RT&A, No 4(80)
Volume 19, December, 2024

REVIEW OF CENSORING SCHEMES: CONCEPTS, DIFFERENT TYPES, MODEL DESCRIPTION, APPLICATIONS AND FUTURE SCOPE

Ninan P Oommen[1], Jiju Gillariose [2]

•

[1,2]Department of Statistics and Data Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
ninan.oommen@res.christuniversity.in, jijugillariose@yahoo.com

Abstract

Survival analysis is one of the key techniques utilized in the domains of reliability engineering, statistics, and medical domains. It focuses on the period between the initialization of an experiment and a subsequent incident. Censoring is one of the key aspects of survival analysis, and the techniques created in this domain are designed to manage various censoring schemes with ease, ensuring accurate and insightful time-to-event data analysis. The statistical efficiency of parameter estimates is improved by accurately incorporating censoring information by making use of the available data. This paper reviews the concepts, model descriptions, and applications of conventional and hybrid censoring schemes. The introduction of new censoring schemes from conventional censoring schemes has evolved by rectifying the drawbacks of the previous schemes, which are explained in detail in this study. The evolution of hybrid censoring schemes through the combination of various conventional censoring schemes, the data structures, concepts, methodology, and existing literature works of hybrid censoring schemes are reviewed in this work.

Keywords: bibliometric analysis, hybrid censoring, left and right censoring, random censoring, survival analysis

1. Introduction

In reliability engineering, Survival analysis plays a major role in determining how long an event is estimated to take place, such as the death of a biological creature or a failure of the mechanical system. A collection of statistical methods used for time-to-event data is called survival analysis. The period of time until an event occurs is the relevant outcome variable. Generally, survival data addresses death as an event, but it may also address any occurrence that happens to a person, such as disease, recurrence following remission, and cure. Survival data describes the amount of time that occurs between an origin and an endpoint of interest. The time is considered as the number of years, months, weeks, or days preceding the start of the study enrollment. The survivor function and the hazard function are two related functions typically used to describe and model survival data. The survival function determines the probability that a person will survive from their point of origin to a point after time. Given survival up to that point in time, the hazard function provides the instantaneous potential for an occurrence at that point. Its main applications are in the fields of diagnostics and mathematical model specification for survival analysis.

Investigating the literature within the Scopus database with the search term (TITLE-ABS-KEY ("censoring scheme*") AND TITLE-ABS-KEY ("survival analysis*")) produced 55 documents in English language in the area of decision science, mathematics, statistics, computer science, and

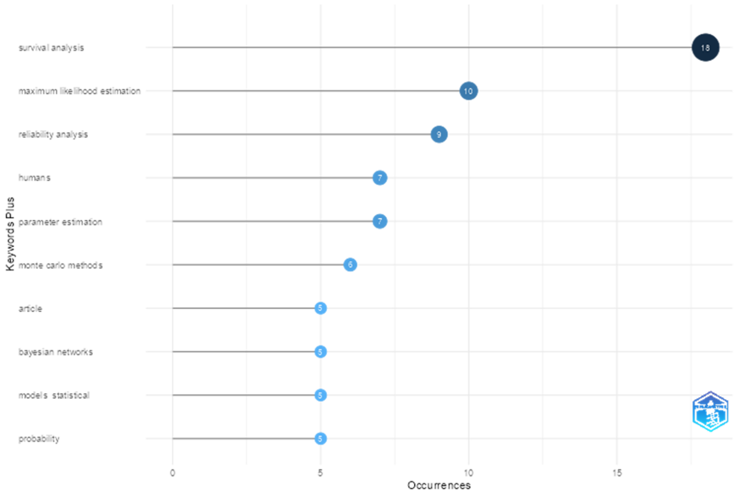

Figure 1: *Most relevant words*

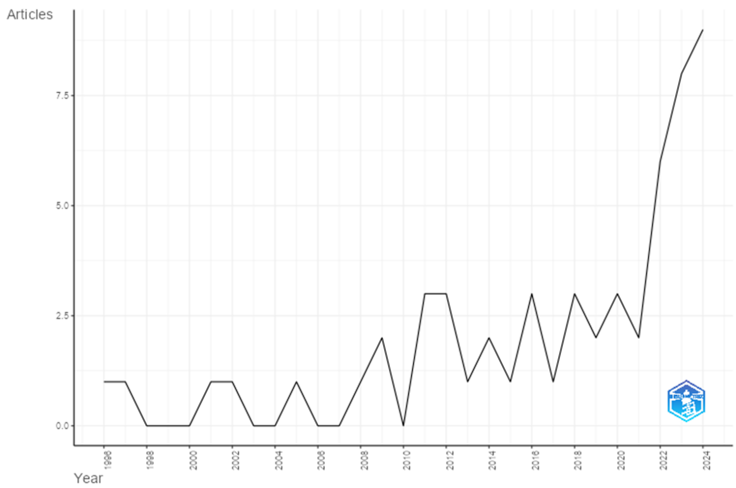

Figure 2: *Annual scientific production*

engineering. The analysis using VOS viewer generated 290 keywords with at least a single occurrence. Furthermore, the visualization of keywords explored using the overlay visualization and the results show that "survival analysis" is the most occurring keyword in the existing literature (Figure 1). The progress of works from 1996 to 2024 is visualized using overlay visualization, and the result interprets that the progress of articles in survival analysis after 2021 is very high (Figure 2). These two results show the importance of survival analysis and its applications in this current scenario.

In survival analysis, censoring schemes are essential to collect the censored data or failure time data of products. The evolution process of various censoring schemes by overcoming the limitations of previous censoring schemes is as follows. By overcoming the limitations of conventional censoring schemes, many hybrid censoring schemes have evolved in the existing literature. Right censoring, left censoring, interval censoring, and random censoring are different wide categories of censoring schemes. Type–I and type–II censoring schemes are the two major conventional censoring schemes. The hybrid censoring scheme, which is a mixture of type–I and type–II censoring schemes, has been introduced by [17]. In the case of type-I hybrid censoring scheme, from the experimenter's perspective, this scheme has the benefit of pre-fixed termination time. The limitation of type-I hybrid and conventional type-I censoring schemes is that, the inferential results must be developed under the assumption that there is at least one observed failure. In addition, very few failures may occur up to the pre-fixed time T, which leads to low efficiency for the estimators of the model parameters. The type-II hybrid censoring scheme can ensure the observation of a specific number of failures after the experiment compared to the type-I hybrid censoring scheme, which can result in more efficient inferential results. The termination time of the type-II hybrid censoring scheme is a random variable, which is one of its disadvantages. From the experimenter's point of view, this randomness is a drawback as it adds uncertainty to the duration of the experiment.

In order to overcome these drawbacks, [10] introduced generalized type-I and type-II hybrid censoring schemes and obtained the statistical inferences under the conditions of exponential distribution. This censoring scheme can guarantee the number of failures and can provide a limit on the experimental time. In the case of generalized type-I hybrid censoring scheme, one of the drawbacks is there is no guarantee of observing a minimum number of failures if the termination of life test occurs at or before time T. In the case of generalized type–II hybrid censoring scheme, there is a chance of observing zero failures or observing only a few failures until the prefixed time T_2. By combining all the censoring schemes, [8] introduced unified hybrid censoring scheme and derived the exact distribution of the maximum likelihood estimator and exact confidence intervals for the mean of the exponential distribution.

In all the above-mentioned censoring schemes, the units or products cannot be taken at random before the termination of the experiment process. However, in practice, there may be occasions where the units are withdrawn or lost from the experiment before they break down. The loss may happen due to damage to testing facilities or depletion of funds, etc. Hence, the censoring scheme that is suitable for these circumstances will be progressive hybrid censoring scheme, which allows the withdrawal of units from the experiment before the termination time period as well as after the termination of the experiment. A detailed discussion on the theory, methods, and applications of progressive censoring can be seen in [9]. The number of withdrawing units of the censoring scheme is known in advance in progressively censored experiments, which is one of its crucial assumptions. In real life, due to unavoidable circumstances, if the experimenter changes the censoring numbers during the experiment, it will violate the mentioned assumption and hence affect the correctness and efficiency of statistical inference. Hence, the adaptive progressive type-II censoring scheme is an appropriate model proposed by [30], which considers this adaptation.

This paper is beneficial for many experts and stakeholders who are managing censoring schemes and survival analysis from various sectors. Researchers employ survival analysis in clinical trials to assess the efficacy of novel therapies and interventions and to enhance patient care and treatment procedures. Epidemiologists can utilize these survival analysis and censoring techniques to study disease incidence, distribution, and control, facilitating public health decision-making. In engineering, survival analysis aids in the estimation of product and system lifespans and failure rates, which helps with quality control and design enhancements. This paper can be beneficial for such researchers as it explains basic concepts, methodologies, existing literature works, and applications of different conventional and hybrid censoring schemes.

The paper is organized as follows: The concepts and model descriptions of conventional type censoring schemes are given in Section 2. Right censoring, left censoring, interval censoring, random censoring, and type-I and type-II censoring schemes are explained under the conventional censoring scheme. Section 3 provides aspects of various hybrid censoring schemes and their data structures. Section 4 explains the applications and future scope of different censoring schemes. Conclusions about various censoring techniques and applications are given in Section 5.

2. Conventional censoring schemes

2.1. Right censoring

Censoring is the first and most important challenge that survival analysis attempts to address. Censoring occurs when some experiments fail to view the event of interest prior to the termination of the study. It happens when the researcher is not aware of the precise survival time of the entities but has partial survival information. In survival analysis, the time record of events can be collected in two ways. The actual survival time can be collected from those who completed the event of interest. However, one cannot collect the exact failure time of subjects that exceed the censor time or those who lost to follow up the progress. Hence, on a timeline, their true lifetimes will be on the right of their observed censor times; such entities are referred to as right censored. Here, time determines the duration of monitoring, which is either full or partial observation of survival intervals. Survival analysis models generate consistent parameter values with various estimation methods using data from both censored and uncensored observations.

2.1.1 Methodology

According to [23], the methodology for right censoring can be explained as follows. Let Y be the lifetime of an individual under study and S be the fixed censoring time. Let us assume $Y's$ are independent and identically distributed with probability density function $f(Y)$ and survival function $s(Y)$. The two conditions in which the exact lifetime Y of an individual is known are when Y is less than S or equal to S. If Y is greater than S, then the event time is censored at S, and he or she is a survivor. The data obtained from this experiment can be represented by pairs of random variables (T, λ). Here, λ indicates whether the lifetime Y corresponds to an event of interest ($\lambda = 1$) or is censored ($\lambda = 0$).
i.e. For $i = 1, 2, \cdots, n$

$$\lambda = \begin{cases} 1, & \text{if } Y_i \leq S \text{ (unsensored)} \\ 0, & \text{if } Y_i > S \text{ (censored)} \end{cases}$$

T equals Y if the lifetime is observed and S if it is censored. Hence, the observed times are $T = min(Y_i, S)$ for right censoring.

2.2. Left censoring

The definition of left censoring explained by [23] is as follows. Let S be the left censoring time of an individual having a lifetime of Y. To define left censoring, the event of interest must occur for the subject before that person is observed in the study at time S ($T < S$). In other words, a time Y associated with a specific subject in a study is considered to be left censored if it is less than the censoring time S, where S is the left-censoring time. In this case, we know that the event of interest occurred before time S, but we don't know the exact failure time for these subjects. The exact time will be known if and only if Y is greater than or equal to S. As mentioned in the case of right censoring, the data obtained from this experiment can be represented by pairs of random variables (T, δ), where T is equal to Y if the lifetime is observed and δ indicates whether the lifetime Y corresponds to an event of interest ($\delta = 1$), or is censored ($\delta = 0$).

2.2.1 Methodology

Let the survival variables $Y_1, Y_2, ..., Y_n$ are left censored. If the observed sample consists of the ordered pairs (Y_i, δ_i) for $i = 1, 2, \cdots, n$. For each $i : T_i = max(Y_i, T_i)$

$$\delta = \begin{cases} 1, & \text{if } Y_i \geq S \text{ (uncensored)} \\ 0, & \text{if } Y_i < S \text{ (censored)} \end{cases}$$

T is equal to Y if the lifetime is observed, and it is equal to S if it is censored. Hence, the observed times for left censoring are $T = max(Y_i, S)$.

2.3. Interval censoring

Interval censoring often indicates an imperfect data structure or a sampling scheme. A subject is considered interval censored when studied for a while, disappears for follow-up, reappears, and is further investigated. Interval censoring is the practice of only knowing the precise location of a random variable of interest within an interval rather than observing it exactly. Data with interval censoring could be produced by several clinical trials and longitudinal studies. An instance of interval censoring explained by [24] is given as follows. A prevalent instance arises in health or medical research investigations that require regular monitoring. In this case, a patient scheduled for prearranged observations for a clinically evident change in sickness or health status can miss some visits and return with a different state. The true event time, therefore, contains the real (but unseen) time of the change's occurrence, and we only know that it is greater than the most recent observation time at which the change has not occurred and less than or equal to the initial observation time at which the change has been observed to occur, thus giving an interval of data which contains the real-time of occurrence of the change.

2.3.1 Methodology

Let the intervals $I_1, I_2, ..., I_n$ be the observed data in interval censoring for each $i = 1, 2, \cdots, n$, the m^{th} response lies in the interval I_m. An observed interval that consists of a single point, in this instance, corresponds to an uncensored observation of an observed death. Suppose t_1 is the time when we have performed an experiment or test on a subject, and the result of the subject test is negative. However, the subject tested positive on a further checkup on time t_2. In this case, we know that the person was infected with the virus at some point between t_1 and t_2, but the precise time of exposure is unknown. In a clinical trial, for instance, if the time to remission has been determined, then if the m^{th} patient is in remission at the 6^{th} week of the trial but misses subsequent follow-up visits and reappears and was out of remission by the 10^{th} week, then I_m = (6,10] represents the m^{th} patient's censoring interval or length of remission.

2.4. Random censoring

In all the previous methods, it was not explicitly stated as an assumption that the considered censoring time for an individual is a constant value or a value that is known in advance. However, censoring times are usually not fixed constants but rather random variables in real-world contexts. Therefore, one considers censoring as a random variable and the distributions of two random variables (time to event and time to censoring) and how they relate to each other to obtain the influence of censoring on the inferential part. Two distributions need to be generated for the random censoring data in survival analysis. Survival times without censorship are considered to be the first distribution, and the censorship mechanism is controlled by the second distribution. Based on these elements, the censored event is determined by comparing the censored time created in the second distribution with the uncensored survival time generated in the first distribution.

2.4.1 Methodology

In the case of first distribution, consider a random variable Y that generates times until the event of interest occurs. Let $F_Y(y)$ be the cumulative function of the distribution and $S_Y(y)$ be the survival function derived from $F_Y(y)$, which defines the probability of survival after time y. Hence, $F_Y(y) = P(Y \leq y)$ and $S_Y(y) = 1 - P(Y \leq y) = P(Y \geq y)$ and $F_Y(y)$ be the probability density function of the distribution. In the case of the second distribution, it is recommended to define a continuous distribution in the case of censorship mechanism D. And let $f(D)$ be the probability density function of the second distribution. Here, the two distributions are assumed to be independent of each other. Following is the arrangement for the censored samples. We have (T_y, λ_y) where $T_y = min(Y_y, D_y)$ and $\lambda_y = I(Y_y \leq D_y)$ for each Y^{th} observation. Here, when $\lambda = 1$, the lifetime is observed or is uncensored, and $\lambda = 0$, it is censored.

Specific works in the field of interval censoring can be seen in [20]. This work explains the proportional hazards model with case one interval censored data and its maximum likelihood estimator. While the maximum likelihood estimator for the baseline cumulative hazard function only converges at n rate, it is demonstrated that the maximum likelihood estimator for the regression parameter is asymptotically normal with n convergence rate and achieves the information bound. The asymptotic variance matrix estimation for the maximum likelihood estimator of the regression parameter is also taken into account in this article. The work in [21] provides estimates in interval censoring models, including regression model estimation as well as non-parametric distribution function estimation. The asymptotic characteristics and computational processes of the non-parametric maximum likelihood estimators are discussed in the non-parametric case, and the accelerated failure time semi-parametric regression models, the proportional odds, and the proportional hazards in the regression conditions.

In order to estimate density and hazard rate functions from randomly right-censored data, [6] presented a wavelet approach in the study. A non-parametric technique is adopted by assuming that there is no particular parametric form for the density and hazard rate. Simulation of the estimators and two real-life data examples of survival time data for patients with liver metastases from a colorectal primary tumor without other distant metastases and times of unemployment for women are also explained in the study. The methods to estimate the exponential mean lifetime in a random censoring model with insufficient data has been determined by [16], and the simulation results are also demonstrated. For the exponential model, point and interval estimates for mean lifespan of item are obtained using the maximum likelihood method. Using Monte Carlo simulation, it is shown that the large sample approximation to the log likelihood ratio provides precise confidence intervals, and the mean lifetime is positively biased when estimated using the maximum likelihood method. For the data following a generalized inverted exponential distribution, [18] determined the maximum likelihood estimators of the model parameters, expected fisher information, Bayesian estimators under the squared error loss function using Lindley's approximation, and highest posterior density credible intervals of the parameters under the random censoring scheme. The computation of Bayes estimators of Weibull distribution parameters under random censoring scheme has been done by [14]. Bayes estimators are obtained by using Lindley's approximation, importance sampling, and Gibbs sampling techniques. The credible intervals of the estimators and a real life data analysis are also performed.

2.5. Type-I and type-II censoring

In practice, two major conventional censoring schemes are type-I and type-II censoring schemes. Type I censoring data is frequently seen in many applications related to engineering and medical research. The experiment is continuous up to a pre-determined time in the case of type-I censoring scheme, while the experiment is continuous till a pre-determined number of failures of units occurs in the type-II censoring scheme. In the case of a type-I censoring scheme, zero failure units can be produced at times. The stopping time of experiments might be large in the case of a type-II censoring scheme.

2.5.1 Methodology

In type–I, a random number of failure units $x \in (0, n)$ is used in the test, where n is the total number of observations, and a fixed prior test time t is considered. Since many of the products are highly reliable, the experiment may take years to complete, and the product may become outdated. Hence, to overcome this drawback, the final time (t) of the experiment is defined initially, and the products with survival time greater than the fixed time would be termed as a censored observation. In type–II, the test time is randomly monitored, and n units are used in the test until the prefixed m^{th} failure occurs. Initially, the number m of observations is defined out of the total n of observations. After generating the n samples from a distribution, the observations are sorted in increasing order, and the first $n-m$ values are termed as the observed data. The remaining m values correspond to the value of the $m+1$ position. Hence, we assume the survival time of these observations is equal to the largest one observed since these observations did not experience the event of interest. The major difference between type–I and type–II is that the number of failures is random in the type-I censoring scheme, and the experimental time is random in the type-II censoring scheme.

Some of the works that have been done in the field of type–I and type–II censoring schemes are mentioned below. Bayes and classical estimators have been derived by [34] for the two-parameter exponentiated-Weibull distribution when the data is taken from a type-II censoring scheme. Using non-informative priors, Bayes estimators have been generated under both the squared error loss function and the LINEX loss function. A comparison of the proposed estimators was also made on the basis of their simulated risks, which were obtained under the two loss functions. The Bayesian estimation procedure under the type-II censoring scheme has been developed by [33] for the data following flexible Weibull distribution. Jeffrey's scale invariant is considered the non informative prior, and gamma prior is considered the informative prior for the model parameters. Using a Monte Carlo simulation analysis, the efficiency of the Bayes estimators has also been compared with the classical estimators of the model parameters. For illustrative purposes, a real data set showing the intervals between secondary reactor pump failures has been examined in the study. The estimation of model parameters of Chen lifetime distributions under partial step stress accelerated life tests has been done by [3] based on the type-II censoring scheme. The parametric bootstrap and the two asymptotic distributions are applied to construct each confidence interval of the model parameters. A Monte Carlo simulation study is used to evaluate the precision outcomes.

3. Hybrid censoring

A mixture of type-I and type-II censoring schemes is known as the hybrid censoring scheme. Assume $Y_{1:n}, ..., Y_{n:n}$, where n be the ordered lifetimes of the sample units, which are independently and identically distributed random variables. The experiment is terminated when a prefixed time T has reached or when a prefixed number $r < n$ has failed. To overcome the disadvantages of the two conventional censoring schemes, [17] introduced type-I hybrid censoring scheme where the experiment gets terminated at time $T* = min(T, Y_{m:n})$ where $Y_{m:n}$ is the failure time of the m^{th} unit. And [13] introduced a type-II hybrid censoring scheme, explaining that the experiment terminates at time $T* = max(T, Y_{m:n})$.

3.0.1 Data form of type–I hybrid censoring

First, we assume the lifetimes of the units are independent and identically distributed random variables. There are two cases in which the data can be type–I hybrid censored. That is, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} \leq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$ if $Y_{d:n} > T$

Here, d represents the number of failures that occurred before time T.

3.0.2 Data form of type–II hybrid censoring

There are two cases in which the data is type II hybrid censored. That is, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} \geq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$ if $Y_{d:n} < T$

Here, $r \leq d \leq n$ represents the number of failures that occurred before time T. As assumed earlier, the lifetime of units are independent and identically distributed random variables.

Many studies have been done in the field of censoring schemes in reliability theory so far. The introduction of type-I hybrid censoring scheme, analysis of data under the condition of exponential life distribution of the experimental units, and two-sided confidence interval of the unknown parameter (without formal proof) have been developed by [17]. Using the conditional moment generating function method, [11] determined the exact distribution of the conditional maximum likelihood estimator of θ and applied it to derive an exact lower confidence bound for θ. A simplified but identical form of the exact distribution of the maximum likelihood estimator of θ as derived by [11] has been obtained by [13]. For a Weibull distributed data, [25] derived the maximum likelihood estimators and approximate maximum likelihood estimators of the distribution parameters under hybrid censoring scheme. The approximate confidence intervals are obtained using the asymptotic distribution of the maximum likelihood estimators and the Bayes estimates of the distribution parameters using Gibbs sampling procedures.

3.1. Generalized hybrid censoring

To overcome the drawbacks of above mentioned censoring schemes, [10] introduced generalized type-I and type-II hybrid censoring schemes and obtained the statistical inferences under the conditions of the exponential distribution.

3.1.1 Methodology – generalized type – I hybrid censoring

Let n be the number of units placed on a life test before the initialization of the experiment. Define $r, k \in 1, 2, \cdots, n$ with $k \leq r \leq n$, and time $T \in (0, \infty)$. The experiment terminates at $min(Y_{r:n}, T)$ if the k^{th} failure occurs before time T. And the experiment terminates at $Y_{k:n}$ if the k^{th} failure occurs after time T. If only a few failures had been reported up to time T, it is clear that this generalized hybrid censoring alters the type-I hybrid censoring by enabling the experiment to continue. Note that the experimenter is equipped to consider a minimum of k failures but prefers to observe r failures under this censoring method.

3.1.2 Data form of generalized type – I hybrid censoring

There are three cases if the data are generalized type–I hybrid censored. That is, the observed data consists of one of the three kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{k:n}\}$ if $Y_{k:n} > T$
Case-II: $\{Y_{1:n} < \cdots < Y_{k:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} < T$
Case-III: $\{Y_{1:n} < \cdots < Y_{k:n} < \cdots < Y_{D:n}\}$ if $T < Y_{r:n}$

3.1.3 Methodology – generalized type – II hybrid censoring

According to [10], fix T_1, $T_2 \in (0,\infty)$ as time points with $T_1 < T_2$ and $r \in 1, 2, \cdots, n$. The experiment terminates at T_1 if the r^{th} failure occurs before time T_1. The experiment terminates at $y_{r:n}$ if the r^{th} failure occurs between T_1 and T_2. Finally, terminate the experiment at T_2 if the r^{th} failure occurs at T_2. This hybrid censoring scheme modifies the type-II hybrid censoring scheme by ensuring that the experiment will be terminated by time T_2. As a result, T_2 is the maximum duration of time the experiment is allowed to complete.

3.1.4 Data form of generalized type – II hybrid censoring

Three cases exist if the data are generalized type–II hybrid censored. That is, the observed data consists of one of the three kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{k:n}\}$ if $Y_{1:n} < \cdots < Y_{D1:n} < T_1$
Case-II: $\{Y_{1:n} < \cdots < Y_{D1:n} < \cdots < Y_{r:n}\}$ if $T_1 < Y_{r:n} < T_2$
Case-III: $\{Y_{1:n} < \cdots < Y_{D2:n} < T_2\}$ if $Y_{r:n} \geq T_2$

Here, D_1 and D_2 represents the number of failures that occur before T_1 and T_2, respectively.

Several works have been done in the field of generalized hybrid censoring, and mentioning a few of them here. The computation of maximum likelihood estimators and its approximate confidence intervals for the unknown parameters of Weibull distribution has been done by [5] under type-I generalized hybrid censoring scheme. They also derived its Bayes estimates using the Markov chain Monte Carlo (MCMC) technique. To compute the statistical inferences of the competing risks model with partially observed causes of failure, [4] derived the maximum likelihood estimators, associated confidence intervals, Bayes estimators, and credible intervals of the model parameters when the latent failure times satisfy the assumptions of Lomax life distribution under type-II generalized hybrid censoring scheme. In the presence of partial constant stress acceleration, [4] computed the statistical inference when the Gompertz life products undergo the accelerated life testing experiment. The maximum likelihood and its corresponding confidence intervals for the two samples of exponential distribution have been derived by [2], and hence obtained the Bayes estimators relative to both symmetric and asymmetric loss functions using gamma conjugate priors.

3.2. Unified hybrid censoring

Unified hybrid censoring scheme has been introduced by [8] after combining all the censoring schemes. They derived the exact distribution of the maximum likelihood estimator and exact confidence intervals for the mean of the exponential distribution.

3.2.1 Methodology

According to [8], the methodology for unified hybrid censoring scheme is as follows. Define $r, k \in 1, 2, \cdots, n$ with $k < r$, and fix T_1, $T_2 \in (0,\infty)$ as time points with $T_1 < T_2$. The experiment terminates at $min(max(y_{r:n}, T_1), T_2)$, if the k^{th} failure occurs in prior time T_1. The experiment terminates at $min(y_{r:n}, T_2)$, if the k^{th} failure occurs between T_1 and T_2. Finally, terminate the experiment at $y_{k:n}$ if the k^{th} failure occurs after time point T_2. Therefore, we are certain that the experiment will be completed at most in time T_2 with at least k failures under this censoring scheme. If this is not the case, we can guarantee precisely k failures.

3.2.2 Data form of unified hybrid censoring

There are six cases if the data are unified hybrid censored. That is, the observed data consists of one of the six kinds of observations listed below:

Case-I: $0 < Y_{k:n} < Y_{r:n} < T_1$, the test is terminated at T_1
Case-II: $0 < Y_{k:n} < T_1 < Y_{r:n} < T_2$, the test is terminated at $Y_{r:n}$
Case-III: $0 < Y_{k:n} < T_1 < T_2 < Y_{r:n}$, the test is terminated at T_2
Case-IV: $0 < T_1 < Y_{k:n} < Y_{r:n} < T_2$, the test is terminated at $Y_{r:n}$
Case-V: $0 < T_1 < Y_{k:n} < T_2 < Y_{r:n}$, the test is terminated at T_2
Case-VI: $0 < T_1 < T_2 < Y_{k:n} < Y_{r:n}$, the test is terminated at $Y_{k:n}$

Several works in the field of unified hybrid censoring scheme are as follows. When the data satisfies the assumptions of Burr type-XII distribution, [32] developed maximum likelihood estimators for estimating the unknown parameters using the expectation maximization algorithm (EM) under the for unified hybrid censoring data and obtained Bayesian estimates using Lindley's approximation and MCMC method under the assumption of independent gamma prior and hence constructed the highest posterior density credible interval. Point and interval estimations of the unknown parameters of the inverse Weibull distribution based on a unified hybrid censoring scheme have been studied by [7]. The Bayesian estimations have been obtained based on the squared error loss function and Linex loss function using the MCMC method. Hence, a $(1 - \tau) \times 100\%$ approximate, bootstrap-p, credible, and highest posterior density confidence intervals for the parameters have been constructed. In Rayleigh distribution, [22] computed the maximum likelihood estimators of the scale parameter under unified hybrid censoring scheme. Bayes estimator using mean and mode of the posterior distribution and confidence interval, credible interval, and highest posterior density credible interval have also been computed for the scale parameter. By satisfying the assumptions of Gompertz distribution, [28] computed the statistical inference of unified hybrid censored data under a constant-stress partially accelerated life test model. The stochastic EM algorithm is used to compute the maximum likelihood estimate and shows that it exists uniquely. Bootstrap-p and bootstrap-t methods are used to construct the asymptotic confidence intervals and confidence intervals.

3.3. Progressive hybrid censoring

The introduction of type-I progressive hybrid censoring scheme has been done by [26] and [12].

3.3.1 Methodology - progressive type-I hybrid censoring

According to [26] and [12], the type-I progressive hybrid censoring scheme can be described as follows. Suppose n is the total number of observations throughout the experiment. $(D_1, D_2, ..., D_n$) be the prefixed progressive censoring scheme. Under this censoring scheme, D_1 of the $n-1$ survival units are removed randomly at the time of the first failure from the experiment. Then D_2 of the $n-D_1-2$ surviving units are removed at the time of the second failure, and so on. Finally, $D_m = n - D_1 - \cdots D_{m-1} - m$ surviving units are removed from the life test at the time of the m_{th} failure. Let us denote the failure times as $Y_{1:n} < \cdots < Y_{m:n}$ since the $D_i's$ are already fixed. In type-I progressive hybrid censoring, the experiment terminates at $T* = min(y_{m:n}, T)$. If $y_{m:n}$ occurs before time point T, the experiment will stop at $y_{m:n}$ otherwise, it will stop at time point T. The experimental length clearly cannot be longer than T, and this sampling technique will likely provide additional information regarding the tail of the lifetime distribution because of progressive censoring.

3.3.2 Data form of progressive type-I hybrid censoring

Under this censoring scheme, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{m:n}\}$, if $Y_{m:n} \leq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$, if $Y_{m:n} > T$

Where d is the number of failures that occur prior to time point T.

3.3.3 Methodology - progressive type-II hybrid censoring

The type-II progressive hybrid censoring scheme has been introduced by [12], which overcomes the drawback that the maximum likelihood estimate may not always exist for type-I progressive hybrid censoring scheme. In type-II progressive hybrid censoring, the experiment terminates at time $T* = max(Y_{m:n}, T)$. The experiment terminates at m^{th} failure if $Y_{m:n} > T$ with units being withdrawn following each failure in accordance with the predetermined progressive censoring scheme $(D_1, D_2, ..., D_n)$. However, if $Y_{m:n} < T$, then instead of ending the experiment by withdrawing all remaining D units after the m^{th} failure, the experiment will continue to record failures up to time T without withdrawing any units. Thus, in this situation, we have $R_m = R_{m+1} = \cdots = R_d = 0$ where d is the number of failures that occur prior to time point T.

3.3.4 Data form of progressive type-II hybrid censoring

Under this censoring scheme, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{m:n} < Y_{m+1:n} < \cdots < Y_{d:n}\}$, if $Y_{m:n} < T$
Case-II: $\{Y_{1:n} < \cdots < Y_{m:n}\}$, if $Y_{m:n} \geq T$

Where d is the number of failures that occur prior to time point T.

Some important works under the progressive hybrid censoring scheme are mentioned as follows. The statistical inferences for type-II progressive hybrid censored data under the assumptions of exponential distribution have been derived and analyzed by [26] and [12]. In the presence of constant stress partial acceleration, [15] obtained the maximum likelihood and Bayesian estimates of model parameters for the Nadarajah-Haghighi distribution under progressive type-II censoring and derived the asymptotic confidence intervals through asymptotic variance and covariance matrix and also obtained the Bayesian credible intervals. For Burr XII distribution, [31] derived the estimates using classical and Bayesian approach under progressive type-II hybrid censoring scheme. The EM algorithm is used to compute maximum likelihood estimators, and Lindley's approximation and MCMC techniques are used to compute Bayes estimators. In the case of exponentiated exponential distribution, [1] developed the progressive stress model for type-II progressive hybrid censoring. The Bayes estimates and maximum likelihood estimates have been compared, and normal approximation and bootstrap confidence intervals for the unknown model parameters are developed.

3.4. Adaptive progressive censoring scheme

A mixture of type-I and type-II progressive censoring schemes is known as the adaptive progressive censoring scheme introduced by [30]. In this case, the progressive scheme and effective sample size k are predetermined; however, the number of items gradually eliminated from the experiment at each failure may vary during the experiment being conducted. The experiment can be terminated as soon as possible by adjusting the number of items progressively withdrawn from the experiment at each failure in such a way that the intended level of efficiency of the estimate can be achieved. This will happen if the experimental time exceeds a pre-fixed time T, but the number of observed failures has not yet reached k.

3.4.1 Methodology

Fix integers k, n such that $k < n$ where k is the effective sample size, and n is the total number of observations in a life test. Let $(D_1, D_2, ..., D_n\ )$ be the progressive censoring scheme, which is fixed in advance before the test, but the values of some $D_i's$ may change in an adaptive way during the test procedure. Let time T be the ideal time of the experiment or the time fixed by the researcher to end the process. Let k be the completely observed failure times by $Y_{i:k:n}$ where $i = 1, \cdots, k$. The experiment is terminated at time $Y_{k:k:n}$ if the k^{th} progressively censored failure time occurs prior to time point T. That is, $Y_{i:m:n} < T$. If not, end the experiment as soon as possible to meet a certain inferential efficiency requirement after the experimental duration has passed time T, but the number of observed failures has not yet reached k. As a result, this arrangement can be considered as a design where the optimal situation would be to have k observed failure times for the efficiency of the inference and, at the same time, ensure the overall test time is slightly similar to the ideal test period T.

Some important works under the adaptive progressive censoring scheme are mentioned as follows. The statistical inferences under adaptive progressive hybrid censoring schemes for the model parameters of exponential and Weibull distributions have been computed by [30] and [27], respectively. In the case of two-parameter exponentiated Weibull distribution, [35] constructed the maximum likelihood estimates, reliability and hazard functions, approximate confidence intervals, Bayes estimates, and credible intervals under adaptive progressive type II censoring samples. Under the inverse Weibull distribution, [29] estimated both the frequentist and Bayesian estimates for the scale parameter λ and shape parameter β using an adaptive type-II progressive hybrid censoring scheme. The statistical inference for log-normal distribution under adaptive type-II progressive hybrid censoring scheme has been derived and analyzed by [19].

4. Applications and future scope

Censored data or various censoring schemes in survival analysis are applicable in a diverse range of fields and practical scenarios. In engineering and industrial scenarios, censoring techniques are used to analyze the reliability of systems or components that are subjected to different stress levels. This helps determine the causes of failure and product lifetimes. Censoring schemes are also used for acceptance sampling and quality control. This allows for the development of the best possible sample plans and appropriate choices based on only partially observed data. Censoring techniques are used in clinical trials and medical research to examine patient survival times under various treatment regimens, which helps assess the efficacy of treatments and patient outcomes. In genetic research, censoring techniques are used to assess the effects of penalized or censored data on genetic parameters, breeding values, and computational effectiveness for duration indicators. In cardiovascular research, censoring techniques are used to evaluate the correlation between offspring and parental age at the onset of cardiovascular events. In order to examine the sustainability and resilience of the environment, censoring schemes are used in environmental studies to analyze the lifespan of ecological systems or environmental components under changing conditions.

5. Conclusions

The purpose of life testing in reliability theory is to examine a wide range of items and collect data relevant to some or all of their lifetimes, which facilitates quality assurance and risk assessment and hence, thereby, improves product satisfaction and warranty estimations. This, in turn, satisfies a customer or buyer who would expect the items to operate as intended for an appropriate length of time. In this paper, the concepts, data forms, methodology, and applications of various censoring schemes are explained in detail. Censoring schemes can improve the quality of goods in various businesses by providing several benefits to customers. Partially observed data can be analyzed through censoring algorithms, which is useful when complete failure data is not obtainable. This renders more accurate statistical analysis and inference on the performance and dependability of products. Censoring techniques can be used to evaluate competing risks and determine the effect of each component on product quality in scenarios where multiple risk variables can lead to product failure. More specific plans for quality enhancement can result from this approach. During the testing process, adaptive censoring methods, including adaptive progressive censoring, enable dynamic changes based on observed data. This flexibility can result in quicker, better-informed decisions that improve final output quality. By utilizing censoring schemes effectively, manufacturers can collect substantial information about product performance, make well-informed decisions to improve quality, and eventually provide customers with more reliable and durable products.

References

[1] Alaa H Abdel-Hamid and Tahani A Abushal. Inference on progressive-stress model for the exponentiated exponential distribution under type-ii progressive hybrid censoring. *Journal of Statistical Computation and Simulation*, 85(6):1165–1186, 2015.

[2] Osama Abo-Kasem and Ahmed Elshahhat. A new two sample generalized type-ii hybrid censoring scheme. *American Journal of Mathematical and Management Sciences*, 41(2):170–184, 2022.

[3] Hanaa H Abu-Zinadah and Neveen Sayed-Ahmed. Competing risks model with partially step-stress accelerate life tests in analyses lifetime chen data under type-ii censoring scheme. *Open Physics*, 17(1):192–199, 2019.

[4] Tahani A Abushal, AA Soliman, and GA Abd-Elmougod. Inference of partially observed causes for failure of lomax competing risks model under type-ii generalized hybrid censoring scheme. *Alexandria Engineering Journal*, 61(7):5427–5439, 2022.

[5] Ali Algarni, Abdullah M Almarashi, and Gamal A Abd-Elmougod. Joint type-i generalized hybrid censoring for estimation two weibull distributions. *Journal of Information Science & Engineering*, 36(6), 2020.

[6] Anestis Antoniadis, Gérard Grégoire, and Guy Nason. Density and hazard rate estimation for right-censored data by using wavelet methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(1):63–84, 1999.

[7] Saieed F Ateya. Estimation under inverse weibull distribution based on balakrishnan's unified hybrid censored scheme. *Communications in Statistics-Simulation and Computation*, 46(5):3645–3666, 2017.

[8] N Balakrishnan, Abbas Rasouli, and N Sanjari Farsipour. Exact likelihood inference based on an unified hybrid censored sample from the exponential distribution. *Journal of Statistical Computation and Simulation*, 78(5):475–488, 2008.

[9] Narayanaswamy Balakrishnan and Rita Aggarwala. *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media, 2000.

[10] B Chandrasekar, A Childs, and N Balakrishnan. Exact likelihood inference for the exponential distribution under generalized type-i and type-ii hybrid censoring. *Naval Research Logistics (NRL)*, 51(7):994–1004, 2004.

[11] Shu-Mei Chen and Gouri K Bhattacharyya. Exact confidence bounds for an exponential parameter under hybrid censoring. *Communications in Statistics-Theory and Methods*, 16(8):2429–2442, 1987.

[12] A Childs, B Chandrasekar, and N Balakrishnan. Exact likelihood inference for an exponential parameter under progressive hybrid censoring schemes. *Statistical models and methods for biomedical and technical systems*, pages 319–330, 2008.

[13] A Childs, B Chandrasekar, N Balakrishnan, and D Kundu. Exact likelihood inference based on type-i and type-ii hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics*, 55:319–330, 2003.

[14] Muhammad Yameen Danish and Muhammad Aslam. Bayesian inference for the randomly censored weibull distribution. *Journal of Statistical Computation and Simulation*, 84(1):215–230, 2014.

[15] Sanku Dey, Liang Wang, and Mazen Nassar. Inference on nadarajah–haghighi distribution with constant stress partially accelerated life tests under progressive type-ii censoring. *Journal of Applied Statistics*, 49(11):2891–2912, 2022.

[16] TI Elperin and IB Gertsbakh. Estimation in a random censoring model with incomplete information: exponential lifetime distribution. *IEEE Transactions on Reliability*, 37(2):223–229, 1988.

[17] Benjamin Epstein. Truncated life tests in the exponential case. *The Annals of Mathematical Statistics*, pages 555–564, 1954.

[18] Renu Garg, Madhulika Dube, Kapil Kumar, and Hare Krishna. On randomly censored generalized inverted exponential distribution. *American Journal of Mathematical and Management Sciences*, 35(4):361–379, 2016.

[19] Fariba Hemmati and Esmaile Khorram. Statistical analysis of the log-normal distribution under type-ii progressive hybrid censoring schemes. *Communications in Statistics-simulation and Computation*, 42(1):52–75, 2013.

[20] Jian Huang. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568, 1996.

[21] Jian Huang and Jon A Wellner. Interval censored survival data: a review of recent progress. In *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, pages 123–169. Springer, 1997.

[22] Young Eun Jeon and Suk-Bok Kang. Estimation of the rayleigh distribution under unified hybrid censoring. *Austrian Journal of Statistics*, 50(1):59–73, 2021.

[23] John P Klein, Melvin L Moeschberger, et al. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

[24] John P Klein, Hans C Van Houwelingen, Joseph George Ibrahim, and Thomas H Scheike. *Handbook of survival analysis*. CRC Press Boca Raton, FL:, 2014.

[25] Debasis Kundu. On hybrid censored weibull distribution. *Journal of Statistical Planning and Inference*, 137(7):2127–2142, 2007.

[26] Debasis Kundu and Avijit Joarder. Analysis of type-ii progressively hybrid censored data. *Computational Statistics & Data Analysis*, 50(10):2509–2528, 2006.

[27] Chien-Tai Lin, Hon Keung Tony Ng, and Ping Shing Chan. Statistical inference of type-ii progressively hybrid censored data with weibull lifetimes. *Communications in Statistics—Theory and Methods*, 38(10):1710–1729, 2009.

[28] Showkat Ahmad Lone and Hanieh Panahi. Estimation procedures for partially accelerated life test model based on unified hybrid censored sample from the gompertz distribution. *Eksploatacja i Niezawodność*, 24(3), 2022.

[29] Mazen Nassar and OE Abo-Kasem. Estimation of the inverse weibull parameters under adaptive type-ii progressive hybrid censoring scheme. *Journal of computational and applied mathematics*, 315:228–239, 2017.

[30] Hon Keung Tony Ng, Debasis Kundu, and Ping Shing Chan. Statistical analysis of exponential lifetimes under an adaptive type-ii progressive censoring scheme. *Naval Research Logistics (NRL)*, 56(8):687–698, 2009.

[31] M Noori Asl, R Arabi Belaghi, and H Bevrani. On burr xii distribution analysis under progressive type-ii hybrid censored data. *Methodology and Computing in Applied Probability*, 19:665–683, 2017.

[32] Hanieh Panahi and Abdolreza Sayyareh. Estimation and prediction for a unified hybrid-censored burr type xii distribution. *Journal of Statistical Computation and Simulation*, 86(1):55–73, 2016.

[33] Sanjay Kumar Singh, Umesh Singh, and Vikas Kumar Sharma. Bayesian estimation and prediction for flexible weibull model under type-ii censoring scheme. *Journal of Probability and Statistics*, 2013(1):146140, 2013.

[34] Umesh Singh, Pramod K Gupta, and Satyanshu K Upadhyay. Estimation of parameters for exponentiated-weibull family under type-ii censoring scheme. *Computational statistics & data analysis*, 48(3):509–523, 2005.

[35] Mashail M AL Sobhi and Ahmed A Soliman. Estimation for the exponentiated weibull model with adaptive type-ii progressive censored schemes. *Applied Mathematical Modelling*, 40(2):1180–1192, 2016.