

A Sequential Sampling Approach for Discriminating Log-normal, Weibull, and Log-logistic Distributions

Biplab Paul¹, Shyamal K. De¹ and Debasis Kundu²

¹School of Mathematical Sciences, National Institute of Science Education and Research, HBNI, Odisha, India

²Department of Mathematics and Statistic, Indian Institute of Technology Kanpur, Uttar Pradesh, India

Abstract: Log-normal, Weibull, and log-logistic distributions are widely used in modeling nonnegative skewed data. We develop sequential methodologies to discriminate between any two of these three distributions as well as to discriminate between these three distributions. Discriminating three or more distributions having similar shapes often requires large sample size (Marshall et al. (2001)). Sequential procedures allow early stopping which in turn reduce the sample size needed for discrimination. Proposed methods yield high probabilities of correct selection that are shown to converge to 1 asymptotically. Asymptotic behavior of expected sample size and error probabilities are studied as stopping boundaries tend to infinity. Extensive simulation study validates finite sample performances of the proposed procedures requiring significantly fewer samples on average. These methods are applied to three benchmark datasets on cancer trials and are shown to select the correct model with high probability.

Keywords: Error probabilities; Multihypothesis testing; Ratio of maximized likelihood; Probability of correct selection; Sequential testing; Stopping rule

1 Introduction

Log-normal (LN), Weibull (WE), and log-logistic (LL) distributions are frequently used in modeling non-negative valued skewed data appearing in many branches of science such as medical science, reliability theory, environmental studies, hydrology, economics, among others. See, for example, Dey and Kundu (2009b), Lee and Wang (2003), Overduin (2004)

Address correspondence to Biplab Paul, School of Mathematical Sciences, National Institute of Science Education and Research, Odisha 752050, India; E-mails: biplab.paul@niser.ac.in; sde@niser.ac.in; kundu@iitk.ac.in

and Tai et al. (2003). For parameters $\sigma > 0$, $\eta > 0$, $\gamma > 0$, $\theta > 0$, $\xi > 0$, $\lambda > 0$, let us denote the log-normal, Weibull, and log-logistic distributions as $LN(\sigma, \eta)$, $WE(\gamma, \theta)$, and $LL(\xi, \lambda)$ respectively. The probability density functions (PDF) of log-normal, Weibull, and log-logistic distributions are given as:

$$\text{Log-normal: } f_{LN}(x; \sigma, \eta) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \ln \eta)^2} \mathbb{1}_{\{x>0\}} \quad (1)$$

$$\text{Weibull: } f_{WE}(x; \gamma, \theta) = \gamma \theta^\gamma x^{\gamma-1} e^{-(\theta x)^\gamma} \mathbb{1}_{\{x>0\}} \quad (2)$$

$$\text{Log-logistic: } f_{LL}(x; \xi, \lambda) = \frac{1}{\lambda x} \frac{e^{(\ln x - \ln \xi)/\lambda}}{\{1 + e^{(\ln x - \ln \xi)/\lambda}\}^2} \mathbb{1}_{\{x>0\}} \quad (3)$$

respectively. Even though shapes of these distributions are different in general, interestingly, for certain parameter values, PDFs of these distributions can have very similar shapes making it difficult to distinguish between their corresponding distributions. Figure 1(a), illustrates such a case where cumulative distribution functions (CDF) of LN, WE and LL distributions are very close to each other. Although, shapes of these distributions are quite similar, hazard functions corresponding to these distributions may be very different (see Figure 1(b)) which was also observed in Figure 2 of Raqab et al. (2018). Moreover, even though Figure 1(a) shows remarkable similarity in shapes of CDFs, tail parts of these probability distributions may be different. Table 1 presents the lower 1st and 99th percentiles of these three distributions.

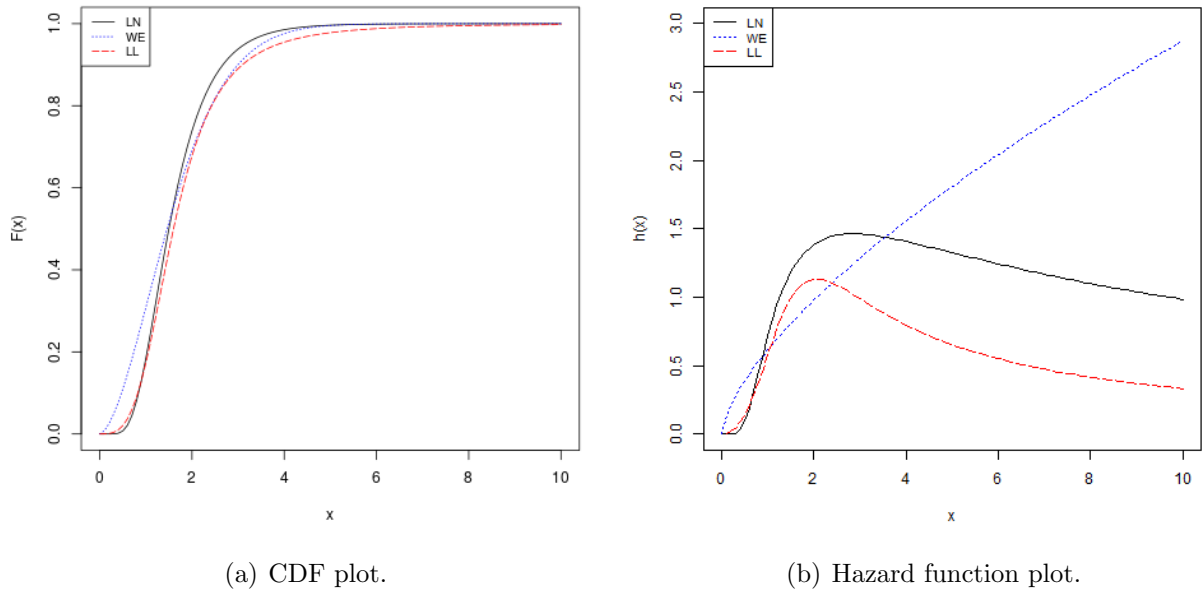


Figure 1: CDF and hazard function plot of $LN(0.45, 1.50)$, $WE(1.67, 0.55)$, and $LL(1.60, 0.30)$.

In this article, we are interested in the following question. If the observed data is known

	$LN(0.45, 1.50)$	$WE(1.67, 0.55)$	$LL(1.60, 0.30)$
1 st percentile	0.5265	0.1157	0.4031
99 th percentile	4.2730	4.5372	6.3505

Table 1: 1st and 99th percentiles of $LN(0.45, 1.50)$, $WE(1.67, 0.55)$ and $LL(1.60, 0.30)$.

to come from log-normal or Weibull or log-logistic distribution, how can we identify the most appropriate distribution that would fit the data best? Correct identification of a probability distribution including its tail parts is very crucial as model misspecification may lead to wrong inference of hazard functions or percentile points or even computing p-values that depend on tail parts of distributions.

Correct model selection from two competing distributions becomes difficult when the distributions have similar shapes and available sample size is small. The problem becomes even more challenging when the correct model is to be selected from three or more similar shaped distributions and available sample size is not large enough. Many authors including Marshall et al. (2001), Dey and Kundu (2009a) and Raqab et al. (2018) discussed the challenges of discriminating three (or more) distributions. For instance, in the context of discriminating gamma, Weibull, log-normal and geometric extreme exponential distributions using maximized likelihood ratio (MLR) criteria, Marshall et al. (2001) shows in Figures 1-6 that the probability of correct selection (PCS) may be unexpectedly low for moderate (for even as large as 400) sample sizes. Similar challenges have also been reported by Raqab et al. (2018) for discriminating Weibull, log-normal and log-logistic distributions and Dey and Kundu (2009a) for discriminating Weibull, log-normal and generalized exponential distributions. This paper makes a modest attempt to overcome some of these issues by adopting a sequential sampling approach that allows early stopping, and thus, allows to have less average sample size without sacrificing PCS.

The literature on discrimination of two or more populations are well developed when the sample size is fixed, hereafter referred as fixed-sample-size procedures. Some notable fixed-sample-size discrimination procedures are given in Marshall et al. (2001), Gupta and Kundu (2003), Gupta and Kundu (2004), Kundu and Manglick (2004), Dey and Kundu (2009a), Dey and Kundu (2009b), Raqab (2013), Elsherpieny et al. (2013), Raqab et al. (2018), among others. To the best of our knowledge, methodologies are yet to be developed when the data is observed sequentially.

The contribution of this article is as follows:

(1) We develop discrimination procedures when observations are collected sequentially or in groups. Fixed-sample-size procedures can be considered as a special case of sequential procedures, and hence, our methods extend the ideas of fixed-sample-size methods to the sequential domain. Unlike fixed-sample-size procedures, sequential sampling offers an addi-

tional flexibility that, in each step, it can decide whether to stop or to continue sampling according to some prefixed boundaries. One of the competing populations is selected at stopping once enough evidence is accrued in its favor.

(2) We show that the proposed procedures require much smaller average sample sizes compared to their fixed-sample-size analogs, without sacrificing their probability of correct selections. Tables 2-9 show that savings in the average sample sizes are mostly quite significant ranging from roughly 12% to 70%. Moreover, Section 4 provides some recommendations on selecting stopping boundaries (see Figure 3(a)) to ensure high PCS of the proposed sequential methods.

(3) We establish theoretical properties of the proposed methodologies, namely (i) sequential sampling is shown to stop in finite time, (ii) PCS is shown to converge to unity as stopping boundaries tend to infinity, (iii) asymptotic behavior of stopping times and expected stopping times are studied, (iv) in the context of discriminating two distributions, probabilities of type I and type II errors are shown to converge to zero.

The paper is organized as follows. Section 2 develops sequential procedures to discriminate between two populations, namely log-normal and Weibull, log-normal and log-logistic, and Weibull and log-logistic. Section 3 develops sequential methods to discriminate between three populations, namely log-normal, Weibull, and log-logistic. Simulation results are presented in Section 4 validating effectiveness of proposed procedures. Three benchmark datasets on survival times of patients suffering from cancer are analyzed in Section 5. In Section 6, we briefly discuss a general sequential procedure for discriminating among $M(> 1)$ number of distributions. Finally, we provide some concluding remarks and future research directions in Section 7.

2 Discriminating Between Two Distributions

This section describes sequential procedures for discriminating between two competing populations. Three cases of discrimination are considered, namely between log-normal and Weibull, Weibull and log-logistic, and log-logistic and log-normal.

2.1 Discriminating Log-normal and Weibull Distribution

Consider a sequence of independent and identically distributed (i.i.d.) random variables X_1, X_2, \dots that are known to be observed from either log-normal or Weibull distribution. As the data X_1, X_2, \dots is observed sequentially, our interest is to determine the distribution from which the sequence is actually observed. This problem can be formulated as a hypothesis testing problem where the null hypothesis $H_0^{(1)} : X_1, X_2, \dots \sim WE(\gamma, \theta)$ is tested against

the alternative $H_A^{(1)} : X_1, X_2, \dots \sim LN(\sigma, \eta)$.

For testing two simple hypotheses, Wald (1945) introduced the celebrated sequential probability ratio test (SPRT) based on likelihood ratios (LR). Unfortunately, $H_0^{(1)}$ and $H_A^{(1)}$ are not simple as the parameters γ , θ , σ and η are unknown. For testing two composite hypotheses sequentially, a popular approach is to replace LR by generalized likelihood ratio statistic where unknown parameters are replaced by its maximum likelihood estimators (MLE). This approach, typically known as generalized sequential likelihood ratio test (GSLRT), is adaptive and considered by many authors (see, e.g., Schwarz (1962), Lai et al. (1988), Tartakovsky (2014)). We adopt the idea of GSLRT in the context of discriminating between two distributions.

Sequential Discrimination Method and Related Results

The proposed discrimination method is based on the logarithm of ratio of maximized likelihood (RML) defined as

$$\Lambda_n^{(1)} = \ln \left(\frac{\prod_{i=1}^n f_{LN}(X_i; \hat{\sigma}_n, \hat{\eta}_n)}{\prod_{i=1}^n f_{WE}(X_i; \hat{\gamma}_n, \hat{\theta}_n)} \right), \quad (4)$$

where $\hat{\sigma}_n, \hat{\eta}_n, \hat{\gamma}_n, \hat{\theta}_n$ are MLEs of $\sigma, \eta, \gamma, \theta$ respectively based on X_1, \dots, X_n . Fix some pilot sample size $k (> 1)$ and boundaries $b < 0 < a$. The purely sequential procedure for testing $H_0^{(1)}$ against $H_A^{(2)}$ is as follows:

Stage 0: Draw i.i.d. samples X_1, \dots, X_k and compute $\Lambda_k^{(1)}$ defined in 4. Stop sampling if $\Lambda_k^{(1)} \geq a$ or $\Lambda_k^{(1)} \leq b$. Else, proceed to the next stage.

Stage 1: Collect one more observation X_{k+1} independent of X_1, \dots, X_k and compute $\Lambda_{k+1}^{(1)}$ based on X_1, \dots, X_{k+1} . Stop sampling if $\Lambda_{k+1}^{(1)} \geq a$ or $\Lambda_{k+1}^{(1)} \leq b$. Else, proceed to the next stage.

We continue sampling until a stopping criteria is reached, i.e., until

$$T_1 = \inf\{n \geq k : \Lambda_n^{(1)} \notin (b, a)\}. \quad (5)$$

At stopping time T_1 , a decision is made whether to accept or reject $H_0^{(1)}$ as follows. If $\Lambda_{T_1}^{(1)} \geq a$, then reject $H_0^{(1)}$, i.e., conclude that the observed data sequence is coming from log-normal distribution and if $\Lambda_{T_1}^{(1)} \leq b$, accept $H_0^{(1)}$, i.e., conclude that the data sequence is from Weibull distribution.

Three types of stopping boundaries (b, a) such that $b < 0 < a$ are considered in this article. Wald (1945) provides the SPRT stopping boundaries $(b_w = \ln(\frac{\beta}{1-\alpha}), a_w = \ln(\frac{1-\beta}{\alpha}))$

where $0 < \alpha < 1$ and $0 < \beta < 1$ are some prefixed levels at which probabilities of type I and type II errors are to be controlled respectively. These boundaries are shown to be optimal in the sense of requiring the least expected sample size among all simple versus simple test procedures that control both type I and type II error rates at some specified levels (Wald et al. (1948)). Even though the hypotheses $H_0^{(1)}$ and $H_A^{(1)}$ are not simple, we consider this popular choice of boundaries and hope to benefit from the optimality properties of SPRT. A mathematical proof that these boundaries control probabilities of type I and type II errors at levels α and β respectively for finite sample remains an open problem (see, e.g., Tartakovsky (2014)).

A slightly more conservative choice of stopping boundaries ($b_c = \ln \beta, a_c = -\ln \alpha$), derived from Wald's boundaries, is also considered in this article. The third type of stopping boundaries (b_s, a_s) are obtained from Monte Carlo simulation and are chosen in each case to ensure that the probability of correct selection attains certain prefixed value.

Remark 1. If observed $\{X_i\}$ is from LN or WE or LL distribution, then corresponding distribution of $Y_i = \ln X_i$ is normal or extreme-value or logistic distribution that belong to location-scale family. Dumonceaux and Antle (1973) and Dumonceaux et al. (1973), showed that the distribution of RML for two location-scale family distributions does not depend on unknown parameter values. Therefore, the distribution of $\Lambda_n^{(1)}$ is same for all $(\sigma, \eta, \gamma, \theta)$ if the data $\{X_i\}$ follow $LN(\sigma, \eta)$ or $WE(\gamma, \theta)$. This is particularly helpful in designing stopping boundaries (b_s, a_s) from Monte Carlo simulation.

We remark that the stopping variable T_1 is a Markov stopping time and it is also proper. This is a very desirable property of any sequential procedure as it ensures that sampling will stop in finite time.

Lemma 1. *The stopping time T_1 is proper, i.e., $P(T_1 < \infty) = 1$ under both H_0 and H_A .*

Proof. see Appendix. □

Asymptotic Properties

Asymptotic behavior of error probabilities and expected sample size of Wald's SPRT are studied in Berk (1973). Below, we establish asymptotic properties of the proposed sequential method for selecting the correct model from log-normal and Weibull.

From Lemma 1 and 2 of Kundu and Manglick (2004) $\hat{\sigma}_n \xrightarrow{a.s.} \sigma, \hat{\eta}_n \xrightarrow{a.s.} \eta, \hat{\gamma}_n \xrightarrow{a.s.} \tilde{\gamma}$ and $\hat{\theta}_n \xrightarrow{a.s.} \tilde{\theta}$ where $E_{H_A^{(1)}}[\ln f_{LN}(X; \sigma, \eta)] = \max_{\bar{\sigma}, \bar{\eta}} E_{H_A^{(1)}}[\ln f_{LN}(X; \bar{\sigma}, \bar{\eta})]$ and $E_{H_A^{(1)}}[\ln f_{WE}(X; \tilde{\gamma}, \tilde{\theta})] = \max_{\bar{\gamma}, \bar{\theta}} E_{H_A^{(1)}}[\ln f_{WE}(X; \bar{\gamma}, \bar{\theta})]$ under the assumption that $H_A^{(1)}$ is true. When $H_0^{(1)}$ is true, $\hat{\sigma}_n \xrightarrow{a.s.} \tilde{\sigma}, \hat{\eta}_n \xrightarrow{a.s.} \tilde{\eta}, \hat{\gamma}_n \xrightarrow{a.s.} \gamma$ and $\hat{\theta}_n \xrightarrow{a.s.} \theta$ where $E_{H_0^{(1)}}[\ln f_{LN}(X; \tilde{\sigma}, \tilde{\eta})] = \max_{\bar{\sigma}, \bar{\eta}} E_{H_0^{(1)}}[\ln f_{LN}(X; \bar{\sigma}, \bar{\eta})]$

Algorithm 1 Discrimination procedure between two distributions

- 1: Start with a sample of fixed size k (> 1), fixed upper boundary $a > 0$ and lower boundary $b < 0$ and set $j = 0$.
 - 2: **while** $j < 1$ **do**
 - 3: Calculate the logarithm of RML, Λ , as in 4.
 - 4: **if** $\Lambda \geq a$ or $\Lambda \leq b$ **then**
 - 5: Stop sampling.
 - 6: **else**
 - 7: Draw one observation and augment with the existing samples.
 - 8: **end if**
 - 9: **end while**
-

and $E_{H_0^{(1)}}[\ln f_{WE}(X; \gamma, \theta)] = \max_{\bar{\gamma}, \bar{\theta}} E_{H_0^{(1)}}[\ln f_{WE}(X; \bar{\gamma}, \bar{\theta})]$. The above results can also be derived from Theorem 1 of White (1982).

Lemma 2. Let $\mu_{LN} = E_{H_A^{(1)}}\left[\ln\left(\frac{f_{LN}(X; \sigma, \eta)}{f_{WE}(X; \bar{\gamma}, \bar{\theta})}\right)\right]$, $\mu_{WE} = E_{H_0^{(1)}}\left[\ln\left(\frac{f_{LN}(X; \bar{\sigma}, \bar{\eta})}{f_{WE}(X; \gamma, \theta)}\right)\right]$. Then

$$(i) \frac{\Lambda_n^{(1)}}{n} \xrightarrow{a.s.} \mu_{LN} \text{ when } H_A^{(1)} \text{ is true and } \frac{\Lambda_n^{(1)}}{n} \xrightarrow{a.s.} \mu_{WE} \text{ when } H_0^{(1)} \text{ is true } n \rightarrow \infty.$$

$$(ii) \Lambda_n^{(1)} \xrightarrow{a.s.} \infty \text{ when } H_A^{(1)} \text{ is true and } \Lambda_n^{(1)} \xrightarrow{a.s.} -\infty \text{ when } H_0^{(1)} \text{ is true as } n \rightarrow \infty.$$

Proof. see Appendix. □

The following theorem shows that the error probabilities are asymptotically zero as stopping boundaries grow to infinity and also provides information about the asymptotic behavior of stopping time T_1 and expected stopping time $E[T_1]$ under both of null and alternative hypotheses. Following result is obtained along the lines of Berk (1973).

Theorem 1. (a) Suppose $H_0^{(1)}$ is true, i.e., data follows Weibull distribution. Then

$$(i) \lim_{a \rightarrow \infty} I(\Lambda_{T_1}^{(1)} \geq a) = \lim_{a \rightarrow \infty} P(\Lambda_{T_1}^{(1)} \geq a) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{b \rightarrow -\infty} \left(\frac{T_1}{b}\right) = \frac{1}{\mu_{WE}} \text{ with probability 1 and } \limsup_{b \rightarrow -\infty} E\left[\frac{T_1}{b}\right] \leq \frac{1}{\mu_{WE}}.$$

(b) Suppose $H_A^{(1)}$ is true, i.e., data follows log-normal distribution. Then

$$(i) \lim_{b \rightarrow -\infty} I(\Lambda_{T_1}^{(1)} \leq b) = \lim_{b \rightarrow -\infty} P(\Lambda_{T_1}^{(1)} \leq b) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{a \rightarrow \infty} \left(\frac{T_1}{a}\right) = \frac{1}{\mu_{LN}} \text{ with probability 1 and } \liminf_{a \rightarrow \infty} E\left[\frac{T_1}{a}\right] \geq \frac{1}{\mu_{LN}}.$$

Proof. See Appendix. □

The above theorem holds when either Wald boundaries ($b_w = \ln(\frac{\beta}{1-\alpha})$, $a_w = \ln(\frac{1-\beta}{\alpha})$) or conservative Wald boundaries ($b_c = \ln \beta$, $a_c = -\ln \alpha$) are used and the prefixed levels from error probabilities $\alpha, \beta \rightarrow 0$. PCS is defined as the probability of accepting $H_0^{(1)}$ (or $H_A^{(1)}$) when $H_0^{(1)}$ (or $H_A^{(1)}$) is true. As $\delta := \min\{|b|, a\} \rightarrow \infty$, an immediate conclusion from Theorem 1 can be made as follows:

$$\text{If } H_0^{(1)} \text{ is true, } PCS_\delta = P_{H_0^{(1)}}(\text{select } H_0^{(1)}) = 1 - P_{H_0^{(1)}}(\Lambda_{T_1}^{(1)} \geq a) \xrightarrow{a.s.} 1.$$

Similar result holds when $H_A^{(1)}$ is true. Hence, Corollary 1 follows.

Corollary 1. *Under both $H_0^{(1)}$ and $H_A^{(1)}$, $PCS_\delta \xrightarrow{a.s.} 1$ as $\delta = \min\{|b|, a\} \rightarrow \infty$.*

2.2 Discriminating Weibull and Log-logistic Distribution

Suppose X_1, X_2, \dots is a sequence of random variables from either Weibull or log-logistic distribution. As the data comes sequentially, we identify the correct distribution by testing

$$H_0^{(2)} : X_1, X_2, \dots \sim LL(\xi, \lambda) \text{ versus } H_A^{(2)} : X_1, X_2, \dots \sim WE(\gamma, \theta).$$

The test is based on the logarithm of RML

$$\Lambda_n^{(2)} = \ln \left(\frac{\prod_{i=1}^n f_{WE}(X_i; \hat{\gamma}_n, \hat{\theta}_n)}{\prod_{i=1}^n f_{LL}(X_i; \hat{\xi}_n, \hat{\lambda}_n)} \right), \quad (6)$$

where $\hat{\gamma}_n, \hat{\theta}_n, \hat{\xi}_n, \hat{\lambda}_n$ are maximum likelihood estimators of $\gamma, \theta, \xi, \lambda$ respectively based on n i.i.d. observations. The distribution of $\Lambda_n^{(2)}$ does not depend on the model parameters by Remark 1. The sequential discrimination procedure for Weibull and log-logistic is same as that of the log-normal and Weibull distribution where only $\Lambda_n^{(1)}$ is replaced by $\Lambda_n^{(2)}$. Here, sampling stops at time

$$T_2 = \inf\{n \geq k : \Lambda_n^{(2)} \notin (b, a)\} \text{ for some fixed } a > 0 \text{ and } b < 0. \quad (7)$$

In light of the arguments given in Lemma 1, we conclude that T_2 is finite with probability 1 under both $H_0^{(2)}$ and $H_A^{(2)}$.

Asymptotic Properties

If $H_A^{(2)}$ is true, following Lemma 1 and 2 of Elsherpieny et al. (2013), we have $\hat{\gamma}_n \xrightarrow{a.s.} \gamma$, $\hat{\theta}_n \xrightarrow{a.s.} \theta$, $\hat{\xi}_n \xrightarrow{a.s.} \tilde{\xi}$ and $\hat{\lambda}_n \xrightarrow{a.s.} \tilde{\lambda}$, where $E_{H_A^{(2)}}[\ln f_{WE}(X; \gamma, \theta)] = \max_{\tilde{\gamma}, \tilde{\theta}} E_{H_A^{(2)}}[\ln f_{WE}(X; \tilde{\gamma}, \tilde{\theta})]$ and $E_{H_A^{(2)}}[\ln f_{LL}(X; \tilde{\xi}, \tilde{\lambda})] = \max_{\tilde{\xi}, \tilde{\lambda}} E_{H_A^{(2)}}[\ln f_{LL}(X; \tilde{\xi}, \tilde{\lambda})]$. When $H_0^{(2)}$ is true, we have $\hat{\gamma}_n \xrightarrow{a.s.} \tilde{\gamma}$,

$\hat{\theta}_n \xrightarrow{a.s.} \tilde{\theta}$, $\hat{\xi}_n \xrightarrow{a.s.} \xi$ and $\hat{\lambda}_n \xrightarrow{a.s.} \lambda$ where $E_{H_0^{(2)}}[\ln f_{WE}(X; \tilde{\gamma}, \tilde{\theta})] = \max_{\tilde{\gamma}, \tilde{\theta}} E_{H_0^{(2)}}[\ln f_{WE}(X; \tilde{\gamma}, \tilde{\theta})]$ and $E_{H_0^{(2)}}[\ln f_{LL}(X; \xi, \lambda)] = \max_{\xi, \lambda} E_{H_0^{(2)}}[\ln f_{LL}(X; \xi, \lambda)]$.

Theorem 2. Let $\mu_{WE} = E_{H_A^{(2)}} \left[\ln \left(\frac{f_{WE}(X; \gamma, \theta)}{f_{LL}(X; \xi, \lambda)} \right) \right]$ and $\mu_{LL} = E_{H_0^{(2)}} \left[\ln \left(\frac{f_{WE}(X; \tilde{\gamma}, \tilde{\theta})}{f_{LL}(X; \xi, \lambda)} \right) \right]$.

(a) Assume $H_0^{(2)}$ is true, i.e., data follows log-logistic distribution. Then

$$(i) \lim_{a \rightarrow \infty} I(\Lambda_{T_2}^{(2)} \geq a) = \lim_{a \rightarrow \infty} P(\Lambda_{T_2}^{(2)} \geq a) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{b \rightarrow -\infty} \left(\frac{T_2}{b} \right) = \frac{1}{\mu_{LL}} \text{ w.p. 1 and } \limsup_{b \rightarrow -\infty} E \left[\frac{T_2}{b} \right] \leq \frac{1}{\mu_{LL}}.$$

(b) Assume $H_A^{(2)}$ is true, i.e., data follows Weibull distribution. Then

$$(i) \lim_{b \rightarrow -\infty} I(\Lambda_{T_2}^{(2)} \leq b) = \lim_{b \rightarrow -\infty} P(\Lambda_{T_2}^{(2)} \leq b) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{a \rightarrow \infty} \left(\frac{T_2}{a} \right) = \frac{1}{\mu_{WE}} \text{ with probability 1 and } \liminf_{a \rightarrow \infty} E \left[\frac{T_2}{a} \right] \geq \frac{1}{\mu_{WE}}.$$

Corollary 2. Under both $H_0^{(2)}$ and $H_A^{(2)}$, $PCS_\delta \xrightarrow{a.s.} 1$ as $\delta = \min\{|b|, a\} \rightarrow \infty$.

Proofs of Theorem 2 and Corollary 2 are similar to the proofs of Theorem 1 and Corollary 1, and thus, omitted.

2.3 Discriminating Log-normal and Log-logistic Distribution

Suppose X_1, X_2, \dots is a sequence of observations from either log-normal or log-logistic distribution. As the data comes sequentially, we would like to test

$$H_0^{(3)} : X_1, X_2, \dots \sim LL(\xi, \lambda) \text{ versus } H_A^{(3)} : X_1, X_2, \dots \sim LN(\sigma, \eta).$$

The test is based on the logarithm of RML

$$\Lambda_n^{(3)} = \ln \left(\frac{\prod_{i=1}^n f_{LN}(X_i; \hat{\sigma}_n, \hat{\eta}_n)}{\prod_{i=1}^n f_{LL}(X_i; \hat{\xi}_n, \hat{\lambda}_n)} \right), \quad (8)$$

where $\hat{\sigma}_n, \hat{\eta}_n, \hat{\xi}_n, \hat{\lambda}_n$ are the maximum likelihood estimators of $\sigma, \eta, \xi, \lambda$ respectively based on X_1, \dots, X_n . In light of Remark 1, the distribution of $\Lambda_n^{(3)}$ is same for all $\sigma, \eta, \xi, \lambda$. The sequential discrimination procedure for log-normal and log-logistic is same as that of the other two discriminating procedures with stopping time

$$T_3 = \inf\{n \geq k : \Lambda_n^{(3)} \notin (b, a)\} \text{ for some fixed } a > 0 \text{ and } b < 0. \quad (9)$$

Again, we can argue along the same lines of Lemma 1 that sampling will stop in finite time, that is, $P(T_3 < \infty) = 1$ under both $H_0^{(3)}$ and $H_A^{(3)}$.

Asymptotic Properties

When $H_A^{(3)}$ is true, following the arguments of Lemma 3.1 and 3.2 of Dey and Kundu (2009a), we have $\hat{\sigma}_n \xrightarrow{a.s.} \sigma$, $\hat{\eta}_n \xrightarrow{a.s.} \eta$, $\hat{\xi}_n \xrightarrow{a.s.} \tilde{\xi}$ and $\hat{\lambda}_n \xrightarrow{a.s.} \tilde{\lambda}$, where $E_{H_A^{(3)}}[\ln f_{LN}(X; \sigma, \eta)] = \max_{\bar{\sigma}, \bar{\eta}} E_{H_A^{(3)}}[\ln f_{LN}(X; \bar{\sigma}, \bar{\eta})]$ and $E_{H_A^{(3)}}[\ln f_{LL}(X; \tilde{\xi}, \tilde{\lambda})] = \max_{\bar{\xi}, \bar{\lambda}} E_{H_A^{(3)}}[\ln f_{LL}(X; \bar{\xi}, \bar{\lambda})]$. When $H_0^{(3)}$ is true, $\hat{\sigma}_n \xrightarrow{a.s.} \tilde{\sigma}$, $\hat{\eta}_n \xrightarrow{a.s.} \tilde{\eta}$, $\hat{\xi}_n \xrightarrow{a.s.} \xi$ and $\hat{\lambda}_n \xrightarrow{a.s.} \lambda$ where $E_{H_0^{(3)}}[\ln f_{LN}(X; \tilde{\sigma}, \tilde{\eta})] = \max_{\bar{\sigma}, \bar{\eta}} E_{H_0^{(3)}}[\ln f_{LN}(X; \bar{\sigma}, \bar{\eta})]$ and $E_{H_0^{(3)}}[\ln f_{LL}(X; \xi, \lambda)] = \max_{\bar{\xi}, \bar{\lambda}} E_{H_0^{(3)}}[\ln f_{LL}(X; \bar{\xi}, \bar{\lambda})]$.

Theorem 3. Let $\mu_{LN} = E_{H_A^{(3)}} \left[\ln \left(\frac{f_{LN}(X; \sigma, \eta)}{f_{LL}(X; \xi, \lambda)} \right) \right]$ and $\mu_{LL} = E_{H_0^{(3)}} \left[\ln \left(\frac{f_{LN}(X; \tilde{\sigma}, \tilde{\eta})}{f_{LL}(X; \xi, \lambda)} \right) \right]$.

(a) Assume $H_0^{(3)}$ is true, i.e., data follows log-logistic distribution. Then

$$(i) \lim_{a \rightarrow \infty} I(\Lambda_{T_3}^{(3)} \geq a) = \lim_{a \rightarrow \infty} P(\Lambda_{T_3}^{(3)} \geq a) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{b \rightarrow -\infty} \left(\frac{T_3}{b} \right) = \frac{1}{\mu_{LL}} \text{ with probability 1 and } \limsup_{b \rightarrow -\infty} \left[\frac{T_3}{b} \right] \leq \frac{1}{\mu_{LL}}.$$

(b) Assume $H_A^{(3)}$ is true, i.e., data follows log-normal distribution. Then

$$(i) \lim_{b \rightarrow -\infty} I(\Lambda_{T_3}^{(3)} \leq b) = \lim_{b \rightarrow -\infty} P(\Lambda_{T_3}^{(3)} \leq b) = 0 \text{ with probability 1.}$$

$$(ii) \lim_{a \rightarrow \infty} \left(\frac{T_3}{a} \right) = \frac{1}{\mu_{LN}} \text{ with probability 1 and } \liminf_{a \rightarrow \infty} E \left[\frac{T_3}{a} \right] \geq \frac{1}{\mu_{LN}}.$$

Corollary 3. Under both $H_0^{(3)}$ and $H_A^{(3)}$, $PCS_\delta \xrightarrow{a.s.} 1$ as $\delta = \min\{|b|, a\} \rightarrow \infty$.

Proofs of Theorem 3 and Corollary 3 are similar to the proofs of Theorem 1 and Corollary 1, and hence, omitted.

3 Discriminating Among Three Distributions

In this section, a sequential procedure for discriminating three competing populations, namely log-normal, Weibull, and log-logistic is described. Suppose X_1, X_2, \dots is a sequence of i.i.d. observations from either log-normal or Weibull or log-logistic distribution. As the data is observed sequentially, the goal is to identify the correct model by testing multiple hypotheses:

$$H_1 : X_1, X_2, \dots \sim LN(\sigma, \eta) \text{ vs } H_2 : X_1, X_2, \dots \sim WE(\gamma, \theta), \text{ vs } H_3 : X_1, X_2, \dots \sim LL(\xi, \lambda).$$

To our knowledge, earliest work on sequential procedures for choosing one hypothesis from more than two alternative hypotheses were discussed in Sobel and Wald (1949) and Armitage (1950). The idea of M-ary (or matrix) sequential probability ratio test (MSPRT)

for testing more than two simple hypotheses was introduced by Baum and Veeravalli (1994). Later, Dragalin et al. (1999) established its asymptotic optimality properties. The following discrimination procedure is developed in the spirit of MSPRT replacing likelihood ratios by log RMLs.

Sequential Discrimination Method and Related Results

A natural idea for multihypothesis testing is to select the hypothesis having maximum likelihood. The sequential procedure proposed here is based on log RML functions $\Lambda_n^{(1)}$, $\Lambda_n^{(2)}$ and $\Lambda_n^{(3)}$ defined in (4), (6) and (8) respectively. Suppose $b_i < 0 < a_i$ for $i = 1, 2, 3$, are some prefixed lower and upper boundaries to be compared with log RMLs $\Lambda_n^{(i)}$ for $i = 1, 2, 3$, and $k (> 1)$ is some prefixed pilot sample size.

Stage 0: Draw i.i.d. samples X_1, \dots, X_k and compute $\Lambda_k^{(1)}$, $\Lambda_k^{(2)}$, $\Lambda_k^{(3)}$. Stop sampling if $\{\Lambda_k^{(1)} \geq a_1 \text{ and } \Lambda_k^{(3)} \geq a_3\}$ or $\{\Lambda_k^{(1)} \leq b_1 \text{ and } \Lambda_k^{(2)} \geq a_2\}$ or $\{\Lambda_k^{(2)} \leq b_2 \text{ and } \Lambda_k^{(3)} \leq b_3\}$. Else, proceed to the next stage.

Stage 1: Draw a new observation X_{k+1} independent of X_1, \dots, X_k and compute $\Lambda_{k+1}^{(1)}$, $\Lambda_{k+1}^{(2)}$ and $\Lambda_{k+1}^{(3)}$. Stop sampling if $\{\Lambda_{k+1}^{(1)} \geq a_1 \text{ and } \Lambda_{k+1}^{(3)} \geq a_3\}$ or $\{\Lambda_{k+1}^{(1)} \leq b_1 \text{ and } \Lambda_{k+1}^{(2)} \geq a_2\}$ or $\{\Lambda_{k+1}^{(2)} \leq b_2 \text{ and } \Lambda_{k+1}^{(3)} \leq b_3\}$. Else, proceed to the next stage.

We continue sampling one observation at a time until

$$T = \inf_{n \geq k} \{ \{ \Lambda_n^{(1)} \geq a_1 \cap \Lambda_n^{(3)} \geq a_3 \} \cup \{ \Lambda_n^{(1)} \leq b_1 \cap \Lambda_n^{(2)} \geq a_2 \} \cup \{ \Lambda_n^{(2)} \leq b_2 \cap \Lambda_n^{(3)} \leq b_3 \} \}. \quad (10)$$

At stopping time T , we select H_1 (log-normal) if $\{\Lambda_T^{(1)} \geq a_1 \text{ and } \Lambda_T^{(3)} \geq a_3\}$, select H_2 (Weibull) if $\{\Lambda_T^{(1)} \leq b_1 \text{ and } \Lambda_T^{(2)} \geq a_2\}$, select H_3 (log-logistic) if $\{\Lambda_T^{(2)} \leq b_2 \text{ and } \Lambda_T^{(3)} \leq b_3\}$. In other words, the stopping time $T = \min\{\tau_1, \tau_2, \tau_3\}$ where

$$\tau_1 = \inf_{n \geq k} \{ \Lambda_n^{(1)} \geq a_1, \Lambda_n^{(3)} \geq a_3 \}, \tau_2 = \inf_{n \geq k} \{ \Lambda_n^{(1)} \leq b_1, \Lambda_n^{(2)} \geq a_2 \}, \tau_3 = \inf_{n \geq k} \{ \Lambda_n^{(2)} \leq b_2, \Lambda_n^{(3)} \leq b_3 \}.$$

The hypothesis H_i is accepted if $T = \tau_i$ for $i = 1, 2, 3$. The stopping criteria in (10) is achieved in finite time which is proved in Appendix.

Lemma 3. $P(T < \infty) = 1$ under H_1, H_2 and H_3 .

Selection of stopping boundaries is crucial for good performance of our procedure. We consider two types of stopping boundaries in this article. To justify the choice of boundaries used here, let us first discuss some results for MSPRT from Tartakovsky et al. (2014). If the true values of the model parameters $\sigma, \eta, \gamma, \theta, \xi$ and λ were known, H_1, H_2, H_3 would

be simple hypotheses and matrix SPRT would be the optimal (oracle) procedure, where log RMLs in the stopping rule (10) would be replaced by log likelihood ratios (LLR)

$$\Lambda_{1n} = \ln \left(\prod_{i=1}^n \frac{f_{LN}(X_i; \sigma, \eta)}{f_{WE}(X_i; \gamma, \theta)} \right), \Lambda_{2n} = \ln \left(\prod_{i=1}^n \frac{f_{WE}(X_i; \gamma, \theta)}{f_{LL}(X_i; \xi, \lambda)} \right), \Lambda_{3n} = \ln \left(\prod_{i=1}^n \frac{f_{LN}(X_i; \sigma, \eta)}{f_{LL}(X_i; \xi, \lambda)} \right).$$

From Lemma 4.1.1 of Tartakovsky et al. (2014), the error probabilities and PCS under H_1, H_2, H_3 is obtained as follows:

- (i) $P_{H_1}(H_1 \text{ rejected}) \leq e^{b_1} + e^{b_3}$. Hence, under H_1 , $PCS_{H_1} \geq 1 - e^{b_1} - e^{b_3}$
- (ii) $P_{H_2}(H_2 \text{ rejected}) \leq e^{-a_1} + e^{b_2}$. Hence, under H_2 , $PCS_{H_2} \geq 1 - e^{-a_1} - e^{b_2}$
- (iii) $P_{H_3}(H_3 \text{ rejected}) \leq e^{-a_2} + e^{-a_3}$. Hence, under H_3 , $PCS_{H_3} \geq 1 - e^{-a_2} - e^{-a_3}$.

Clearly, under any hypotheses H_1, H_2 or H_3 , we can guarantee for fixed level $\alpha \in (0, 1)$,

$$PCS_{H_i} \geq 1 - \alpha, \quad \text{provided } b_i = \ln\left(\frac{\alpha}{2}\right) \text{ and } a_i = -\ln\left(\frac{\alpha}{2}\right) \text{ for } i = 1, 2, 3. \quad (11)$$

Dragalin et al. (1999) showed that MSPRT is nearly optimal in the sense that, while controlling error probabilities, it also minimizes the expected stopping time asymptotically under all hypotheses under consideration.

Our first choice of stopping boundaries is the equal and symmetric boundaries in (11). Since the stopping rule (10) mimics the oracle optimal MSPRT stopping rule (at least in large sample), we hope to benefit from the optimality properties of MSPRT and achieve $PCS_{H_i} \geq 1 - \alpha$ for $i = 1, 2, 3$ while the average sample size $E_{H_i}[T]$ remains smaller than existing fixed-sample-size procedures. With boundaries (11), we expect (asymptotically) that $P_{H_i}(H_j \text{ accepted}) \leq \alpha/2$ and $P_{H_i}(H_i \text{ rejected}) \leq \alpha$ for $i = 1, 2, 3$ and $j \neq i$. Using stopping boundaries in (11) for all LLR statistics can be justified if we treat all hypotheses equally and no prior information about $H_i, i = 1, 2, 3$ is available. However, if some prior information about H_i is available or H_i is more important than others, some weights should be assigned appropriately to the boundaries to accommodate this extra information (see for example, Tartakovsky et al. (2014)) and unequal boundaries should be adopted in (10). Since the distributions of $\Lambda_n^{(i)}, i = 1, 2, 3$ do not depend on model parameters by Remark 1, we also consider simulated symmetric stopping boundaries $(-a_s, a_s)$ for $a_s > 0$ similar to the one discussed in subsection 2.1.

Asymptotic Properties

Let $\mu_{LN}^{(1)} = E_{H_1} \left[\ln \left(\frac{f_{LN}(X; \sigma, \eta)}{f_{WE}(X; \tilde{\gamma}, \tilde{\theta})} \right) \right]$ and $\mu_{LN}^{(3)} = E_{H_1} \left[\ln \left(\frac{f_{LN}(X; \sigma, \eta)}{f_{LL}(X; \tilde{\xi}, \tilde{\lambda})} \right) \right]$ where $\tilde{\gamma}, \tilde{\theta}, \tilde{\xi}$ and $\tilde{\lambda}$ are same as in Section 2. An application of Jensen's inequality yields $\mu_{LN}^{(1)}, \mu_{LN}^{(3)} > 0$. Following

Algorithm 2 Discrimination procedure among three distributions.

- 1: Start with a sample of fixed size k (> 1), fixed threshold $a_i > 0$, $b_i < 0$ for $i = 1, 2, 3$, and $m = 0$.
 - 2: **while** $m < 1$ **do**
 - 3: Calculate log RMLs, $\Lambda_k^{(1)}$, $\Lambda_k^{(2)}$ and $\Lambda_k^{(3)}$ using 4, 6 and 8 respectively.
 - 4: **if** $(\Lambda_k^{(1)} \geq a_1$ and $\Lambda_k^{(3)} \geq a_3)$ or $(\Lambda_k^{(1)} \leq b_1$ and $\Lambda_k^{(2)} \geq a_2)$ or $(\Lambda_k^{(2)} \leq b_2$ and $\Lambda_k^{(3)} \leq b_3)$ **then**
 - 5: Stop sampling.
 - 6: **else**
 - 7: Draw one observation and augment with existing samples.
 - 8: **end if**
 - 9: **end while**
-

the arguments of Lemma 2, it can be shown that $\Lambda_n^{(1)}/n \rightarrow \mu_{LN}^{(1)}$ and $\Lambda_n^{(3)}/n \rightarrow \mu_{LN}^{(3)}$ almost surely as $n \rightarrow \infty$ when H_1 is true. Similar results hold when H_2 or H_3 is true. The following theorem ensures that the error probabilities $\{P_{H_i}(H_j \text{ accepted}) : i, j = 1, 2, 3, i \neq j\}$ tend to zero as boundaries go to infinity.

Theorem 4. *As $a_i \rightarrow \infty$, $b_i \rightarrow -\infty$ for $i = 1, 2, 3$, the error probabilities are given by:*

$$(i) \text{ Under } H_1, \lim_{b_1 \rightarrow -\infty} P(\Lambda_T^{(1)} \leq b_1) = \lim_{b_3 \rightarrow -\infty} P(\Lambda_T^{(3)} \leq b_3) = 0.$$

$$(ii) \text{ Under } H_2, \lim_{a_1 \rightarrow \infty} P(\Lambda_T^{(1)} \geq a_1) = \lim_{b_2 \rightarrow -\infty} P(\Lambda_T^{(2)} \leq b_2) = 0.$$

$$(iii) \text{ Under } H_3, \lim_{a_2 \rightarrow \infty} P(\Lambda_T^{(2)} \geq a_2) = \lim_{a_3 \rightarrow \infty} P(\Lambda_T^{(3)} \geq a_3) = 0.$$

Proof of Theorem 4 follows along the lines of Theorem 1, and hence, omitted. Let $PCS_\delta(H_i) = P_{H_i}(\text{select } H_i)$ for $i = 1, 2, 3$. Theorem 4 implies that, under H_1 ,

$$PCS_\delta(H_1) = 1 - P_{H_1}(\Lambda_T^{(1)} \leq b_1) - P_{H_1}(\Lambda_T^{(3)} \leq b_3) \rightarrow 1 \text{ as } \delta = \min\{|b_1|, |b_3|\} \rightarrow \infty.$$

This result holds when H_2 or H_3 is true for appropriately defined δ .

Corollary 4. *For $i = 1, 2, 3$, under H_i , $PCS_\delta(H_i) \rightarrow 1$ as $\delta = \min_{i=1,2,3}\{a_i, |b_i|\} \rightarrow \infty$.*

4 Numerical Study

We use R version 3.6.1 for all numerical studies of this article. In the context of discriminating Weibull and log-normal, Table 2 compares the performances of fixed-sample-size discrimination procedure of Kundu and Manglick (2004) with the sequential method of Section 2.1. $\widehat{E}(T_1)$, $\widehat{sd}(T_1)$ and \widehat{PCS}_S denote the estimated expected sample size (i.e., $\widehat{E}(T_1) = \bar{T}_1$ is the

average obtained from 10^4 Monte Carlo runs), standard error of \bar{T}_1 , and the estimated PCS for the sequential procedure. Using conservative Wald boundaries ($b_c = \ln \beta, a_c = -\ln \alpha$) for some prefixed values of $\alpha, \beta \in (0, 1)$, \widehat{PCS}_S turns out to be *at least* $(1 - \alpha)$ or $(1 - \beta)$ in all cases, as conjectured in Section 2.1. In Table 2, \widehat{PCS}_F represents the estimated PCS produced by the fixed-sample-size procedure of Kundu and Manglick (2004) using n_F many observations. Prefixed sample size n_F is chosen in each case such that \widehat{PCS}_F is approximately equal to \widehat{PCS}_S . To achieve same PCS, the sequential method requires much lower expected sample size than that of the method of Kundu and Manglick (2004). Last columns of Tables 2 - 7 report savings in the average (expected) sample sizes as $(n_F - \widehat{E}(T_i))/n_F \times 100\%$ for $i = 1, 2, 3$. We observe remarkable savings in average sample sizes ranging from 38.2% to 59.1% in Table 2 .

Table 2: Comparing the method of Kundu and Manglick (2004) with sequential procedure in (5) with boundaries $(\ln \beta, -\ln \alpha)$ and $k = 10$ when data follows Weibull or log-normal ($\gamma = \theta = \sigma = \eta = 1$). \widehat{PCS}_S is computed first and n_F is determined so that \widehat{PCS}_F matches with \widehat{PCS}_S .

True distribution	boundary ($\ln \beta, -\ln \alpha$)	\widehat{PCS}_S	\widehat{PCS}_F	$\widehat{E}(T_1)$	$\widehat{sd}(T_1)$	n_F	Saving in avg sample size %
Weibull	($\ln 0.01, -\ln 0.01$)	0.9924	0.9922	65.3804	0.5817	160	59.1
	($\ln 0.05, -\ln 0.05$)	0.9630	0.9649	44.41	0.4405	90	50.65
	($\ln 0.1, -\ln 0.1$)	0.9236	0.9244	34.35	0.3399	60	42.7
log-normal	($\ln 0.01, -\ln 0.01$)	0.9904	0.9912	71.77	0.5877	150	52.16
	($\ln 0.05, -\ln 0.05$)	0.9594	0.9596	48.65	0.4478	85	42.76
	($\ln 0.1, -\ln 0.1$)	0.9194	0.9172	37.10	0.3557	60	38.2

Table 3 continues to compare the fixed-sample-size procedure of Kundu and Manglick (2004) with our sequential procedure with symmetric simulated boundaries $(-a_s, a_s)$. In the context of discriminating between two distributions, we consider equal loss for both type I or type II errors, and hence, symmetric stopping boundaries are considered here. We first determine the sample sizes, n_F , in order to attain certain values of \widehat{PCS}_F such as 0.99 and 0.95 for the fixed-sample-size method. Exploiting the idea in Remark 1, we determine boundaries $(-a_s, a_s)$ from Monte Carlo simulations such that \widehat{PCS}_S is approximately same as \widehat{PCS}_F . Table 3 illustrates that the sample size savings are really significant ranging from 38.7% to 58.09%.

The fixed-sample-size procedure of Elsherpieny et al. (2013) for discriminating between Weibull and log-logistic distributions is compared in Tables 4 and 5 with the sequential analog of Section 2.2 having fixed boundaries (b_c, a_c) and simulated boundaries $(-a_s, a_s)$

Table 3: Comparing the method of Kundu and Manglick (2004) with sequential rule in (5) with simulated boundaries $(-a_s, a_s)$ when data follows Weibull or log-normal ($\gamma = \theta = \sigma = \eta = 1$). \widehat{PCS}_F is computed first for given n_F , and a_s is determined so that \widehat{PCS}_S matches with \widehat{PCS}_F .

True distribution	Fixed-sample-size procedure	k	Simulated boundary	$\widehat{E}(T_1)$	$\widehat{sd}(T_1)$	\widehat{PCS}_S	Saving in avg sample size %
Weibull	$n_F = 140$ $\widehat{PCS}_F = 0.9902$	10	$(-4.12, 4.12)$	58.92	0.5376	0.9870	57.9
		20	$(-4.08, 4.08)$	58.67	0.5261	0.9862	58.09
Weibull	$n_F = 76$ $\widehat{PCS}_F = 0.9490$	10	$(-2.72, 2.72)$	40.42	0.4042	0.9498	46.8
		20	$(-2.66, 2.66)$	41.53	0.3838	0.9512	45.4
log-normal	$n_F = 144$ $\widehat{PCS}_F = 0.9900$	10	$(-4.46, 4.46)$	69.92	0.5770	0.9880	51.4
		20	$(-4.34, 4.34)$	68.66	0.5678	0.9880	52.3
log-normal	$n_F = 74$ $\widehat{PCS}_F = 0.9516$	10	$(-2.80, 2.80)$	45.37	0.4222	0.9506	38.7
		20	$(-2.66, 2.66)$	44.66	0.4000	0.9516	39.65

Table 4: Comparing the method of Elsherpieny et al. (2013) with sequential procedure in (7) with boundaries $(\ln \beta, -\ln \alpha)$ and $k = 10$ when data follows Weibull or log-logistic ($\gamma = \theta = \xi = \lambda = 1$). \widehat{PCS}_S is computed first and n_F is determined so that \widehat{PCS}_F matches with \widehat{PCS}_S .

True distribution	boundary $(\ln \beta, -\ln \alpha)$	\widehat{PCS}_S	\widehat{PCS}_F	$\widehat{E}(T_1)$	$\widehat{sd}(T_1)$	n_F	Saving in avg sample size %
Weibull	$(\ln 0.01, -\ln 0.01)$	0.9944	0.9926	76.26	0.5516	152	49.8
	$(\ln 0.05, -\ln 0.05)$	0.9708	0.9714	48.94	0.4170	85	42.40
	$(\ln 0.1, -\ln 0.1)$	0.9442	0.9424	36.41	0.3280	58	37.2
log-logistic	$(\ln 0.01, -\ln 0.01)$	0.9900	0.9888	60.80	0.5704	150	59.5
	$(\ln 0.05, -\ln 0.05)$	0.9410	0.9452	42.58	0.4275	81	47.4
	$(\ln 0.1, -\ln 0.1)$	0.8914	0.8912	33.62	0.3331	56	39.9

respectively. Similarly, the fixed-sample-size discrimination procedure of Dey and Kundu (2009a) for selecting between log-logistic and log-normal distributions is compared in Tables 6 and 7 with the corresponding sequential procedure of Section 2.3 having fixed boundaries (b_c, a_c) and simulated boundaries $(-a_s, a_s)$ respectively. In each table, we observe PCS exceeding the desired level whereas savings in expected sample sizes are quite remarkable.

Tables 8 and 9 compare the performance of the fixed-sample-size procedure of Raqab et al. (2018) for discriminating Weibull, log-normal and log-logistic distributions with our

Table 5: Comparing the method of Elsherpieny et al. (2013) with sequential rule in (7) with simulated boundaries $(-a_s, a_s)$ when data follows Weibull or log-logistic ($\gamma = \theta = \xi = \lambda = 1$). \widehat{PCS}_F is computed first for given n_F , and a_s is determined so that \widehat{PCS}_S matches with \widehat{PCS}_F .

True distribution	Fixed-sample-size procedure	k	Simulated boundary	$\widehat{E}(T_2)$	$\widehat{sd}(T_2)$	\widehat{PCS}_S	Saving in avg sample size %
Weibull	$n_F = 140$ $\widehat{PCS}_F = 0.9914$	10	$(-4.50, 4.50)$	74.42	0.5462	0.9926	46.84
		20	$(-4.40, 4.40)$	72.77	0.5457	0.9930	48.02
Weibull	$n_F = 76$ $\widehat{PCS}_F = 0.9632$	10	$(-2.70, 2.70)$	43.60	0.3820	0.9634	42.63
		20	$(-2.52, 2.52)$	41.65	0.3471	0.9634	45.20
log-logistic	$n_F = 140$ $\widehat{PCS}_F = 0.9872$	10	$(-4.40, 4.40)$	58.71	0.5583	0.9872	58.06
		20	$(-4.38, 4.38)$	58.40	0.5497	0.9868	58.29
log-logistic	$n_F = 76$ $\widehat{PCS}_F = 0.9322$	10	$(-2.70, 2.70)$	38.61	0.3820	0.9340	49.20
		20	$(-2.60, 2.60)$	39.33	0.3512	0.9294	48.25

Table 6: Comparing the method of Dey and Kundu (2009a) with sequential procedure in (9) with boundaries $(\ln \beta, -\ln \alpha)$ and $k = 10$ when data follows log-normal or log-logistic ($\sigma = \eta = \xi = \lambda = 1$). \widehat{PCS}_S is computed first and n_F is determined so that \widehat{PCS}_F matches with \widehat{PCS}_S .

True distribution	boundary $(\ln \beta, -\ln \alpha)$	\widehat{PCS}_S	\widehat{PCS}_F	$\widehat{E}(T_3)$	$\widehat{sd}(T_3)$	n_F	Saving in avg sample size %
log-logistic	$(\ln 0.01, -\ln 0.01)$	0.9850	0.9842	380.12	3.7161	900	57.8
	$(\ln 0.05, -\ln 0.05)$	0.9386	0.9356	244.15	2.5972	465	47.0
	$(\ln 0.1, -\ln 0.1)$	0.8892	0.8890	180.70	1.9751	330	45.24
log-normal	$(\ln 0.01, -\ln 0.01)$	0.9970	0.9906	474.42	3.6078	900	47.3
	$(\ln 0.05, -\ln 0.05)$	0.9678	0.9674	290.35	2.5404	540	46.23
	$(\ln 0.1, -\ln 0.1)$	0.9418	0.9414	207.97	1.9716	390	46.7

sequential procedure in Section 3. As earlier, $\widehat{E}(T)$, $\widehat{sd}(T)$ and \widehat{PCS}_S denote the estimated expected sample size, standard error of $\widehat{E}(T)$, and the estimated PCS for the sequential procedure. For simulation study in Table 8, the stopping rule T of (10) uses fixed boundaries $\{(b_i = \log(\alpha/2), a_i = -\log(\alpha/2)) : i = 1, 2, 3\}$. Interestingly, \widehat{PCS}_S exceeds the desired level $(1 - \alpha)$ in all cases as it was conjectured in Section 3. For fair comparison between expected sample sizes of sequential and fixed-sample-size procedures, we first determine appropriate sample size n_F such that \widehat{PCS}_F (corresponding to fixed-sample-size method of Raqab et al.

Table 7: Comparing the method of Dey and Kundu (2009a) with sequential rule in (9) with simulated boundaries $(-a_s, a_s)$ when data follows log-normal or log-logistic ($\sigma = \eta = \xi = \lambda = 1$). \widehat{PCS}_F is computed first for given n_F , and a_s is determined so that \widehat{PCS}_S matches with \widehat{PCS}_F .

True distribution	Fixed-sample-size procedure	k	Simulated boundary	$\widehat{E}(T_3)$	$\widehat{sd}(T_3)$	\widehat{PCS}_S	Saving in avg sample size %
log-logistic	$n_F = 350$ $\widehat{PCS}_F = 0.8990$	10	$(-2.52, 2.52)$	205.55	2.2363	0.9086	41.27
		20	$(-2.42, 2.42)$	197.57	2.1487	0.9018	43.55
log-logistic	$n_F = 250$ $\widehat{PCS}_F = 0.8514$	10	$(-2, 2)$	156.26	1.7275	0.8542	37.49
		20	$(-2, 2)$	156.49	1.7223	0.8548	37.40
log-normal	$n_F = 250$ $\widehat{PCS}_F = 0.9050$	10	$(-1.76, 1.76)$	145.97	1.4571	0.9068	41.61
		20	$(-1.74, 1.74)$	143.62	1.4298	0.9072	42.55
log-normal	$n_F = 144$ $\widehat{PCS}_F = 0.8554$	10	$(-1.25, 1.25)$	88.05	0.9234	0.8522	38.86
		20	$(-1.20, 1.20)$	82.36	0.8521	0.8528	42.81

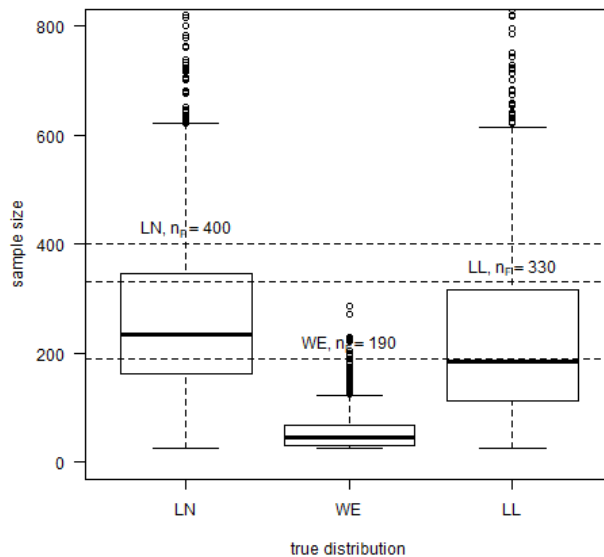


Figure 2: Boxplots of stopping time T in (10) with boundaries $\{(b_i = \log(0.05), a_i = -\log(0.05)) : i = 1, 2, 3\}$ and $k = 25$ when the true data distribution is LN, WE and LL. Horizontal dashed lines represent fixed sample sizes n_F such that \widehat{PCS}_F matches with \widehat{PCS}_S as given in Table 8.

(2018)) is approximately same as \widehat{PCS}_S . Table 8 illustrates that, to attain same PCS, the sequential procedure requires much less average sample size (savings between 12.13% to

Table 8: Comparing the method of Raqab et al. (2018) with sequential procedure in (10) with symmetric boundaries $\{(b_i = \log(\alpha/2), a_i = -\log(\alpha/2)) : i = 1, 2, 3\}$ when data follows Weibull or log-normal or log-logistic ($\sigma = \eta = \gamma = \theta = \xi = \lambda = 1$). \widehat{PCS}_S is computed first and n_F is determined so that \widehat{PCS}_F matches with \widehat{PCS}_S .

True distribution	Boundaries $ b_i = a_i = \log(\frac{\alpha}{2}) $	k	\widehat{PCS}_S	\widehat{PCS}_F	n_F	$\widehat{E}(T)$	$\widehat{sd}(T)$	Saving in avg sample size %
log-normal	$ \log(0.05) $	25	0.9350	0.9350	400	281.36	5.7490	29.66
		50	0.9470	0.9490	440	286.05	5.7673	34.99
	$ \log(0.1) $	25	0.9000	0.9040	275	202.39	4.5133	26.40
		50	0.9130	0.9170	330	207.39	4.5040	37.15
Weibull	$ \log(0.05) $	25	0.9960	0.9930	190	55.70	1.1259	70.68
		50	0.9950	0.9930	190	66.32	0.9594	65.09
	$ \log(0.1) $	25	0.9860	0.9840	150	46.38	1.0180	69.08
		50	0.9890	0.9900	170	60.65	0.7880	64.32
log-logistic	$ \log(0.05) $	25	0.8880	0.8830	330	238.85	5.7711	27.62
		50	0.9040	0.9070	360	243.58	5.6217	32.34
	$ \log(0.1) $	25	0.8090	0.8100	200	175.75	4.3228	12.13
		50	0.8510	0.8540	270	185.72	4.4240	31.21

70.68%) than the competing fixed-sample-size method.

To further illustrate sample size comparison, we present three boxplots in Figure 2 for distributions of sample size T in (10) with pilot sample size $k = 25$ and boundaries $\{(b_i = \log(0.05), a_i = -\log(0.05)) : i = 1, 2, 3\}$ when the observations come from LN, WE and LL distributions. In each case of H_1, H_2 and H_3 , \widehat{PCS}_S is calculated first and then sample size n_F for the procedure of Raqab et al. (2018) is determined such that \widehat{PCS}_F matches with \widehat{PCS}_S . We observe that n_F is larger than the 75th percentile of the distribution of T which confirms savings in sample size.

In Table 9, we use symmetric simulated boundaries $\{(b_i = -A_s, a_i = A_s) : i = 1, 2, 3\}$ in the stopping rule T of (10). Clearly, as A_s increases, both T and PCS increases. We first determine sample sizes n_F required for the method of Raqab et al. (2018) to attain certain values of \widehat{PCS}_F , namely 0.94, 0.89, and 0.86. Then the boundary A_s is determined from Monte Carlo simulation to ensure that \widehat{PCS}_S matches with \widehat{PCS}_F in Table 9. We are able to compute A_s from simulation since the distributions of $\Lambda_n^{(i)}, i = 1, 2, 3$, do not depend on model parameters. As noted earlier, the sequential method brings huge savings in the expected sample size (between 23.4% to 64.54%).

In Figure 3(a), the sequential stopping rule of (10) is implemented with symmetric bound-

Table 9: Comparing the method of Raqab et al. (2018) with sequential rule in (10) with symmetric simulated boundaries $\{(b_i = -A_s, a_i = A_s) : i = 1, 2, 3\}$ when data follows Weibull or log-normal or log-logistic ($\sigma = \eta = \gamma = \theta = \xi = \lambda = 1$). \widehat{PCS}_F is computed first for given n_F , and A_s is determined so that \widehat{PCS}_S matches with \widehat{PCS}_F .

True distribution	Fixed-sample-size procedure	k	Boundaries $ b_i = a_i = A_s$	$\widehat{E}(T)$	$\widehat{sd}(T)$	\widehat{PCS}_S	Saving in avg sample size %
log-normal	$n_F = 430$	25	3.00	284.64	5.7257	0.9390	33.80
	$\widehat{PCS}_F = 0.9410$	50	2.90	274.27	5.6152	0.9370	36.22
log-normal	$n_F = 247$	25	2.20	189.204	4.3065	0.8810	23.4
	$\widehat{PCS}_F = 0.8860$	50	1.90	159.36	3.4899	0.8860	35.48
Weibull	$n_F = 83$	25	1.40	35.73	0.6452	0.9470	56.95
	$\widehat{PCS}_F = 0.9490$	50	1.34	54.67	0.4322	0.9450	34.13
Weibull	$n_F = 50$	10	0.88	17.73	0.3607	0.8900	64.54
	$\widehat{PCS}_F = 0.8880$	25	0.86	29.26	0.3026	0.8880	41.48
log-logistic	$n_F = 495$	25	3.58	296.83	6.9183	0.9450	40.03
	$\widehat{PCS}_F = 0.9420$	50	3.44	281.21	6.4949	0.9450	43.19
log-logistic	$n_F = 295$	25	2.52	198.46	4.9409	0.8620	32.73
	$\widehat{PCS}_F = 0.8630$	50	2.42	190.94	4.5306	0.8680	35.28

aries (sb) $\{(b_i = -|sb|, a_i = |sb|) : i = 1, 2, 3\}$ and \widehat{PCS}_S (when true data distribution is LN, WE and LL) is plotted for different values of $|sb|$. This graph aims to guide a practitioner to select stopping boundaries to achieve a specific value of PCS. Figure 3(b) plots \widehat{PCS}_S for sequential procedure with symmetric boundaries versus expected average sample numbers (easn) $\widehat{E}(T)$ obtained using different values of $|sb|$. Clearly, if the true data distribution is LL, we need wider boundaries and larger sample size to achieve certain PCS compared to the case when the true data distribution is LN or WE.

5 Data Analysis

In this section, we analyze three benchmark datasets using the proposed sequential method (10) to select the most appropriate distribution among Weibull, log-normal and log-logistic. The sequential method selects the distribution (with highest probability) that has the smallest KolmogorovSmirnov (K-S) distance from the empirical distribution.

Dataset 1: This dataset, available in Andersen et al. (2012) and R package “timereg”

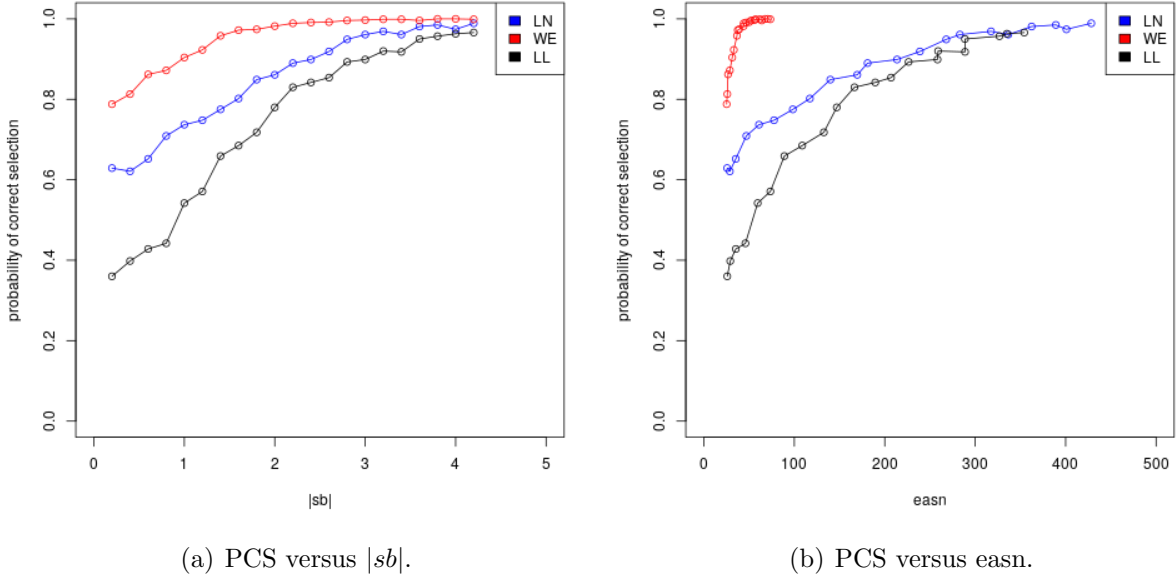


Figure 3: PCS for sequential procedure in (10) with $k = 25$ and symmetric boundaries (sb) $\{(b_i = -|sb|, a_i = |sb|) : i = 1, 2, 3\}$ are plotted against $|sb|$ and estimated average sample number ($easn$) $\hat{E}(T)$ when true data distribution is LN, WE and LL.

Scheike (2019), contains survival times (in years) of $n_F = 205$ patients after surgery for malignant melanoma (skin cancer) between the periods 1962 to 1977 collected at the University Hospital of Odense, Denmark by K.T. Drzewiecki. Andersen et al. (2012) suggests using a *Weibull* model for this dataset which is summarized in Table 10. Histogram of dataset 1 in

min	1st quantile	median	mean	3rd quantile	max	skewness coefficient
0.0274	4.1781	5.4932	5.8981	8.3342	15.2466	0.3278

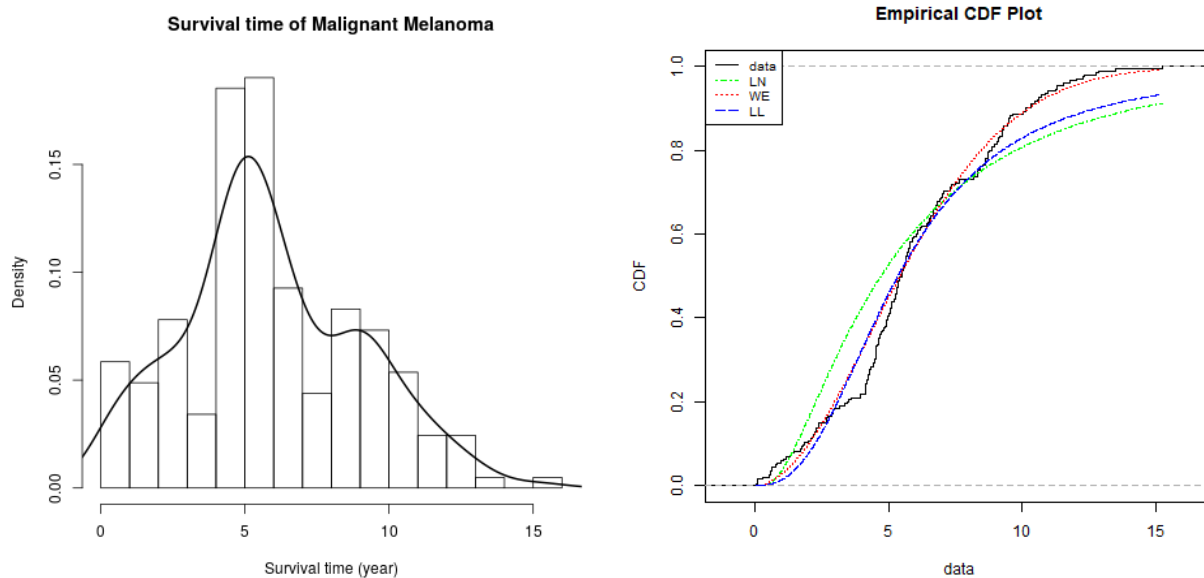
Table 10: Summary of survival time data on malignant melanoma patients.

Figure 4(a) appears to be right-skewed. Figure 4(b) shows the empirical CDF and CDFs fitted using Weibull, log-normal and log-logistic distributions where unknown parameter values are replaced by corresponding maximum likelihood estimates. Table 11 represents K-S

	log-normal	Weibull	log-logistic
K-S distance	0.2187	0.1176	0.1207
K-S p-value	$6.096 * 10^{-09}$	0.0069	0.0051
CVM p-value	$1.392 * 10^{-05}$	0.1116	0.0306

Table 11: K-S distances and p-values of goodness-of-fit tests for dataset 1.

distances between the empirical CDF and the fitted distribution functions corresponding to Weibull, log-normal, log-logistic as well as p-values for corresponding goodness-of-fit tests.



(a) Histogram.

(b) Empirical and fitted distribution functions.

Figure 4: Survival time of malignant melanoma patients.

We also report p-values corresponding to Cramer-Von Mises (CVM) goodness-of-fit tests in the last row of Table 11. Based on these K-S distances and p-values, Weibull distribution seems to provide the best fit among these three distributions. As the fixed-sample-size procedure of Raqab et al. (2018) is applied on the dataset 1 after resampling with replacement (non-parametric bootstrap), we observe that this procedure always (100% of the times) chooses Weibull distribution among the three distributions. The sequential procedure (10) with pilot sample size $k = 10$ and boundaries $\{(b_i = -4, a_i = 4) : i = 1, 2, 3\}$ (after resampling with replacement) also chooses Weibull distribution 99.8% of the times. However, it is remarkable that the estimated average sample size $\hat{E}(T)$ is only 40.10 which is much less than $n_F = 205$ giving 80.44% savings in the average sample size.

Dataset 2: This dataset, obtained from Overduin (2004), consists of survival times (monthly) of $n_F = 184$ patients having limited-stage small-cell lung cancer (LC). Tai et al. (2003) and Overduin (2004) both fit *log-normal* model for this dataset. Histogram in Figure

min	1st quantile	median	mean	3rd quantile	max	skewness coefficient
4.0400	12.5300	18.0000	19.7459	23.7000	64.6400	1.6635

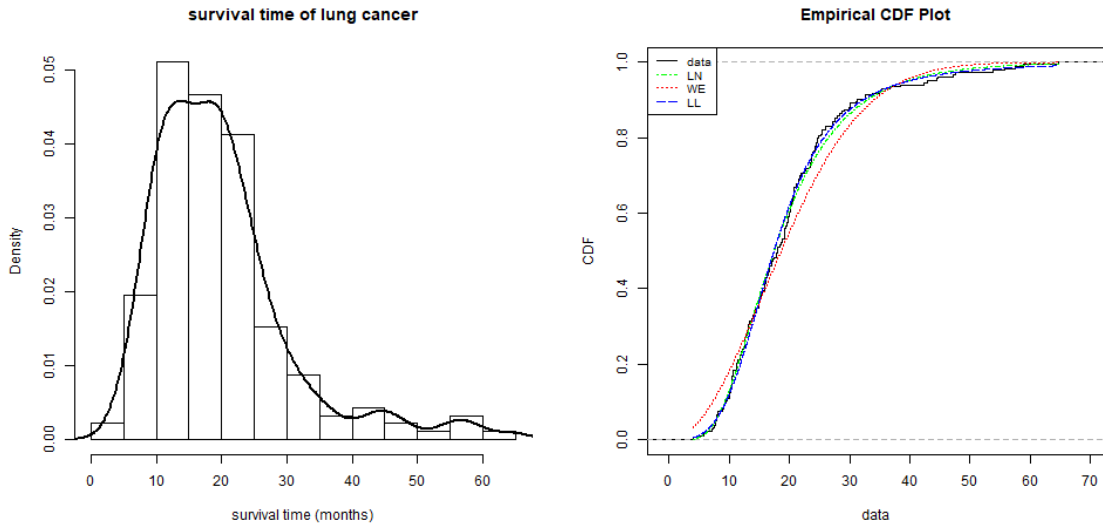
Table 12: Summary of survival times of lung cancer patients.

5(a) and skewness coefficient in Table 12 suggests that the distribution of dataset 2 is right-skewed. The empirical and fitted distribution functions, presented in Figure 5(b), indicates that both log-normal and log-logistic distributions may fit the data well. K-S distances and

p-values of goodness-of-fit tests in Table 13 also supports this fact though, in this case, one may consider log-normal to be more appropriate model than log-logistic. Fixed-sample-size

	log-normal	Weibull	log-logistic
K-S distance	0.0434	0.0979	0.0496
p-value	0.8776	0.0587	0.7558

Table 13: K-S distances and p-values of goodness-of-fit tests for dataset 2.



(a) Histogram.

(b) Empirical and fitted distribution functions.

Figure 5: Survival times (monthly) of lung cancer patients.

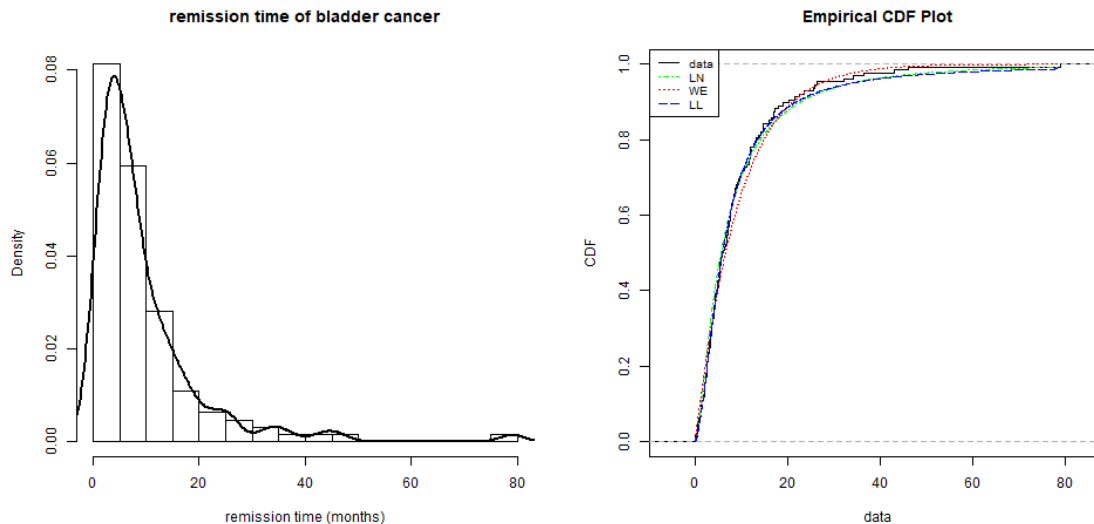
procedure of Raqab et al. (2018) (after resampling with replacement) selects log-normal and log-logistic distribution with probabilities 0.596 and 0.404 respectively and it never selects Weibull distribution. The sequential procedure (10) with pilot sample size $k = 10$ and boundaries $\{(b_i = -1.5, a_i = 1.5) : i = 1, 2, 3\}$ selects log-normal, log-logistic and Weibull with probabilities 0.524, 0.296 and 0.18 respectively. Performance of the sequential discrimination approach is good in the sense that it selects (with highest probability) the distribution that has smallest K-S distance from the empirical CDF. Moreover, the estimated average sample size $\hat{E}(T)$ is about 133.33 which is much less than $n_F = 184$ giving 27.54% savings in the average sample size. If we wish to select the best model between only two distributions, namely log-normal and log-logistic, the sequential procedure 7 with pilot sample size $k = 10$ and boundaries $(b = -1.5, a = 1.5)$ chooses log-normal and log-logistic model with probabilities 0.60 and 0.40 respectively and the estimated average sample size is 137.42.

Dataset 3: The third dataset, adapted from Lee and Wang (2003), consists of monthly remission times of $n_F = 128$ patients having bladder cancer. Al-Shomrani et al. (2016) provides the full data (summarized in Table 14) and suggests fitting a *log-logistic* distribution using maximum likelihood estimates. Histogram in Figure 6(a) and skewness coefficient in

min	1st quantile	median	mean	3rd quantile	max	skewness coefficient
0.0800	3.2950	6.0500	9.2875	11.8375	79.0500	3.2404

Table 14: Summary of remission times of bladder cancer patients.

Table 14 indicates that a positively skewed distribution may fit the data well. Figure 6(b) illustrates that all three fitted distribution functions are close to the empirical distribution function. Hence, selecting the best model among the three competing models becomes quite challenging. Based on K-S distances and corresponding p-values of goodness-of-fit tests in



(a) Histogram.

(b) Empirical and fitted distribution functions.

Figure 6: Remission times (monthly) of bladder cancer patients.

Table 15, one may prefer the log-logistic model than the other two. The sequential procedure

	log-normal	Weibull	log-logistic
K-S distance	0.0635	0.0721	0.0393
p-value	0.6802	0.5187	0.9889

Table 15: K-S distances and p-values of goodness-of-fit tests for dataset 3.

(10) with $k = 10$ and boundaries $\{(b_i = -0.8, a_i = 3.5) : i = 1, 2, 3\}$ selects log-normal,

Weibull and log-logistic with probabilities 0.092, 0.356 and 0.552 respectively whereas the fixed-sample-size method of Raqab et al. (2018) selects log-normal, Weibull and log-logistic with probabilities 0.128, 0.34 and 0.532 respectively. Similar to the cases of datasets 1 and 2, the sequential procedure selects the distribution (with highest probability) that has the smallest K-S distance from the empirical distribution. Even in this difficult case, the estimated average sample size $\widehat{E}(T)$ is about 105.87 which is less than the fixed sample size $n_F = 128$ leading to 17.29% savings in the average sample size.

6 Discussion: Discriminating M Distributions

In this section, sequential procedures of Sections 2 and 3 are extended to the problem of discriminating $M(> 2)$ number of distributions. Suppose X_1, X_2, \dots is a sequence of i.i.d. random variables from one of the distributions $\{f_i(\cdot; \underline{\theta}_i) : i = 1, 2, \dots, M\}$. As the data is observed sequentially, our goal is to select the distribution that fits the observed data sequence best. This can be done by simultaneously testing

$$H_i : X_1, X_2, \dots \sim f_i(\cdot; \underline{\theta}_i), \quad i = 1, 2, \dots, M. \quad (12)$$

Borrowing the ideas from MSPRT (Baum and Veeravalli (1994), Tartakovsky et al. (2014)), a sequential discrimination procedure can be developed based on log RML functions

$$\Lambda_{ij}^n = \ln \left(\frac{\prod_{k=1}^n f_i(X_k; \widehat{\underline{\theta}}_{in})}{\prod_{k=1}^n f_j(X_k; \widehat{\underline{\theta}}_{jn})} \right) \quad i, j = 1, 2, \dots, M, i \neq j,$$

where $\widehat{\underline{\theta}}_{in}$ is the maximum likelihood estimator of $\underline{\theta}_i$ based on X_1, \dots, X_n . Suppose $A = [a_{ij}]$ is a prefixed boundary matrix where $a_{ij} > 0$ for all $i \neq j$ and $k > 1$ is some prefixed pilot sample size.

Let $T_i = \inf\{n \geq k : \Lambda_{ij}^n \geq a_{ij} \text{ for all } j \neq i\}$ be the first time when we observe enough evidence in favor of H_i to accept it. Therefore, we need to continue sampling one observation at a time until the stopping time

$$T^* = \inf \{n \geq k : \text{for some } i, \Lambda_{ij}^n \geq a_{ij} \text{ for all } j \neq i\} = \min\{T_1, \dots, T_M\}.$$

At stopping T^* , accept H_i if $T^* = T_i$. Under the finiteness of second moments of RML functions, the stopping time T^* can be shown to be finite with probability 1.

In M -hypotheses testing, we encounter error probabilities $P(\text{accept } H_j | H_i \text{ true})$ for all $i, j = 1, 2, \dots, M$ and $i \neq j$. Fix levels $\{\alpha_{ij} \in (0, 1) : i, j = 1, \dots, M, i \neq j\}$ and $\{\alpha_i : i = 1, \dots, M\}$. In the context of MSPRT, if boundaries $\{a_{ij} = \ln(1/\alpha_{ij}), i \neq j\}$ are used,

we can control $P(\text{accept } H_j | H_i \text{ true})$ at level α_{ij} for all $i \neq j$. Alternatively, if boundaries $\{a_{ij} = \ln\left(\frac{M-1}{\alpha_i}\right) : i, j = 1, \dots, M, i \neq j\}$ are used, $P(\text{reject } H_i | H_i \text{ true})$ is controlled at level α_i for all $i = 1, \dots, M$. MSPRT with either of these thresholds is first-order asymptotically optimal for iid case and also for some non-iid cases (Tartakovsky et al. (2014)). Thus we can adopt these boundaries for testing (12), and hope to have PCS_{H_i} (probability of accepting H_i when H_i is true) at least $(1 - \sum_{j \neq i} \alpha_{ij})$ for all $i = 1, \dots, M$ for the first choice of boundaries and $PCS_{H_i} \geq 1 - \alpha_i$ for the second choice of boundaries. However, a mathematical proof of this fact for finite sample is an open problem.

7 Concluding Remarks

Likelihood based sequential testing approaches are proposed to select the most appropriate model from two or three competing models, namely log-normal, Weibull and log-logistic. Stopping boundaries used for our methods are adapted from well-known sequential tests GSLRT and MSPRT that are proven to be asymptotically optimal (Lai et al. (1988), Dragalin et al. (1999)) as error probabilities tend to zero (equivalently, boundaries tend to infinity). We prove consistency of the proposed sequential approaches by showing that attained PCS is asymptotically one. However, providing an exact (non-asymptotic) upper bound on error probabilities or an exact lower bound on PCS with the proposed boundaries and finite sample size is extremely challenging, if not impossible (see, e.g., Tartakovsky (2014)). Therefore, finite sample properties of sequential methods with the prescribed Wald-type boundaries are studied through Monte Carlo simulations, and it is shown that PCS attains or exceeds the desired level in all cases considered. More importantly, to attain same PCS, sequential procedures require much fewer (sometimes even three times fewer) sample sizes on average compared to their fixed-sample-size competitors available in literature leading to huge savings in sampling costs. Selection of appropriate boundaries are crucial for savings in sample sizes. If symmetric boundaries are used in (10), Figure 3(a) may guide practitioners about attained PCS (roughly within some interval) for some chosen boundary $|sb|$.

Instead of using frequentist approach via RML functions, one can possibly develop sequential discrimination methods using Bayesian framework for multi-hypotheses as given in Tartakovsky et al. (2014). To avoid the inconvenience of purely sequential sampling, two-stage methodologies in the spirit of Stein's two-stage method Stein (1945) may be designed for discrimination problems. Such methodologies may be investigated in future.

Appendix

Proof of Lemma 1

Proof. We prove this lemma under $H_0^{(1)}$. The proof for the other case is similar. Note that

$$\begin{aligned} P(T_1 = \infty) &= \lim_{n \rightarrow \infty} P(T_1 > n) \leq \lim_{n \rightarrow \infty} P(b < \Lambda_n^{(1)} < a) \\ &= \lim_{n \rightarrow \infty} \left[\Phi \left(\frac{a/n - E(\Lambda_n^{(1)})/n}{\frac{1}{\sqrt{n}} \sqrt{\text{Var}(\Lambda_n^{(1)})/n}} \right) - \Phi \left(\frac{b/n - E(\Lambda_n^{(1)})/n}{\frac{1}{\sqrt{n}} \sqrt{\text{Var}(\Lambda_n^{(1)})/n}} \right) \right] \end{aligned} \quad (13)$$

Since, from Theorem 2 of Kundu and Manglick (2004), under the assumption that $H_0^{(1)}$ is true, the distribution of $\Lambda_n^{(1)}$ is asymptotically normal with mean $E(\Lambda_n^{(1)})$ and variance $\text{Var}(\Lambda_n^{(1)})$. Here, Φ denotes the CDF of standard normal random variable. Kundu and Manglick (2004) shows that $\lim_{n \rightarrow \infty} E(\Lambda_n^{(1)})/n$ and $\lim_{n \rightarrow \infty} \text{Var}(\Lambda_n^{(1)})/n$ are negative and positive constants (finite) respectively. Therefore, (13) yields $P(T = \infty) = 1 - 1 = 0$. This completes the proof. \square

Proof of Lemma 2

Proof. (i) Assume $H_A^{(1)}$ is true. Let $\underline{P} = (\sigma, \eta, \tilde{\gamma}, \tilde{\theta})$, $\hat{\underline{P}} = (\hat{\sigma}_n, \hat{\eta}_n, \hat{\gamma}_n, \hat{\theta}_n)$, $g(X_i|\hat{\underline{P}}) = \ln \left(\frac{f_{LN}(X_i; \hat{\sigma}_n, \hat{\eta}_n)}{f_{WE}(X_i; \hat{\gamma}_n, \hat{\theta}_n)} \right)$. Using Taylor expansion at \underline{P} we have,

$$\begin{aligned} g(X_i|\hat{\underline{P}}) &= g(X_i|\underline{P}) + (\hat{\sigma}_n - \sigma) \frac{\partial}{\partial \sigma} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\sigma=\sigma_0} + (\hat{\eta}_n - \eta) \frac{\partial}{\partial \eta} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\eta=\eta_0} \\ &\quad + (\hat{\gamma}_n - \tilde{\gamma}) \frac{\partial}{\partial \tilde{\gamma}} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\tilde{\gamma}=\tilde{\gamma}_0} + (\hat{\theta}_n - \tilde{\theta}) \frac{\partial}{\partial \tilde{\theta}} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\tilde{\theta}=\tilde{\theta}_0}, \end{aligned} \quad (14)$$

where $\sigma_0 = p_1 \hat{\sigma}_n + (1-p_1)\sigma$, $\eta_0 = p_2 \hat{\eta}_n + (1-p_2)\eta$, $\tilde{\gamma}_0 = p_3 \hat{\gamma}_n + (1-p_3)\tilde{\gamma}$, $\tilde{\theta}_0 = p_4 \hat{\theta}_n + (1-p_4)\tilde{\theta}$ and $p_i \in (0, 1)$ for $i = 1, 2, 3, 4$. It is easy to check that that $\frac{\partial}{\partial \sigma} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\sigma=\sigma_0}$ converges to finite random variable $\frac{\partial}{\partial \sigma} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})$. Therefore, $\hat{\sigma}_n \xrightarrow{a.s.} \sigma$ yields,

$$(\hat{\sigma}_n - \sigma) \frac{\partial}{\partial \sigma} g(X_i|\sigma, \eta, \tilde{\gamma}, \tilde{\theta})|_{\sigma=\sigma_0} \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty.$$

This holds for all others terms of RHS of (14). Hence, $g(X_i|\hat{\underline{P}}) - g(X_i|\underline{P}) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. Thus, the Cesaro mean

$$\frac{\Lambda_n^{(1)}}{n} - \frac{1}{n} \sum_{i=1}^n g(X_i|\underline{P}) = \frac{1}{n} \sum_{i=1}^n \left(g(X_i|\hat{\underline{P}}) - g(X_i|\underline{P}) \right) \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty. \quad (15)$$

Since $\frac{1}{n} \sum_{i=1}^n g(X_i | \underline{P}) \xrightarrow{a.s.} E_{H_A^{(1)}}[g(X | \underline{P})] = \mu_{LN}$ as $n \rightarrow \infty$ by strong law of large number,

we conclude from (15) that $\frac{\Lambda_n^{(1)}}{n} \xrightarrow{a.s.} \mu_{LN}$. The proof is similar when $H_0^{(1)}$ is true.

(ii) An application of Jensen's inequality yields $\mu_{WE} < 0 < \mu_{LN}$. Using part (i) of Lemma 2, we have $\Lambda_n^{(1)} \xrightarrow{a.s.} \infty$ or $-\infty$ when $H_A^{(1)}$ or $H_0^{(1)}$ is true as $n \rightarrow \infty$. \square

Proof of Theorem 1

Proof. (a) Assume that $H_0^{(1)}$ is true, i.e., data follows Weibull distribution.

(i) Note that if $\Lambda_n^{(1)}$ is a random variable then $\sup_n \Lambda_n^{(1)}$ is also a random variable. Following the argument given in Theorem 2.1 of Berk (1973), $\sup_n \Lambda_n^{(1)}$ is finite with probability 1. Therefore,

$$I(\Lambda_{T_1}^{(1)} \geq a) \leq I(\sup_n \Lambda_n^{(1)} \geq a) \rightarrow 0 \text{ as } a \rightarrow \infty \text{ with probability 1.}$$

Thus, $\lim_{a \rightarrow \infty} I(\Lambda_{T_1}^{(1)} \geq a) = 0$ with probability 1. By Lebesgue's dominated convergence theorem

$$\lim_{a \rightarrow \infty} P(\Lambda_{T_1}^{(1)} \geq a) = \lim_{a \rightarrow \infty} E[I(\Lambda_{T_1}^{(1)} \geq a)] = E\left[\lim_{a \rightarrow \infty} I(\Lambda_{T_1}^{(1)} \geq a)\right] = 0.$$

(ii) Since $\frac{\Lambda_n^{(1)}}{n} \xrightarrow{a.s.} \mu_{WE}$ as $n \rightarrow \infty$ and $T_1 \xrightarrow{a.s.} \infty$ as $b \rightarrow -\infty$, using Theorem 2.1 of Gut (2009), $\frac{\Lambda_{T_1}^{(1)}}{T_1} \xrightarrow{a.s.} \mu_{WE}$ as $b \rightarrow -\infty$. From definition of T_1 ,

$$\Lambda_{T_1-1}^{(1)} I(\Lambda_{T_1}^{(1)} \leq b) \geq b I(\Lambda_{T_1}^{(1)} \leq b) \geq \Lambda_{T_1}^{(1)} I(\Lambda_{T_1}^{(1)} \leq b).$$

Dividing all parts of this inequality by T_1 and taking $b \rightarrow -\infty$

$$\lim_{b \rightarrow -\infty} \frac{\Lambda_{T_1-1}^{(1)}}{T_1} I(\Lambda_{T_1}^{(1)} \leq b) \geq \lim_{b \rightarrow -\infty} \frac{b}{T_1} I(\Lambda_{T_1}^{(1)} \leq b) \geq \lim_{b \rightarrow -\infty} \frac{\Lambda_{T_1}^{(1)}}{T_1} I(\Lambda_{T_1}^{(1)} \leq b).$$

Using part (i) of Theorem 1, both extreme terms are μ_{WE} with probability 1. Therefore,

$\lim_{b \rightarrow -\infty} \frac{b}{T_1} I(\Lambda_{T_1}^{(1)} \leq b) = \mu_{WE}$ implies that, $\lim_{b \rightarrow -\infty} \frac{b}{T_1} = \mu_{WE}$, since $\lim_{b \rightarrow -\infty} I(\Lambda_{T_1}^{(1)} \leq b) = 1$ (using part (i)). Therefore,

$$\lim_{b \rightarrow -\infty} \frac{T_1}{b} = \frac{1}{\mu_{WE}} \text{ w.p.1.}$$

By Fatou's lemma,

$$\liminf_{b \rightarrow -\infty} E \left[\frac{T_1}{-b} \right] \geq -\frac{1}{\mu_{WE}} \text{ implies that } \limsup_{b \rightarrow -\infty} E \left[\frac{T_1}{b} \right] \leq \frac{1}{\mu_{WE}}.$$

(b) Proof is similar as part (a), and hence, omitted. \square

Proof of Lemma 3

Proof. Let us provide the proof assuming H_1 , that is, the data is coming from log-normal distribution.

$$\begin{aligned} P_{H_1}(T = \infty) &= \lim_{n \rightarrow \infty} P_{H_1}(T > n) \\ &\leq \lim_{n \rightarrow \infty} P_{H_1}(\{\Lambda_n^{(1)} < a_1 \cup \Lambda_n^{(3)} < a_3\} \cap \{\Lambda_n^{(1)} > b_1 \cup \Lambda_n^{(2)} < a_2\} \cap \{\Lambda_n^{(2)} > b_2 \cup \Lambda_n^{(3)} > b_3\}) \\ &\leq \lim_{n \rightarrow \infty} P_{H_1}(\{\Lambda_n^{(1)} < a_1 \cup \Lambda_n^{(3)} < a_3\}) \\ &\leq \lim_{n \rightarrow \infty} P_{H_1}(\Lambda_n^{(1)} < a_1) + \lim_{n \rightarrow \infty} P_{H_1}(\Lambda_n^{(3)} < a_3) = 0. \end{aligned}$$

As the data is coming from log-normal distribution, it is easy to verify that $\Lambda_n^{(1)} \rightarrow \infty$ and $\Lambda_n^{(3)} \rightarrow \infty$ with probability 1 following the arguments of Lemma 2. Hence, $P_{H_1}(\Lambda_n^{(1)} < a_1) \rightarrow 0$ and $P_{H_1}(\Lambda_n^{(3)} < a_3) \rightarrow 0$ for any fixed $a_1, a_3 > 0$ as $n \rightarrow \infty$. Proof is similar when the data is coming from Weibull or log-logistic distribution, and hence, omitted. \square

References

- Ali A Al-Shomrani, AI Shawky, Osama H Arif, and Muhammad Aslam. Log-logistic distribution for survival data analysis using mcmc. *SpringerPlus*, 5(1):1774, 2016.
- Per K Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.
- P Armitage. Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):137–144, 1950.
- Carl W Baum and Venugopal V Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6), 1994.
- Robert H Berk. Some asymptotic aspects of sequential analysis. *The Annals of Statistics*, pages 1126–1138, 1973.

- Arabin Kumar Dey and Debasis Kundu. Discriminating between the log-normal and log-logistic distributions. *Communications in Statistics-Theory and Methods*, 39(2):280–292, 2009a.
- Arabin Kumar Dey and Debasis Kundu. Discriminating among the log-normal, weibull, and generalized exponential distributions. *IEEE Transactions on reliability*, 58(3):416–424, 2009b.
- VP Dragalin, Alexander G Tartakovsky, and Venugopal V Veeravalli. Multihypothesis sequential probability ratio tests. i. asymptotic optimality. *IEEE Transactions on Information Theory*, 45(7):2448–2461, 1999.
- Robert Dumonceaux and Charles E Antle. Discrimination between the log-normal and the weibull distributions. *Technometrics*, 15(4):923–926, 1973.
- Robert Dumonceaux, Charles E Antle, and Gerald Haas. Likelihood ratio test for discrimination between two models with unknown location and scale parameters. *Technometrics*, 15(1):19–27, 1973.
- AE Elsherpieny, NS Ibrahim, and UN Radwan. Discriminating between weibull and log-logistic distributions. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(8):3358–3371, 2013.
- Rameshwar D Gupta and Debasis Kundu. Discriminating between weibull and generalized exponential distributions. *Computational statistics & data analysis*, 43(2):179–196, 2003.
- Rameshwar D Gupta and Debasis Kundu. Discriminating between gamma and generalized exponential distributions. *Journal of Statistical Computation & Simulation*, 74(2):107–121, 2004.
- Allan Gut. *Stopped random walks*. Springer, 2009.
- Debasis Kundu and Anubhav Manglick. Discriminating between the weibull and log-normal distributions. *Naval Research Logistics (NRL)*, 51(6):893–905, 2004.
- Tze Leung Lai et al. Nearly optimal sequential tests of composite hypotheses. *The Annals of Statistics*, 16(2):856–886, 1988.
- Elisa T Lee and John Wang. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.
- AW Marshall, Juan C Meza, and Ingram Olkin. Can data recognize its parent distribution? *Journal of Computational and Graphical Statistics*, 10(3):555–580, 2001.

- Stephen Overduin. *Use of the lognormal distribution for survival data: inference and robustness*. PhD thesis, Department of Statistics and Actuarial Science-Simon Fraser University, 2004.
- Mohammad Z Raqab. Discriminating between the generalized rayleigh and weibull distributions. *Journal of Applied Statistics*, 40(7):1480–1493, 2013.
- Mohammad Z Raqab, Shafiqah A Al-Awadhi, and Debasis Kundu. Discriminating among weibull, log-normal, and log-logistic distributions. *Communications in Statistics-Simulation and Computation*, 47(5):1397–1419, 2018.
- Maintainer Thomas Scheike. Package timereg. 2019.
- Gideon Schwarz. Asymptotic shapes of bayes sequential testing regions. *The Annals of mathematical statistics*, pages 224–236, 1962.
- Milton Sobel and Abraham Wald. A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *The annals of mathematical statistics*, 20(4):502–522, 1949.
- Charles Stein. A two-sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics*, 16(3):243–258, 1945.
- Patricia Tai, Jon Tonita, Edward Yu, and David Skarsgard. Twenty-year follow-up study of long-term survival of limited-stage small-cell lung cancer and overview of prognostic and treatment factors. *International Journal of Radiation Oncology, Biology, Physics*, 56(3):626–633, 2003.
- Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: hypothesis testing and changepoint detection*. Chapman and Hall/CRC, 2014.
- Alexander G Tartakovsky. Nearly optimal sequential tests of composite hypotheses revisited. *Proceedings of the Steklov Institute of Mathematics*, 287(1):268–288, 2014.
- Abraham Wald. Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, 16(2):117–186, 1945.
- Abraham Wald, Jacob Wolfowitz, et al. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.
- Halbert White. Regularity conditions for cox’s test of non-nested hypotheses. *Journal of Econometrics*, 19(2-3):301–318, 1982.