		B	[0.0070,1.1644]	89.00	[-0.5026,0.7848]	84.00	[-0.1238,1.0195]	96.00
		C	[-0.0094,1.1157]	89.00	[-0.4585,0.7523]	85.00	[-0.0304,1.1134]	97.00
50	24	A	[0.1303, 0.8443]	76.00	[-0.1245,0.7731]	87.00	[0.1256,0.8730]	95.00
		B	[0.1278,0.8214]	78.00	[-0.1401,0.7520]	86.00	[0.1333,0.8486]	97.00
		C	[0.1254,0.8880]	88.00	[-0.1364,0.7265]	83.00	[0.1304,0.8432]	96.00

According to the simulation results presented in Table 3, the following conclusions can be drawn:

(1) From the overall performance of the interval estimation methods, all three approaches can provide effective interval estimates for the true parameter. However, differences exist in terms of coverage probabilities and interval lengths. In general, the HPD credible interval achieves coverage probabilities closest to the nominal confidence level (95%) and exhibits relatively stable performance.

The Bootstrap interval generally provides higher coverage probabilities than the ACI, demonstrating certain advantages in improving interval coverage performance.

(2) The ACI method tends to have relatively lower coverage probabilities under small sample sizes or high censoring proportions. Although its performance improves with increasing sample size, the coverage probability remains slightly below the nominal level overall. In contrast, the HPD credible interval not only maintains higher coverage probabilities but also generally produces shorter interval lengths in most cases, indicating higher interval estimation efficiency.

(3) Regarding different censoring schemes, the interval estimation results are affected to some extent by the censoring mechanism under the same sample size. However, the overall differences are not significant. All three censoring schemes can provide relatively stable interval estimation results. In some cases, Scheme C exhibits slightly higher coverage probabilities, but the overall patterns remain largely consistent.

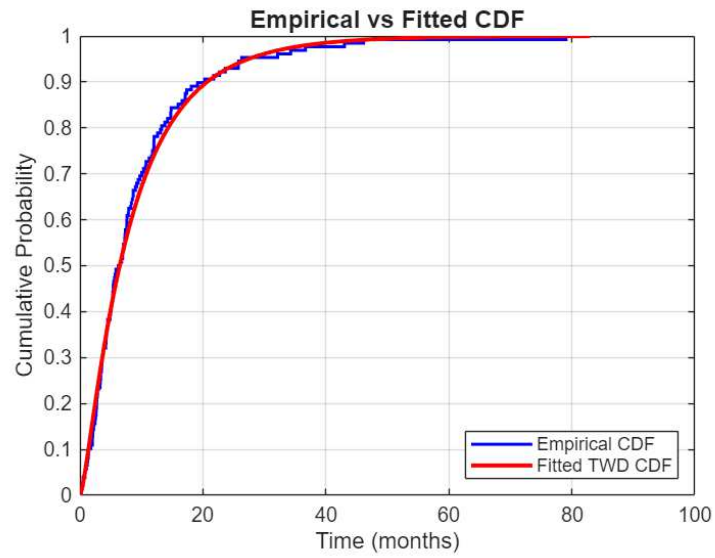
(4) As the sample size increases, the coverage probabilities of all interval estimation methods gradually become more stable, while the interval lengths generally decrease. This indicates that increasing sample information improves estimation accuracy and that the interval estimation methods possess desirable asymptotic properties.

By combining the results of point estimation and interval estimation, it can be concluded that the Bayesian methods generally outperform the MLE under APHC-T2 censoring. For point estimation, Bayesian estimators usually achieve smaller Bias and MSE values, with more pronounced advantages under small sample sizes or high censoring proportions. For interval estimation, the HPD credible interval provides relatively shorter interval lengths while maintaining high coverage probabilities, demonstrating higher estimation efficiency. Meanwhile, the Bootstrap interval shows improved coverage performance compared with the ACI. Therefore, Bayesian methods exhibit good applicability and robustness under APHC-T2 censoring and can serve as a preferable approach for the statistical inference of entropy parameters.

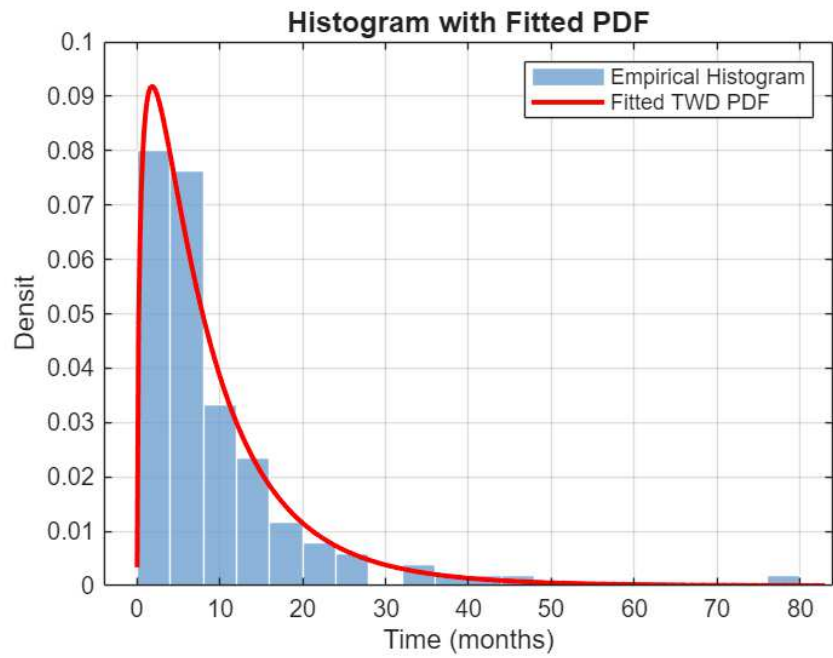
Empirical Analysis

Several criteria have been proposed in the existing literature to evaluate the performance of entropy statistical inference methods. In this study, Bias, MSE, and the coverage probability of confidence intervals are adopted as evaluation criteria to assess the effectiveness of the proposed estimation methods. For a given dataset, the method that simultaneously achieves the smallest Bias and MSE values and provides coverage probabilities closest to the nominal confidence level is regarded as the optimal estimation method. To validate the performance of the proposed Bayesian estimation method, the MLE is selected as the benchmark for comparison. The dataset used in this study consists of randomly selected remission times (in months) of 128 bladder cancer patients.(Lee and Wang, 2003[46]).

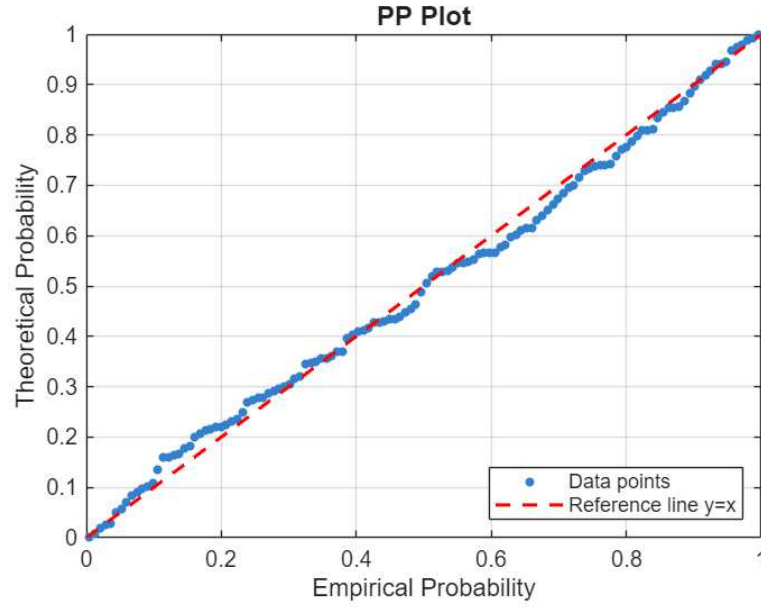
To further validate the goodness-of-fit performance of the TWD for the remission time data of bladder cancer patients, a more intuitive visualization approach is adopted in this study. The results are presented in Fig. 4.



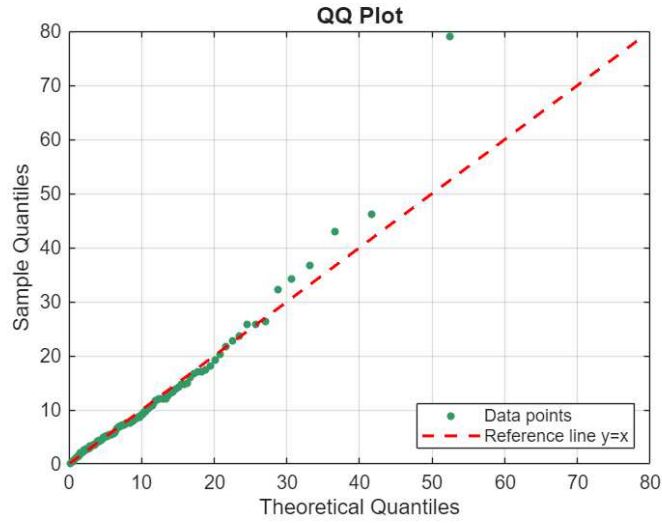
(a)



(b)



(c)



(d)

Figure 4. Goodness-of-fit plot of the TWD model for the remission time data of bladder cancer patients

To further validate and illustrate the performance of the proposed statistical inference method in practical applications, three different APHC-T2 censoring schemes, namely Schemes A, B, and C, are considered. The corresponding results are presented in Tables 4–8.

Table 4. MLEs of the parameters and the corresponding entropy estimates for the cancer dataset under different APHC-T2 censoring schemes

CS	a	b	λ	Shannon 熵
A	1.1254	6.0841	-0.6105	2.9447
B	1.1371	6.1161	-0.5963	3.4150
C	1.1484	6.1463	-0.5826	3.2363

Table 5. Parameter estimates and the corresponding entropy estimates obtained under the SEL function for the cancer dataset under different APHC-T2 censoring schemes

CS	a	b	λ	Shannon 熵
A	1.2752	8.8294	0.0482	2.9730
B	0.9372	10.9309	-0.0945	3.3712
C	0.9737	10.2741	-0.0361	3.2320

Table 6. Parameter estimates and the corresponding entropy estimates obtained under the LL function LL(c=1) for the cancer dataset under different APHC-T2 censoring schemes

CS	a	b	λ	Shannon 熵
A	1.2566	6.7606	-0.1102	2.9659
B	0.9220	6.7596	-0.2976	3.3623
C	0.9582	6.1441	-0.2996	3.2244

Table 7. Parameter estimates and the corresponding entropy estimates obtained under the LL function LL(c=-1) for the cancer dataset under different APHC-T2 censoring schemes

CS	a	b	λ	Shannon 熵
A	1.2930	12.5930	0.2033	2.9803
B	0.9521	19.6409	0.1079	3.3805
C	0.9890	17.3667	0.2195	3.2398

Table 8. Confidence intervals of the entropy estimates obtained by MLE and Bayesian methods for the cancer dataset

CS	ACI	Bootstrap	HPD
A	[2.7369, 3.2033]	[2.7355, 3.2041]	[2.2333, 3.5469]
B	[2.7514, 3.2217]	[2.7499, 3.2174]	[1.7629, 2.4375]
C	[2.7645, 3.2299]	[2.7535, 3.2191]	[2.2253, 2.7122]

To assess whether the Markov chains generated during the MCMC sampling process have reached satisfactory convergence, the bladder cancer patient dataset under different censoring schemes is selected for convergence diagnostics. Under the specified gamma and uniform prior distributions, Fig. 5 presents the trace plots of the parameters during the sampling process.

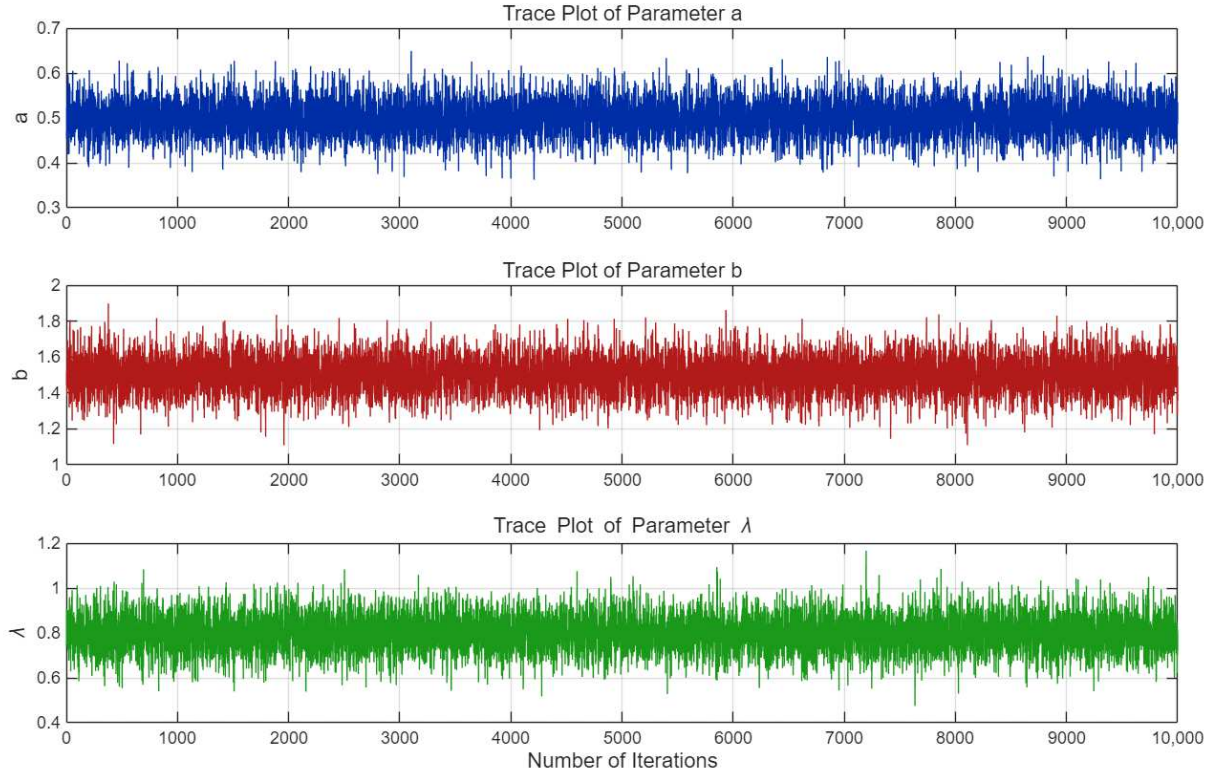


Figure 5. Trace plots of the parameters during the MCMC sampling process for the sample.

Conclusion

This paper develops a systematic statistical inference framework for the Shannon entropy of the TWD under APHC-T2 samples. Within the frequentist framework, the MLEs of the model parameters and Shannon entropy are obtained through numerical optimization based on the log-likelihood function. The ACI is constructed using the Delta method, and the Bootstrap method is further introduced to improve the interval coverage performance under small sample conditions. Within the Bayesian framework, posterior point estimates and HPD credible intervals of the Shannon entropy are obtained by employing the hybrid Gibbs sampling-based MCMC method under the SEL and LL functions.

The Monte Carlo simulation results demonstrate that Bayesian estimators generally outperform the MLE in terms of point estimation accuracy, with more pronounced advantages under small sample sizes or high censoring proportions. Among them, the Bayesian estimator under the SEL function achieves the smallest bias. Regarding censoring schemes, relatively balanced censoring structures generally correspond to smaller MSE values, indicating that balanced censoring mechanisms can contribute to improving estimation accuracy. For interval estimation, the Bootstrap interval exhibits the most stable coverage probability performance, while the HPD credible interval provides a better balance between coverage probability and interval length. The ACI performs well under large sample sizes but shows certain fluctuations in small-sample scenarios. As the sample size or the amount of effective failure information increases, the estimation accuracy of all methods improves, confirming their asymptotic effectiveness.

The proposed inference framework is mainly developed for Shannon entropy under a single failure mechanism. Further extensions remain possible in more general settings. Future research may consider extending the framework to generalized entropy measures and competing failure models, as well as exploring computationally more efficient approximate Bayesian methods.

References

1. Elmahdy, E.E. and A.W. Aboutahoun, A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied mathematical modelling*, 2013. 37(4): p. 1800-1810.
2. Cousineau, D., Fitting the three-parameter weibull distribution: review and evaluation of existing and new methods. *IEEE transactions on dielectrics and electrical insulation*, 2009. 16(1): p. 281-288.
3. Aryal, G.R., Tsokos, C.P. (2011). Transmuted Weibull Distribution: A Generalization of the Weibull Probability Distribution. *European Journal of Pure and Applied Mathematics*, 4(2), 89-102.
4. Ahmad K, Ahmad S P, Ahmed A. Structural properties of transmuted Weibull distribution[J]. *Journal of Modern Applied Statistical Methods*, 2015, 14(2): 141–158.
5. Pobočíková I, Sedláčková Z, Michalková M. Transmuted Weibull distribution and its applications[C]// *MATEC Web of Conferences*, 2018, 157: 08007.
6. Yousaf R, Ali S, Aslam M. Bayesian estimation of transmuted Weibull distribution under different loss functions[J]. *Journal of Reliability and Statistical Studies*, 2020, 13(2): 287–324.
7. Yousaf R, Ali S, Aslam M. On the Bayesian analysis of two-component mixture of transmuted Weibull distribution[J]. *Scientia Iranica*, 2021, 28(3): 1711–1735.
8. Othman S A, Omar K M T, Abdulazeez S T. Reliability analysis and statistical fitting for the transmuted Weibull model in R[J]. *Mathematical Statistician and Engineering Applications*, 2023, 72(2): 161–177.
9. Jeslin J, Radhika A, Haripriya M. Bayesian parameter estimation for transmuted Weibull distribution with censoring rates and various loss functions[J]. *Reliability: Theory & Applications*, 2025, 20(1): 855–864.
10. Zhang F, Dai L S. Statistical analysis for progressive stress accelerated life test with transmutation Weibull distribution[J]. *Communications in Statistics—Simulation and Computation*, 2026, 55(3): 919–935.
11. Escobar, L.A. and W.Q. Meeker, THE ASYMPTOTIC EQUIVALENCE OF THE FISHER INFORMATION MATRICES FOR TYPE I AND TYPE II CENSORED DATA FROM LOCATION-SCALE FAMILIES. *Communications in Statistics - Theory and Methods*, 2001. 30(10): p. 2211-2225.
12. Cohen A C, Clifford J. Progressively censored samples in life testing[J]. *Technometrics*, 1963, 5(3): 327–339.
13. Balakrishnan N, Aggarwala R. *Progressive Censoring: Theory, Methods, and Applications*[M]. Boston: Birkhäuser, 2000.
14. Kundu D, Joarder A. Analysis of Type-II progressively hybrid censored data[J]. *Computational Statistics*

and Data Analysis, 2006, 50(10): 2509–2528.

15. Cho Y, Sun H, Lee K. Exact likelihood inference for an exponential parameter under generalized progressive hybrid censoring scheme[J]. *Statistical Methodology*, 2015, 23: 18–34.
16. Ng H K T, Kundu D, Chan P S. Statistical analysis of exponential lifetimes under an adaptive Type-II progressive censoring scheme[J]. *Naval Research Logistics*, 2009, 56(8): 687–698.
17. Lv Q, Tian Y, Gui W. Statistical inference for Gompertz distribution under adaptive type-II progressive hybrid censoring[J]. *Journal of Applied Statistics*, 2024, 51(3): 451-480.
18. Elshahhat A, Nassar M. Bayesian survival analysis for adaptive Type-II progressive hybrid censored Hjorth data[J]. *Computational Statistics*, 2021, 36(3): 1965–1990.
19. Mustafa A, Ahmed S M. Estimation of power Lomax distribution for censored data with applications[J]. *Journal of Mathematics*, 2022, 2022: 3682098.
20. Hussam E, ALMetwally E M. Statistical inference on progressive-stress accelerated life testing for the Perk distribution under adaptive type-II hybrid censoring scheme[J]. *Journal of Statistical Theory and Applications*, 2025, 24(1): 113-138.
21. Shannon C.E. A mathematical theory of communication[J]. *The Bell system technical journal*, 1948, 27(3): 379-432.
22. Cover T M, Thomas J A. Elements of Information Theory[M]. Hoboken: John Wiley & Sons, 1999.
23. Ebrahimi N, Soofi E S, Soyer R. Information measures in perspective[J]. *International Statistical Review*, 2010, 78(3): 383-412.
24. Tenreiro Machado J A, Lopes A M. Entropy analysis of human death uncertainty[J]. *Nonlinear Dynamics*, 2021, 104(4): 3897-3911.
25. Ali A, Anam S, Ahmed M M. Shannon entropy in artificial intelligence and its applications based on information theory[J]. *J. Appl. Emerg. Sci*, 2023, 13(1): 9-17.
26. Huang R, Hanif M F, Siddiqui M K, et al. Analyzing boron oxide networks through Shannon entropy and Pearson correlation coefficient[J]. *Scientific Reports*, 2024, 14(1): 26552.
27. Kehinde T O, Chung S H, Olanrewaju O A. An integrated approach to stock selection and ranking: Combining Shannon entropy technique, DEA, and inverse DEA[J]. *IEEE access*, 2025, 13, 138851-138866.
28. Gong Q, Yin B. Statistical inference of entropy functions of generalized inverse exponential model under progressive type-II censoring test[J]. *PLOS ONE*, 2024, 19(9): e0311129.
29. Hassan A S, Alsadat N, Balogu O S, Helmy B A. Bayesian and non-Bayesian estimation of some entropy measures for a Weibull distribution[J]. *AIMS Mathematics*, 2024, 9(11): 32646–32673.
30. Alotaibi R, Nassar M, Khan Z A, Alajlan W A, Elshahhat A. Entropy evaluation in inverse Weibull unified hybrid censored data with application to mechanical components and head-neck cancer patients[J]. *AIMS Mathematics*, 2025, 10(1): 1085–1115.
31. Ren H, Zhang Z, Gong Q. Estimation of Shannon entropy of the inverse exponential Rayleigh model under progressively Type-II censored test[J]. *AIMS Mathematics*, 2025, 10(4): 9378–9414.

32. Chakraborty S, Bhattacharya R, Pradhan B. Cumulative entropy of progressively Type-II censored order statistics and associated optimal life testing plans[J]. *Statistics*, 2023, 57(1): 161–174.
33. Gong Q, Yin B. Statistical inference of entropy functions of generalized inverse exponential model under progressive type-II censoring test[J]. *PLOS ONE*, 2024, 19(10): e0311888.
34. Ahmed S M, Ismail G M. Entropy estimation of Lomax distribution under adaptive type-II progressive censoring[J]. *European Journal of Pure and Applied Mathematics*, 2025, 18(1): 5737.
35. Kayal S, Maurya S K, Kundu D. Statistical inference on the Shannon and Rényi entropy measures of generalized exponential distribution under the progressive censoring[J]. *Sankhya B*, 2023, 85(1): largest available.
36. Khan, M.S.; King, R.; Hudson, I.L. Transmuted Weibull distribution: Properties and Estimation. *Commun. Stat.-Theory Methods* **2017**, *46*, 5394–5418. <https://doi.org/10.1080/03610926.2015.1100744>
37. Shaw W T, Buckley I R C. The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map[EB/OL]. (2009-01-05)[2025-02-01]. <https://arxiv.org/abs/0901.0434>.
38. Balakrishnan N, Mitra D. Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*, 2012, 56(12): 4011-4025.
39. Afify, A.Z., et al., The Transmuted Weibull Lomax Distribution: Properties and Application. *Pakistan journal of statistics and operation research*, 2015. 11(1): p. 135.
40. Elsherpieny, E.A., Muhammed, H.Z., and Radwan, N.U. (2015). Discriminating Between Weibull and Log-Logistic Distributions in Presence of Progressive Type II Censoring. *International Journal of Innovative Research in Science, Engineering and Technology*, 4(8), 7281–7292. DOI: 10.15680/IJIRSET.2015.0408107.
41. Efron B. Bootstrap methods: Another look at the jackknife[J]. *The Annals of Statistics*, 1979, 7(1): 1-26.
42. Nassar M, Abo-Kasem O, Zhang C, Dey S. Analysis of Weibull distribution under adaptive type-II progressive hybrid censoring scheme[J]. *Journal of the Indian Society for Probability and Statistics*, 2018, 19(1): 25–65.
43. Xu R, Gui W. Entropy estimation of inverse Weibull distribution under adaptive type-II progressive hybrid censoring schemes[J]. *Symmetry*, 2019, 11(12): 1463.
44. Chen S, Gui W. Statistical analysis of a lifetime distribution with a bathtub-shaped failure rate function under adaptive progressive type-II censoring[J]. *Mathematics*, 2020, 8(5): 670.
45. Zhao J, Gui W. Estimation for entropy and parameters of generalized Bilal distribution under adaptive type-II progressive hybrid censoring scheme[J]. *Entropy*, 2021, 23(2): 206.
46. Lee, E.T., and Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*, Third ed. Wiley, New York.

Acknowledgements

The authors would like to sincerely thank the Editor and the anonymous reviewers for their valuable time, careful evaluation, and constructive comments, which have helped us improve the quality and presentation of this manuscript. We also sincerely appreciate the Editorial Office for its patience and assistance throughout the submission and review process.

Funding

This work was supported by the 2026 Xiamen Municipal Natural Science Foundation Project (Grant No. 3502Z202674035) and the 2025 Xiamen Municipal Natural Science Foundation Project (Grant No. 3502Z202573102).

Data Availability Statement

The dataset analyzed in this study is a publicly available dataset.

Author contributions

Conceptualization, S.X.; methodology, S.X., X.W., and Y.Z.; software, X.W.; formal analysis, S.X. and X.W.; investigation, X.W.; validation, X.H. and C.H.; writing—original draft preparation, X.W.; writing—review and editing, S.X., Y.Z., X.H., and C.H.; supervision, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no conflict of interest.

Table 9. Design Table

Question	Hypothesis (if applicable)	Sampling plan (e.g. power analysis)	Analysis Plan	Interpretation given to different outcomes
Can the Shannon entropy of the Transmuted Weibull Distribution be accurately estimated under	The maximum likelihood estimator provides reliable estimates of the Shannon entropy under APHC-T2 censoring, and its	Monte Carlo simulations will be conducted using 1,000 independent replications under multiple sample sizes, parameter	Maximum likelihood estimates will be obtained using the Newton–Raphson algorithm. Estimation	If Bias and MSE decrease with increasing sample size while CP remains close to the nominal confidence level

APHC-T2 censoring using the maximum likelihood approach?	estimation accuracy improves as the sample size increases and the censoring level decreases.	settings, and three APHC-T2 censoring schemes (Schemes A–C) specified in the study.	performance will be evaluated using Bias, Mean Squared Error (MSE), Coverage Probability (CP), and Average Interval Length (AIL) of the asymptotic confidence interval (ACI) and Bootstrap confidence interval (BCI).	and AIL decreases, then the MLE procedure will be regarded as providing satisfactory statistical inference under APHC-T2 censoring. Otherwise, its limitations under the corresponding simulation settings will be discussed.
Does Bayesian estimation outperform maximum likelihood estimation for Shannon entropy under APHC-T2 censoring?	Bayesian estimators under the squared error loss (SEL) and LINEX loss (LL) functions are expected to produce lower Bias and MSE and more efficient interval estimates than the maximum likelihood estimator, particularly under small sample sizes and heavy censoring.	The same Monte Carlo settings will be adopted as those used for the MLE. Bayesian estimation will be implemented using a hybrid Gibbs–Metropolis–Hastings MCMC algorithm with 10,000 iterations, of which the first 5,000 iterations will be discarded as burn-in.	Bayesian point estimates under the SEL and LL functions will be obtained together with Highest Posterior Density (HPD) credible intervals. Performance will be compared with the MLE using Bias, MSE, CP, and AIL.	If Bayesian estimators exhibit lower Bias and MSE while maintaining CP close to the nominal confidence level and shorter AIL than the MLE, then Bayesian inference will be considered superior under the corresponding simulation setting. Otherwise, the relative strengths and limitations of the two estimation methods will be discussed.
Are the proposed estimation methods applicable to real lifetime data?	The proposed estimation procedures provide reliable and consistent estimates of the Shannon entropy when applied to real lifetime data.	A real bladder cancer remission-time dataset will be analysed using the APHC-T2 framework described in this study. No additional sampling will be performed.	The TWD model will be fitted to the real dataset, and the Shannon entropy will be estimated using both the classical and Bayesian approaches. Point estimates and interval estimates will be reported and compared.	If the proposed estimation procedures provide reasonable point and interval estimates for the real dataset, then they will be considered applicable to practical reliability and survival analysis. Otherwise, the possible reasons and

				limitations will be discussed.
--	--	--	--	--------------------------------