Distinguishing between the Exponential and Lindley distributions. An Illustration from Experimental Psychology

Shovan Chowdhury [1], Marco Marozzi [2], Freddy Hernández-Barajas [3], and Fernando Marmolejo-Ramos [4]

[1]Quantitative Methods and Operations Management Area, Indian Institute of Management Kozhikode, Kozhikode, Kerala, India. Email: meetshovan@gmail.com
[2]Department of Mathematics and Computer Science, University of Ferrara, Ferrara, Italy. Email: marco.marozzi@unife.it
[3]Department of Statistics, National University of Colombia, Medellín, Colombia. Email: fhernanb@unal.edu.co
[4]College of Education, Psychology, and Social Work, Flinders University, Adelaide, Australia. Email: fernando.marmolejoramos@flinders.edu.au

January 8, 2025

Abstract

The exponential distribution has been a popular choice for modeling positively skewed data in experimental psychology. However, the lesser-known Lindley distribution, despite not being typically used for this purpose, shares similar density and cumulative shapes with the exponential distribution. This similarity suggests that the Lindley distribution could be a strong candidate for modeling such data types. While the probability density and cumulative distribution functions of these two one-parameter distributions can be quite similar, their hazard rate functions differ. Therefore, selecting the most appropriate distribution significantly impacts the interpretation of the hazard rate function. To aid in this selection, we introduce a method that distinguishes between the exponential and Lindley distributions by examining the ratio of their maximized likelihood functions. This method is versatile, as it can also be applied to type I censored data, enhancing its practical appeal. We conducted a simulation study to demonstrate the method's effectiveness, even with small sample sizes. Furthermore, we illustrate the method's application using a published dataset from experimental psychology and provide an implementation as an R function.

Keywords: Censored data; Exponential distribution; Hazard rate function; Likelihood ratio test; Lindley distribution; distributional regression.

1 Introduction

In distributional regression modeling [19] (e.g. Generalized Additive Models for Location, Scale, and Shape (GAMLSS) modeling), two different distributions with the same number of parameters may fit the response variable equally well (e.g. their Akaike Information Criterion [AIC] or Bayesian Information Criterion [BIC] estimates may be too close, and their probability density functions [pdf] and cumulative distribution functions [cdf] will have a large overlap), making it problematic to decide which distribution is best. While differences between distributions are usually examined via their pdfs or cdfs, it has recently been suggested that examining the hazard rate functions (hrf) of the distributions may help to better discriminate between distributions and provide novel information [30].

This study emphasizes the critical need to choose the appropriate distribution for accurately interpreting hrfs and introduces an innovative method based on the ratio of the likelihood functions of different distributions. Our study tackles the crucial issue of distinguishing between the exponential distribution, widely used for modeling positively skewed psychological data such as response times (see e.g. [14]), and the less commonly employed but potentially significant Lindley distribution. The proposed discrimination method demonstrates effective handling of type I censored data, making it very useful in practice given the occurrence of this type of data in this context.

The paper is structured as follows. First, we discuss the importance of choosing an appropriate probability distribution and its impact on interpreting hazard rates. Second, we briefly explore the role of hrfs in psychological data analysis. Third, we examine the statistical properties of the exponential and Lindley distributions. Fourth, we introduce a novel discriminating method based on the logarithm of the ratio of maximized likelihoods (RML). We then describe the selection procedure for distinguishing between the exponential and Lindley distributions, accompanied by a comprehensive simulation study. The following section demonstrates the proposed method with a practical application to cognitive tasks. Concluding remarks are provided in the final section, and the Appendix contains the asymptotic results. The supplementary material includes the implementation of the proposed method in R and additional examples and simulation results (see https://github.com/fhernanb/T_exp_lin).

2 The influence of probability distributions on hazard rates

In linear regression, the dependent variable's location parameter is modeled through the mean of the conditional distribution and the normal distribution is used by default. That is, in linear regression, it is assumed that the residuals are normally distributed. In practice, however, continuous dependent variables often deviate from normality, so linear regression may not guarantee normally distributed residuals. Generalized linear models (GLM) extend linear regression by accommodating response variables with various error distribution models

from the exponential family, including normal, binomial, Poisson, gamma, and exponential distributions. Generalized additive models for location, scale, and shape (GAMLSS) enhance the capabilities of traditional regression models, such as linear regression and GLMs, by offering several advantages. These include the ability to utilize a broader range of distributions beyond those typically employed in GLMs, thereby providing increased flexibility for different types of data [19, 34]. That is, GAMLSS allows the selection of an optimal distribution that best fits the response variable. Furthermore, GAMLSS modeling encourages researchers to explore beyond "mean differences" [20, 21] and to examine all parameters of probability distributions in their statistical analyses [36].

GAMLSS modeling may reveal scenarios where different distributions with the same number of parameters fit a dataset equally well, as indicated by goodness-of-fit metrics like AIC or BIC. In such cases, the choice of distribution may have minimal practical impact on regression estimates, with the similarity visually confirmed by comparing their pdfs or cdfs (e.g. see Figures 2 and 3). For example, the probability of a number being smaller than 1 in an exponential distribution with $\lambda = 1.944$ and a Lindley distribution with $\theta = 2.5$ are 0.8568 and 0.8592 respectively (i.e. $\approx 86\%$); these probabilities are very similar (see the plot in row 1, column 3 in Figure 3). However, their hrfs convey quite different narratives (see the plot in row 1, column 3 in Figure 4).

The hazard rate function (hrf) or failure rate, quantifies the instantaneous risk or rate at which an event of interest (such as a response, failure, or occurrence) happens at a specific time t, given that the event has not occurred before time t. In other words, it represents the likelihood that an event will occur in the next instant, given that it has not occurred up to that point. The hazard rate function $h(t)$ is defined as the ratio of the probability density function $f(t)$ to the survival function $S(t)$. The survival function $S(t)$ gives the probability that the event has not occurred by time t, and it is defined as $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function (i.e. for any given value x, the cdf gives the probability that the random variable X will take on a value less than or equal to x; $F(x) = P(X \leq x)$). That is, $h(t) = \frac{f(t)}{S(t)}$. It is thus clear that while the pdf influences the hazard rate by determining the instantaneous likelihood of the event at a given time (e.g., a higher pdf value at time t generally leads to a higher hazard rate at that time, assuming the survival function does not change drastically), the cdf affects the hazard rate through the survival function (e.g., as the cdf increases [meaning the probability of the event having occurred by time t increases], the survival function decreases. This, in turn, generally leads to an increase in the hazard rate). All of this underscores that the choice of a probability distribution directly impacts the hazard rate function and its interpretation.

Hazard rate functions can exhibit various shapes (see Figure 4 and also Figure 1A in Panis et al. [30]), which in turn influence their interpretation. Some of these shapes include a constant hazard rate, indicating that the risk of an event remains consistent over time, suggesting a random failure pattern (e.g., exponential and Poisson distributions); an increasing

hazard rate, implying that the risk of an event increases with time, as seen in wear-out processes (e.g., Weibull [with shape parameter > 1] and Lindley distributions); a decreasing hazard rate, indicating that the risk of an event decreases over time, suggesting improvement (e.g., Gamma [with shape parameter < 1] and Pareto distributions); a bathtub-shaped hazard rate, which combines an initial decreasing hazard rate, followed by a constant period, and then an increasing hazard rate (e.g., Hjorth and Chen distributions); and a hump-shaped hazard rate, characterized by an initial increasing hazard rate that reaches a peak and then declines (e.g., Birnbaum-Saunders and log-normal distributions). Note also that certain probability distributions can display various hrfs shapes. For instance, when the shape parameters in the two-parameter Weibull and Gamma distributions are greater than 1, equal to 1, or less than 1, they exhibit increasing, constant, and decreasing hrfs, respectively.

There could be various scenarios where these hazard rate functions are applicable. For instance, in a simple reaction time task with random intervals, participants press a button when a light appears, and the probability of responding remains constant once they are ready. Since this task involves pure detection processes without strategic components, a constant hazard rate shape could result. A different hrf can occur in a working memory task where participants must retain information in memory, and the risk of error increases as cognitive load accumulates. Such tasks can lead to fatigue or resource depletion effects, and thus an increasing hazard rate would be expected. When learning a new motor skill, initial errors are common but decrease with practice, so the risk of mistakes is highest at the start and then improves. This performance demonstrates learning and adaptation and could be represented by a decreasing hazard rate. In a complex decision-making task, an initially high error rate is expected during the task understanding phase, followed by stable performance during the optimal phase, and then an increase in errors due to fatigue. This task exhibits learning, stability, and fatigue phases, and thus a bathtub-shaped hazard rate could describe such a process. Finally, in a response priming task, there is an initial increase in response probability, a peak at the optimal response time, and then a decline as the response window closes. This task reveals the temporal dynamics of response activation, and its hazard rate function could have a hump shape.

The units of the hazard rate are typically expressed as "events per unit of time." While the hazard rate is always non-negative (i.e., there cannot be a negative risk), there is no upper limit on its potential value. Consider a scenario where participants undergo a "digit recognition memory task." In this task, participants are shown a sequence of single digits (0-9) and must quickly identify if a digit has appeared previously in the sequence, with a response time limit of 2,000 milliseconds. One experimental condition, termed "immediate recognition," requires participants to compare the current digit only with the immediately preceding trial. As the working memory load remains constant in this condition, a constant hazard rate is expected, reflecting the relatively simple comparison process. In contrast, the "cumulative recognition" condition demands that participants compare the current digit against all previously presented

digits in the sequence. This leads to an increasing memory load as the sequence progresses. Consequently, an increasing hazard rate is expected due to the growing demands of memory search. This phenomenon is illustrated in the first row of Figure 1.

In terms of interpreting the hazard rate function plots, it is evident that the instantaneous rate of response occurrence differs between the "immediate recognition" and "cumulative recognition" tasks. For instance, if the response time has reached 1,000ms, the hazard rate is higher in the "immediate recognition" condition (represented by a Gamma(1,100) distribution with a rate of approximately 0.01 events per millisecond) compared to the "cumulative recognition" condition (represented by a Gamma(1.5,100) distribution with a rate of approximately 0.0096 events per millisecond). As illustrated in the plot in the first row, second column of Figure 1, the probability of responding in the "immediate recognition" task stays relatively constant over time, assuming no response has occurred. Conversely, the "cumulative recognition" task shows an increasing probability of response over time, signaling a growing certainty of a response as the memory search becomes more challenging. The second row in Figure 1 illustrates this phenomenon using a different time scale (seconds) for the same mock task, further emphasizing the impact of memory load on the hazard rate function.

3 What hazard rate functions offer to data analysis

Analysis of variance (ANOVA) is the most common method to analyze response times in the fields of psychology and cognitive science, despite being widely criticized for its shortcomings when used to analyze such data [30]. Some scholars propose that event history analysis (EHA), an hrf-based method, offers a superior approach compared to ANOVA. As emphasized by Panis et al. (2020) [30], despite EHA already being commonplace in many fields of psychology, like developmental, clinical and social psychology, it is still rather unfamiliar in other fields like experimental psychology. EHA has several advantages over ANOVA when dealing with response times (a ubiquitous metric in experimental psychology), or more generally when the occurrence of an event is of interest:

- The hrf of response occurrence is much more useful in describing the data than pdf or cdf, being able to emphasize important differences between two situations that are not clear when comparing the corresponding pdfs or cdfs. Two hrfs can be sharply different while the corresponding pdfs or cdfs look very alike [16] (see Figures 2, 3, and 4).

- Townsend (1990) [35] showed that the comparison of hrfs allows stronger conclusions than the comparison of distributions or means, because a complete ordering on the hazard functions implies a complete ordering on the cumulative distribution and survival functions that implies an ordering on the two distribution means, whereas the reverse is not always true. Therefore, an ordering on the two distribution means may happen in the presence of either a complete ordering on the hazard functions or crossing hazard functions. In

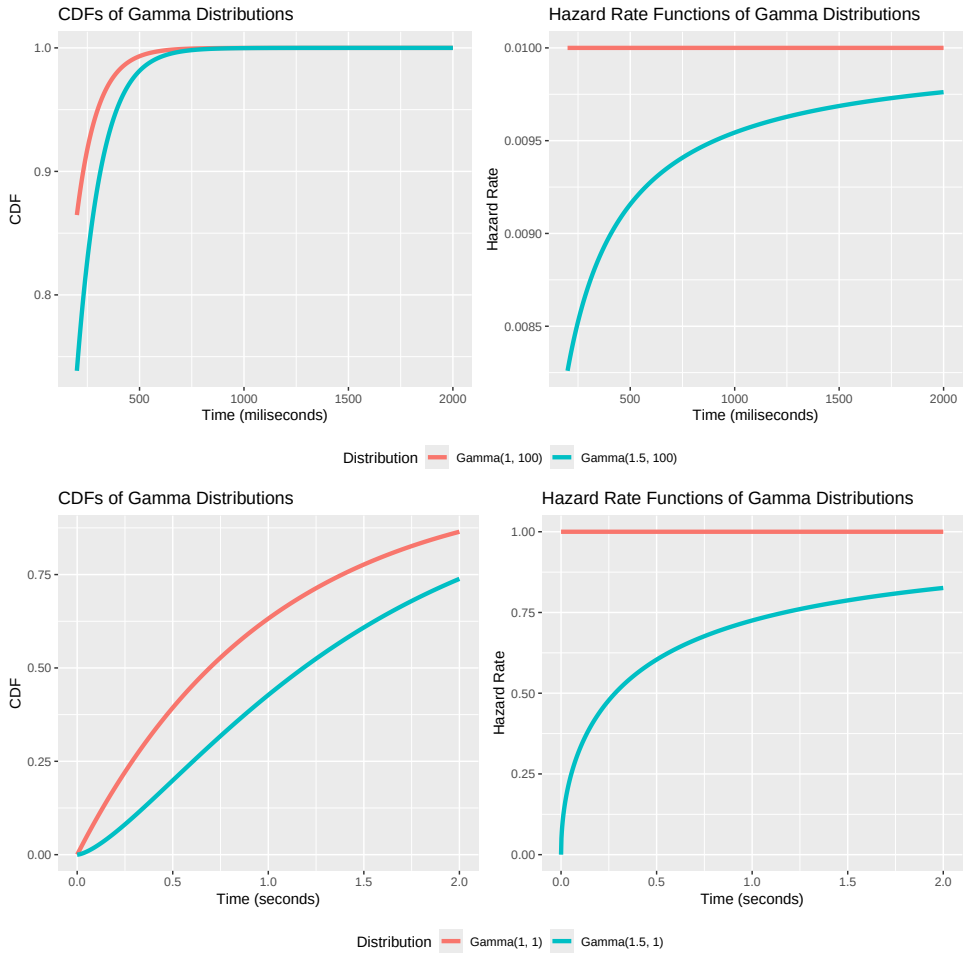

Figure 1: Reaction time distributions modeled with gamma distributions. The distributions have positive skews (first column) but different hazard rates (second column). The first row shows measurements in milliseconds, and the second row shows measurements in seconds.

other words, comparing the means may give a distorted picture of what is really going on.

- Right-censored observations are very common when dealing with response times and can be incorporated into EHA, giving important information often overlooked when using ANOVA.

- Time-varying explanatory covariates are commonly encountered in experimental psychology (e.g. biomarkers related to respondent autonomic nervous system) and they are easier to incorporate into EHA than into ANOVA.

- The hazard function itself is a very important measure in many psychological experiments, measuring the instantaneous failure rate: the instantaneous subject's ability to complete

a task given that the subject has not yet completed the task before. EHA assesses hazard functions directly whereas ANOVA is far less suitable to assess them directly.

The advantages of hrfs in dealing with reaction times are very clear. Still, we believe that hrfs are also very suitable to examine other positively skewed data and in general to compare distributions with different shapes. We thus echo Panis et al. (2020) [30] in that this analytical approach needs to be incorporated into current data modeling practices; particularly, in the framework of distributional regression modeling. Figure 4 shows the hazard for various Lindley distributions and shows that it is very different from the hazard for the exponential distribution, which is constant.

4 The Lindley and exponential distributions

The Lindley distribution with scale parameter θ, written as $Lin(\theta)$, having probability density function (pdf) and cumulative distribution function (cdf)

$$f_L(x;\theta) = \frac{\theta^2}{1+\theta}\,(1+x)\,e^{-\theta x};\ x,\theta > 0, \tag{4.1}$$

and

$$F_L(x;\theta) = 1 - \frac{\theta+1+\theta x}{1+\theta}e^{-\theta x};\ x,\theta > 0, \tag{4.2}$$

respectively, was introduced by Lindley [26]. The pdf is decreasing for $\theta \geq 1$ and is unimodal for $\theta < 1$. It is also known that the hazard function and the mean residual life (MRL) function of the distribution increase for all θ. Several aspects of this distribution are studied in detail by Ghitany *et al.* [13]. It has been found that many of the mathematical properties of the Lindley distribution are more flexible than those of the Exponential distribution. As pointed out in Ghitany *et al.* [13], due to the popularity of the Exponential distribution in statistics and many applied areas, the Lindley distribution has not been very well explored in the literature. Still, the Lindley distribution has found its place in many applications where it is found to be preferable over other competitor distributions. For example, this distribution has been used to model lifetime data, especially in applications relating to stress-strength reliability [26, 27]. Mazucheli and Achcar [29] used the Lindley distribution to model lifetime data relating to competing risks and Ghitany *et al.* [13] used it to model the waiting time of bank customers. That is, the Lindley distribution seems suitable to fit data akin to lifetime data; i.e., survival times, failure times, or fatigue-life data.

The exponential distribution is among the most popular univariate continuous distributions with several significant statistical properties, most importantly, its characterization through lack of memory property. The exponential distribution with scale parameter λ, written as $Exp(\lambda)$, has pdf and cdf given by

$$f_E(x;\lambda) = \lambda e^{-\lambda x};\ x,\lambda > 0 \tag{4.3}$$

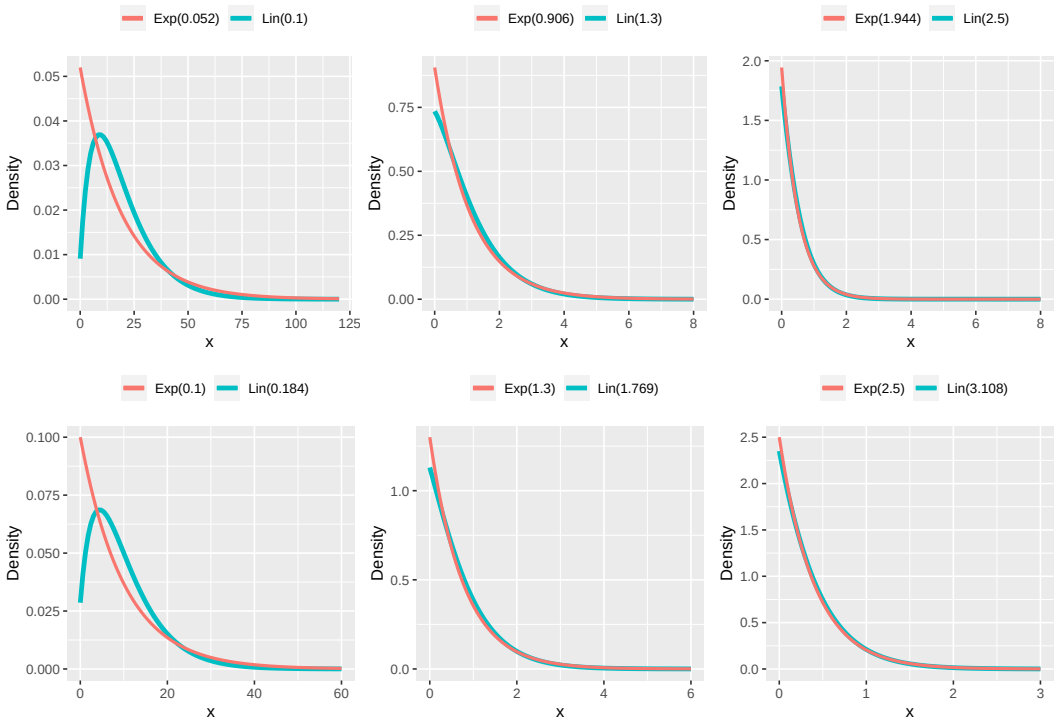

Figure 2: PDF of Lindley (Lin) and exponential (Exp) distributions for different parameters.

and

$$F_E(x;\lambda) = 1 - e^{-\lambda x};\ x, \lambda > 0. \tag{4.4}$$

While the pdf of the exponential is decreasing for all λ, the hazard and MRL functions are constant. This exhibits one of the distinguishable characteristics of exponential and Lindley distributions. However, both one-parameter distributions can be very similarly effective in analyzing positively skewed data. Moreover, the pdfs or cdfs of both distributions can be very close to each other for certain ranges of parameter values. Figures 2 and 3 illustrate the similarity between the distributions when their parameter take different values. Such similarity translates in that while data could come from the *Exp* or *Lin* distribution, the other distribution could provide a very good fit too.

In Table 4 we present a summary with key properties of the exponential and Lindley distributions. Although the two models may provide a similar fit for small or moderate sample sizes, it is still important to select the best fit for a given data set, and much more importantly to select the correct model in those cases where the difference between a constant and a monotonically increasing hazard function has practical consequences.

The problem of selecting the correct distribution is not new in the statistics literature. The

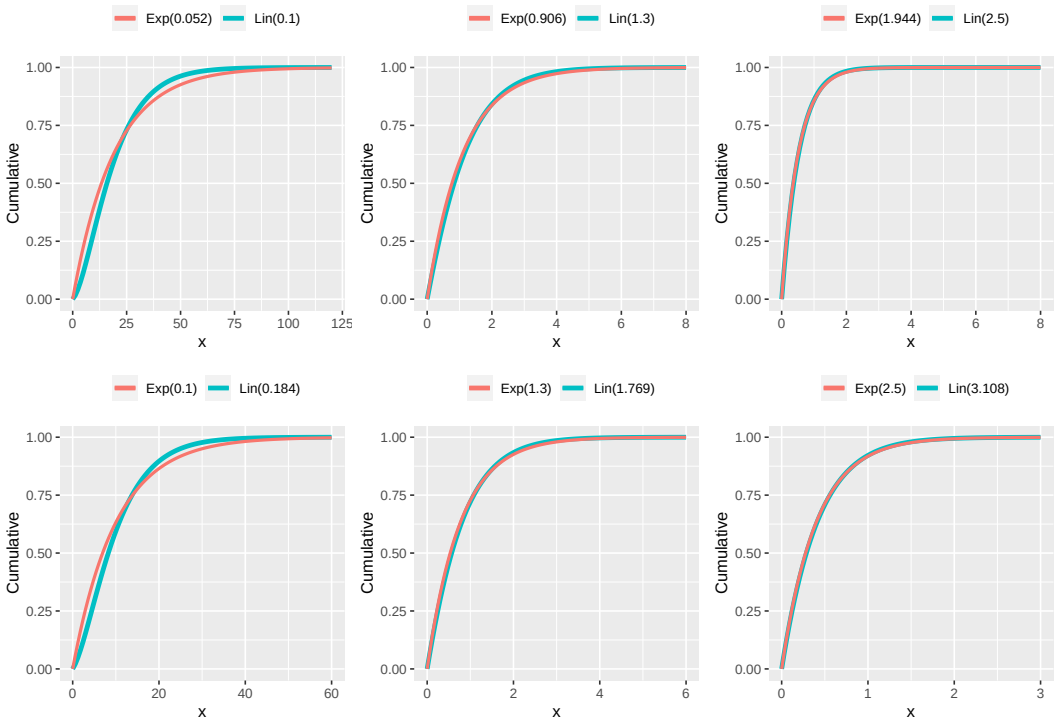

Figure 3: CDF of Lindley (Lin) and Exponential (Exp) distributions for different parameters.

Distribution	$f(x)$	$F(x)$	$S(x)$	$H(x)$	Mean	Median	Variance
Exponential	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$e^{-\lambda x}$	λ	$\frac{1}{\lambda}$	$\frac{ln2}{\lambda}$	$\frac{1}{\lambda^2}$
Lindley	$\frac{\theta^2}{1+\theta}(1+x)e^{-\theta x}$	$1-\frac{\theta+1+\theta x}{\theta+1}e^{-\theta x}$	$\frac{\theta+1+\theta x}{\theta+1}e^{-\theta x}$	$\frac{\theta^2(1+x)}{\theta+1+\theta x}$	$\frac{\theta+2}{\theta(\theta+1)}$	through iteration	$\frac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2}$

Table 1: Key properties of the exponential and Lindley distributions. $f(x)$, $F(x)$, $S(x)$ and $H(x)$ represent the pdf, cdf, survival and hazard functions respectively.

problem of discriminating between two non-nested models for a complete data set was first considered by Cox [7, 8] and was later studied by other scholars including Bain and Engelhardt [3], Chen [6] and Fearn and Nebenzahl [12]. Due to the increasing applicability of distributions suitable to fit lifetime-like data, special attention has been paid to selecting between the Weibull and Log-Normal distributions ([22], [17]), the Gamma and Log-Normal distributions ([23]), the generalized Rayleigh and Log-Normal distribution ([24]), the generalized Rayleigh and Weibull distribution ([33], [1]), the log-normal and generalized exponential distributions ([25]), the Weibull and generalized exponential distributions ([15]), the log-normal, Weibull, and generalized exponential distributions ([11]), the exponential-Poisson and gamma distributions ([4]), the exponential and Lindley distributions ([9], [39]), and the Poisson and geometric distributions ([31]). On the other hand, related literature in the censored data set-up is limited and includes papers for selecting between Weibull, lognormal and gamma distributions for

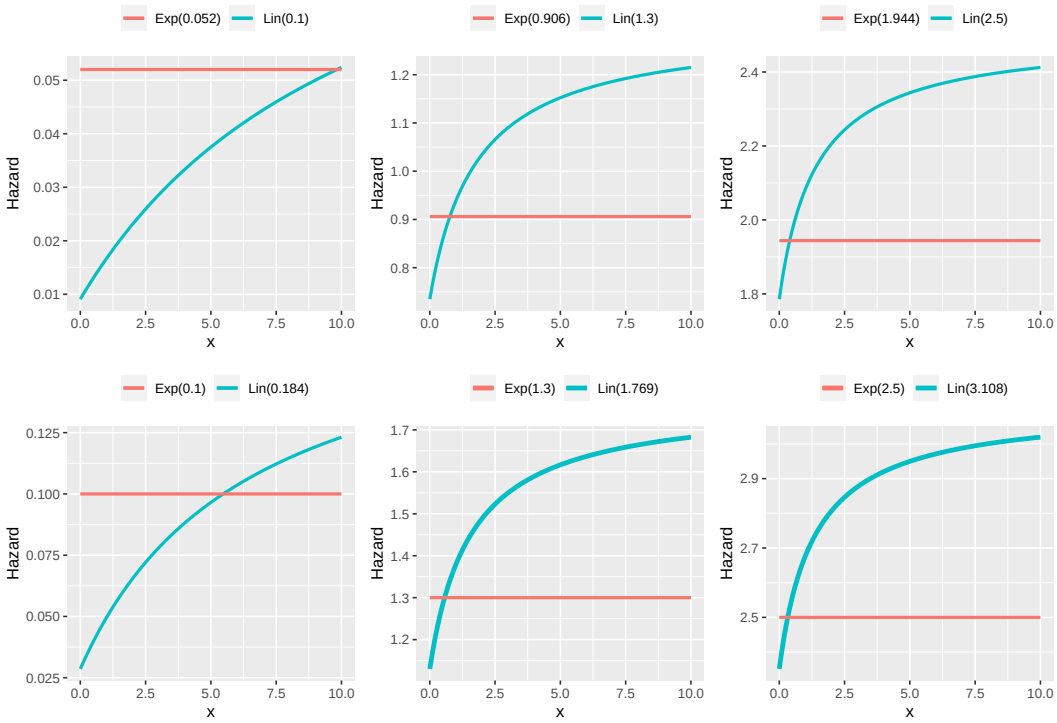

Figure 4: HRF of Lindley (Lin) and exponential (Exp) distributions for different parameters.

Type-I censored samples ([32]), and Weibull and lognormal distributions for Type II censored samples ([5], [18], [10]).

5 Discrimination Procedure

Data analyzed in psychology, as well as in many other fields such as biology and ecology, can be affected by censoring or truncation. While censoring can reduce information either by not having access to some observations because they are out of range or simply not known, truncation affects data by removing existing observations. Censoring can also lead to skewed distributions of the response variable, with an excess of either low (floor effect) or high (ceiling effect) values, and analysis of such data using traditional methods can lead to biased results [2, 28]. There are two types of censoring. Type I censoring occurs when a study is designed to end at a fixed time determined by the researcher, and type II censoring occurs when a study ends when a fixed number of events have been observed. In traditional cross-sectional experimental psychology studies, censorship of this kind is rare, although it may occur. Type I censoring may occur when a researcher sets a maximum allowable response time (RT) to a pre-specified value, such as 2000 ms. In such cases, responses not made within that time frame are recorded as 2000 ms, and the subjects are thus partially observed. In the same type of

study, type II censoring could occur when the researcher decides to stop the study when a pre-specified number of subjects have responded within 2000 ms.

A hypothetical scenario in a longitudinal cognitive neuroscience study may help to clarify this type of censoring. Suppose that 10 patients with a traumatic brain injury are followed in a similar cognitive rehabilitation treatment. The aim is to assess the effectiveness of the treatment in improving cognitive function. The patients' reaction times to a cognitive test are recorded periodically (e.g. once a week). Some patients may drop out of the study and others may stay. The trial cannot go on indefinitely, so the data will eventually have to be censored so that a complete data set cannot be used. The following types of censoring can be considered:
Type I: The researchers set a censoring period of six months, i.e. the study continues for six months and the reaction times of the patients are recorded up to six months. For patients continuing beyond six months, only their six-month reaction time will be used.
Type II: The researchers stop the experiment when three patients drop out, i.e. the study is stopped when the third patient drops out. The only usable data is from the remaining seven participants.

In this section, we describe the selection procedure for the most common right censoring scheme, known as Type I censoring. The scheme is briefly described as follows. Let us consider n items being observed in a particular experiment. Also suppose that $x_{i:n}, i = 1, 2, ..., n$ be the ith ordered observation. In the conventional Type I censoring scheme, the experiment is aborted at a pre-determined time t_0 such that $x_{t_0} < x_{n:n}$.

Let us first discuss the case of the complete data setting. It is assumed that the data is generated from one of the $Exp(\lambda)$ or $Lin(\theta)$ distributions and the corresponding likelihood functions for the complete data set are respectively

$$L_E(\lambda) = \prod_{i=1}^{n} f_E(x_i; \lambda) \quad \text{and} \quad L_L(\theta) = \prod_{i=1}^{n} f_L(x_i; \theta).$$

The ratio of maximized likelihoods (RML) is defined as $L = \frac{L_E(\hat{\lambda})}{L_L(\hat{\theta})}$; and $\hat{\lambda}$ and $\hat{\theta}$ are the maximum likelihood estimators of λ and θ respectively. Hence the logarithm of the RML, written as, $T = \log L = l_E(\hat{\lambda}) - l_L(\hat{\theta})$ is obtained as

$$T = n \log \left(\frac{\hat{\lambda}\left(\hat{\theta}+1\right)}{\hat{\theta}^2} \right) + (\hat{\theta} - \hat{\lambda}) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(1 + x_i). \tag{5.1}$$

In case of the exponential distribution, $\hat{\lambda}$ can be easily obtained as

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}. \tag{5.2}$$

Note that equation 5.2 also equals $1/\bar{x}$. Similarly, $\hat{\theta}$, the ML estimator of the Lindley distribution can be obtained as

$$\hat{\theta} = \frac{-(\bar{x}-1) + \sqrt{(\bar{x}-1)^2 + 8\bar{x}}}{2\bar{x}}. \tag{5.3}$$

The natural model selection criterion will be to choose the exponential distribution, if $T > 0$, otherwise, choose Lindley distribution. Next, we discuss the selection procedure for the right censored sample.

For type I censoring, the likelihood functions are given as

$$L_E^*(\lambda) = \prod_{i=1}^{n} \{f_E(x_i;\lambda)\}^{\delta_i} \{1 - F_E(t_0;\lambda)\}^{1-\delta_i} \quad \text{and} \quad L_L^*(\theta) = \prod_{i=1}^{n} \{f_L(x_i;\theta)\}^{\delta_i} \{1 - F_L(t_0;\theta)\}^{1-\delta_i}.$$

where

$$\begin{aligned} \delta_i &= 1 \quad \text{if} \quad x_i \le t_0 \\ \delta_i &= 0 \quad \text{if} \quad x_i > t_0. \end{aligned} \tag{5.4}$$

The loglikelihood function of λ (ignoring constants) is obtained as

$$l_E^*(\lambda) = \log\lambda \sum_{i=1}^{n} \delta_i - \lambda \sum_{i=1}^{n} \delta_i x_i - \lambda t_0 \sum_{i=1}^{n} (1-\delta_i).$$

Assuming d observations fall below the censoring time t_0, the MLE of λ is obtained as

$$\hat{\lambda^*} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} \delta_i x_i + t_0 \sum_{i=1}^{n} (1-\delta_i)} = \frac{d}{\sum_{i=1}^{d} x_i + (n-d)t_0}.$$

Similarly, the loglikelihood function of θ is obtained as

$$l_L^*(\theta) = 2\log\theta \sum_{i=1}^{n} \delta_i - \theta \sum_{i=1}^{n} \delta_i x_i + \log(1+\theta+\theta t_0) \sum_{i=1}^{n} (1-\delta_i) - n\log(1+\theta) - \theta t_0 \sum_{i=1}^{n} (1-\delta_i) + \sum_{i=1}^{n} \delta_i \log(1+x_i)$$

.

and the MLE of θ is obtained iteratively from the following equation

$$\frac{2d}{\hat{\theta^*}} + \frac{(n-d)(1+t_0)}{1+\hat{\theta^*}+\hat{\theta^*}t_0} - \frac{n}{1+\hat{\theta^*}} = \sum_{i=1}^{d} x_i + (n-d)t_0.$$

The logarithm of the RML for the right censored sample in a similar line can be found as

$$T^* = \sum_{i=1}^{n} \left[\delta_i \left\{ \log f_E(x_i, \hat{\lambda^*}) - \log f_L(x_i, \hat{\theta^*}) \right\} + (1-\delta_i) \left\{ \log \bar{F}_E(t_0, \hat{\lambda^*}) - \log \bar{F}_L(t_0, \hat{\theta^*}) \right\} \right]. \tag{5.5}$$

The results concerning the asymptotic distribution of T and T^* along with the asymptotic mean and variance are detailed in the Appendix for two different cases, namely when the data come from $Exp(\lambda)$ and when the data come from a $Lin(\theta)$.

6 Selection procedure and simulation results

In this section, we present the selection procedure for discriminating between the exponential and Lindley distributions. We determine the minimum sample size for a given probability of correct selection (PCS) and tolerance limits which can be measured through the distance between two cdfs. Practically, the tolerance limit measures the closeness between two cdfs. It is obvious that if the distance between two cdfs is very small, one needs a very large sample size to discriminate between them for a given PCS. On the other hand, if the cdfs are far apart, a moderate to small sample size may be sufficient to discriminate between the two for a given PCS. Here we use Kolmogrov−Smirnov (K−S) distance to discriminate between the two cdfs with K−S distance being defined as $\sup_x |F(x) - G(x)|$, where F and G are the cdf of exponential and Lindley distributions respectively. One may use other distance measures with the same selection criterion. So, the minimum sample size can be determined based on the given PCS (p, say) and the tolerance limit (D, say) as described in the next subsection.

6.1 Determination of sample size

In view of Theorem 1 in the Appendix A, T is asymptotically normally distributed with mean $E_E(T)$ and variance $V_E(T)$. The PCS for selecting exponential distribution is given by

$$PCS(\lambda) = P(T > 0 \mid \lambda) \approx \Phi\left(\frac{E_E(T)}{\sqrt{V_E(T)}}\right) = \Phi\left(\frac{\sqrt{n}AM_E(T)}{\sqrt{AV_E(T)}}\right),$$

where Φ denotes the cdf of the standard normal random variable (AM and AV stand for asymptotic mean and asymptotic variance respectively). Sample size can be determined by equating the PCS(λ) to the given protection level p as given by

$$\Phi\left(\frac{\sqrt{n}AM_E(T)}{\sqrt{AV_E(T)}}\right) = p$$

to get

$$n = \frac{z_p^2 AV_E(T)}{\left(AM_E(T)\right)^2}.$$

Here z_p is the $100p$ percentile point of a standard normal distribution. For different values of λ, $p = 0.6, 0.7, 0.8$ and K−S distance, sample sizes n are reported in Table 2. From this table, it is evident that as λ increases, the K−S distance decreases, necessitating a larger sample size. Additionally, it is noticeable that as the probability of correct selection p increases, a larger sample size becomes necessary, as anticipated.

Proceeding similarly and using Theorem 2 in Appendix A, sample sizes when the true distribution is Lindley, can be determined by the next expression:

$$n = \frac{z_p^2 AV_L(T)}{\left(AM_L(T)\right)^2}.$$

Table 3 shows the minimum sample size n to differentiate between the distributions for different values of θ, $p = 0.6, 0.7, 0.8$ and K−S distance. Similar patterns are observed in Table 2 and 3.

$\lambda \rightarrow$	0.1	0.5	0.9	1.3	1.5	2.0	2.5
n $(p = 0.6)$	2	8	20	37	49	88	144
n $(p = 0.7)$	9	36	84	159	208	375	617
n $(p = 0.8)$	23	93	216	408	536	967	1590
K-S	0.106	0.054	0.034	0.027	0.021	0.018	0.012

Table 2: Values for n given λ, p and K−S distance between $Exp(\lambda)$ and $Lin(\widetilde{\theta})$ distributions.

$\theta \rightarrow$	0.1	0.5	0.9	1.3	1.5	2.0	2.5
n $(p = 0.6)$	1	4	9	18	24	45	78
n $(p = 0.7)$	5	17	40	77	103	194	332
n $(p = 0.8)$	13	45	103	199	265	499	856
K-S	0.120	0.070	0.049	0.036	0.031	0.022	0.015

Table 3: Values for n given θ, p and K−S distance between $Lin(\theta)$ and $Exp(\widetilde{\lambda})$ distributions.

Similarly, the values for n can be found out for the censored data using Theorem 3 in Appendix B. In Tables 4 and 5 we have the sample size n given λ (or θ), p and K−S distance with 10% of censored observations. These tables illustrate that as the percentage of censored observations increases, the required sample size also increases.

$\lambda \rightarrow$	0.1	0.5	0.9	1.3	1.5	2.0	2.5
n $(p = 0.6)$	2	11	27	53	72	137	236
n $(p = 0.7)$	10	46	114	229	309	588	1011
n $(p = 0.8)$	26	117	294	590	975	1514	2603
K-S	0.106	0.054	0.034	0.027	0.021	0.018	0.012

Table 4: Values for n given λ, p and K−S distance between $Exp(\lambda)$ and $Lin(\widetilde{\theta})$ distributions with 10% of censored observations.

We shall now discuss how to use the PCS and the tolerance level to discriminate between exponential and Lindley models. Suppose that data come from an exponential population. Further, suppose that the tolerance level is based on the K−S distance and is fixed at 0.054, additionally, suppose that the protection level is $p = 0.8$. Here tolerance level $D = 0.054$ means that the practitioner wants to discriminate between exponential and Lindley cdfs only when their K−S

$\theta \rightarrow$	0.1	0.5	0.9	1.3	1.5	2.0	2.5
n $(p = 0.6)$	2	5	13	26	35	70	126
n $(p = 0.7)$	6	22	54	111	152	301	541
n $(p = 0.8)$	15	57	139	285	390	776	1393
K-S	0.120	0.070	0.049	0.036	0.031	0.022	0.015

Table 5: Values for n given θ, p and K$-$S distance between $Lin(\theta)$ and $Exp(\widetilde{\lambda})$ distribution with 10% of censored observations.

distance is more than 0.054. Table 2 shows that one needs to take a sample of size 93 for $p = 0.8$ to discriminate the exponential and Lindley distributions. On the other hand, if the data come from a Lindley population, and for an approximate $D = 0.054$ and $p = 0.8$, Table 3 indicates an approximate value of n as 103. Therefore, to meet the protection level $p = 0.8$ with a given tolerance level of $D = 0.054$, one needs a sample size of $max\{93, 103\} = 103$ to satisfy the requirements for both cases simultaneously. In Table 3, the value $D = 0.054$ is not directly tabulated; however, values of 0.070 and 0.049 are provided, suggesting that n should lie close to 103.

For the case of censored data, a similar interpretation can be obtained from Tables 4 and 5. For example, if we set $D = 0.054$ with $p = 0.8$ and believe that a maximum of 10% of the observations collected will be censored, then a sample size of $\max\{117, 139\} = 139$ is required to meet the requirements for both cases simultaneously. The sample sizes from Tables 4 and 5 are slightly higher than the values for the case without censored observations given in Tables 2 and 3.

6.2 Computation of PCS for finite sample sizes

In this subsection, we show that the asymptotic results derived in the Appendix work well for finite sample sizes.

We compute the probability of correct selection (PCS) using the asymptotic results and Monte Carlo simulations. Samples of sizes $n = 20, 40, 60, 80, 100$, and 200 were taken for the findings. First, we consider the case of a complete sample. Assuming the null distribution as exponential and the alternative as Lindley, the results obtained by using the asymptotic theory are shown in Table 6 for various choices of the scale parameter of exponential distribution $\lambda = 0.1, 0.5, 0.9, 1.3, 1.5, 2.0, 2.5$. Similarly, we obtain the results for the same choice of n and the scale parameter of Lindley distribution θ when the null distribution is Lindley and the alternative is exponential. The results are reported in Table 7.

From Tables 6 and 7, clear patterns emerge. As the sample size increases, the asymptotic PCS also increases, as expected. Additionally, it is observed that the PCS increases as the values of λ and θ decrease. Moreover, asymptotic results perform well even with a small sample

$\lambda\downarrow\ n\rightarrow$	20	40	60	80	100	200
0.1	0.783	0.864	0.912	0.941	0.959	0.993
	(0.712)	(0.840)	(0.899)	(0.936)	(0.958)	(0.995)
0.5	0.652	0.709	0.750	0.782	0.808	0.891
	(0.552)	(0.646)	(0.705)	(0.749)	(0.785)	(0.884)
0.9	0.601	0.641	0.671	0.696	0.716	0.791
	(0.494)	(0.566)	(0.613)	(0.646)	(0.675)	(0.770)
1.3	0.574	0.604	0.626	0.645	0.661	0.722
	(0.451)	(0.517)	(0.562)	(0.589)	(0.611)	(0.692)
1.5	0.567	0.590	0.612	0.634	0.648	0.701
	(0.438)	(0.502)	(0.538)	(0.573)	(0.582)	(0.667)
2.0	0.553	0.575	0.584	0.608	0.617	0.655
	(0.410)	(0.469)	(0.512)	(0.529)	(0.547)	(0.607)
2.5	0.537	0.553	0.564	0.575	0.583	0.617
	(0.403)	(0.454)	(0.484)	(0.500)	(0.523)	(0.584)

Table 6: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Exp(\lambda)$ distribution for different values of λ and n.

size of 20 in both cases, across all possible parameter ranges. The simulated PCS, reported in parentheses, exhibits the same pattern as the asymptotic PCS. Furthermore, the simulated PCS aligns with the asymptotic PCS.

We conducted a secondary analysis of the probability of correct selection (PCS), considering random samples with censored observations. Tables 8 and 9 present the PCS for 10% and 20% of censored observations, respectively, when the random sample is drawn from an Exponential distribution. Similarly, Tables 10 and 11 display the PCS for 10% and 20% of censored observations, respectively, when the random sample is drawn from a Lindley distribution.

To generate censored random samples, we employed a two-step process: Firstly, we determined a value t_0 corresponding to a percentile to ensure that a certain percentage (10% or 20%) of observations exceeded t_0. Secondly, observations surpassing t_0 were substituted with t_0, with the indicator δ set to 0, while observations equal to or below t_0 were assigned the indicator $\delta = 1$.

When comparing Tables 6, 8, and 9 for the Exponential case, it's evident that an increasing percentage of censored observations leads to a decrease in PCS (both asymptotic and simulated), indicating a trend toward lower values. A similar pattern is observed when comparing Tables 7, 10, and 11 in the Lindley case. Additionally, the simulated PCS are close to asymptotic PCS for each percentage of censored observations.

Another notable pattern emerges when comparing the PCS for Exponential and Lindley

$\theta \downarrow \ n \rightarrow$	20	40	60	80	100	200
0.1	0.850	0.930	0.966	0.983	0.988	0.990
	(0.884)	(0.933)	(0.963)	(0.979)	(0.988)	(0.998)
0.5	0.714	0.786	0.832	0.861	0.891	0.965
	(0.780)	(0.828)	(0.861)	(0.889)	(0.908)	(0.963)
0.9	0.644	0.704	0.748	0.773	0.809	0.882
	(0.733)	(0.759)	(0.780)	(0.801)	(0.819)	(0.894)
1.3	0.601	0.656	0.683	0.702	0.727	0.804
	(0.706)	(0.710)	(0.723)	(0.751)	(0.763)	(0.825)
1.5	0.598	0.632	0.655	0.683	0.702	0.776
	(0.698)	(0.699)	(0.707)	(0.722)	(0.731)	(0.791)
2.0	0.577	0.591	0.613	0.634	0.657	0.708
	(0.684)	(0.674)	(0.674)	(0.689)	(0.692)	(0.735)
2.5	0.555	0.579	0.594	0.606	0.615	0.669
	(0.670)	(0.651)	(0.660)	(0.664)	(0.667)	(0.695)

Table 7: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Lin(\theta)$ distribution for different values of θ and n.

distributions under the same percentage of censored observations (complete, 10%, or 20%). Interestingly, the PCS tends to be higher when the random sample is drawn from a Lindley population compared to the Exponential case.

7 Application to Experimental Psychology

Posada-Quintero and Bolkhovsky [38] had 16 participants undergoing three cognitive tasks; a psychomotor vigilance task (PVT), an auditory working memory task (n-back), and a visual search task (SS). A condition that did not require any psychomotor activity was used as a baseline task (BL). The researchers measured several biomarkers related to the body's autonomic nervous system (ANS). The participants' electrodermal activity related to the ANS' sympathetic component (EDASymp), time-varying index of sympathetic tone (TVSymp), and low- and high-frequency components of heart rate variability (HRVLF and HRVHF) were measured over 12 trials in each task. We have identified two conditional data sets from Posada-Quintero and Bolkhovsky [38] suitable for Lindley and exponential fits; TVSymp|BL (i.e. time-varying index of sympathetic tone during the baseline task) and TVSymp|n-back (i.e. time-varying index of sympathetic tone during the auditory working memory task). See supplementary material for details. Table 12 shows the results of the fits for the data sets. The p-values of the Kolmogorov-Smirnov test indicate that there is not enough evidence that the distributions

$\lambda\downarrow\ n\rightarrow$	20	40	60	80	100	200
0.1	0.769	0.850	0.898	0.929	0.950	0.990
	(0.704)	(0.817)	(0.882)	(0.921)	(0.949)	(0.989)
0.5	0.636	0.689	0.726	0.757	0.782	0.864
	(0.530)	(0.629)	(0.669)	(0.715)	(0.747)	(0.844)
0.9	0.587	0.622	0.648	0.670	0.688	0.756
	(0.478)	(0.538)	(0.582)	(0.62)	(0.643)	(0.724)
1.3	0.562	0.587	0.606	0.622	0.635	0.688
	(0.440)	(0.506)	(0.533)	(0.568)	(0.582)	(0.640)
1.5	0.553	0.575	0.591	0.605	0.617	0.664
	(0.431)	(0.491)	(0.528)	(0.544)	(0.564)	(0.623)
2.0	0.539	0.554	0.567	0.577	0.586	0.620
	(0.409)	(0.47)	(0.497)	(0.512)	(0.529)	(0.578)
2.5	0.529	0.542	0.551	0.559	0.566	0.592
	(0.408)	(0.448)	(0.472)	(0.496)	(0.510)	(0.552)

Table 8: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Exp(\lambda)$ distribution for different values of λ, n, and 10% of censored observations.

$\lambda\downarrow\ n\rightarrow$	20	40	60	80	100	200
0.1	0.757	0.838	0.886	0.918	0.941	0.986
	(0.660)	(0.791)	(0.857)	(0.900)	(0.922)	(0.984)
0.5	0.625	0.674	0.710	0.739	0.763	0.844
	(0.522)	(0.604)	(0.641)	(0.679)	(0.706)	(0.802)
0.9	0.579	0.610	0.634	0.654	0.671	0.735
	(0.459)	(0.532)	(0.571)	(0.587)	(0.609)	(0.684)
1.3	0.555	0.578	0.595	0.609	0.621	0.669
	(0.442)	(0.483)	(0.525)	(0.541)	(0.564)	(0.622)
1.5	0.547	0.567	0.581	0.594	0.605	0.646
	(0.430)	(0.473)	(0.504)	(0.524)	(0.554)	(0.596)
2.0	0.534	0.548	0.559	0.568	0.576	0.606
	(0.416)	(0.463)	(0.485)	(0.509)	(0.509)	(0.552)
2.5	0.526	0.537	0.545	0.552	0.558	0.581
	(0.411)	(0.453)	(0.471)	(0.488)	(0.501)	(0.531)

Table 9: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Exp(\lambda)$ distribution for different values of λ, n, and 20% of censored observations.

$\theta\downarrow\ n\rightarrow$	20	40	60	80	100	200
0.1	0.833	0.914	0.953	0.973	0.985	0.999
	(0.869)	(0.918)	(0.951)	(0.974)	(0.983)	(0.998)
0.5	0.691	0.76	0.807	0.841	0.868	0.943
	(0.769)	(0.807)	(0.84)	(0.862)	(0.887)	(0.945)
0.9	0.625	0.674	0.710	0.738	0.762	0.843
	(0.706)	(0.739)	(0.753)	(0.78)	(0.796)	(0.858)
1.3	0.588	0.624	0.650	0.672	0.691	0.759
	(0.689)	(0.689)	(0.706)	(0.714)	(0.728)	(0.779)
1.5	0.576	0.606	0.629	0.648	0.665	0.727
	(0.680)	(0.676)	(0.687)	(0.687)	(0.706)	(0.757)
2.0	0.554	0.576	0.593	0.606	0.619	0.665
	(0.666)	(0.650)	(0.648)	(0.666)	(0.665)	(0.685)
2.5	0.540	0.557	0.569	0.580	0.589	0.625
	(0.651)	(0.638)	(0.630)	(0.639)	(0.636)	(0.653)

Table 10: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Lin(\theta)$ distribution for different values of θ, n and 10% of censored observations.

$\theta\downarrow\ n\rightarrow$	20	40	60	80	100	200
0.1	0.820	0.902	0.943	0.966	0.980	0.998
	(0.855)	(0.913)	(0.945)	(0.965)	(0.975)	(0.997)
0.5	0.677	0.742	0.787	0.821	0.848	0.927
	(0.754)	(0.784)	(0.813)	(0.839)	(0.862)	(0.92)
0.9	0.613	0.658	0.691	0.718	0.740	0.819
	(0.702)	(0.709)	(0.720)	(0.748)	(0.763)	(0.832)
1.3	0.579	0.611	0.635	0.655	0.672	0.735
	(0.669)	(0.664)	(0.677)	(0.688)	(0.698)	(0.749)
1.5	0.567	0.595	0.615	0.632	0.647	0.704
	(0.659)	(0.65)	(0.662)	(0.668)	(0.679)	(0.725)
2.0	0.547	0.567	0.582	0.594	0.605	0.647
	(0.653)	(0.642)	(0.631)	(0.639)	(0.645)	(0.662)
2.5	0.535	0.550	0.561	0.570	0.578	0.610
	(0.634)	(0.613)	(0.611)	(0.63)	(0.618)	(0.638)

Table 11: PCS based on asymptotic and simulated (on parenthesis) results when the data come from $Lin(\theta)$ distribution for different values of θ, n and 20% of censored observations.

Data Set	Distribution	MLE	Max log-likelihood	AIC	BIC	KS (*p-value)*
TVSymp | BL	Exponential	2.251	-36.257	74.513	77.772	0.0543 (0.622)
	Lindley	2.837	-36.640	75.279	78.537	0.0500 (0.724)
TVSymp | n-back	Exponential	2.979	17.560	-33.120	-29.863	0.0449 (0.724)
	Lindley	3.623	17.648	-33.296	-30.038	0.0433 (0.846)

Table 12: Goodness of fit metrics of the distributions fitted to two data sets

underlying the data differ from both the Lindley and exponential distributions.

Now, we illustrate the application of our proposed statistical test using complete and censored data sets. Although the original data sets are complete, we intentionally censor them to measure the performance of the test. Each data set of size 16 is censored in two ways.
Case 1. The data is censored at the 15th observation so that the 15th and 16th data become identical.
Case 2. The data is censored at the 14th observation so that the 14th-16th data becomes identical.
It is also interesting to study the asymptotic results for such a small sample size of 16. The results for each data set are described below and are shown in Table 13.

TVSymp | BL data

When exponential and Lindley distributions are used to fit the complete data, maximized log-likelihood functions are given as $l_E(2.251) = -36.257$ and $l_L(2.837) = -36.640$ respectively resulting in $T = l_E(2.251) - l_L(2.837) = 0.382 > 0$ which indicates to choose the exponential model. Let us compare this finding with the asymptotic results as also shown in Table 13. Assuming the data come from exponential cdf, the asymptotic mean and variance are obtained as $AM_E(2.251) = 0.00120$ and $AV_E(2.251) = 0.0026$ along with the $PCS = 0.629$ yielding an estimated risk around 37% to choose the wrong model. Similarly, assuming that the data are from Lindley cdf, we compute $AM_L(2.837) = -0.0011$ and $AV_L(2.837) = 0.0022$, with the $PCS = 0.633$ to yield an estimated risk nearly 27% to choose the wrong model. Therefore, the PCS is at least $min(0.629, 0.633) = 0.629$. The PCS attains the maximum when the data are coming from the Lindley distribution and hence we should choose the Lindley distribution to fit the data.

TVSymp | n-back data

Data Set	Distribution	Censoring	Statistic (T)	AM	AV	PCS
TVSymp | BL	Exponential	No	0.38243	0.00120	0.00255	0.629
	Lindley	No	0.38243	-0.00114	0.00216	0.633
	Exponential	Case 1	0.56303	0.00303	0.00439	
	Lindley					
	Exponential	Case 2	0.52071	0.00280	0.00368	
	Lindley					
TVSymp | n-back	Exponential	No	-0.0877	0.00062	0.00130	0.594
	Lindley	No	-0.0877	-0.00059	0.00113	0.596
	Exponential	Case 1	0.04301	0.00273	0.00378	
	Lindley					
	Exponential	Case 2	-0.047211	0.00216	0.00226	
	Lindley					

Table 13: Results from the data analysis

The Lindley and exponential distributions also fit the data well (see Table 12). When exponential and Lindley distributions are used to fit the data, maximized log-likelihood functions are given as $l_E(2.979) = 17.648$ and $l_L(3.623) = 17.648$ respectively resulting in $T = l_E(2.979) - l_L(3.623) = -329 + 319 = -0.088 < 0$ which indicates to choose the Lindley model. Now we compute the PCS based on asymptotic results. Assuming that the data come from exponential cdf, the asymptotic mean and variance are obtained as $AM_E(2.979) = 0.00062$ and $AV_E(2.979) = 0.00129$ along with the $PCS = 0.594$ yielding an estimated risk around 40% to choose the wrong model. Similarly, assuming that the data are from Lindley cdf, we compute $AM_L(3.623) = -0.00059$ and $AV_L(3.623) = 0.0011$, with the $PCS = 0.596$ to yield an estimated risk nearly similar to what obtained using exponential cdf. Therefore, the PCS is at least $min(0.594, 0.596) = 0.594$ in this case. The PCS attains the maximum when the data are coming from a Lindley distribution and hence we should choose Lindley distribution to fit the data.

It is noteworthy that, in both datasets, the Lindley distribution demonstrates the most favorable fit according to our methodology. However, based on the AIC, the exponential distribution emerges as the optimal fit in the first dataset, whereas the Lindley distribution exhibits the best fit in the second dataset. While it may be coincidental that the Lindley distribution provides the best fit for both datasets using our method, this consistency simplifies the interpretation and comparison between the two datasets. In the Posada-Quintero and Bolkhovsky study, [38] data were analyzed via ANOVA, task classification analysis (via k-nearest neighbor classifiers, support vector machines, decision trees, and discriminant analysis), and classifier performance

(via leave-one-subject-out cross-validation). We believe that Lindley distributional regression modeling via GAMLSS [19] would have been helpful to investigate the potential effects of the explanatory variables on the rate scale parameter. In addition, Lindley time-dependent hrf plots would have been informative for estimating the likelihood of an event increasing over time. Specifically, only Lindley hrf plots could have estimated the likelihood of ANS sympathetic tone increasing over time given a cognitive task. If an exponential distribution were the best fit, there would instead be a constant hazard rate, indicating that a cognitive task has a monotonic likelihood of affecting ANS sympathetic tone over time. Thus, the data example visited above signals that choosing the correct probability distribution has relevant applied consequences, that is, choosing the Lindley or exponential distributions would lead to radically opposite conclusions in an hrf analysis.

8 Concluding Remarks

In this paper, we highlight the role of hazard rate functions in discriminating between distributions and show the practical significance of discriminating between two distributions with very similar shapes like the exponential and Lindley ones, suitable to model positively skewed data, for example when analyzing response time data. We propose a method to address this problem based on the ratio of the likelihood functions. The method is flexible and can be applied to Type I censored data, making it very useful in experimental psychology as well as in other fields where censored data are quite common.

It was mentioned earlier that GAMLSS is a framework promoting careful consideration of probability distributions for data interpretation and prediction. As previously explained, hrfs are directly dependent on the pdf and the cdf through the survival function, and they represent an underutilized yet informative approach to analyzing data (see [30]). We believe that while GAMLSS offers a powerful tool for selecting optimal distributions, our proposed method provides an additional layer of refinement, particularly when dealing with distributions that have an equal number of parameters and similar pdfs and cdfs. By carefully selecting the most suitable distribution, we can optimize the shape of the hrf and enhance the accuracy of our analysis.

References

[1] Ahmad, M. A., Raqab, M. Z., and Kundu, D. (2017). Discriminating between the generalized Rayleigh and Weibull distributions: Some comparative studies. *Communications in Statistics-Simulation and Computation*, **46(6)**, 4880-4895.

[2] Ahmadi H,., Granger, D. A., Hamilton, K. R., Blair, C., and Riis, J. L. (2021). Censored

data considerations and analytical approaches for salivary bioscience data. *Psychoneuroendocrinology*, **129(105274)**. doi: 10.1016/j.psyneuen.2021.105274.

[3] Bain, L. J. and Engelhardt, M. (1980). Probability of correct selection of Weibull versus gamma based on likelihood ratio. *Communications in Statistics-Theory and Methods*, **9 (4)**, 375-381.

[4] Barreto-Souza, W., and Silva, R. B. (2015). A likelihood ratio test to discriminate exponential–Poisson and gamma distributions. *Journal of Statistical Computation and Simulation*, **85(4)**, 802-823.

[5] Cain, S. R. (2002). Distinguishing between lognormal and Weibull distributions. *IEEE Transactions on Reliability*, **51(1)**, 32-38.

[6] Chen, W. W. (1980). On the tests of separate families of hypotheses with small sample size, *Journal of Statistical Computations and Simulations*, **2**, 183-187.

[7] Cox, D. R. (1961). Tests of separate families of hypotheses, Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley, University of California Press, 105-123.

[8] Cox, D. R. (1962). Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society, Ser. B*, **24**, 406-424.

[9] Chowdhury, S. (2019). Selection between Exponential and Lindley distributions (No. 316).IIMK/WPS/316/QM&OM/2019/07.

[10] Dey, A. K. and Kundu, D. (2012). Discriminating between the Weibull and log-normal distributions for Type-II censored data. *Statistics*, **46(2)**, 197-214.

[11] Dey, A. K. and Kundu, D. (2009). Discriminating among the log-normal, weibull, and generalized exponential distributions. *IEEE Transactions on reliability*, **58 (3)**, 416-424.

[12] Fearn, D. H. and Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and Gamma distributions, *Communications in Statistics-Theory and Methods*, **20**, 579-593.

[13] Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, **78(4)**, 493-506.

[14] Gordon, L. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18(5)**, 883-914. doi: 10.1037/0278-7393.18.5.883

[15] Gupta, R. D. and Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis*, **43 (2)**, 179-196.

[16] Holden, J. G., Van Orden, G. C., and Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, **116(2)**, 318–342.

[17] Kim, J. S. and Yum, B. J. (2008). Selection between Weibull and lognormal distributions: A comparative simulation study. *Computational Statistics and Data Analysis*, **53(2)**, 477-485.

[18] Kim, D. H., Lee, W. D., and Kang, S. G. (2000). Bayesian model selection for life time data under type II censoring. *Communications in Statistics-Theory and Methods*, **29(12)**, 2865-2878.

[19] Klein, N. (2024). Distributional Regression for Data Analysis. *Annual Review of Statistics and Its Application*, **11(1)**. doi: 10.1146/annurev-statistics-040722-053607.

[20] Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, **13(4)**, 275-303. doi:10.1177/1471082X13494159

[21] Kneib, T., Silbersdorff, A., and Säfken, B. (2023). Rage Against the Mean – A Review of Distributional Regression Approaches. *Econometrics and Statistics*, **26**, 99-123. doi:10.1016/j.ecosta.2021.07.006

[22] Kundu, D. and Manglick, A. (2004). Discriminating between the Weibull and Log-Normal distributions, *Naval Research Logistics*, **51**, 893-905.

[23] Kundu, D. and Manglick, A. (2005). Discriminating between the Log-Normal and gamma distributions, *Journal of the Applied Statistical Sciences*, **14**, 175-187.

[24] Kundu, D. and Raqab, M. Z. (2007). Discriminating between the generalized Rayleigh and log-normal distribution. *Statistics*, **41(6)**, 505-515.

[25] Kundu, D., Gupta, R. D., and Manglick, A. (2005). Discriminating between the log-normal and generalized exponential distributions. *Journal of Statistical Planning and Inference*, **127(1-2)**, 213-227.

[26] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102-107.

[27] Lindley, D. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference, Cambridge University, Press, New York.

[28] Liu, Q., and Wang, L. (2021). t-Test and ANOVA for data with ceiling and/or floor effects. *Behavior Reseach Methods*, **53(1)**, 264-277. doi: 10.3758/s13428-020-01407-2

[29] Mazucheli, J., and Achcar, J. A. (2011). The Lindley distribution applied to competing risks lifetime data. *Computer Methods and Programs in Biomedicine*, **104(2)**, 188-192.

[30] Panis, S., Schmidt, F., Wolkersdorfer, M. P., and Schmidt, T. (2020). Analyzing Response Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, **11(6)**.

[31] Pradhan, B., and Kundu, D. (2016). A Choice Between Poisson and Geometric Distributions. *Journal of the Indian Society for Probability and Statistics*, **17(2)**, 111-123.

[32] Quesenberry, C. P. (1982). Selecting among Weibull, lognormal and gamma distributions using complete and censored samples. *Naval Research Logistics Quarterly*, **29(4)**, 557-569.

[33] Raqab, M. Z. (2013). Discriminating between the generalized Rayleigh and Weibull distributions. *Journal of Applied Statistics*, **40(7)**, 1480-1493.

[34] Stasinopoulos, M., Kneib, T., Klein, N., Mayr, A., and Heller, G. (2024). *Generalized Additive Models for Location, Scale and Shape: A Distributional Regression Approach, with Applications*. Cambridge University Press.

[35] Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, **108(3)**, 551–567.

[36] Trafimow, D., Wang, T., and Wang, C. (2018). Means and standard deviations, or locations and scales? That is the question! *New Ideas in Psychology*, **50**, 34-37. doi: 10.1016/j.newideapsych.2018.03.001

[37] White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses. *Econometrica*, **19**, 301–318.

[38] Posada-Quintero, H. F., and Bolkhovsky, J. B. (2019). Machine Learning models for the Identification of Cognitive Tasks using Autonomic Reactions from Heart Rate Variability and Electrodermal Activity. *Behavioral Sciences*, **9(45)**. doi: 10.3390/bs9040045.

[39] Vaidyanathan, V., and Varghese, A. (2019). Discriminating Between Exponential and Lindley Distributions. *Journal of Statistical Theory and Applications*, **18(3)**, 295-302.

Appendix

For the Borel measurable function $h(X)$, $E_E(h(X))$ and $V_E(h(X))$ will denote the mean and variance of $h(X)$ under the assumptions that X follows exponential distribution. Similarly, we define $E_L(h(X))$ and $V_L(h(X))$ as mean and variance of $h(X)$ under the assumption that X follows Lindley distribution. Moreover, if $g(X)$ and $h(X)$ are two Borel measurable functions, we define $Cov_E(g(U), h(U)) = E_E(g(U)h(U)) - E_E(g(U))E_E(h(U))$, and $Cov_L(g(U), h(U)) =$

$E_L(g(U)h(U)) - E_L(g(U))E_L(h(U))$. Almost sure convergence will be denoted by $a.s.$ throughout the paper. We define the following notation for few integrals used in the next subsections.

$$\Lambda(i,j,k,l) = \int_0^\infty (\log(1+x))^i (1+x)^j e^{-(k\lambda+l\theta)x} dx.$$

A: Asymptotic results for complete sample

A1: Exponential distribution as the null hypothesis

We begin this section with the following Lemma. The proof of the Lemma follows using similar arguments of White [37] and hence is omitted.

Lemma 1 Suppose the data are from $Exp(\lambda)$ distribution. Then, as $n \to \infty$ we have that

(i) $\hat{\lambda} \to \lambda$ $a.s.$

(ii) $\hat{\theta} \to \widetilde{\theta}$ $a.s.$, where

$$E_E\left[\log f_L(x;\widetilde{\theta})\right] = \max_\theta E_E\left[\log f_L(x;\theta)\right],$$

where $\widetilde{\theta}$ is the quasi-likelihood estimators of θ

(iii) If $T_* = l_E(\hat{\lambda}) - l_L(\widetilde{\theta})$, then $n^{-1/2}(T - E_E T)$ is asymptotically equivalent to $n^{-1/2}(T_* - E_E T_*)$.

The following theorem follows from the central limit theorem and Lemma 1 (iii), and hence its proof is omitted.

Theorem 1 If the data come from $Exp(\lambda)$ distribution, then T is approximately normally distributed with mean $E_E(T)$ and variance $V_E(T)$.

Now, we discuss how to obtain $\widetilde{\theta}, E_E(T)$ and $V_E(T)$. Let us define

$$\begin{aligned}\Psi_E(\theta) &= E_E\left[\log f_L(x;\theta)\right] \\ &= 2\log\theta - \log(1+\theta) + \lambda\Lambda(1,0,1,0) - \frac{\theta}{\lambda}.\end{aligned}$$

Differentiating $\Psi_E(\theta)$ with respect to θ (> 0), we get $\widetilde{\theta}$ after solving the following quadratic equation

$$\theta^2 + (1-\lambda)\theta - 2\lambda = 0.$$

We observe that $\lim_{n\to\infty} \frac{E_E(T)}{n}$ and $\lim_{n\to\infty} \frac{V_E(T)}{n}$ exist. Next we obtain the asymptotic mean and variance of T under $Exp(\lambda)$ distribution which are denoted by $AM_E \approx \frac{E_E(T)}{n}$ and $AV_E \approx \frac{V_E(T)}{n}$ respectively and are derived as follows.

$$\begin{aligned}AM_E &= E_E\left[\log f_E(x;\lambda) - \log f_L(x;\widetilde{\theta})\right] \\ &= \log\lambda - 2\log\widetilde{\theta} + \log(1+\widetilde{\theta}) + \frac{\widetilde{\theta}}{\lambda} - \lambda\Lambda(1,0,1,0) - 1,\end{aligned}$$

and

$$\begin{aligned} AV_E &= V_E\left[\log f_E(x;\lambda) - \log f_L(x;\widetilde{\theta})\right] \\ &= \left(\widetilde{\theta}-\lambda\right)^2 V_E(X) + V_E\left(\log(1+X)\right) - 2(\widetilde{\theta}-\lambda)Cov_E\left(X, \log(1+X)\right) \\ &= \left(\frac{\widetilde{\theta}}{\lambda} - 1\right)^2 + \lambda\Lambda(2,0,1,0) - \left(\lambda\Lambda(1,0,1,0)\right)^2 - 2(\widetilde{\theta}-\lambda)\left[\lambda\Lambda(1,1,1,0) - (1+\lambda)\Lambda(1,0,1,0)\right]. \end{aligned}$$

A2: Lindley distribution as the null hypothesis

We state the following results along the same lines as Lemma 1 and Theorem 1.

Lemma 2 Suppose the data are from $Lin(\theta)$ distribution. Then, as $n \to \infty$ we have that

(i) $\hat{\theta} \to \theta$ *a.s.*

(ii) $\hat{\lambda} \to \widetilde{\lambda}$ *a.s.* where

$$E_L\left[\log f_E(x;\widetilde{\lambda})\right] = \max_{\lambda} E_L\left[\log f_E(x;\lambda)\right],$$

where $\widetilde{\lambda}$ is the quasi-likelihood estimators of λ

(iii) If $T_{**} = l_E(\widetilde{\lambda}) - l_L(\hat{\theta})$, then $n^{-1/2}(T - E_L T)$ is asymptotically equivalent to $n^{-1/2}(T_{**} - E_L T_{**})$.

As mentioned earlier, the following theorem follows from the central limit theorem and Lemma 2(iii), and hence its proof is omitted.

Theorem 2 If the data come from $Lin(\theta)$ distribution, then T is approximately normally distributed with mean $E_L(T)$ and variance $V_L(T)$.

To obtain $\widetilde{\lambda}, E_L(T)$ and $V_L(T)$. Let us define

$$\begin{aligned} \Psi_L(\lambda) &= E_L\left[\log f_E(x;\lambda)\right] \\ &= \log\lambda - \frac{\lambda(\theta+2)}{\theta(\theta+1)}. \end{aligned}$$

Differentiating $\Psi_L(\lambda)$ with respect to λ, we get

$$\widetilde{\lambda} = \frac{\theta(\theta+1)}{\theta+2}.$$

As before, we obtain the expressions of asymptotic mean and variance of T under $Lin(\theta)$ distribution as

$$\begin{aligned} AM_L &= E_L\left[\log f_E(x;\widetilde{\lambda}) - \log f_L(x;\theta)\right] \\ &= \log\widetilde{\lambda} - (\widetilde{\lambda}-\theta)\frac{\theta+2}{\theta(\theta+1)} + \log\left(\frac{1}{\theta} + \frac{1}{\theta^2}\right) - \frac{\theta^2}{\theta+1}\Lambda(1,1,0,1), \end{aligned}$$

λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	0.0773	0.1952	0.184
0.5	0.0175	0.0401	0.781
0.9	0.0073	0.0162	1.292
1.3	0.0038	0.0083	1.769
1.5	0.0028	0.0062	2.000
2.0	0.0016	0.0033	2.561
2.5	0.0009	0.0020	3.108

Table 14: Estimated values of $AM_E(\lambda)$, $AV_E(\lambda)$ and $\widetilde{\theta}$ for different choices of λ.

θ	$AM_L(\theta)$	$AV_L(\theta)$	$\widetilde{\lambda}$
0.1	-0.0802	0.1203	0.052
0.5	-0.0275	0.0479	0.300
0.9	-0.0124	0.0223	0.589
1.3	-0.0066	0.0121	0.906
1.5	-0.0049	0.0092	1.071
2.0	-0.0026	0.0049	1.500
2.5	-0.0016	0.0029	1.944

Table 15: Estimated values of $AM_L(\theta)$, $AV_L(\theta)$ and $\widetilde{\lambda}$ for different choices of θ.

and

$$\begin{aligned} AV_L &= V_L\left[\log f_E(x;\widetilde{\lambda}) - \log f_L(x;\theta)\right] \\ &= \left(\theta-\widetilde{\lambda}\right)^2 V_L(X) + V_L\left(\log(1+X)\right) - 2(\theta-\widetilde{\lambda})Cov_L\left(X, \log(1+X)\right) \\ &= \left(\theta-\widetilde{\lambda}\right)^2 \frac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2} + \frac{\theta^2}{\theta+1}\Lambda(2,1,0,1) - \left(\frac{\theta^2}{\theta+1}\Lambda(1,1,0,1)\right)^2 - \\ &2(\theta-\widetilde{\lambda})\left(\left(\frac{\theta^2}{\theta+1}\left(\Lambda(1,2,0,1) - \Lambda(1,1,0,1)\right)\right) - \frac{\theta(\theta+2)}{(\theta+1)^2}\Lambda(1,1,0,1)\right). \end{aligned}$$

Table 14 shows the estimated values for the asymptotic $AM_E(\lambda)$, $AV_E(\lambda)$ and $\widetilde{\theta}$ for different choices of λ. From this table we can observe that as λ increases, $AM_E(\lambda)$, $AV_E(\lambda)$ go to zero.

Table 15 shows the estimated values for the asymptotic $AM_L(\theta)$, $AV_L(\theta)$ and $\widetilde{\lambda}$ for different choices of θ. From this table we can observe that as θ increases, $AM_L(\theta)$, $AV_L(\theta)$ go to zero.

B: Asymptotic results for censored sample

The asymptotic results for the censored data setting can be easily obtained from the corre-

sponding results for the complete data setting in the previous section.

Theorem 3

If the data come from $Exp(\lambda)$ ($Lin(\theta)$) distribution, then T^* is approximately normally distributed with mean $E_E(T^*)$ ($E_L(T^*)$) and variance $V_E(T^*)$ ($V_L(T^*)$), where the asymptotic means and variances of T^* assuming exponential and Lindley as null distributions can be given by

$$AM_E = \lim_{n\to\infty} \frac{E_E(T^*)}{n} = E_E\left[\delta_i \left\{\log f_E(x_i;\lambda) - \log f_L(x_i;\theta)\right\} + (1-\delta_i)\left\{\log \bar{F}_E(t_0;\lambda) - \log \bar{F}_L(t_0;\theta)\right\}\right],$$

$$AV_E = \lim_{n\to\infty} \frac{V_E(T^*)}{n} = V_E\left[\delta_i \left\{\log f_E(x_i;\lambda) - \log f_L(x_i;\theta)\right\} + (1-\delta_i)\left\{\log \bar{F}_E(t_0;\lambda) - \log \bar{F}_L(t_0;\theta)\right\}\right],$$

$$AM_L = \lim_{n\to\infty} \frac{E_L(T^*)}{n} = E_L\left[\delta_i \left\{\log f_E(x_i;\lambda) - \log f_L(x_i;\theta)\right\} + (1-\delta_i)\left\{\log \bar{F}_E(t_0;\lambda) - \log \bar{F}_L(t_0;\theta)\right\}\right],$$

$$AV_L = \lim_{n\to\infty} \frac{V_L(T^*)}{n} = V_L\left[\delta_i \left\{\log f_E(x_i;\lambda) - \log f_L(x_i;\theta)\right\} + (1-\delta_i)\left\{\log \bar{F}_E(t_0;\lambda) - \log \bar{F}_L(t_0;\theta)\right\}\right],$$

respectively.

Table 16 presents the estimated values for the asymptotic metrics $AM_E(\lambda)$, $AV_E(\lambda)$, and $\widetilde{\theta}$ for various values of λ. From this table, it can be observed that as λ increases, both $AM_E(\lambda)$ and $AV_E(\lambda)$ approach zero. Similarly, Table 17 displays the estimated values for the asymptotic metrics $AM_L(\theta)$, $AV_L(\theta)$, and $\widetilde{\lambda}$ for different choices of θ. It can be noted from this table that as θ increases, both $AM_L(\theta)$ and $AV_L(\theta)$ also approach zero.

λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	0.0651	0.1570	0.1844
0.5	0.0130	0.0280	0.7808
0.9	0.0050	0.0105	1.2926
1.3	0.0025	0.0051	1.7694
1.5	0.0018	0.0038	2.0000
2.0	0.0010	0.0019	2.5616
2.5	0.0006	0.0011	3.1085

Table 16: Estimated values of $AM_E(\lambda)$, $AV_E(\lambda)$ and $\widetilde{\theta}$ for different choices of λ with 10% of censored observations.

θ	$AM_L(\theta)$	$AV_L(\theta)$	$\widetilde{\lambda}$
0.1	-0.0710	0.1084	0.0524
0.5	-0.0226	0.0409	0.3000
0.9	-0.0096	0.0181	0.5897
1.3	-0.0048	0.0092	0.9061
1.5	-0.0035	0.0068	1.0714
2.0	-0.0018	0.0035	1.5000
2.5	-0.0010	0.0020	1.9444

Table 17: Estimated values of $AM_L(\theta)$, $AV_L(\theta)$ and $\widetilde{\lambda}$ for different choices of θ with 10% of censored observations.

Exponential case with complete data			
λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	0.0773	0.1952	0.184
0.5	0.0175	0.0401	0.781
0.9	0.0073	0.0162	1.292
1.3	0.0038	0.0083	1.769
1.5	0.0028	0.0062	2.000
2.0	0.0016	0.0033	2.561
2.5	0.0009	0.0020	3.108
Lindley case with complete data			
θ	$AM_L(\theta)$	$AV_L(\theta)$	$\widetilde{\lambda}$
0.1	-0.0802	0.1203	0.052
0.5	-0.0275	0.0479	0.300
0.9	-0.0124	0.0223	0.589
1.3	-0.0066	0.0121	0.906
1.5	-0.0049	0.0092	1.071
2.0	-0.0026	0.0049	1.500
2.5	-0.0016	0.0029	1.944
Exponential case with 10% of type-I censored data			
λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	0.0651	0.1570	0.1844
0.5	0.0130	0.0280	0.7808
0.9	0.0050	0.0105	1.2926
1.3	0.0025	0.0051	1.7694
1.5	0.0018	0.0038	2.0000
2.0	0.0010	0.0019	2.5616
2.5	0.0006	0.0011	3.1085
Lindley case with 10% of type-I censored data			
θ	$AM_L(\theta)$	$AV_L(\theta)$	$\widetilde{\lambda}$
0.1	-0.0710	0.1084	0.0524
0.5	-0.0226	0.0409	0.3000
0.9	-0.0096	0.0181	0.5897
1.3	-0.0048	0.0092	0.9061
1.5	-0.0035	0.0068	1.0714
2.0	-0.0018	0.0035	1.5000
2.5	-0.0010	0.0020	1.9444

Table 18: Values of AM, AV and estimated $\widetilde{\lambda}$ (or $\widetilde{\theta}$) for different choices of λ (or θ) with a certain of censored observations.

Exponential case with complete data				Exponential case with 10% of type-I censored data			
λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$	λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	0.0773	0.1952	0.184	0.1	0.0651	0.1570	0.1844
0.5	0.0175	0.0401	0.781	0.5	0.0130	0.0280	0.7808
0.9	0.0073	0.0162	1.292	0.9	0.0050	0.0105	1.2926
1.3	0.0038	0.0083	1.769	1.3	0.0025	0.0051	1.7694
1.5	0.0028	0.0062	2.000	1.5	0.0018	0.0038	2.0000
2.0	0.0016	0.0033	2.561	2.0	0.0010	0.0019	2.5616
2.5	0.0009	0.0020	3.108	2.5	0.0006	0.0011	3.1085
Lindley case with complete data				Lindley case with 10% of type-I censored data			
λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$	λ	$AM_E(\lambda)$	$AV_E(\lambda)$	$\widetilde{\theta}$
0.1	-0.0802	0.1203	0.052	0.1	-0.0710	0.1084	0.0524
0.5	-0.0275	0.0479	0.300	0.5	-0.0226	0.0409	0.3000
0.9	-0.0124	0.0223	0.589	0.9	-0.0096	0.0181	0.5897
1.3	-0.0066	0.0121	0.906	1.3	-0.0048	0.0092	0.9061
1.5	-0.0049	0.0092	1.071	1.5	-0.0035	0.0068	1.0714
2.0	-0.0026	0.0049	1.500	2.0	-0.0018	0.0035	1.5000
2.5	-0.0016	0.0029	1.944	2.5	-0.0010	0.0020	1.9444

Table 19: Values of AM, AV and estimated $\widetilde{\lambda}$ (or $\widetilde{\theta}$) for different choices of λ (or θ) with a certain of censored observations.