JIRSS (2018)
Vol. 17, No. 1, pp 1-31
DOI: 10.29252/jirss.17.1.1

Model Selection Based on Tracking Interval Under Unified Hybrid Censored Samples

Abdolreza Sayyareh[1] **and Hanieh Panahi**[2]

[1] Faculty of Mathematics, K.N. Toosi University of Technology, Tehran, Iran
[2] Department of Mathematics and Statistics, Lahijan Branch, Islamic Azad University, Lahijan, Iran

Received: 26/09/2016, Revision received: 09/09/2017, Published online: 19/05/2018

Abstract. The aim of statistical modeling is to identify the model that most closely approximates the underlying process. Akaike information criterion (AIC) is commonly used for model selection but the precise value of AIC has no direct interpretation. In this paper we use a normalization of a difference of Akaike criteria in comparing between the two rival models under unified hybrid censoring scheme. Asymptotic properties of maximum likelihood estimator based on the missing information principle are derived. Also, asymptotic distribution of the normalized difference of AICs is obtained and it is used to construct an interval, say tracking interval, for comparing the two competing models. Monte Carlo simulations are performed to examine how the proposed interval works for different censoring schemes. Two real datasets have been analyzed for illustrative purposes. The first is selecting between Weibull and generalized exponential distributions for main component of spearmint essential oil purification data. The second is the choice between models of the lifetimes of 20 electronic components.

Keywords. Asymptotic distribution, Kullback-Leibler risk, Missing information

Abdolreza Sayyareh (asayyareh@kntu.ac.ir)
Corresponding Author: Hanieh Panahi (panahi@liau.ac.ir)

principle, Model selection, Tracking interval, Unified hybrid censoring, Vuong's test.

MSC: 62N01;62N03;62E20.

1 Introduction

Model selection is an important process in identifying the closest model to the true model. Several articles have been published on model selection based on complete data, for example, Kundu *et al.* (2005) compared the log-normal and generalized exponential distribution using maximized likelihood method, Dey and Kundu (2009) considered the problem of discriminating among the log-normal, Weibull and generalized exponential distributions, Cox (1961) improved the classical hypothesis testing to compare the non-nested hypothesis, Vuong (1989) tested the equivalence of the two models using the expectation of the log-likelihood ratio of the two candidate models. The results in Vuong have been extended and applied in a number of ways, including, Vuong and Wang (1993), Lien (1987), Commenges *et al.* (2008), Sayyareh *et al.* (2011), Commenges *et al.* (2012) and Panahi (2016). Akaike (1973) introduced a criterion to select the best model under parsimony. One problem with AIC is that its values depend on the number of observations. For this reason, Commenges *et al.* (2008) and Sayyareh (2012) considered the normalized difference of AIC as an estimate of a difference of KL risks between two models and then constructed the tracking interval to verify the equivalence of two rival models. However, in testing experiments, experimenters may not always be in a position to observe the times of all inspected items in the test. This may be because of time limitation and/or other restrictions (such as money and material resources, etc) on data collection. Data obtained from such experiments are called censored data. The problem of choosing the closest rival model to the true model becomes more difficult if the data are censored. Because the distance between the two censored rival models can be very small, and it may be very difficult to discriminate between them. The new interval say tracking interval based on normalized difference of AIC may be used to discriminate between them. The common hybrid censoring schemes are Type I and Type II hybrid censoring schemes. Both these censoring schemes have some disadvantages. To overcome these problems, Chandrasekar *et al.* (2004) proposed two generalized hybrid censoring schemes. These schemes may be considered as an extension of Type I and Type II hybrid censoring schemes in some sense. In generalized Type I hybrid censoring scheme, one prefixes $k, r \in \{1, ..., n\}$ and $T \in (0, \infty)$ such that $k < r$. If the k^{th} failure occurs before time T, the experiment terminates at $\min(X_{r:n}, T)$.

If the k^{th} failure occurs after time T, the experiment terminates at $X_{k:n}$. In generalized Type II hybrid censoring scheme, one prefixes $r \in \{1, ..., n\}$ and $T_1, T_2 \in (0, \infty)$ such that $T_1 < T_2$. If the r^{th} failure occurs before time T_1, the experiment terminates at T_1; if the r^{th} failure occurs between T_1 and T_2, the experiment terminates at $X_{r:n}$; otherwise, the experiment terminates at T_2. This hybrid censoring scheme guarantees that the experiment time will not exceed T_2. Although these censoring schemes are improvements over Type I and Type II hybrid censoring schemes but they have some drawbacks. To overcome these problems, Balakrishnan *et al.* (2008) proposed a unified hybrid censoring scheme (UHCS). For more about unified hybrid censored samples, the reader is referred to Mohie El-Din *et al.* (2017), Hasaballah (2016), Rastogi and Tripathi (2013), Balakrishnan and Kundu (2013), Habibi Rad and Izanlo (2011), Huang and Yang (2010), Panahi (2017). Although some articles have been done on the unified hybrid censoring scheme but we have not come across any article on the behavior of the two rival models under unified hybrid censoring scheme. Thus, in this paper, we want to decide whether or not the two candidate models are two equivalent models. For this purpose, we obtain the asymptotic normality of the maximum likelihood estimator under unified hybrid censored sample. We also propose a test statistic that converges in distribution to the normal distribution and use it to test the null hypothesis that the rival models are equally close to the data generating model against the alternative hypothesis that one model is closer. Moreover, based on the Kullback asymmetric we obtain the tracking interval with a pre-specified probability. This interval helps us to evaluate proposed models in comparison with each other. Monte Carlo simulations are performed to study the behavior of the proposed interval and two real datasets are analyzed for illustrative purposes. The remainder of the paper is organized as follows: In Section 2, we first describe the unified hybrid censoring scheme and introduce the Kullback-Leibler divergence. In Section 3, we provide the asymptotic results based on unified hybrid censoring scheme. Tracking interval for the difference of the expected KL divergence of two non-nested rival models under unified hybrid censoring scheme is presented in Section 4. Monte Carlo simulations results and the analysis of two real datasets are provided in Section 5 and finally we conclude the paper in Section 6. Some acronyms are presented in the Appendix for quick references.

2 Model Description and Kullback-Leibler (KL) Divergence

2.1 Unified Hybrid Censoring Scheme (UHCS)

Suppose that n identical units are put on a test, with the lifetimes and ordered lifetimes of the n items are denoted by $X_1, ..., X_n$ and $Y_1, ..., Y_n$ respectively. Fix $k, r \in \{0, ..., n\}$ and $T_1 < T_2 \in (0, \infty)$, such that $k < r$. If k^{th} failure occurs before time T_1, the experiment terminate at $\min\{\max(Y_r, T_1),\ T_2\}$, if the k^{th} failure occurs between T_1 and T_2, the experiment terminate at $\min\{\ Y_r,\ T_2\}$ and if the k^{th} failure occurs after time T_2, then the experiment terminates at Y_k. Under this censoring scheme, we can guarantee that the experiment would be completed at most in time T_2 with at least k failures and if not, we can guarantee exactly k failures.

Therefore, under this censoring scheme we can observe the following six Types of observations as:

Case 1: $0 < Y_k < Y_r < T_1 < \mathrm{T}_2$ the experiment terminate at T_1
Case 2: $0 < Y_k < T_1 < Y_r < T_2$ the experiment terminate at Y_r
Case 3: $0 < Y_k < T_1 < T_2 < Y_r$ the experiment terminate at T_2
Case 4: $0 < \mathrm{T}_1 < Y_k < Y_r < T_2$ the experiment terminate at Y_r
Case 5: $0 < \mathrm{T}_1 < Y_k < T_2 < Y_r$ the experiment terminate at T_2

$$\textit{Case 4}:\ 0 < T_1 < T_2 < Y_k < Y_r \qquad \text{the experiment terminates at } Y_k \tag{2.1}$$

2.2 KL Divergence

Consider a sample of independently identically distributed (*i.i.d.*) random variables, $X_1, ..., X_n$ having probability density function $h(.)$. Let us consider two rival models:

$$F^{\alpha} = \{f^{\alpha}(.),\ \ \alpha \in \mathrm{M} \subset R^p\} \quad \text{and} \quad G^{\beta} = \left\{g^{\beta}(.),\ \ \beta \in \mathrm{B} \subset R^q\right\}.$$

Definition 2.1. (i) (f) and (g) are non-overlapping if $(f) \bigcap (g) = \phi$; (ii) (f) is nested in (g) if $(f) \subset (g)$; (iii) (f) is well- specified if there is a value $\alpha_0 \in \mathrm{M}$ such that $f^{\alpha_0}(.) = h$; otherwise it is misspecified. If the model is well-specified then $\alpha_0 = \alpha_*$, where $\alpha_* = arg\ max_{\alpha \in M} E_h(L_n^f(\alpha))$ and refer to as the pseudo-true value of the α. We consider the $f^{\alpha}(.)$ as a proposed model, then quasi-log-likelihood function is given by $L_n^f(\alpha) = \sum_{i=1}^n \log f^{\alpha}(x_i)$. Under the following condition, $\hat{\alpha}_n$ is a quasi-maximum likelihood estimator (QMLE):

$$L_n^f(\hat{\alpha}_n) = sup_{\alpha \in M} L_n^f(\alpha)$$

The KL information in favor of $h(x)$ against $f^{\alpha}(.)$ is defined in Kullback and Leibler (1951) to be(Pardo, 2005):

$$D_{KL(h,f^{\alpha})} = \int_{-\infty}^{\infty} h(x) \log \frac{h(x)}{f^{\alpha}(x)}\, dx = E_h \left(\log \frac{h(X)}{f^{\alpha}(X)} \right)$$

We can say that (f) is closer to h than (g) if $KL(h,\ f^{\alpha_*}) < KL(h,\ g^{\beta_*})$. We cannot estimate $KL(h, f^{\alpha_*})$ because the entropy of h, $E_h(\ \log h(X)\ )$, cannot be correctly estimated. However, we can estimate the difference of risks $\Delta_{USCH}(f^{\alpha_*}, g^{\beta_*}) = KL(h, f^{\alpha_*}) - KL(h, g^{\beta_*})$, a quantitative measure of the difference of misspecification by $[-n^{-1}(L_n^f(\hat{\alpha}_n) - L_n^g(\hat{\beta}_n))]$. This result may not be completely satisfactory in practice if n is not very large because the distribution we will use is $f^{\hat{\alpha}_n}$ rather than f^{α_*}. Thus it is reasonable to consider the risk $E_h \left\{ \log(h(X)/f^{\hat{\alpha}_n}(X)) \right\}$ that we call the expected KL risk and denote by $EKL(h, f^{\hat{\alpha}_n})$.

3 Asymptotic Results

In this section, we consider the difference quasi-log-likelihood functions of the $UHCS(r, k, T_1, T_2)$. From (2.1), the quasi-log-likelihood function of combined Cases 1-6 can be written as:

$$L_n^f(\alpha) = \sum_{i=1}^{d} \log f^{\alpha}(y_i) + (n-d) \log \overline{F}^{\alpha}(s)$$

Here, $s(s \in \{T_1, T_2, y_k, y_r\})$ denotes the stopping point and $d(d \in \{d_1, d_2, k, r\})$ is the number of failures that occur before time point s. On the other word,

$$d = \begin{cases} \breve{d} & case1 \\ r & case2 \ \& \ case4 \\ d_2 & case3 \ \& \ case5 \\ k & case6 \end{cases}$$

and

$$d = \begin{cases} T_1 & case1 \\ y_r & case2 \ \& \ case4 \\ T_2 & case3 \ \& \ case5 \\ y_k & case6 \end{cases}$$

where, $\breve{d} = d_1 = d_2$. Therefore, the differences of the quasi-log-likelihood functions of the two rival models can be obtained as:

$$L_n^{f/g}(\hat{\alpha}_n, \hat{\beta}_n) = L_n^f(\hat{\alpha}_n) - L_n^g(\hat{\beta}_n) = \sum_{i=1}^{d} \log \frac{f^{\hat{\alpha}_n}(y_i)}{g^{\hat{\beta}_n}(y_i)} + (n-d)\log \frac{\overline{F}^{\hat{\alpha}_n}(s)}{\overline{G}^{\hat{\beta}_n}(s)}. \tag{3.1}$$

where, $\hat{\alpha}_n$ is the quasi-maximum likelihood estimator for the parameter α. Now, we first prove the asymptotic normality property of MLE under unified hybrid censoring scheme.

The minimum assumptions, Q, for non-degenerate interval M are:
Q_1: The parameter space M is an open interval in R.
Q_2: $(\partial/\partial\alpha)\, f^{\alpha}(x)$ is a strictly monotone function on M for each x.
Q_3: For all $\alpha \in M$, the partial derivative $(\partial/\partial\alpha)\, f^{\alpha}(x)$, is integrable on R, the partial derivative $(\partial/\partial\alpha)\, F^{\alpha}(x)$, exists for $x \in \chi$,; and satisfies

$$(\partial/\partial\alpha)\, F^{\alpha}(x) = \int_{-\infty}^{x} (\partial/\partial\alpha)\, f^{\alpha}(u)\ du$$

Q_4: For every α, we have,

$$\mid \frac{\partial}{\partial\alpha} f^{\alpha}(x) \mid \leq K_1, \ \ \mid \frac{\partial^2}{\partial\alpha^2} f^{\alpha}(x) \mid \leq K_3, \ \ \mid \frac{\partial^3}{\partial\alpha^2} f^{\alpha}(x) \mid \leq K_3,$$

where, $\int K_i d\mu(x) < \infty; \quad i = 1, 2, 3.$
Q_5: For every α, $\frac{1}{\overline{F}^{\alpha}(x)}$ is bounded by $v(x)$, where, $E(v(X)) \leq C$; C is positive constant.
Q_6: For every α,we have, $\wp = \int \left(\frac{\partial}{\partial\alpha} \ln f^{\alpha}(x)\right)^2 f^{\alpha}(x) d\mu(x) < \infty$.

Lemma 3.1. *Based on the missing information principle, the normalized forms of the* $\frac{\partial}{\partial\alpha} \log L_n^f(\alpha)$ *and* $\frac{\partial^2}{\partial\alpha\partial\alpha'} \log L_n^f(\alpha)$ *under unified hybrid censored sample are:*

$$U_1 = \frac{1}{n}\left\{ (\sum_{i=1}^{n} \frac{\partial}{\partial\alpha_0} \log f^{\alpha}(w_i)) - \sum_{i=1}^{n-d} \frac{\partial}{\partial\alpha_0} \log f^{\alpha}(z_i) + (n-d)\frac{\partial}{\partial\alpha_0} \log(\overline{F}(s)) \right\}$$

and

$$U_2 = \frac{1}{n}\left\{ (\sum_{i=1}^{n} \frac{\partial^2}{\partial\alpha_0\partial\alpha'_0} \log f^{\alpha}(w_i)) - \sum_{i=1}^{n-d} \frac{\partial^2}{\partial\alpha_0\partial\alpha'} \log f^{\alpha}(z_i) + (n-d)\frac{\partial^2}{\partial\alpha_0\partial\alpha'} \log(\overline{F}(s)) \right\}$$

Proof. The Taylor expansion of $n^{-1}\frac{\partial L_n^f(\alpha)}{\partial\alpha}$ around $\alpha = \alpha_0$ gives:

$$\begin{aligned} n^{-1}\frac{\partial L_n^f(\alpha)}{\partial\alpha} &= n^{-1}\frac{\partial L_n^f(\alpha)}{\partial\alpha}\Big|_{\alpha=\alpha_0} + n^{-1}(\alpha-\alpha_0)\frac{\partial^2 L_n^f(\alpha)}{\partial\alpha\,\partial\alpha'}\Big|_{\alpha=\alpha_0} + o_p(1) \\ &= U_1 + U_2(\alpha-\alpha_0) + o_p(1) \end{aligned} \tag{3.2}$$

where,

$$U_1 = \frac{1}{n}\left\{(\sum_{i=1}^{d}\frac{\partial}{\partial\alpha}\log f^{\alpha}(y_i)) + (n-d)\frac{\partial}{\partial\alpha}\log(\overline{F}(s))\right\}_{\alpha=\alpha_0}$$

$$U_2 = \frac{1}{n}\left\{(\sum_{i=1}^{d}\frac{\partial^2}{\partial\alpha\,\partial\alpha'}\log f^{\alpha}(y_i)) + (n-d)\frac{\partial^2}{\partial\alpha\,\partial\alpha'}\log(\overline{F}(s))\right\}_{\alpha=\alpha_0}$$

Now, using the missing information principle (Louis, 1982; Lin and Balakrishnan, 2011), the observed information under unified hybrid censoring scheme is

$$\sum_{i=1}^{d}\log f^{\alpha}(y_i) = \sum_{i=1}^{n}\log f^{\alpha}(w_i) - \sum_{i=1}^{n-d}\log f^{\alpha}(z_i \mid Y) \tag{3.3}$$

where, $W = (w_1, ..., w_n)$ stands for the complete data and $Z = (z_1, ..., z_{n-d})$ represents the complete data of size $n - d$, from the left truncated population with density function:

$$h^* = \frac{f^{\alpha}(z)}{\overline{F}^{\alpha}(s)};\ z > s.$$

Note that, the sequences of random variables W's and Z's are independent. For simplicity, we use $f^{\alpha}(z_i)$ instead of $f^{\alpha}(z_i \mid Y)$ in what follows. Thus, U_1 can be rewritten as

$$\begin{aligned} U_1 &= \frac{1}{n}\left\{(\sum_{i=1}^{n}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(w_i)) - \sum_{i=1}^{n-d}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z_i) + (n-d)\frac{\partial}{\partial\alpha_0}\log(\overline{F}(s))\right\} \\ &\equiv \frac{1}{n}(U_1^* - U_1^{**}) \end{aligned} \tag{3.4}$$

where, $\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(.)$ means that $\frac{\partial}{\partial\alpha}\log f^{\alpha}(.)\big|_{\alpha=\alpha_0}$. Also, $U_2 \equiv \frac{1}{n}(U_2^* - U_2^{**})$ can be obtained similarly. ◻

Theorem 3.1 (Asymptotic distribution of the Maximum Likelihood Estimator)**.** *Assume that $f^{\alpha}(.)$ is a well specified model satisfying conditions Q_1-Q_6 and ($\hat{\alpha}_n = \max L_n^f(\alpha)$) . Then as $n \to \infty$, the asymptotic distribution of the MLE, $\sqrt{n}\,(\hat{\alpha}_n - \alpha_0)$ is, $N(0, J^{-1}_{f_{UHCS}})$, where, $J_{f_{UHCS}} \equiv \wp + (1-\breve{p}_i)\xi$ and*

$$\text{for } i = 1,2,3,4;\ \breve{p}_i = \lim_{n\to\infty}\frac{d}{n} = \begin{cases} \lim_{n\to\infty}\frac{\breve{d}}{n} & if \quad d = \breve{d} \text{ (Case 1)} \\ \lim_{n\to\infty}\frac{r}{n} & if \quad d = r \text{ (Cases 2 \& 4)} \\ \lim_{n\to\infty}\frac{d_2}{n} & if \quad d = d_2 \text{ (Cases 3 \& 5)} \\ \lim_{n\to\infty}\frac{k}{n} & if \quad d = k \text{ (Cases 6)} \end{cases}$$

and

$$\xi = \begin{cases} \xi_{\breve{d}} & if \quad d = \breve{d} \text{ (Case 1)} \\ \xi_r & if \quad d = r \text{ (Cases 2 \& 4)} \\ \xi_{d_2} & if \quad d = d_2 \text{ (Cases 3 \& 5)} \\ \xi_k & if \quad d = k \text{ (Cases 6)} \end{cases}$$

Proof. Using Lemma 3.1 and the Cramér (1946), $\frac{1}{n}U_1^* = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(w_i) \xrightarrow{P} 0$ and we will prove that

$$\frac{1}{n}U_1^{**} = \frac{1}{n}\sum_{i=1}^{n-d}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z_i) - (n-d)\frac{\partial}{\partial\alpha_0}\log(\overline{F}(s)) \xrightarrow{P} 0$$

We can rewrite U_1^{**} as

$$U_1^{**} = \sum_{i=1}^{n-d}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z_i) - \sum_{i=1}^{n-d}E\left(\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(Z_i)\right)$$

$$+\sum_{i=1}^{n-d}E\left(\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(Z_i)\right) - (n-d)\frac{\partial}{\partial\alpha_0}\log(\overline{F}(s))$$

From Q_3, we have

$$E\left(\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(Z)\right) = \int_s^{\infty}\left(\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z)\right)\frac{f^{\alpha}(z)}{\overline{F}^{\alpha}(s)}\,d\mu(z)$$

$$= \frac{\frac{\partial}{\partial \alpha_0}\overline{F}^{\alpha}(s)}{\overline{F}^{\alpha}(s)} = \frac{\partial}{\partial \alpha_0}\log(\overline{F}^{\alpha}(s)) \tag{3.5}$$

So, we get

$$\sup\left\{\left|\frac{\sum_{i=1}^{n-d}\frac{\partial}{\partial \alpha_0}\log f^{\alpha}(z_i)}{n-d} - \frac{\sum_{i=1}^{n-d}E\left(\frac{\partial}{\partial \alpha_0}\log f^{\alpha}(Z_i)\right)}{n-d}\right|\right\} \xrightarrow{P} 0$$

Thus, $\frac{U_1^{**}}{n} \xrightarrow{P} 0$. Now, by using Slutskys theorem, the result follows ($U_1 \xrightarrow{P} 0$). Similarly, we can write

$$U_2^{*} = \sum_{i=1}^{n}\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(w_i)$$

and

$$U_2^{**} = \sum_{i=1}^{n-d}\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(z_i) - (n-d)\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log(\overline{F}(s))$$

We know that, $\frac{U_2^{*}}{n} \xrightarrow{P} -\wp$ and,

$$\frac{U_2^{**}}{n} = \frac{n-d}{n}\left\{\frac{1}{n-d}\left(\sum_{i=1}^{n-d}\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(z_i) - \sum_{i=1}^{n-d}E\left(\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(Z_i)\right)\right)\right\}$$

$$-\frac{1}{n}\left\{(n-d)\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log(\overline{F}(s)) - \sum_{i=1}^{n-d}E\left(\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(Z_i)\right)\right\} \tag{3.6}$$

The first term in (3.6) converges in probability to zero. So, based on (3.5) and after some simplification, we obtain

$$\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log \overline{F}^{\alpha}(s) = \frac{\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\overline{F}^{\alpha}(s)}{\overline{F}^{\alpha}(s)} - \left[E\left(\frac{\partial}{\partial \alpha_0}\log f^{\alpha}(Z)\right)\right]^2 \tag{3.7}$$

and

$$E\left(\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\log f^{\alpha}(Z)\right) = \frac{\frac{\partial^2}{\partial \alpha_0 \partial \alpha'_0}\overline{F}^{\alpha}(s)}{\overline{F}^{\alpha}(s)} - \int_s^{\infty}\left(\frac{\partial}{\partial \alpha_0}\ln f^{\alpha}(z)\right)^2 \frac{f^{\alpha}(z)}{\overline{F}^{\alpha}(s)}d\mu(z) \tag{3.8}$$

Thus, from (3.6), (3.7) and (3.8), we have

$$\frac{1}{n-d}\sum_{i=1}^{n-d}\left\{\frac{\partial^2}{\partial\alpha_0\partial\alpha'_0}\log(\overline{F}(s)) - E\left(\frac{\partial^2}{\partial\alpha_0\partial\alpha'_0}\log f^{\alpha}(Z_i)\right)\right\}$$

$$= \frac{1}{n-d}\sum_{i=1}^{n-d} Var\left(\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(Z_i)\right) = B_i^*;\ i = 1,2,3,4$$

where, $B_i^*(for\ i = 1,2,3,4)$ converge to some bounded value, say $(\xi \in (\xi_{\breve{d}}, \xi_r, \xi_{d_2}, \xi_k))$. Thus, $-\frac{U_2^{**}}{n} \xrightarrow{P} (1-\breve{p}_i)\xi$, and combining these results gives, $U_2 = \frac{1}{n}(U_2^* - U_2^{**}) \xrightarrow{P} -J_{f_{UHCS}}$, where

$$J_{f_{UHCS}} \equiv \wp + \left(1-\breve{p}_i\right)\xi \tag{3.9}$$

Now, from (3.2) and (3.3), we have

$$\sqrt{nJ_{f_{UHCS}}}(\hat{\alpha}_n - \alpha_0) = \frac{\sqrt{n}U_1/\sqrt{J_{f_{UHCS}}}}{-U_2/J_{f_{UHCS}}}$$

$$= \frac{\left(nJ_{f_{UHCS}}\right)^{-1/2}\left(\sum\limits_{i=1}^{n}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(w_i)) - \sum\limits_{i=1}^{n-d}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z_i) + (n-d)\frac{\partial}{\partial\alpha_0}\log(\overline{F}(s))\right)}{-U_2/J_{f_{UHCS}}} \tag{3.10}$$

where, $-U_2/J_{f_{UHCS}} \xrightarrow{P} 1$. So, it suffices to show that the numerator is asymptotically $N(0,1)$. Using (3.5) and Slutsky Theorem, we have

$$\frac{\sqrt{n-d}}{\sqrt{n}}\left\{\frac{1}{\sqrt{n-d}}\left(\sum_{i=1}^{n-d}\varpi(z_i;\alpha) - \sum_{i=1}^{n-d}E(\varpi(Z_i;\alpha))\right)\right\} \xrightarrow{D} N\left(0,(1-\breve{p}_i)\xi\right). \tag{3.11}$$

Now, using Slutsky theorem again, we obtain,

$$\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(w_i),\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n-d}\{\varpi(z_i;\alpha) - E(\varpi(Z_i;\alpha))\}\right] \xrightarrow{D} (C_1, C_2),$$

where, $\varpi(z;\alpha) = \frac{\partial}{\partial\alpha_0}\log f^{\alpha}(z)$ and $C_1 = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(w_i) \sim N(0,\wp)$ and $C_2 \sim N\left(0,(1-\breve{p}_i)\xi\right)$ are independent. Now, using the Continuous Mapping Theorem and (3.10) and (3.11), we conclude that

$$\sum_{i=1}^{n-d}\frac{\partial}{\partial\alpha_0}\log f^{\alpha}(y_i) + (n-d)\frac{\partial}{\partial\alpha_0}\log(\overline{F}(s)) \xrightarrow{D} N(0,\ \wp + (1-\breve{p}_i)\xi)$$

and the proof is complete. □

In the previous theorem based on the unified hybrid censored data, we proved that $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ =O$_p$(1). So, for mis-specified models, we can conclude that, $\sqrt{n}(\hat{\alpha}_n - \alpha_*)$ =O$_p$(1), (see Vuong, 1989). Now based on the following lemma, we consider the asymptotic distribution of the difference quasi-log-likelihood function for mis-specified models.

Remark 1. Suppose that $Y_1, ..., Y_d$ are distributed as the order statistics of a random sample of size d from truncated distribution at s by probability density function (pdf) h^*. Now, if $\frac{r}{n} \to p$ and $\frac{k}{n} \to p*$ as $n \to \infty$ such that $Y_r \xrightarrow{P} \zeta_p$ and $Y_k \xrightarrow{P} \zeta_{p*}$, the p^{th} and $p*^{th}$ percentile of true distribution respectively, then from Vuong (1989) and the property of Continuous Mapping, we have

$$\frac{1}{n}\sum_{i=1}^{d} \log \frac{f^{\hat{\alpha}_n}(y_i)}{g^{\hat{\beta}_n}(y_i)} \xrightarrow{P} \breve{p}E_{h*}\left[\log \frac{f^{\alpha_*}(Y)}{g^{\beta_*}(Y)}\right];$$

and

$$\frac{1}{n}(n-d)\log \frac{\overline{F}^{\hat{\alpha}_n}(s)}{\overline{G}^{\hat{\beta}_n}(s)} \xrightarrow{P} \left(1-\breve{p}\right)\log \frac{\overline{F}^{\alpha_*}(\breve{\zeta}_p)}{\overline{G}^{\beta_*}(\breve{\zeta}_p)}$$

where, for simplicity, we replace $\breve{p}_i;\ i = 1, ..., 4$ by $\breve{p}$ through the paper and

$$\breve{\zeta}_p = \begin{cases} \zeta_{F(T_1)}; & if \quad F(T_1) > p \\ \zeta_p; & if \quad p* < F(T_1) < p < F(T_2) \\ \zeta_{F(T_2)}; & if \quad p* < F(T_1) < F(T_2) < p \\ \zeta_p; & if \quad F(T_1) < p* < p < F(T_2) \\ \zeta_{F(T_2)}; & if \quad F(T_1) < p* < F(T_2) < p \\ \zeta_{p*}; & if \quad F(T_1) < F(T_2) < p* < p \end{cases}$$

Then the difference quasi-log-likelihood function of two rival models converges in probability as below:

$$\frac{1}{n}L_n^{f/g}\left(\hat{\alpha}_n, \hat{\beta}_n\right) \xrightarrow{P} \left\{\breve{p}E_{h*}\left[\log \frac{f^{\alpha_*}(Y)}{g^{\beta_*}(Y)}\right] + \left(1-\breve{p}\right)\log \frac{\overline{F}^{\alpha_*}(\breve{\zeta}_p)}{\overline{G}^{\beta_*}(\breve{\zeta}_p)}\right\}$$

where,

$$\alpha_* = \arg\max\left\{\breve{p}E_{h*}\left[\log f^{\alpha}(Y)\right] + \left(1-\breve{p}\right)\log \overline{F}^{\alpha}(\breve{\zeta}_p)\right\}$$

$$\beta_* = \arg\max\left\{\breve{p}E_{h*}\left[\log g^{\beta}(Y)\right] + \left(1-\breve{p}\right)\log\overline{G}^{\beta}(\breve{\zeta}_p)\right\}$$

are pseudo-true values of α and β, respectively. Also quasi-maximum likelihood estimator of α say $\hat{\alpha}_n$, can be obtained as a solution of $\frac{\partial}{\partial\alpha}L_n^f(\alpha) = 0$.

Theorem 3.2 (Asymptotic Distribution of the $L_n^{f/g}(\hat{\alpha}_n, \hat{\beta}_n)$ Statistic). *Under suitable regularity conditions of Vuong (1989), suppose that the proposed model is mis-specified and $f^{\alpha*} \neq g^{\beta*}$, then,*

$$\sqrt{n}\left(\frac{1}{n}L_n^{f/g}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \breve{p}E_{h*}\left[\log\frac{f^{\alpha_*}(Y)}{g^{\beta_*}(Y)}\right] - \left(1-\breve{p}\right)\log\frac{\overline{F}^{\alpha_*}(\breve{\zeta}_p)}{\overline{G}^{\beta_*}(\breve{\zeta}_p)}\right)$$

$$\xrightarrow{D} N\left(0, \omega^2_{*UHCS}\right) \tag{3.12}$$

where,

$$\omega^2_{*UHCS} = Var_h\left(\log\frac{f^{\alpha_*}(W)}{g^{\beta_*}(W)}\right) + (1-\breve{p})Var_{h*}\left(\log\frac{f^{\alpha_*}(Z)}{g^{\beta_*}(Z)}\right) \tag{3.13}$$

and $w = (w_1, ..., w_n)$, $z = (z_1, ..., z_{n-d})$ are defined as before in (3.3).

Proof. From the Taylor expansion of $L_n^f(\alpha_*)$ around $\hat{\alpha}_n$, we can write

$$L_n^f(\hat{\alpha}_n) = L_n^f(\alpha_*) + \frac{n}{2}(\hat{\alpha}_n - \alpha_*)' U_2(\hat{\alpha}_n - \alpha_*) + o_p(1)$$

and

$$L_n^g(\hat{\beta}_n) = L_n^f(\beta_*) + \frac{n}{2}(\hat{\beta}_n - \beta_*)' \breve{U}_2(\hat{\beta}_n - \beta_*) + o_p(1)$$

where, U_2 is defined as before for rival model $f^{\alpha}(y)$ and similarly $\breve{U}_2$ is given by

$$\breve{U}_2 = \frac{1}{n}\left\{(\sum_{i=1}^{n}\frac{\partial^2}{\partial\beta\,\partial\beta'}\log g^{\beta}(y_i)) - \sum_{i=1}^{n-d}\frac{\partial^2}{\partial\beta\,\partial\beta'}\log g^{\beta}(z_i)) \right.$$

$$\left. +(n-d)\frac{\partial^2}{\partial\beta\,\partial\beta'}\log(\overline{F}(s))\right\}$$

Thus,

$$L_n^{f/g}(\hat{\alpha}_n, \hat{\beta}_n) = L_n^{f/g}(\alpha_*, \beta_*) + \frac{n}{2}(\hat{\alpha}_n - \alpha_*)' U_2(\hat{\alpha}_n - \alpha_*)$$

$$-\frac{n}{2}(\hat{\beta}_n - \beta_*)' \breve{U}_2(\hat{\beta}_n - \beta_*) + o_p(1)$$

From (3.9), $U_2 \xrightarrow{P} -J_{f_{UHCS}}$. Similarly, $\breve{U}_2 \xrightarrow{P} -J_{g_{UHCS}}$. Also, it is known that, $\sqrt{n}(\hat{\alpha}_n - \alpha_*)$ and $\sqrt{n}(\hat{\beta}_n - \beta_*)$ are $O_p(1)$. So, we have

$$\sqrt{n}\left(\frac{1}{n}L_n^{f/g}\left(\hat{\alpha}_n, \hat{\beta}_n\right) - \breve{p}E_{h*}\left[\log\frac{f^{\alpha_*}(Y)}{g^{\beta_*}(Y)}\right] - \left(1-\breve{p}\right)\log\frac{\overline{F}^{\alpha_*}(\breve{\zeta}_p)}{\overline{G}^{\beta_*}(\breve{\zeta}_p)}\right)$$

$$= \sqrt{n}\left\{\frac{1}{n}L_n^{f/g}\left(\alpha_*, \beta_*\right) - \breve{p}E_{h*}\left[\log\frac{f^{\alpha_*}(Y)}{g^{\beta_*}(Y)}\right] - \left(1-\breve{p}\right)\log\frac{\overline{F}^{\alpha_*}(\breve{\zeta}_p)}{\overline{G}^{\beta_*}(\breve{\zeta}_p)}\right\} + o_p(1)$$

However, from the multivariate central limit theorem, the first term in the right hand side converges in distribution to $N(0, \omega^2_{*UHCS})$. It now suffices to show that

$$\omega^2_{*UHCS} = Var_h\left(\log\frac{f^{\alpha_*}(W)}{g^{\beta_*}(W)}\right) + (1-\breve{p})Var_{h_{1*}}\left(\log\frac{f^{\alpha_*}(Z)}{g^{\beta_*}(Z)}\right).$$

From (3.3) we can write

$$\omega^2_{*UHCS} = \frac{1}{n}Var\left(\sum_{i=1}^{d}\log\frac{f^{\alpha_*}(Y_i)}{g^{\beta_*}(Y_i)} + (n-d)\log\frac{\overline{F}^{\alpha_*}(s)}{\overline{G}^{\beta_*}(s)}\right)$$

$$= \frac{1}{n}Var\left[\left(\sum_{i=1}^{n}\log\frac{f^{\alpha_*}(W_i)}{g^{\beta_*}(W_i)} - \sum_{i=1}^{n-d}\log\frac{f^{\alpha_*}(Z_i)}{g^{\beta_*}(Z_i)} + (n-d)\log\frac{\overline{F}^{\alpha_*}(s)}{\overline{G}^{\beta_*}(s)}\right)\right]$$

Now, If, $\frac{n-d}{n} \to 1-\breve{p}$ as $n \to \infty$ such that $s \to \breve{\xi}_p$ in probability, then using Continuous Mapping Theorem

$$\omega^2_{*UHCS} = Var_h\left(\log\frac{f^{\alpha_*}(W)}{g^{\beta_*}(W)}\right) + (1-\breve{p})Var_{h_{1*}}\left(\log\frac{f^{\alpha_*}(Z)}{g^{\beta_*}(Z)}\right)$$

Hence, we propose the following statistic:

$$\hat{\omega}^2_{UHCS} = \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)\right)^2$$

$$+(\frac{n-d}{n})\left[\frac{1}{n-d}\sum_{i=1}^{n-d}\left(\log\frac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)^2 - \left(\frac{1}{n-d}\sum_{i=1}^{n-d}\left(\log\frac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)\right)^2\right] \tag{3.14}$$

□

4 Tracking Interval for a Difference of KL Divergences

In this section, we propose the tracking interval for $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) = EKL(h, f^{\hat{\alpha}_n}) - EKL(h, g^{\hat{\beta}_n})$, which should contain the difference of risks with a given probability. Voung test is useful for testing the two non-nested models, but the confidence intervals are equivalent to encapsulating the results of many hypotheses tests. So, we propose an interval, say tracking interval, for comparing the rival models which contain the acceptable hypotheses. This interval is not a usual confidence interval because $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ changes with n. Although it converges toward $\Delta_{UHCS}(f^{\alpha_*}, g^{\beta_*})$, we wish to approach $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ for values of n for which the Akaike correction is not negligible. We can say that the expected KL risk, $EKL(h, f^{\hat{\alpha}_n})$, is the sum of the mis-specification risk $KL(h, f^{\alpha_*})$ plus the statistical risk $\frac{1}{2\text{n}}\, Tr(\Sigma_f \Omega_f^{-1})$ as (Linhart and Zucchini, 1986):

$$EKL\ (h, f^{\hat{\alpha}_n}) = KL\ (h, f^{\alpha_*}) + \frac{1}{2\text{n}}\ Tr\ (\Sigma_f \Omega_f^{-1})\Big|_{\alpha_*} + o(n^{-1})$$

where, $\Sigma_f = E\left(\frac{\partial \log f^{\alpha}(Y)}{\partial \alpha}.\frac{\partial \log f^{\alpha}(Y)}{\partial \alpha'}\right)$ and $\Omega_f = E\left(\frac{\partial^2 \log f^{\alpha}(Y)}{\partial \alpha \partial \alpha'}\right)$. Note that if (f) is well-specified, we have $KL(h, f^{\alpha_*}) = 0$, $EKL\ (h, f^{\hat{\alpha}_n}) = \frac{p}{2n} + o(n^{-1})$ and $\Sigma_f = \Omega_f = \wp$. Also, based on Commenges *et al.* (2008), we have

$$EKL(h, f^{\hat{\alpha}_n}) = KL(f^{\alpha_*}, h) + \frac{1}{2n} Tr(\Sigma_f \Omega_f^{-1}) + o_p(n^{-1})$$

After estimating $E_h(\log\ f^{\alpha_*}(X))$ by $E_h(\frac{1}{n} L_n^f(\hat{\alpha}_n))$, we can write,

$$EKL(h, f^{\hat{\alpha}_n}) = -E_h(n^{-1} L_n^f(\hat{\alpha}_n)) + F(h) + \frac{1}{n} Tr(\Sigma_f \Omega_f^{-1}) + o_p(n^{-1}) \qquad (4.1)$$

Here, because of the overestimation bias, the factor 1/2 in the last term disappears. Akaike criterion follows from (4.1) by multiplying by $2n$, deleting the constant term, $F(h)$, which we cannot estimate, and replacing the expected value of the normalized version of maximized likelihood function by its empirical version. Thus, we can estimate the difference of risks $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ as:

$$\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) = E_h \left\{ -\frac{1}{n} \left[L_n^{f/g}\ (\hat{\alpha}_n, \hat{\beta}_n) - Tr(\Sigma_f \Omega_f^{-1}) + Tr(\Sigma_g \Omega_g^{-1}) \right] \right\}$$

Thus, using the Akaike approximation, $Tr(\Sigma_f \Omega_f^{-1}) \approx p$, the simple estimator of $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ is

$$D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) = \frac{1}{2n} \left[AIC(f^{\hat{\alpha}_n}) - AIC(g^{\hat{\beta}_n}) \right]$$

$$= -\frac{1}{n}\left[\, L_n^{f/g}\,(\hat{\alpha}_n, \hat{\beta}_n) - (p-q)\,\right]$$

where, p and q are the number of parameters in two rival models. Also

$$E_h\left[D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) - \Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})\right]$$

is an $o(n^{-1})$. Thus, in contrast with AIC, $D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ has an interpretation since its expectation tracks the quantity of main interest $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ with a pretty good accuracy. Now, we emphasis on the case where $f^{\alpha_*} \neq g^{\beta_*}$. Thus using Theorem 2, we have

$$n^{1/2}\left(D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) - \Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})\right) \xrightarrow{d} N(0, \omega^2_{UHCS})$$

From this the tracking interval for $\Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$ is given by

$$[A_n, B_n] \tag{4.2}$$

where,

$$\mathrm{A_n} = D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) - n^{-1/2}\, z_{\alpha/2}\, \hat{\omega}_{\,UHCS}$$

$$\mathrm{B_n} = D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) + n^{-1/2}\, z_{\alpha/2}\, \hat{\omega}_{\,UHCS}.$$

This interval has the property

$$P_h\left[\, \mathrm{A_n} < \Delta_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) < \mathrm{B_n}\,\right] \to 1-\alpha$$

where P_h represents the probability with density h. Tracking interval helps us to evaluate proposed models in comparison with each other. In other words, if the calculated distance includes zero, it can be concluded that based on the predetermined confidence, both proposed models are equivalent.

5 Monte Carlo Simulations and Data Analysis

5.1 Simulations

In this section, we present some simulation results to examine the behavior of the two rival models using the tracking interval for different choices of k, r, T_1 and T_2 values. All the programs are written in *R*. we consider an *i.i.d.* sample of size n from Weibull distribution $W(\theta = 2, \lambda = 1)$ which plays the role of the true distribution h. Recently, Gupta and Kundu (2003) observed that the generalized exponential (GE) and Weibull (W) distributions provide a very similar data fit. So, we consider the two non-nested rival models as W (say g), $g_W^{\beta=(\theta,\lambda)} = \theta\lambda x^{\lambda-1}e^{-\theta x^{\lambda}}$, and GE (say f), $f_{GE}^{\alpha=(p,b)} = pbe^{-bx}(1-e^{-bx})^{p-1}$, where g and f are the well-specified and mis-specified models respectively. Since model (g) is well-specified, we know that $g^{\beta_0=(\theta_0,\lambda_0)}(.) = h$, that the miss-specification error $KL(h, g^{\beta_0=(\theta_0,\lambda_0)})$ is zero and that $Tr(\Sigma_g\Omega_g^{-1}) \approx 2$. As for the model ($f$), we must compute the quantities of interest by simulation. We generate 10^4 Monte-Carlo data-sets of size n=50 from a $W(\theta = 2, \lambda = 1)$ and then find the closest rival model to the true model. We consider different k, r, T_1 and T_2 values. For each case, we estimate the unknown parameters of different rival models using the maximum likelihood method under UHCS. Then we compute the $D_{UHCS}(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n})$, $\hat{\omega}_{UHCS}$ and construct a 0.95 tracking interval from (4.2). The results are reported in Tables 1 and 2 respectively. From the Tables 1 and 2 the following general observations can be made. From Table 1, it is observed that for r=19, k=11 and T_1=0.5, the tracking interval contain zero, which indicates that the W and GE are equal or observationally equal to estimate the true model and for other values of r, k, T_1, as expected, both limits of tracking intervals are positive, which indicates that the W is better than the GE distribution to estimate the true model. Note that, we say one model is better than the other one when the tracking interval does not contain zero. Similarly, from Table 2, it is clear that the W is better than the GE distribution to estimate the true model as r increases from 19 for all values of r, k, T_2. Moreover, for fixed r, k, T_2, the length of the tracking interval decrease as T_1 increases. Similarly, for fixed r, k, T_1, it is observed that as T_2 increases the length of the tracking interval decreases, *i.e.* the distance between them increases. Similarly, we obtain the results when the two rival models are $GE(\hat{p}, \hat{b})$ and $W(\theta = 0.5, \lambda = 1)$. In this case we choose the same sample size n and k, r, T_1 and T_2 values. We plot the fitted probability distribution function and the relative histogram for the above rival models under (T_2=3,T_1=1.5, r=19,k=33) and (T_2=3.5,T_1=0.25, r=46,k=43) in Figure 5.1 and Figure 5.1 respectively (similar results have been observed for other cases). From Figure 5.1 and Figure 5.1, it is observed that, $W(\theta = 0.5, \lambda = 1)$ is not a preferable fitted model. Therefore, it is expected that the GE will be chosen as the closest model to the true model by tracking interval. The results are reported in Tables 3 and 4 respectively. From Tables 3 and 4, it is quite clear that the GE is better than the postulated Weibull density to estimate the true model in almost all cases.

From the simulation study, it is recommended that the tracking interval can be used quite effectively even when r and k are small for all possible choices of T_1 and T_2 values. So, this study justifies the use of this interval to compare the two rival models under unified hybrid censored samples.

Table 1: Choice between $GE(\hat{p}, \hat{b})$ and $W(\hat{\theta}, \hat{\lambda})$ models using tracking interval.

$n = 50$ $T_2 = 3$					
r	k		T_1		
			0.5	1.5	2.5
19	11	Lower	-0.1379899	0.03562426	0.02190979
		Upper	1.5273790	0.07671079	0.03343329
		Length	1.6653689	0.04108653	0.01152350
33	19	Lower	0.1569406	0.03681046	0.02256008
		Upper	0.3387795	0.07609499	0.03431710
		Length	0.1818389	0.03928453	0.01175702
33	27	Lower	0.1544227	0.03486241	0.02287221
		Upper	0.3385860	0.07602677	0.03468798
		Length	0.1841633	0.04116437	0.01181578
33	31	Lower	0.1529401	0.03505620	0.02238116
		Upper	0.3433301	0.07569645	0.03411655
		Length	0.1903900	0.04064025	0.01173539
46	21	Lower	0.02485793	0.02390466	0.01850428
		Upper	0.05416066	0.05103080	0.03027675
		Length	0.02930273	0.02712614	0.01177246
46	43	Lower	0.02394826	0.02301148	0.02347404
		Upper	0.05455199	0.05056110	0.03603083
		Length	0.03060373	0.02754962	0.01255679

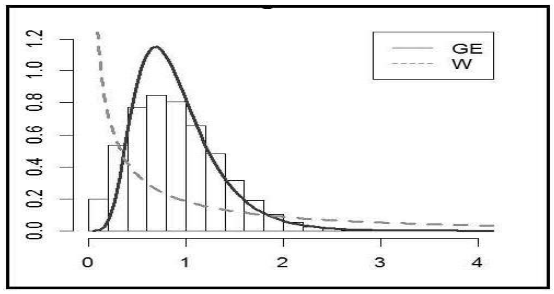

Figure 1: The two fitted rival models for $T_2 = 3, T_1 = 1.5$, r=19,k=33, Weibull (dashed line) and GE (solid line).

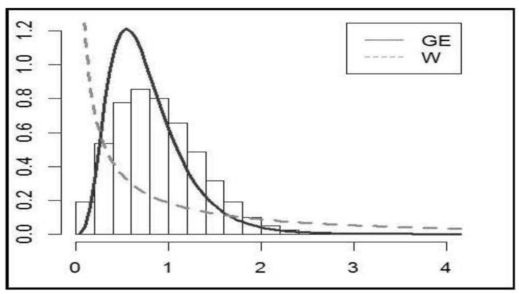

Figure 2: The two fitted rival models for $T_2 = 3.5, T_1 = 0.25$, r=46,k=43, Weibull (dashed line) and GE (solid line).

5.2 Data Analysis

Data-Set 1: The data-set consists of the main component of spearmint essential oil purification obtained by the experimental pilot plant. For illustrative purpose, we will be considering the purification of the main component of spearmint essential oil with sample size $n = 35$ (see Panahi and Sayyareh, 2014). Before progressing further, we have first fitted the Burr XII distribution to the complete data-set. The q-q plot and the fitted probability distribution function (pdf) and the relative histogram of this data are

Table 2: Choice between rival models $GE(\hat{p}, \hat{b})$ and $W(\hat{\theta}, \hat{\lambda})$ using tracking interval.

			$n = 50$ $T_1 = 0.25$		
r	k		T_2		
			0.7	1.75	3.5
19	11	Lower	-2.485056	-0.2635314	-0.040380
		Upper	4.646851	1.6412690	1.422563
		Length	7.131907	1.9048004	1.462943
33	19	Lower	0.2547185	0.1501532	0.1559652
		Upper	1.0866718	0.3473781	0.3382471
		Length	0.8319533	0.1972249	0.1822819
33	27	Lower	0.2077799	0.1569685	0.1569654
		Upper	0.6077219	0.3386884	0.3379971
		Length	0.3999420	0.1817199	0.1810317
33	31	Lower	0.1804022	0.1553873	0.1577691
		Upper	0.4111724	0.3404237	0.3351119
		Length	0.2307701	0.1850364	0.1773428
46	21	Lower	0.2532064	0.02431987	0.02422751
		Upper	0.9614531	0.05485407	0.05302312
		Length	0.7082466	0.03053419	0.02879561
46	43	Lower	0.04456699	0.02662280	0.02323126
		Upper	0.09242408	0.05551849	0.05186485
		Length	0.04785709	0.02889569	0.02863359

Table 3: Choice between $GE(\hat{p}, \hat{b})$ and $W(0.5, 1)$ models using tracking interval.

$n = 50$ $T_2 = 3$					
r	k		T_1		
			0.5	1.5	2.5
19	11	Lower	-0.2402059	-0.7639319	-0.9119689
		Upper	0.8269709	-0.5969956	-0.7761555
		Length	1.0671767	0.1669363	0.1358134
33	19	Lower	-0.4397094	-0.7702901	-0.9151429
		Upper	-0.1385257	-0.6030702	-0.7819565
		Length	0.3011838	0.1672199	0.1331864
33	27	Lower	-0.4382440	-0.7599943	-0.9112897
		Upper	-0.1450885	-0.5942433	-0.7785755
		Length	0.2931556	0.1657510	0.1327142
33	31	Lower	-0.4339584	-0.7658978	-0.9113656
		Upper	-0.1438790	-0.6024916	-0.7866310
		Length	0.2900794	0.1634061	0.1247346
46	21	Lower	-0.8106185	-0.8214801	-0.9140763
		Upper	-0.6540409	-0.6667408	-0.7782674
		Length	0.1565776	0.1547394	0.1358088
46	43	Lower	-0.8224319	-0.8127864	-0.9189451
		Upper	-0.6692899	-0.6600130	-0.7851536
		Length	0.1531420	0.1527734	0.1337915

Table 4: Choice between $GE(\hat{p}, \hat{b})$ and $W(0.5, 1)$ models using tracking interval.

$n = 50$ $T_1 = 0.25$					
r	k		T_2		
			0.7	1.75	3.5
19	11	Lower	-0.2055429	-0.2539600	-0.2390979
		Upper	0.9700580	0.8553028	0.8154895
		Length	1.1756008	1.1092628	1.0545874
33	19	Lower	-0.1765929	-0.4374647	-0.4253552
		Upper	0.6783682	-0.1436175	-0.1373006
		Length	0.8549611	0.2938472	0.2880546
33	27	Lower	-0.2835752	-0.4285991	-0.4309602
		Upper	0.1806446	-0.1356765	-0.1461760
		Length	0.4642198	0.2929226	0.2847842
33	31	Lower	-0.38126949	-0.4324393	-0.4294791
		Upper	-0.03669811	-0.1409502	-0.1421811
		Length	0.34457139	0.2914891	0.2872980
46	21	Lower	-0.1958578	-0.7992035	-0.8070888
		Upper	0.5593171	-0.6415557	-0.6501842
		Length	0.7551749	0.1576478	0.1569047
46	43	Lower	-0.7212354	-0.8130397	-0.8124846
		Upper	-0.5503377	-0.6537241	-0.6560068
		Length	0.1708977	0.1593156	0.1564778

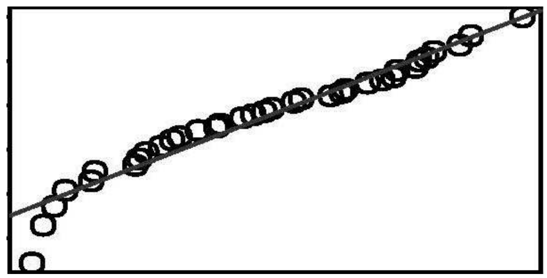

Figure 3: The q-q plot of data-set 1.

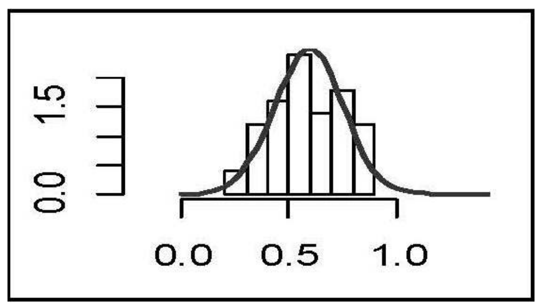

Figure 4: The fitted pdf and the relative histogram for data-set 1.

presented in Figure 5.1 and Figure 5.1 respectively.

These plots show a strong relationship supporting the appropriateness of the Burr XII distribution. For comparison purposes, we fit Burr XII (BXII), Weibull (W), generalized exponential (GE) and Burr III (BIII) distributions to the complete observation. The estimated parameter values, AIC values, Kolmogorov-Smirnov (K-S) distances and the corresponding p-values are presented in Table 5.
From the K-S distances, AIC values and p-values of Table 5, it is quite clear that the Burr XII model with estimated parameters $p = 3.755 \times 10^2$ and $b = 10.3203$ provides much better fit than other distributions. The BXII was originally proposed by Burr (1942) and received more attention by the researchers due to its broad applications in the study of engineering, reliability, life testing, and several industrial and economic experiments. See for example, Ali Mousa and Jaheen (2002); Wu and Yu (2005); Rastogi and Tripathi (2012).

Now, we want to decide which of the two rival models is closer to the true model

Table 5: Estimated parameters, K-S distances and AIC values for different distribution functions of data-set 1.

Distribution	Estimated parameters	K-S (p-value)	AIC
W	$p = 3.719 \times 10^2, b = 10.3074$	0.0926 (0.8978)	-94.0252
BXII	$p = 3.755 \times 10^2, b = 10.3203$	0.0924 (0.8987)	-94.0265
BIII	$p = 5927 \times 10^{-5}, b = 26.844$	0.4994 (small)	-2.49693
GE	$p = 1.179 \times 10^4, b = 18.489$	0.109 (0.760)	-92.0116

$f_W^{(p,b)} = pbx^{b-1}e^{-px^b}$, $f_{BXII}^{(p,b)} = pbx^{b-1}(1+x^b)^{-p-1}$, $f_{BIII}^{(p,b)} = pbx^{-b-1}(1+x^{-b})^{-p-1}$ and $f_{GE}^{(p,b)} = pbe^{-bx}(1-e^{-bx})^{p-1}$.

(Burr XII) of this data. For this purpose, we assume that the true model is unknown and compare two rival models using tracking intervals. We assume the following six different cases of censoring schemes:
Case 1: T_1=0.56, T_2=0.61, k=17, r=20.
Case 2: T_1=0.56, T_2=0.61, k=17, r=25.
Case 3: T_1=0.56, T_2=0.61, k=17, r=33.
Case 4: T_1=0.53, T_2=0.61, k=21, r=29.
Case 5: T_1=0.53, T_2=0.60, k=23, r=32.
Case 6: T_1=0.53, T_2=0.59, k=28, r=33.

For all cases of censoring schemes, we consider two different cases of rival models:
A: W and GE distributions (two mis-specified models).
B: BXII and W distributions (BXII and W are well-specified and miss-specified models respectively).

In all the cases we have estimated the unknown parameters using the MLEs and then constructed the tracking intervals. The results are reported in Table 6.

First, we consider W and GE distributions as rival models (case A). For cases 2,3,4,5 and 6, it is observed that zero is well inside these intervals, so there is no confidence that we incur a lower risk using W rather than GE distribution. It is not surprising because the AICs of W and GE are very similar for cases 2, 3, 4, 5 and 6. But case 1, shows that both limits of the tracking intervals are negative, which indicates that the W is better than the GE density to estimate the true model for this data. As expected

Table 6: Tracking intervals for two rival models (A and B) and six censoring schemes (six cases) of data-set 1. The first, second and third rows represent the lowers, uppers and the corresponding lengths.

Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Case A					
-0.5690	-0.2610	-0.0485	-0.0863	-0.1170	-0.1200
-0.0881	0.00354	0.0384	0.0574	0.0551	0.0443
0.4809	0.2645	0.0869	0.1437	0.1721	0.1643
Case B					
-1.25038	-1.51952	-2.13109	-1.82189	-1.71994	-1.75670
-1.16977	-1.45008	-2.04956	-1.75577	-1.65886	-1.68663
0.08061	0.06944	0.08153	0.06612	0.06108	0.07007

for case B, both limits of the tracking intervals are negative, which indicates that the BXII is better than the W density to estimate the true model for all cases. Furthermore, one of the most important applications of the tracking interval is selecting the closer model as a preferred model which is important when the two rival models are very close. Therefore, we present the following data-set.

Data-Set 2: In this example we provide another data analysis for more illustrative purposes. The data have been taken from Murthy *et al.* (2004) and it represents the lifetimes of 20 electronic components. Teimouri and Gupta (2013) observed that three-parameter Weibull distribution works quite well for this data. The data are given below:

0.03 0.12 0.22 0.35 0.73 0.79 1.25 1.41 1.52 1.79

1.80 1.94 2.38 2.40 2.87 2.99 3.14 3.17 4.72 5.09

We fit Burr XII (BXII), Weibull (W), inverse Weibull (IW) and generalized exponential (GE) distributions to the complete observations. The plot of the empirical and the fitted cumulative distribution functions for different distributions are presented in Figure 5.2. The estimated parameter values, AIC values, Kolmogorov-Smirnov (K-S) distances and the corresponding p-values are presented in Table 7. From minimum Kolmogorov distance, minimum AIC value and high p-value, the W distribution function appears

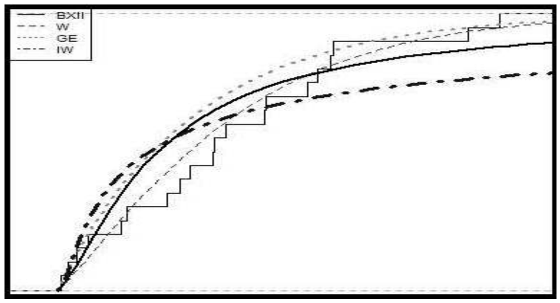

Figure 5: Empirical survival function and the fitted survival functions for data-set 2.

Table 7: Estimated parameters, K-S distances and AIC values for different distribution functions of data-set 2.

Distribution	Estimated parameters	K-S (p-value)	AIC
W	p= 0.42616, b= 1.19617	0.1268 (0.8652)	69.5739
BXII	p= 0.86983, b= 1.45335	0.2302 (0.2053)	77.4807
IW	p= 0.71570, b= 0.62782	0.2368 (0.1803)	87.4368
GE	p= 1.13901, b= 0.55959	0.1573 (0.6488)	70.2204

$$f_W^{(p,b)} = pbx^{b-1}e^{-px^b}, f_{BXII}^{(p,b)} = pbx^{b-1}(1+x^b)^{-p-1}, f_{IW}^{(p,b)} = pbx^{-b-1}e^{-px^{-b}} \text{ and}$$
$$f_{GE}^{(p,b)} = pbe^{-bx}(1-e^{-bx})^{p-1}.$$

to be a more appropriate statistical distribution function in complete case. We also consider a graphical method based on total time on test (TTT) transform. This provides a very good idea about the shape of the hazard function of a distribution. It has been shown that the hazard function of $F(t)$ increases (decreases) if the scaled TTT transform, $\varphi_F(t) = H_F^{-1}(t)/H_F^{-1}(1)$, where $H_F^{-1}(t) = \int_0^{F^{-1}(t)} S(u)du;\ 0 \le u \le 1$, is concave (convex). In addition, for bathtub (unimodal) shaped hazard rate, the TTT transform is first convex (concave) and then concave (convex). The plot of the scaled TTT transform of this data-set, Figure 5.2, indicates that the empirical hazard function is increasing and therefore, the W and GE distributions can be used to analyze the data. So, we consider W (say f) and GE (say g) as rival models. First, we obtain the tracking interval of two rival models for complete data. This interval is (-0.0190596, -0.0131105). It implies that the W is better than the GE density to estimate the true model (three-parameter Weibull). But,

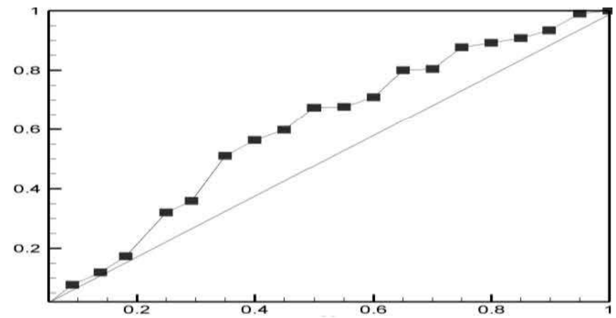

Figure 6: Scaled TTT transform of the lifetimes of 20 electronic components.

in this case the length of the tracking interval is small (as we expected). This indicates that the W and GE distributions are similar in information criteria sense. Therefore, even if the two CDFs are close, the tracking interval works well. Now we consider six different cases of censoring schemes:

Case 1: T_1=1.9, T_2=2.8, k=7, r=10.
Case 2: T_1=1.9, T_2=2.8, k=7, r=13.
Case 3: T_1=1.9, T_2=2.8, k=7, r=17.
Case 4: T_1=1.7, T_2=2.8, k=11, r=14.
Case 5: T_1=1.7, T_2=2.6, k=11, r=18.
Case 6: T_1=1.7, T_2=2.3, k=16, r=19.

For cases 1, 2, 3, 4, 5 and 6, the tracking intervals are (-0.62290, -0.19374), (-0.37699, -0.163914), (-0.312841, -0.158006), (-0.25213, -0.10619), (-0.381619, -0.167256) and (-0.252133, -0.106190) respectively. Interestingly, for all cases, both limits of the tracking intervals are negative, which indicates that the W is better than the GE density to estimate the true model. Also for all cases, the plots of the fitted probability distribution functions (W and GE) and the relative histogram of this data confirm the results of tracking intervals. (These plots are not reported here).

6 Conclusion

In this paper we have considered the problem of comparing between two rival models using the statistic based on the normalization of a difference of Akaike criteria under unified hybrid censoring scheme. We have also obtained the asymptotic distribution of the maximum likelihood estimator under unified hybrid censoring scheme. It is observed that the asymptotic distribution of the maximum likelihood estimator is asymptotically normal. Moreover, using the missing information principle we calculated the variance of the normalized difference of AIC's for constructing an interval say tracking interval. The proposed interval contains the difference of KL risks with a fixed probability. This interval has another interpretation for the use of AIC's. In fact we are not in a situation to detect the best model but we are in search for a model which has relatively less risk compared to other models. Using a Monte Carlo simulation, we compared the two rival models and it is observed that the tracking intervals work quite well for different censoring schemes. Also it is observed that for fixed r, k and T_1 when T_2 increases and for fixed r, k and T_2 when T_1 increases, the length of tracking interval decreases. The statistic DUHCS and the tracking interval for difference of risks are easy to compute and could be useful in a wide variety of applications. Although it may be mentioned that our interval can be extended for other censoring schemes also. More work is needed in these directions.

Appendix (Acronyms)

$EKL(h, f^{\alpha_*})$: The expected Kullback-Leibler risk(or simply Kullback-Leibler risk).
UHCS (r, k, T_1, T_2): The unified hybrid censoring scheme with parameters r, k, T_1 and T_2.
$f^{\alpha}(z_i \mid Y)$: The conditional probability density function of z_i; $i = 1, ..., n-d$ given observed sample.
$o_p(1)$: Convergence in probability to zero.
$O_p(1)$: Bounded in probability as n goes to infinity.
AIC: Akaike information criterion.
KL: Kullback-Leibler divergence.
h : True model.

References

Akaike, H. (1973). Information Theory and an Extension of Maximum Likelihood Principle. *In: Second International Symposium on Information Theory*, Akademia Kiado, 267-281.

Ali Mousa, MAM. and Jaheen, ZF. (2002). Statistical Inference for the Burr model base on Progressively Censored Data.*Comput Math Appl* , **43**, 14411449.

Balakrishnan, N. and Kundu, D. (2013). Hybrid Censoring: Models, Inferential Results and Applications.*Comput Stat Data Analysis*, **57**, 166-209.

Balakrishnan, N., Rasouli, A. and Sanjari Farsipour, N. (2008). Exact Likelihood Inference based on an Unified Hybrid Censored Sample from the Exponential Distribution. *Journal of Statistical Computation and Simulation*, **78**, 475-788.

Burr, IW. (1942). Cumulative Frequency Functions.*Annals Mathematical Statatistics*, **13**, 215232.

Chandrasekar, B., Childs, A. and Balakrishnan, N. (2004). Exact Likelihood Inference for the Exponential Distribution under Generalized Type-I and Type-II Hybrid Censoring. *Naval Research Logistics*, **51**, 994-1004.

Cramér, H. (1946) *Mathematical methods of statistics*. Princeton University Press, Princeton.

Commenges, D., Sayyareh, A., Letenneur, L., Guedj, J. and Bar- Hen, A. (2008). Estimating a Difference of KullbackLeibler Risks Using a Normalized Difference of AIC. *Annals of Applied Statistics*, **2**, 11231142.

Commenges, D., Liquet, B. and Proust-Lima, C. (2012). Choice of prognostic estimators in joint models by estimating differences of expected conditional KullbackLeibler risks. *Biometrics*, **68**, 380-387.

Cox, DR. (1961). Test of Separate Families of Hypothesis. *In: Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**,105123.

Dey, A.K. and Kundu, D. (2009). Discriminating among the Log-Normal, Weibull and Generalized Exponential distributions. *IEEE Transactions on Reliability*, **58**, 416 - 424.

Gupta, RD. and Kundu, D. (2003). Discriminating between the Weibull and the GE distributions. *Computational Statistics and Data Analysis*, **43**, 179-196.

Habibi Rad, A. and Izanlo, M. (2011). An EM Algorithm for Estimating the Parameters of the Generalized Exponential Distribution under Unified Hybrid Censored Data. *Journal of Statistical Research of Iran*, **8**, 149162.

Hasaballah, M. (2016). Exponentiated Rayleigh Distribution: A Bayes Study Using MCMC Approach Based on Unified Hybrid Censored Data.*Journal of Advances in Mathematics*, **12**, 6863-6880.

Huang, WT. and Yang, KC. (2010). A New Hybrid Censoring Scheme and Some of its Properties.*Tamsui Oxford Journal of Mathematical Science*, **26**, 355-367.

Kullback, S. and Leibler, R. (1951). On Information and Sufficiency.*Annals of Mathematical Statistics*, **22**, 79-86.

Kundu, D., Gupta, R.D. and Manglick, A. (2005), Discriminating Between the Log-Normal and Generalized Exponential Distribution. *Journal of the Statistical Planning and Inference*, **127**, 213-227.

Lin, C.T. and Balakrishnan, N. (2011). Asymptotic Properties of Maximum Likelihood Estimators based on Progressive Type-II Censoring.*Metrika*, **74**, 349360.

Lien, D. and Vuong, Q.H. (1987). Selecting the Best Linear Regression Model: A Classical Approach. *Journal of Economica*, **35**,323.

Linhart, H. and Zucchini, W. (1986). *Model Selection*. Wiley, New York, MR0866144.

Louis, T.A. (1982). Finding the Observed Information Matrix When Using the EM Algorithm.*J R Stat Soc Series B*, **44**, 226233.

Mohie El-Din, M.M., Nagy, M. and Shafa, A.R. (2017). Statistical Inference under Unified Hybrid Censoring Scheme, *Journal of Statistics Applications and Probability*, **6**, 149-167.

Murthy, D.N.P., Xie, M. and Jiang, R. (2004). *Weibull Models*. Wiley, New York.

Panahi, H. (2017). Estimation of the Burr type III distribution with application in unified hybrid censored sample of fracture toughness.*Journal of Applied Statistics*, **44**, 2575-2592.

Panahi, H. (2016). Model Selection Test for the Heavy-Tailed Distributions under Censored Samples with Application in Financial Data.*International Journal o f Financial Studies*, **24**, 1-14.

Panahi, H. and Sayyareh, A. (2014). Parameter Estimation and Prediction of Order Statistics for the Burr Type XII Distribution with Type-II Censoring.*Journal of Applied Statistics*, **11**, 403-414.

Pardo, L. (2005). *Statistical inference based on divergence measures*. CRC press.

Rastogi, M.K. and Tripathi, Y.M. (2012). Estimating the Parameters of a Burr Distribution under Progressive Type II Censoring.*Statistical Methodology*, **9**, 381-391.

Rastogi M.K. and Tripathi Y.M. (2013) Inference on unknown parameters of a Burr distribution under hybrid censoring. *Statistical Papers*, **54**, 619-643.

Sayyareh, A. Obeidi, R. and Bar-Hen, A. (2011). Empirical Comparison of Some Model Selection Criteria.*Communication of Statistical Simulation and Computation*, **40**, 7286.

Sayyareh, A. (2012). Tracking Interval for Selecting between Non-nested Models: An Investigation for Type II Right Censored Data.*Journal of Statistical Planning and Inference*, **142**, 32013208.

Teimouri, M. and Gupta, A.K. (2013). On the Three-Parameter Weibull Distribution Shape Parameter Estimation.*Journal of Data Science*, **11**, 403-414.

Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis. *Econometrica*, **57**,307333.

Vuong, Q.H. and Wang, W. (1993). Minimum Chi-Square Estimation and Tests for Model Selection. *Journal of Economica*, **56**,141168.

Wu, J.W. and Yu, H.Y. (2005). Statistical inference about the shape parameter of the Burr type XII distribution under the failure-censored sampling plan.*Applied of Mathematical Computation*, **163**, 443-482.