Optimum Hybrid Censoring Scheme using Cost Function Approach

Ritwik Bhattacharya[1], Biswabrata Pradhan[1], Anup Dewanji[2]

1 *SQC and OR Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata, PIN- 700108, India*

2 *Applied Statistics Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata, PIN- 700108, India*

Correspondence to: Ritwik Bhattacharya (ritwik.bhatta@gmail.com)

Abstract

Hybrid censoring scheme is a combination of Type-I and Type-II censoring schemes. Determination of optimum hybrid censoring scheme is an important practical issue in designing life testing experiments. In this work, we consider determination of optimum hybrid censoring scheme by minimizing the total cost associated with the experiment. It is shown that the proposed cost function is scale invariant for some selected distributions. Optimum solution cannot be obtained analytically. We propose a method for obtaining the optimum solution and consider exponential and Weibull distributions for illustration.

Keywords: Delta method; Fisher information; Optimum censoring scheme; Scale invariant; Type-I hybrid censoring.

1 Introduction

Life testing plans play a significant role in reliability and survival analysis studies. Due to limited number of testing units or highly expensive testing units or time constraint, we cannot continue the experiment until all the failures are observed. So censoring becomes inheritably significant and efficient methodology to estimate model parameters of underlying distributions. The most widely used censoring schemes in practice are Type-I and Type-II censoring schemes. In Type-I censoring scheme, we start the experiment with n experimental units and observe the failures until a pre-fixed censoring time T_0 is reached. The problem with Type-I censoring scheme is that a significantly less number of failures (since the number of failures to be observed is random) may lead us to poor estimate of model parameters of interest. In Type-II censoring scheme, we start with n experimental units and continue the experiment until a pre-fixed number of failures r to be observed. Here, although the fixed number of failures provide a relatively good estimate of model parameters of interest, but the difficulty arises when the censoring time becomes significantly large. As a consequence of these two complications, there is a need for new life testing schemes where both the termination time and the number of failures can be kept at a reasonable level to provide good estimate of the model parameters. Such a scheme is hybrid censoring as described in the following.

Suppose n units are put on a life test. The test is terminated after a prefixed r number of failures or at a prefixed time T_0 whichever is earlier. This is also known as Type-I hybrid censoring scheme in the literature. It is easy to see that Type-I and Type-II censoring schemes are special cases of hybrid censoring scheme. When $r = n$, we get Type-I censoring scheme and when $T_0 = \infty$ it is Type-II censoring scheme. See Epstein [12], Chen and Bhattacharyya [5], Childs *et al.* [6], Draper and Guttman [7], Ebrahimi [10, 11], Fairbanks *et al.* [13], Gupta and Kundu [14] for some early works on hybrid censoring scheme. It is clear that for conducting life testing experiment under hybrid censoring scheme, the values of n, r and T_0 must be known apriori. Then a natural question is whether we should choose n, r and T_0 based on convenience, or by some optimal method. In this work, we choose n, r and T_0 optimally with respect to some optimality criteria.

Determination of optimum censoring scheme is an important practical issue in designing a life testing experiment. It has received considerable attention in the literature. See, for example, Ebrahimi [9], Blight [3], Kundu

[16], Zhang and Meeker [23], Dube *et al.* [8], Ng *et al.* [20], Burkschat *et al.* [4] etc. It may be noted that finding optimum censoring scheme depends upon the proper choice of optimality criterion. Most of the work considered different information measure as the optimality criterion. The main idea in their work is to choose that censoring scheme which provides maximum information on unknown parameters. For instances, Kundu [16] proposed a criteria of choosing (r, T_0) for fixed sample size n in hybrid censoring scheme which maximizes the information measure $I(r, T_0) = [\int_0^1 V((r, T_0)_p)dp]^{-1}$, where $V((r, T_0)_p)$ denotes the asymptotic variance of the estimated pth $(0 < p \leq 1)$ quantile of the life time distribution obtained by using the hybrid censoring scheme (r, T_0). He, however, considered those schemes with a given maximum expected duration.

Although most of the works considered maximum information as the optimality criterion, there are few works which considered minimizing the cost associated with the experiment as the optimality criterion. In industrial setup, total cost associated with the experiment is an important issue while conducting a life testing experiment. Epstein [1] first suggested a cost function to determine the optimum non-replacement and optimum replacement experiments under Type-II censoring as sum of two cost components, namely, total cost of placing items on life test and cost for total experimental duration. Riley [22] used this cost function and showed that non-replacement procedures give less experimental cost than replacement procedures for Type-II censored exponential data. Again Blight [3] considered the same cost function and derived the analytical expressions for the optimum sample size and number of replaced items for Type-II censored exponential data. In the context of Type-I censoring, Lam [18] and Lin *et al.* [19] also considered the same cost function. They have proposed Bayesian variable sampling plan using this cost function.

Ebrahimi [9] first proposed the idea of obtaining optimum values of n and r for known T_0 in hybrid censoring scheme by minimizing the cost of the experiment with some restrictions on n and r. He considered the total cost as the sum of cost for duration of the experiment, cost of items to be tested and an additional overall fixed cost which does not depend on sample size and length of the experiment. Recently, Dube *et al.* [8] introduced a method to obtain the optimum n, r and T_0 in hybrid censored experiment with log-normal distribution in which both the information measure and a cost function associated with the experiment are considered. They used the information measure $I(n,r,T_0) = [\int_0^1 V((n,r,T_0)_p)dW(p)]^{-1}$, introduced by Gupta and Kundu [15], where $V((n,r,T_0)_p)$ denotes the asymptotic variance of the estimated pth quantile of the life time distribution and $W(\cdot)$ be any non-negative weight function such that $\int_0^1 dW(p)=1$. They considered determination of optimum values of n, r, and T_0 by maximizing $I(n,r,T_0)$ subject to $C_1^{'}n + C_2^{'}T_0 = C_0^{'}$, where $C_1^{'}$ is cost per unit of item on test, $C_2^{'}$ is cost per unit of time and $C_0^{'}$ is some prefixed cost. The drawback of their work is that they did not consider cost associated with the number of failures and duration of the experiment for the hybrid censoring scheme.

In this work, we consider hybrid censoring and a cost function with four cost components as given by (1) cost of failed items, (2) cost of the duration of the experiment, (3) cost due to imprecision (variance) of the estimates of the unknown parameters of the lifetime distribution under consideration, and (4) cost of running the experiment with sample size n. The cost component in (3) may be obtained from some prior knowledge on post sale scenario. The cost function is the simple sum of the above four cost components. If n is known, the cost function is the sum of the first three cost components. We consider both the cases. The proposed cost function has an important desirable property. It is shown that the optimum solution obtained by minimizing the cost function is scale invariant in the sense that the optimum n and r remain same if life time is multiplied by some constant k (that is, life time is measured by some other unit). For example, suppose the optimum solution corresponding to life time measured in month is obtained as (n^*, r^*, T_0^*); then, the optimum solution corresponding to the life time measured in days will be (n^*, r^*, kT_0^*), where $k = 30$. We consider exponential and Weibull distributions for illustration. It is easy to see that the objective function is a non-linear function in decision parameters n, r and T_0, where n and r are integers and T_0 is continuous. This is a difficult optimization problem which cannot be solved analytically. We need to apply suitable numerical optimization technique, as described in Section 3.2, to obtain the optimum solution. The effect of different cost components on the optimum solution is also studied.

Rest of the paper is organized as follows. In Section 2, we discuss hybrid censored data and some related results which are required to develop the cost function. We provide different optimum criteria along with the optimum solutions for known n in Section 3. In Section 4, we generalize our results when sample size n is unknown. An alternative approach to obtain optimum design parameters is discussed in Section 5. A sensitivity analysis is carried out in Section 6. Finally, some concluding remarks are made in Section 7.

2 Some Results on Hybrid Censoring Scheme

Let T denote the lifetime with distribution function $F(t;\theta)$ and density function $f(t;\theta)$, where $\theta \in \Theta$, an open subset of $\mathbb{R}$. Let $T_1, T_2, \cdots, T_n$ be the life times of n units with correspond order statistic $T_{1:n} \leq T_{2:n} \leq \cdots \leq T_{n:n}$. In the framework of hybrid censoring, the number of failures and censoring time are denoted by D and $\tau = \min(T_{r:n}, T_0)$, respectively. It is clear that both D and τ are random variables. Thus, the data is represented by $(T_{1:n}, T_{2:n}, \ldots, T_{D:n}, D)$. Note that when $D = 0$ no failure is observed. The distribution of D is given by

$$
\begin{aligned}
P(D=j) &= \binom{n}{j} F(T_0;\theta)^j (1 - F(T_0;\theta))^{n-j}, \quad j = 0, 1, 2, \cdots, r-1, \\
P(D=r) &= \sum_{j=r}^{n} \binom{n}{j} F(T_0;\theta)^j (1 - F(T_0;\theta))^{n-j}.
\end{aligned}
$$

Therefore,

$$\begin{aligned} E[D] &= \sum_{j=0}^{r-1} j \binom{n}{j} F(T_0;\theta)^j (1-F(T_0;\theta))^{n-j} \\ &\quad + r \sum_{j=r}^{n} \binom{n}{j} F(T_0;\theta)^j (1-F(T_0;\theta))^{n-j}. \qquad (1) \end{aligned}$$

$$\begin{aligned} \text{Also, } E[\tau] &= E\left[\min(T_{r:n}, T_0)\right] \\ &= E\left[T_0 | T_0 \leq T_{r:n}\right] P\left(T_0 \leq T_{r:n}\right) + E\left[T_{r:n} | T_{r:n} < T_0\right] P\left(T_{r:n} < T_0\right) \\ &= T_0 P\left(T_0 \leq T_{r:n}\right) + E\left[T_{r:n} | T_{r:n} < T_0\right] P\left(T_{r:n} < T_0\right), \qquad (2) \end{aligned}$$

where $P(T_0 \leq T_{r:n}) = 1 - \sum_{j=r}^{n} \binom{n}{j} F(T_0;\theta)^j (1-F(T_0;\theta))^{n-j}$
and $E\left[T_{r:n} | T_{r:n} < T_0\right] = \frac{r\binom{n}{r}}{P\left(T_{r:n} < T_0\right)} \int_0^{T_0} tF(t;\theta)^{r-1}(1-F(t;\theta))^{n-r} f(t;\theta)dt.$
We now present two useful results which will be needed in the next section to obtain optimal scheme.

Result 1: $E[D]$ is invariant when T is multiplied by some constant k.
Proof: Let us consider the transformation $T^* = kT$ so that T_0 changes to T_0^*. Then, $P[T^* \leq T_0^*] = P[T \leq T_0] = F(T_0;\theta)$. Hence, from (1), $E[D]$ remains the same under the transformed distribution.

Result 2: Suppose τ^* is the censoring time when T is multiplied by some constant k, then $E[\tau^*] = kE[\tau]$.

Proof: Using (2),

$$\begin{aligned} E[\tau^*] &= E\left[\min\left(kT_{r:n}, kT_0\right)\right] \\ &= kT_0 P\left(kT_{r:n} \geq kT_0\right) + E\left[kT_{r:n} | kT_{r:n} < kT_0\right] P(kT_{r:n} < kT_0) \\ &= kT_0 P\left(T_{r:n} \geq T_0\right) + kE\left[T_{r:n} | T_{r:n} < T_0\right] P(T_{r:n} < T_0) \\ &= kE[\tau]. \end{aligned}$$

Note that, for hybrid censored data, the likelihood function can be written as

$$L(\theta) \propto \prod_{i=1}^{d} f(t_{i:n};\theta)(1-F(\tau_0;\theta))^{n-d}, \tag{3}$$

where d and τ_0 denote the observed values of D and τ, respectively. Obtaining the corresponding Fisher information matrix for vector parameter θ, given by

$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\ln L(\theta)\right], \tag{4}$$

is in general difficult. Park *et al.* [21] gave an alternate expression for the Fisher information matrix, as given by

$$\mathcal{I}(\theta) = \int_0^{T_0} \left\{\frac{\partial}{\partial\theta}\ln h(t;\theta)\right\}^T \left\{\frac{\partial}{\partial\theta}\ln h(t;\theta)\right\} \sum_{i=1}^{r} f_{i:n}(t;\theta)dt, \tag{5}$$

where $h(t;\theta)$ is the hazard function of T and $f_{i,n}(t;\theta)$ is the density function of $T_{i:n}$.

In this work, we consider exponential and Weibull distributions for illustration. For exponential distribution with distribution function given by

$$F(t;\lambda) = 1 - e^{-\lambda t},\ \ \lambda, t > 0. \tag{6}$$

The Fisher information can be directly obtained, by using (4), as

$$\mathcal{I}_E(\lambda) = \frac{E[D]}{\lambda^2}. \tag{7}$$

It is easy to see that we get the same expression of Fisher information in exponential case using (5) as well. Now let us consider the Weibull distribution with the distribution function given by

$$F(t;\alpha,\lambda) = 1 - e^{-(\lambda t)^{\alpha}},\ \ \alpha,\lambda,t>0, \tag{8}$$

where α and λ are shape and scale parameters, respectively. The Fisher information matrix for Weibull distribution, obtained by using (5) with $\theta = (\alpha,\lambda)$, is given by

$$\mathcal{I}_W(\theta) = \begin{pmatrix} \mathcal{I}_1 & \mathcal{I}_3 \\ \mathcal{I}_3 & \mathcal{I}_2 \end{pmatrix},$$

where

$$\begin{aligned}
\mathcal{I}_1 &= \int_0^{T_0} \left\{\frac{1}{\alpha} + \ln(\lambda t)\right\}^2 \sum_{i=1}^{r} f_{i:n}(t;\theta)dt, \\
\mathcal{I}_2 &= \left(\frac{\alpha}{\lambda}\right)^2 \int_0^{T_0} \sum_{i=1}^{r} f_{i:n}(t;\theta)dt, \\
\mathcal{I}_3 &= \left(\frac{\alpha}{\lambda}\right) \int_0^{T_0} \left\{\frac{1}{\alpha} + \ln(\lambda t)\right\} \sum_{i=1}^{r} f_{i:n}(t;\theta)dt \\
\text{and} \quad f_{i:n}(t;\theta) &= i\binom{n}{i}\alpha\lambda\,(\lambda t)^{\alpha-1}\, e^{-(\lambda t)^{\alpha}(n-i+1)}(1-e^{-(\lambda t)^{\alpha}})^{i-1}.
\end{aligned}$$

3 Optimum Scheme for Known Sample Size

In this section, we consider the optimal choice of r and T_0 with fixed n for two different cost functions associated with life testing experiment.

3.1 Optimum Criteria

From practical considerations, it is desirable for the experimenter to obtain (n, r, T_0) which minimize the expected cost as well as to achieve a specific precision because both are important factors for business issues in industry houses. Increasing experimental cost can directly affect the business and poor estimation of model parameters may increase post-production cost. Keeping these practical issues in mind, we define the total cost as the sum of three components: (1) cost of failed items, (2) cost of duration of the experiment, and (3) cost due to imprecision of the estimates of unknown parameters. Although the first two cost components can be directly measured, the third component is more abstract and requires ideas about the different issues related to post-production cost. Nevertheless, in our study, we consider a cost associated with per unit of 'imprecision', as described in the following. Our aim is to find optimum values of r and T_0 which minimize the total experimental cost. From (1) and (2), it can be seen that both $E(D)$ and $E(\tau)$ are increasing functions of r and T_0 for fixed n. One could consider only the first two cost components and the total expected cost of the experiment as $C_1E(D) + C_2E(\tau)$, where C_1 and C_2 are the cost per unit of failure and the cost per unit duration of experiment, respectively; but this leads to a trivial solution (i.e., smallest possible values of r and T_0). Hence, we need to incorporate a third component which is decreasing in r and T_0. In view of this we propose to incorporate the variance of the estimates of the parameters representing imprecision. Note that the variance is decreasing in r and T_0 and hence the nature of the proposed cost function is not monotonically increasing or decreasing.

First we consider trace of the inverse of Fisher information matrix as the measure of imprecision, since it essentially gives the sum of variances of the estimated model parameters, commonly referred to as A-optimality [2]. Let $\mathcal{I}$ be the Fisher information matrix. In addition to the costs C_1 and C_2 considered earlier, let C_3 denote the cost per unit of imprecision. Then the total expected cost ($\mathcal{C}_1$) associated with an experiment is

$$\mathcal{C}_1 = C_1 E[D] + C_2 E[\tau] + C_3 Trace(\mathcal{I}^{-1}). \tag{9}$$

Optimum values of r and T_0 are obtained by minimizing (9). Alternatively, one can use determinant of $\mathcal{I}^{-1}$, commonly referred to as D-optimality [2], which provides an overall measure of variability by taking into account the correlations. It is desirable that the proposed cost function should be scale invariant. We check this property for exponential and Weibull distributions. For exponential distribution given by equation (6), the total expected cost ($\mathcal{C}_1$) is

$$\mathcal{C}_1 = C_1 E[D] + C_2 E[\tau] + C_3 \frac{\lambda^2}{E[D]}. \tag{10}$$

Considering the transformation $T^* = kT$, T^* follows exponential distribution with parameter λ^* where $\lambda^* = \lambda/k$. Using Results 1 and 2, we have $E[D^*] = E[D]$ and $E[\tau^*] = kE[\tau]$, where D^* and τ^* denotes the number of failures and duration of the experiment in the transformed time scale. Therefore, the expected total cost ($\mathcal{C}_1^*$) in terms of transformed variable T^* is given by

$$\begin{aligned} \mathcal{C}_1^* &= C_1^* E[D^*] + C_2^* E[\tau^*] + C_3^* \frac{(\lambda^*)^2}{E[D^*]} \\ &= C_1 E[D] + C_2 E[\tau] + C_3 \frac{\lambda^2}{E[D]} \frac{1}{k^2}. \end{aligned} \tag{11}$$

Since $C_1^* = C_1$, $C_2^* = C_2/k$ and $C_3^* = C_3$. Hence, clearly, the optimum solution depends on k. A numerical verification is given in the next section.

For Weibull failure time distribution, given by (8), total expected cost becomes

$$\mathcal{C}_1 = C_1 E[D] + C_2 E[\tau] + C_3 \frac{\mathcal{I}_1 + \mathcal{I}_2}{\mathcal{I}_1 \mathcal{I}_2 - \mathcal{I}_3^2}, \tag{12}$$

where $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ are the elements of Fisher information matrix for Weibull distribution as derived in Section 2. As before, the third component given by $(\mathcal{I}_1 + \mathcal{I}_2)/(\mathcal{I}_1\mathcal{I}_2 - \mathcal{I}_3^2)$ depends on k after the scale transformation of T

to $T^* = kT$. Since this is difficult to prove analytically, the same has been verified numerically. Hence, the cost function (12) is not scale invariant. A numerical study is presented in Section 3.2 .

It is seen that the cost function (9) is not scale invariant due to the third component, $Trace(\mathcal{I}^{-1})$, which depends on the unit of time. To overcome this difficulty, we incorporate a variance measure proposed by Zhang and Meeker [23], given by

$$Var[\ln \hat{T}_p],$$

where $\hat{T}_p$ is the maximum likelihood estimate of the pth quantile of life time distribution. Note that this measure depends on p. In our work, we consider a modified measure, proposed by Kundu [17], as

$$\int_0^1 Var[\ln \hat{T}_p]dp, \tag{13}$$

where the integral over $[0, 1]$ represents some average variance of the log quantile estimates over all quantile points so that it is not a function of p, rather a function of r and T_0. It can be easily shown that this variance measure is scale invariant for some specific distributions for which the linear relationship $\hat{T}_p^* = k\hat{T}_p$ holds good where $\hat{T}_p^*$ is the mle of T_p^*, the pth quantile of T^*. For example, this relationship holds for location-scale family of distributions. A random variable Y belongs to location-scale family of distributions if its cdf can be written as

$$P(Y \leq y) = G\left(\frac{y-\mu}{\sigma}\right),$$

where G does not depend on any unknown parameters. In this case $-\infty < \mu < \infty$ is location parameter and $\sigma > 0$ is scale parameter. The corresponding pth quantile is given by $Y_p = \mu + G^{-1}(p)\sigma$ and maximum likelihood estimate of Y_p is given by $\hat{Y}_p = \hat{\mu} + G^{-1}(p)\hat{\sigma}$ where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates of μ and σ, respectively. It is easy to see that under the transformation $Y^* = kY$, Y^* again belongs to location-scale family of distributions with location and scale parameters $\mu^* = k\mu$ and $\sigma^* = k\sigma$, respectively. So the quantile of transformed variable is given by $Y_p^* = \mu^* + G^{-1}(p)\sigma^* = kY_p$ with the estimate $\hat{Y}_p^* = k\hat{Y}_p$. As a consequence, $\int_0^1 Var[\ln \hat{Y}_p]dp$ preserves the invariance property. Moreover, this invariance property also holds for log-location-scale family of distributions. To prove that consider a random variable Z belongs to log-location-scale family of distributions with cdf

$$P(Z \leq z) = \varnothing \left(\frac{\ln z - \mu}{\sigma} \right),$$

where $\varnothing$ does not depend on any unknown parameters. Here, $-\infty < \mu < \infty$ is location parameter and $\sigma > 0$ is scale parameter also. Therefore, corresponding pth quantile is given by $Z_p = \exp(\mu + \varnothing^{-1}(p)\sigma)$ and maximum likelihood estimate of Z_p is given by $\hat{Z}_p = \exp(\hat{\mu} + \varnothing^{-1}(p)\hat{\sigma})$ where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates of μ and σ, respectively. Under the scale transformation $Z^* = kZ$, Z^* also belongs to the log-location-scale family of distributions with location and scale parameters $\mu^* = \mu + \ln k$ and $\sigma^* = \sigma$, respectively. Now it is an easy exercise to check that the quantile of transformed variable is given by $Z_p^* = \exp(\mu^* + \varnothing^{-1}(p)\sigma^*) = kZ_p$ with the estimate $\hat{Z}_p^* = k\hat{Z}_p$.

Using the variance measure (13), we propose a new cost function as given below.

$$\mathcal{C}_2 = C_1 E[D] + C_2 E[\tau] + C_3 \int_0^1 Var[\ln \hat{T}_p]dp. \tag{14}$$

The expressions of $\int_0^1 Var[\ln \hat{T}_p]dp$ for exponential and Weibull distributions are given in Appendix.

3.2 Optimum Solution

Note that both the cost functions (9) and (14) are non-linear functions of the decision variables r and T_0, where r is discrete and T_0 is continuous.

The optimum solution cannot be obtained in analytic form and so a suitable numerical technique is required to obtain the optimum solution. Since r is discrete taking finite number of value from 1 to n, we first fix r and minimize the cost with respect to T_0. Minimization with respect to T_0 is carried out through "nlminb" function in standard R.2.15.0 software in which we take upper limit of T_0 at 0.99th quantile of corresponding life time distribution. This solution is then considered for all values of r to obtain the global minimum.

For numerical illustrations, we consider $n = 25$ and $C_1 = 5$, $C_2 = 25$, $C_3 = 250$ for both exponential and Weibull distributions. The optimum schemes $(\tilde{r}, \tilde{T_0})$ for both cost functions (9) and (14) are reported in Table 1. Along with the optimum schemes, it also presents the corresponding minimum costs $\tilde{\mathcal{C}_1}$ and $\tilde{\mathcal{C}_2}$, respectively. The parameter values corresponding to the two rows for each distribution refer to the study of invariance under scale transformation, in which the second row represents multiplying lifetime by $k = 2$, whereas the original lifetime T is represented in the first row. As expected, the optimum scheme is invariant for the cost function (14), but not for (9). Note that, when Weibull life time distribution is DFR, optimum cost is higher than when distribution is IFR or constant. Intuitively, when failure rate decreases, T_0 increases and hence experimental cost increases too. Optimum schemes $(\tilde{r}, \tilde{T_0})$ with optimum cost $\tilde{\mathcal{C}_2}$ for different n are reported in Table 2 in order to study the effect of n. It shows that r increases and T_0 decreases as n increases.

We also have investigated the impact of different cost components on the optimum solution by considering the cost function (14) which seems to be more desirable. Although the numerical results are not presented here, following qualitative observations have been noticed. Setting C_2, C_3 as fixed, r decreases with C_1 since failure cost is higher and, at the same time, T_0 also decreases. Again, setting C_1, C_3 as fixed, both r and T_0 decrease as C_2 increases. This is because C_2 directly affects T_0, whereas less number of failures leads to shorter duration of the experiment. Finally, to investigate the effect of the cost component due to imprecision, we set C_1, C_2 as fixed. Intuitively, higher precision cost means requirement of more information about the model parameters; as expected, both r and T_0 increase when C_3 increases.

4 Optimum Scheme for Unknown Sample Size

In the last section, we have obtained optimum (r, T_0) for fixed n. But in practice, the choice of n may also be of interest and we need to determine it optimally. In this section, we consider determination of optimum solution for n, r and T_0. We consider the same optimality criteria of minimizing the cost associated with the experiment. Let C_4 be the cost per unit of item on life test; adding this cost with the cost function (14), the new cost function becomes

$$\mathcal{C} = C_1E[D] + C_2E[\tau] + C_3\int_0^1 Var[\ln \hat{T}_p]dp + C_4n. \tag{15}$$

It is easy to see that this cost model is also scale invariant. To minimize (15) numerically, we used the same algorithmic approach as that of the previous section with n known. For illustration, we have taken cost components $C_1 = 5$, $C_2 = 25$, $C_3 = 250$ and $C_4 = 2$ for both exponential and Weibull distributions. The optimum schemes $(\tilde{n}, \tilde{r}, \tilde{T}_0)$ along with the minimum cost $(\tilde{\mathcal{C}})$ are reported in Table 3. Note that the optimum cost seems to be higher when Weibull life time distribution is DFR, as observed in Section 3.2 also.

5 An Alternative Approach

In the proposed cost functions of Sections 3 and 4, we have considered four types of cost. It may be noted that cost components C_1, C_2 and C_4 are easily available to the experimenter, but in practice, C_3 may not be always known. In view of this, an alternative approach of determining optimum r and T_0, for known sample size n, may be considered as

$$\begin{aligned} &\underset{r,T_0}{\text{minimize}} \int_0^1 Var[\ln \hat{T}_p]dp, \\ &\text{subject to } C_1E[D] + C_2E[\tau] \leq C_0, \end{aligned} \tag{16}$$

where C_1 and C_2 are as defined in Section 3 and C_0 is some prefixed budget amount. Note that our proposed cost function is similar to that of Dube *et al.* [8], but considering the cost of failed items and duration instead of that of n items on test and T_0. Intuitively, the quantity $\int_0^1 Var[\ln \hat{T}_p]dp$ is monotonically decreasing function of both r and T_0. For exponential distribution, this quantity is $1/E[D]$ (See Result A1 in the Appendix) which can be proved to be decreasing in both r and T_0. Note that, for fixed r, $C_1E[D] + C_2E[\tau]$ is a monotonically increasing function in T_0. Hence, for given C_0 and a fixed r, there is a unique choice of T_0 satisfying equality in (16) which minimizes $\int_0^1 Var[\ln \hat{T}_p]dp$. The aim is to choose that pair (r, T_0) as the optimum solution for which $\int_0^1 Var[\ln \hat{T}_p]dp$ is minimum. For numerical illustrations, we take $n = 25$, $C_1 = 5$, $C_2 = 25$ and consider both exponential and Weibull distributions as before. The optimum schemes $(\tilde{r}, \tilde{T}_0)$ for different C_0 are reported in Table 4. As expected, the optimum values of both r and T_0 increase with the budget C_0 for fixed C_1 and C_2. A qualitative analysis of the proposed cost criterion (16) is also done in order to study the effect of the two cost components on the optimum solution. We notice that for fixed C_0, optimum r and T_0 both decrease, when C_1 and C_2 both increase, as expected.

As before, determination of optimum n, r and T_0 may also be considered as

$$\begin{aligned} &\underset{n,r,T_0}{\text{minimize}} \int_0^1 Var[\ln \hat{T}_p]dp, \\ &\text{subject to } C_1E[D] + C_2E[\tau] + C_4n \leq C_0, \end{aligned} \tag{17}$$

where C_4 is as defined in Section 4. For the numerical illustration to obtain optimum schemes $(\tilde{n}, \tilde{r}, \tilde{T}_0)$ for different C_0, we have taken $C_1 = 5$, $C_2 = 25$,

$C_4 = 2$ and considered both exponential and Weibull distributions in Table 5. As in [8], the search for n is restricted to $n = 1, 2, \cdots, n_1$, where $n_1 = \max\{n : C_4 n \leq C_0\}$. It is clear from (16) and (17) that the optimum solutions for both the criteria are scale invariant for exponential and Weibull distributions. In both cases, as expected, the optimum choice of design parameters n, r and T_0 increase with C_0, when C_1, C_2 and C_4 are fixed.

6 Sensitivity Analysis

The proposed method of determining the optimum design parameters depends on the choice of parameters and different cost components. Therefore, in practice, the experimenter needs to specify the values of these parameters and cost components to obtain optimum solution. Hence, a well-planned sensitivity analysis is required to be performed in order to analyze the effect of mis-specification of parameter values and costs on the optimal solution.

First we consider the sensitivity analysis with respect to the parameters of the underlying distribution for a fixed set of cost coefficients. We define the relative efficiency of a set of parameter values θ compared to the set of true parameter values θ_0 based on the cost function is,

$$\mathrm{RE}_1(\theta) = \frac{E_{\theta_0}[\text{Cost for optimum solution under } \theta_0]}{E_{\theta_0}[\text{Cost for optimum solution under } \theta]}.$$

The value of $\mathrm{RE}_1(\theta)$ lies between 0 and 1 with values closer to 1, representing less sensitivity of optimum solution due to mis-specification of model parameters.

We consider the cost function (14) for sensitivity analysis. Note that because of the invariance property of the cost function, sensitivity analysis is not required when lifetime follows exponential distribution. Hence, without loss of generality the parameter of exponential distribution λ can be set to 1. Similarly, for Weibull distribution, the scale parameter λ can be set to 1. We carry out sensitivity analysis for Weibull distribution w.r.t the shape parameter α. We calculate $\text{RE}_1(\theta)$ in the neighborhood of α. Here, we consider $C_1 = 5$, $C_2 = 25$, $C_3 = 25$ and $n = 25$ for illustration. The results are shown in the Table 6. From Table 6, we observe that the optimum schemes based on the cost function (14) is less sensitive to the mis-specification of shape parameter of Weibull distribution.

Next we consider sensitivity analysis w.r.t different cost components. As earlier, for fixed value of distribution parameter θ, we define the relative efficiency of a set of cost values $\mathcal{C}$ compared to the set of true cost values $\mathcal{C}_0$ based on the cost function,

$$\text{RE}_2(\mathcal{C}) = \frac{E_{\mathcal{C}_0}[\text{Cost for optimum solution under } \mathcal{C}_0]}{E_{\mathcal{C}_0}[\text{Cost for optimum solution under } \mathcal{C}]}.$$

Although the numerical results are not presented here, it has been found that the optimum schemes are also less sensitive to the mis-specification of cost components C_1, C_2, C_3 and C_4.

7 Conclusions

In this work, we have considered the design issues in life testing experiment under hybrid censoring scheme. We have proposed a new cost function to obtain optimum values of decision parameters n, r and T_0. The proposed cost function is scale invariant for location-scale and log-location-scale family of distributions, which is a desirable property. Moreover, it can be easily verified that the invariance property of cost function also holds for Birnbaum-Saunders, Rayleigh, generalized Rayleigh, generalized exponential and generalized inverted exponential distributions. It may be noted that this invariance property depends on whether the variance measure used is invariant or not. As an alternative to $\int_0^1 Var[\ln \hat{T}_p]dp$, one can simply consider $Var[\ln \hat{T}_p]$ for some specific p. For example, $Var[\ln \hat{T}_{0.5}]$ may be a good choice.

The proposed methodology described here can be easily extended to Type-II hybrid censoring scheme [6] in which the experiment terminates at a random time, $\tau' = \max(T_{r:n}, T_0)$, where r $(1 \leq r \leq n)$ and T_0 are fixed in advance. The quantity of interests, namely, expected number of failures, expected duration and the variance measure, as Type-I hybrid censoring scheme, can be worked out in similar manner. Considering the cost function similar to that in (14), the invariance property can be established for Type-II hybrid censoring scheme also. It is to be noted that the optimal r and T_0 for this case are generally smaller than those for Type-I hybrid censoring case while the corresponding optimum cost value is also smaller than the Type-I case.

In this article, it is assumed that the parameter values of the life time distribution are known based on some prior knowledge, although the analysis in Section 6 indicates the optimal solutions to be insensitive to little perturbation of the parameter values. In practice, when such knowledge is not available, finding optimum solution in Bayesian framework, or by multi stage procedure, may be considered.

8 Appendix

Result A1: If T follows exponential distribution given by (6), $\int_0^1 Var[\ln \hat{T}_p]dp = 1/E[D]$, where $\hat{T}_p$ is the mle of T_p, the pth quantile of T.

Proof: For exponential distribution, $T_p = -(1/\lambda)\ln(1-p)$. Then, the mle of T_p is given by $\hat{T}_p = -(1/\hat{\lambda})\ln(1-p)$, where $\hat{\lambda}$ is the mle of λ. The asymptotic variance of $\ln \hat{T}_p$ is obtained by delta method and is given by, using (7),

$$Var\left[\ln \hat{T}_p\right] = \left(\frac{1}{\lambda^2}\right)\frac{\lambda^2}{E[D]} = \frac{1}{E[D]}.$$

Hence, $\int_0^1 Var[\ln \hat{T}_p]dp = 1/E[D]$.

Result A2: Suppose that T follows Weibull distribution given by (8) and $\hat{T}_p$ is the mle of T_p, the pth quantile of T. Then we have,

$$\begin{aligned}\int_0^1 Var[\ln \hat{T}_p]dp &= \frac{u_1}{\alpha^4}\int_0^1 \{\ln\{-\ln(1-p)\}\}^2 dp \\ &+\frac{2u_3}{\alpha^2\lambda}\int_0^1 \ln\{-\ln(1-p)\}\,dp + \frac{u_2}{\lambda^2},\end{aligned}$$

where u_1, u_2 and u_3 are the elements of inverse of Fisher information matrix given by $\mathcal{I}_W^{-1} = \left(\begin{smallmatrix} u_1 & u_3 \\ u_3 & u_2\end{smallmatrix}\right)$.

Proof: For Weibull distribution, $T_p = (1/\lambda)\{-\ln(1-p)\}^{\frac{1}{\alpha}}$. Then, the mle of T_p is given by $\hat{T}_p = (1/\hat{\lambda})\{-\ln(1-p)\}^{\frac{1}{\hat{\alpha}}}$, where $\hat{\alpha}$ and $\hat{\lambda}$ be the mle of α and λ, respectively. Again, by delta method, $Var[\ln \hat{T}_p]$ is given by

$$Var[\ln \hat{T}_p] = B\mathcal{I}_W^{-1}B^T,$$

where $B = \big(-(1/\alpha^2)\ln\{-\ln(1-p)\}, -(1/\lambda)\big)$. Now, integrating over p, we get the desired result.

References

[1] B. Epstein, Sampling procedures and tables for life and reliability testing (Based on exponential distribution), Quality control and reliability handbook (Interim) H 108, Office of the assistant secretary of defence (Supply and logistics), 1960.

[2] E. P. Liski, N. K. Mandal, K. R. Shah, and B. K. Sinha, Topics in optimal design (Lecture notes in statistics), Springer, 2002.

[3] B. J. N. Blight, On the most economical choice of a life testing procedure for exponentially distributed data, Technometrics 14 (1972), 613-618.

[4] M. Burkschat, E. Cramer, and U. Kamps, On optimal schemes in progressive censoring, Statistics & Probability Letters 76 (2006), 1032-1036.

[5] S. Chen and G. K. Bhattacharyya, Exact confidence bounds for an exponential parameter under hybrid censoring, Communications in Statistics - Theory and Methods 17 (1988), 1857-1870.

[6] 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 (2003), 319-330.

[7] N. Draper and I. Guttman, Bayesian analysis of hybrid life tests with exponential failure times, Annals of the Institute of Statistical Mathematics 39 (1987), 219-225.

[8] S. Dube, B. Pradhan, and D. Kundu, Parameter estimation of the hybrid censored log-normal distribution, Journal of Statistical Computation and Simulation 81 (2011), 275-287.

[9] N. Ebrahimi, Determining the sample size for a hybrid life test based on the cost function, Naval Research Logistics 35 (1988), 63-72.

[10] N. Ebrahimi, Estimating the parameters of an exponential distribution from a hybrid life test, Journal of Statistical Planning and Inference 14 (1986), 255-261.

[11] N. Ebrahimi, Prediction intervals for future failures in the exponential distribution under hybrid censoring, IEEE Transactions on Reliability 41 (1992), 127-132.

[12] B. Epstein, Truncated life test in the exponential case, Annals of Mathematical Statistics 25 (1954), 555-564.

[13] K. Fairbanks, R. Madsen, and R. Dykstra, A confidence interval for an exponential parameter from a hybrid life test, Journal of the American Statistical Association 77 (1982), 137-140.

[14] R. D. Gupta and D. Kundu, Hybrid censoring schemes with exponential failure distribution, Communications in Statistics - Theory and Methods 27 (1998), 3065-3083.

[15] R. D. Gupta and D. Kundu, On the comparison of Fisher information of the Weibull and GE distributions, Journal of Statistical Planning and Inference 136 (2006), 3130-3144.

[16] D. Kundu, On hybrid censored Weibull distribution, Journal of Statistical Planning and Inference 137 (2007), 2127-2142.

[17] D. Kundu, Bayesian inference and life testing plan for the Weibull distribution in presence of progressive censoring, Technometrics 50 (2008), 144-154.

[18] Y. Lam, Bayesian variable sampling plans for the exponential distribution with Type I censoring, The Annals of Statistics 22 (1994), 696-711.

[19] Y. P. Lin, T. Liang, and W. T. Huang, Bayesian sampling plans for exponential distribution based on Type I censoring data, Annals of the Institute of Statistical Mathematics 54 (2002), 100-113.

[20] H. K. T. Ng, P. S. Chen, and N. Balakrishnan, Optimal progressive censoring plans for the Weibull distribution, Technometrics 46 (2004), 470-481.

[21] S. Park and N. Balakrishnan, On simple calculation of the Fisher information in hybrid censoring schemes, Statistics & Probability Letters 79 (2009), 1311-1319.

[22] J. D. Riley, Comparative cost of two life test procedures, Technometrics 4 (1962), 140-142.

[23] Y. Zhang and W. Q. Meeker, Bayesian life test planning for the Weibull distribution with given shape parameter, Metrika 61 (2005), 237-249.

Table 1: Optimum schemes $(\tilde{r}, \tilde{T}_0)$ with $n = 25$ and $(C_1, C_2, C_3) = (5, 25, 250)$ using cost functions $\mathcal{C}_1$ and $\mathcal{C}_2$ of Section 3.

Distribution	Parameter(s)	Cost Function			
		$\mathcal{C}_1$		$\mathcal{C}_2$	
		$(\tilde{r}, \tilde{T}_0)$	$\tilde{\mathcal{C}}_1$	$(\tilde{r}, \tilde{T}_0)$	$\tilde{\mathcal{C}}_2$
Exponential	$\lambda = 1$	(7, 0.3555)	78.3167	(7, 0.3555)	78.3167
	$\lambda = 0.5$	(4, 0.3319)	38.9095	(7, 0.7110)	78.3167
Weibull	$\alpha = 2,\ \lambda = 1$	(13, 1.9991)	153.7957	(6, 1.2067)	71.6512
	$\alpha = 2,\ \lambda = 0.5$	(12, 3.1647)	149.4188	(6, 2.4037)	71.6512
	$\alpha = 1,\ \lambda = 1$	(11, 2.0001)	118.7021	(10, 2.3355)	122.2269
	$\alpha = 1,\ \lambda = 0.5$	(8, 3.4983)	89.7668	(10, 4.6790)	122.2269
	$\alpha = 0.5,\ \lambda = 1$	(15, 1.1492)	170.5286	(15, 7.9706)	253.3002
	$\alpha = 0.5,\ \lambda = 0.5$	(10, 5.5031)	99.0998	(15, 15.9407)	253.3002

Table 2: Optimum schemes $(\tilde{r}, \tilde{T}_0)$ with (C_1, C_2, C_3)=$(5, 25, 250)$ using cost function $\mathcal{C}_2$ of Section 3 for different n.

Distribution	Parameter(s)	n	$(\tilde{r}, \tilde{T}_0)$	$\tilde{\mathcal{C}}_2$
Exponential	$\lambda = 1$	30	(7, 0.3070)	76.9708
		35	(9, 0.2146)	76.0724
		40	(9, 0.1882)	75.3669
		45	(8, 0.1788)	74.8144
		50	(13, 0.1439)	74.4094
Weibull	$\alpha = 2,\ \lambda = 1$	30	(7, 1.1265)	73.0841
		35	(7, 1.0002)	74.2322
		40	(7, 0.9999)	75.6196
		45	(8, 0.9999)	77.1042
		50	(8, 0.9999)	78.2119
	$\alpha = 1,\ \lambda = 1$	30	(11, 2.3922)	121.9403
		35	(12, 1.4675)	122.6809
		40	(12, 1.8497)	123.8173
		45	(13, 1.6203)	125.2264
		50	(13, 1.0000)	126.8366
	$\alpha = 0.5,\ \lambda = 1$	30	(17, 6.8965)	241.1732
		35	(18, 5.2833)	234.1830
		40	(20, 2.8445)	230.4492
		45	(21, 2.3192)	228.4848
		50	(22, 1.9778)	227.8521

Table 3: Optimum schemes $(\tilde{n}, \tilde{r}, \tilde{T}_0)$ with (C_1, C_2, C_3, C_4)=$(5, 25, 250, 2)$ using cost function $\mathcal{C}$ of Section 4.

Distribution	Parameter(s)	$(\tilde{n}, \tilde{r}, \tilde{T}_0)$	$\tilde{\mathcal{C}}$
Exponential	$\lambda = 1$	(10, 5, 4.4330)	111.1409
	$\lambda = 0.5$	(10, 5, 8.5601)	111.1409
Weibull	$\alpha = 2,\ \lambda = 1$	(8, 4, 1.8862)	90.1416
	$\alpha = 2,\ \lambda = 0.5$	(8, 4, 3.7621)	90.1416
	$\alpha = 1,\ \lambda = 1$	(15, 8, 3.4111)	160.1794
	$\alpha = 1,\ \lambda = 0.5$	(15, 8, 6.8119)	160.1794
	$\alpha = 0.5,\ \lambda = 1$	(25, 15, 7.9688)	303.3002
	$\alpha = 0.5,\ \lambda = 0.5$	(25, 15, 15.9381)	303.3002

Table 4: Optimum schemes $(\tilde{r}, \tilde{T}_0)$ with $n = 25$ and (C_1, C_2)=$(5, 25)$ for different C_0 using criterion (16) of Section 5.

Distribution	Parameter(s)	C_0	$(\tilde{r}, \tilde{T}_0)$
Exponential	$\lambda = 1$	25	(5, 0.2126)
		50	(9, 0.4499)
		100	(16, 1.0987)
		200	(25, 3.4597)
Weibull	$\alpha = 2,\ \lambda = 1$	25	(3, 2.1460)
		50	(7, 2.1460)
		100	(16, 1.0314)
		200	(25, 2.1460)
	$\alpha = 1,\ \lambda = 1$	25	(4, 4.6052)
		50	(8, 4.6052)
		100	(16, 1.0987)
		200	(25, 3.4598)
	$\alpha = 0.5,\ \lambda = 1$	25	(5, 0.9910)
		50	(9, 0.5899)
		100	(16, 1.2737)
		200	(22, 5.1438)

Table 5: Optimum schemes $(\tilde{n}, \tilde{r}, \tilde{T}_0)$ with (C_1, C_2, C_4)=$(5, 25, 2)$ for different C_0 using criterion (17) of Section 5.

Distribution	Parameter(s)	C_0	$(\tilde{n}, \tilde{r}, \tilde{T}_0)$
Exponential	$\lambda = 1$	25	(5, 2, 0.4175)
		50	(9, 4, 0.8435)
		100	(16, 10, 1.0235)
		200	(30, 22, 1.5907)
		500	(72, 62, 2.2209)
Weibull	$\alpha = 2,\ \lambda = 1$	25	(3, 1, 2.1460)
		50	(7, 4, 0.8732)
		100	(17, 9, 2.1460)
		200	(32, 22, 1.2664)
		500	(76, 63, 2.1460)
	$\alpha = 1,\ \lambda = 1$	25	(4, 2, 0.5217)
		50	(8, 4, 0.9660)
		100	(17, 10, 0.9302)
		200	(33, 22, 1.2708)
		500	(80, 61, 1.7974)
	$\alpha = 0.5,\ \lambda = 1$	25	(5, 2, 0.6527)
		50	(10, 5, 0.4750)
		100	(19, 10, 0.9590)
		200	(35, 22, 1.2568)
		500	(85, 59, 21.2076)

Table 6: Optimum schemes based on θ and their relative efficiencies compared to θ_0 with C_1=5, C_2=25,C_3=250 and n=25 for Weibull distribution.

θ_0	θ	$(r, T_0)\|_\theta$	$\mathrm{RE}_1(\theta)$
2	1.7	(7, 1.2406)	0.9899
2	1.8	(7, 1.3798)	0.9899
2	1.9	(6, 2.1556)	1.0000
2	2	(6, 1.2067)	1.0000
2	2.1	(6, 1.2455)	1.0000
2	2.2	(6, 1.1836)	1.0000
2	2.3	(6, 1.1943)	1.0000
0.5	0.3	(17, 9.2615)	0.9802
0.5	0.4	(16, 7.8637)	0.9948
0.5	0.5	(15, 7.9706)	1.0000
0.5	0.6	(14, 4.2207)	0.9951
0.5	0.7	(13, 4.0430)	0.9792
0.5	0.8	(12, 2.7996)	0.9513