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

Shovan Chowdhury [1], Marco Marozzi, corresponding author [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

May 15, 2025

Abstract

The exponential distribution has been used for modeling positively skewed data in biological 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 biological 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

Research in biological psychology frequently encounters data that is positively skewed, meaning the tail of the data distribution extends further to the right, with a concentration of data points at the lower end of the scale and fewer, more extreme values at the higher end. This characteristic is observed in various types of data, including reaction times, certain physiological measures, and scores on psychological scales assessing symptoms or distress. Several factors contribute to the prevalence of positive skew in biological psychology data. For instance, reaction times, a biomarker measure in biological psychology [12], are inherently bounded by zero (responses cannot be faster than instantaneous) and often exhibit a pattern where most responses are relatively quick, but a smaller number of trials elicit significantly slower responses due to various cognitive or motor factors. Similarly, physiological measures or symptom scores in clinical populations may show a floor effect, where many individuals have low or zero scores (indicating absence or low levels of a symptom/measure), while a smaller subset exhibits higher scores reflecting greater severity or activity [32]. Effectively modeling such non-normal and skewed data distributions is crucial for appropriate statistical analysis and interpretation.

To address the characteristics of such positively skewed data, probability distributions are employed. Evidence suggests that the exponential distribution, or models incorporating it, have been widely utilized in biological psychology and related neuroscience fields. A prominent example is the modeling of reaction time data using the ex-Gaussian distribution. The ex-Gaussian distribution is a convolution of a Gaussian (normal) distribution and an exponential distribution. This composite model effectively captures the typical shape of reaction time distributions, with the exponential component accounting for the positively skewed tail [25]. Beyond reaction times, the exponential distribution has been considered in other biological and psychological contexts. Some research indicates that the distributions of psychological distress scores in large populations can exhibit an exponential pattern [33]. Furthermore, in neuroscience, the timing of neuronal firing intervals is sometimes analyzed using exponential distributions, reflecting the stochastic nature of these events [30]. Studies in computational neuroscience also note that brain activities can follow distributions belonging to the exponential family, highlighting the relevance of such distributions in understanding neural processes [4].

While the exponential distribution, as discussed, has proven a useful model for certain types of positively skewed data in biological psychology, the task of identifying the most appropriate distribution for a given dataset can be challenging when multiple models offer a seemingly good fit. This challenge is particularly relevant when considering distributions with the same number of parameters, such as different one-parameter distributions that might both provide a good fit for the response variable in the context of distributional regression modeling [16] (e.g., Generalized Additive Models for Location, Scale, and Shape (GAMLSS) modeling). In such cases, standard model comparison criteria like the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) estimates may be too close, and the probability density

functions (pdfs) and cumulative distribution functions (cdfs) may show significant overlap, making it problematic to definitively 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 [27] [13] [24].

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 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. We begin by discussing the importance of selecting an appropriate probability distribution and its influence on interpreting hazard rates, followed by an exploration of the role of hazard rates in biological psychology data analysis. Next, we examine the statistical properties of the exponential and Lindley distributions. We then introduce our novel discriminating method based on the logarithm of the ratio of maximized likelihoods (RML). Subsequently, we detail the procedure for selecting between the exponential and Lindley distributions using this method, supported by a comprehensive simulation study. A practical application of the proposed method is demonstrated using data from a biological psychology experiment. Concluding remarks are provided in the final section. Asymptotic results are presented in the Appendix. Supplementary material, including the implementation of the proposed method in R and additional examples and simulation results, is available online (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 [16, 31]. 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" [17, 18] and to examine all parameters of probability distributions in their statistical analyses [35].

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 3 and 4). 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 4). However, their hrfs convey quite different narratives (see the plot in row 1, column 3 in Figure 5).

The 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 display various shapes, as illustrated in Figure 1 (also refer to Figure 1A from Panis et al. [27]), which consequently affects 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

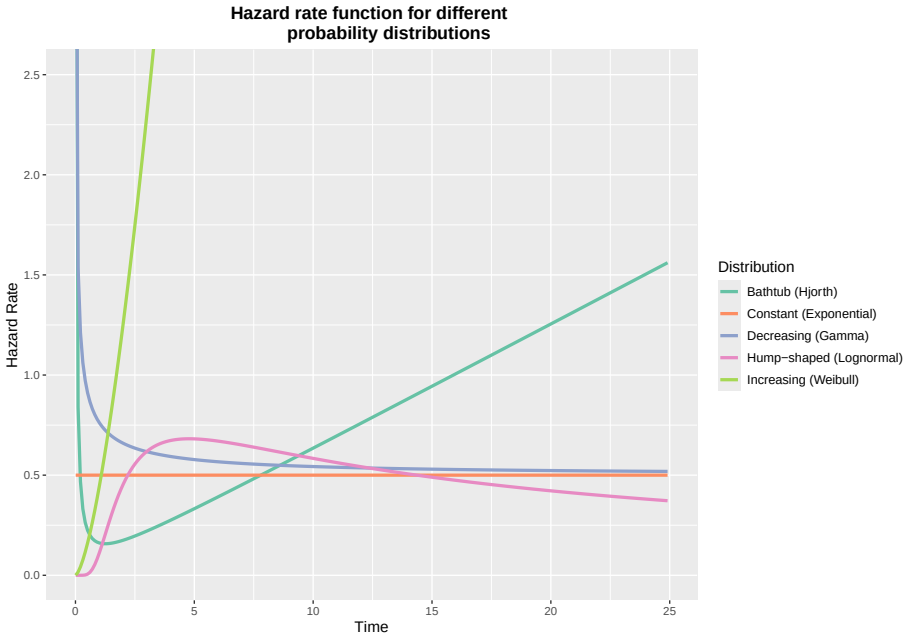

Figure 1: Illustration of hazard rate function shapes corresponding to different probability distributions: Hjorth (Bathtub), Exponential (Constant), Gamma (Decreasing), Lognormal (Hump-shaped), and Weibull (Increasing).

(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.

Examining hrfs offers valuable insights into the changing probability of an event (like a response or an error) over time within biological psychology experiments. Across different tasks that utilize temporal or performance measures, the resulting hazard rate functions can adopt distinct shapes, as the following scenarios illustrate. For instance, consider a task measuring the instantaneous probability of a basic sensory or motor response occurring immediately after a stimulus presentation, where the underlying process is simple and automatic. In such a scenario, involving pure detection or reflexive action without complex decision-making, a constant hazard rate shape could result, indicating a consistent instantaneous likelihood of response over time once processing begins. 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 processing 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. To illustrate these concepts with a specific example from cognitive assessment, 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, under time constraints. 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 2.

In terms of interpreting the hrf plots, it is evident that the instantaneous rate at which the event occurs differs between the 'immediate recognition' and 'cumulative recognition' tasks. For instance, at a given time point (e.g., 1,000 time units), 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 time unit) compared to the cumulative recognition condition (represented by a Gamma(1.5, 100) distribution with a rate of approximately 0.0096 events per time unit). As illustrated in the plot in the first row, second column of Figure 2, the probability of the event occurring in the immediate recognition task stays relatively constant over time, assuming the event has not yet occurred. Conversely, the cumulative recognition task shows an increasing probability of the event occurring over time, signaling a growing certainty of a correct identification as the memory search becomes more challenging. The second row in Figure 2 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.

As the preceding examples illustrate, hrfs offer a powerful means of capturing the temporal dynamics inherent in psychophysiological processes. Because hrfs are fundamentally properties derived from the probability distribution fitted to event-based or time-to-event data, they serve as a versatile tool for investigating temporal structure across various phenomena in biological psychology. This analytical approach is not limited to the specific types of events previously discussed. Many other measures commonly encountered in biological psychology, such as various

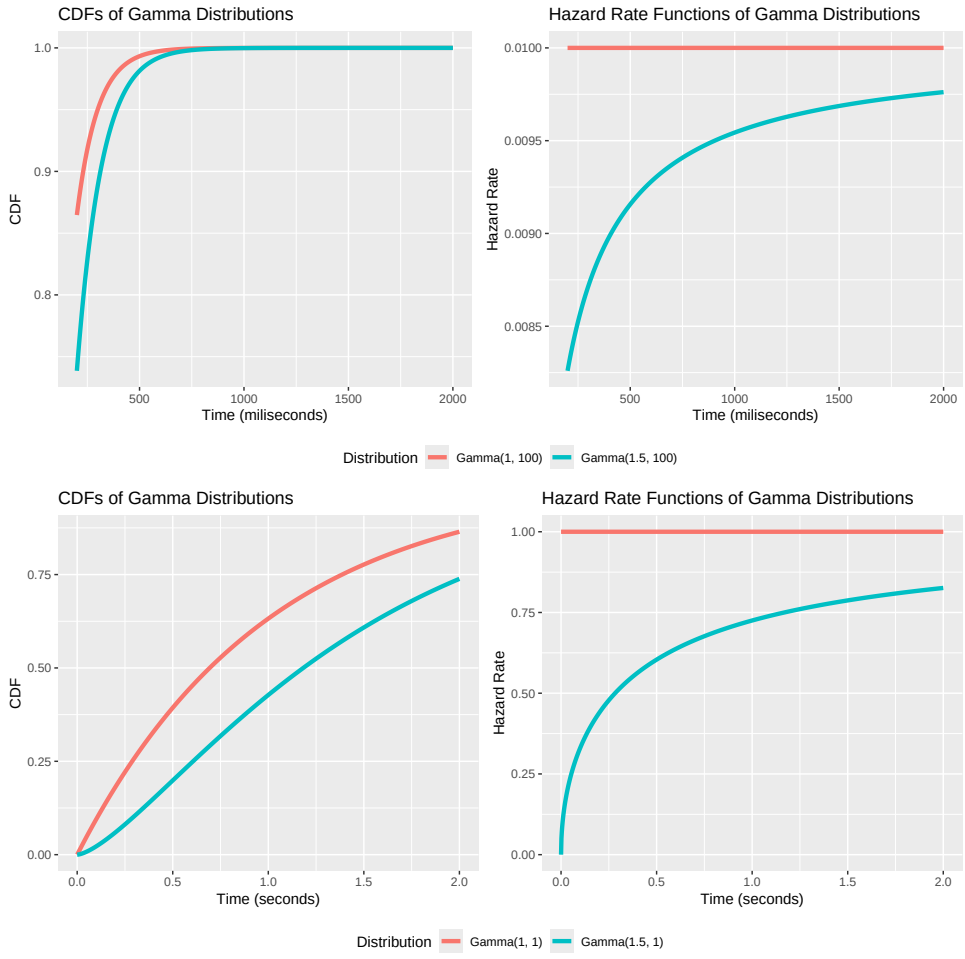

Figure 2: Distributions modeled with gamma distributions. The distributions have positive skews (first column) but different hazard rates (second column). The same data are presented in two analogous metrics.

physiological indices and symptom severities tracked over time, also yield positively skewed data suitable for modeling with similar probability distributions. Consequently, the application and interpretation of hrfs can be broadly extrapolated to these diverse datasets. Analyzing hazard rates thus provides a unique, distribution-informed perspective on the temporal unfolding of biological and psychological events, offering insights that go beyond analyses focused merely on central tendency.

3 What hazard rate functions offer to data analysis

Panis et al. (2020) [27] argue for using hrfs to model positively skewed data, based on the following points:

- 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 [11] (see Figures 3, 4, and 5).

- Townsend (1990) [34] 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 other words, comparing the means may give a distorted picture of what is really going on.

- Right-censored observations are common when dealing with positively skewed data and can be incorporated into hrf analyses, giving important information often overlooked when using other standard methods (e.g. ANOVA).

- Time-varying explanatory covariates are commonly encountered in biological psychology (e.g. biomarkers related to respondent autonomic nervous system) and they are easier to incorporate into hrf-based analyses than into standard methods (e.g. ANOVA).

- The hazard function is a critical measure in many biological psychology experiments. It quantifies the instantaneous rate at which an event of interest occurs (e.g., a response or task completion) at a given moment in time, provided the event has not occurred previously. Analyzing hrfs allows for a direct assessment of this instantaneous rate of occurrence over time, whereas standard statistical methods (such as ANOVA) are typically less suited to directly evaluate these specific temporal dynamics.

We thus believe that hrfs are very suitable to examine positively skewed data and in general to compare distributions with different shapes. We thus echo Panis et al. (2020) [27] in that this analytical approach needs to be incorporated into current data modeling practices, particularly in the framework of distributional regression modeling. Figure 5 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 [21]. 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.* [10]. 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.* [10], 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 [21, 22]. Mazucheli and Achcar [26] used the Lindley distribution to model lifetime data relating to competing risks and Ghitany *et al.* [10] used it to model the waiting time of bank customers. That is, the Lindley distribution seems suitable for fitting 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

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

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 3 and 4 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.

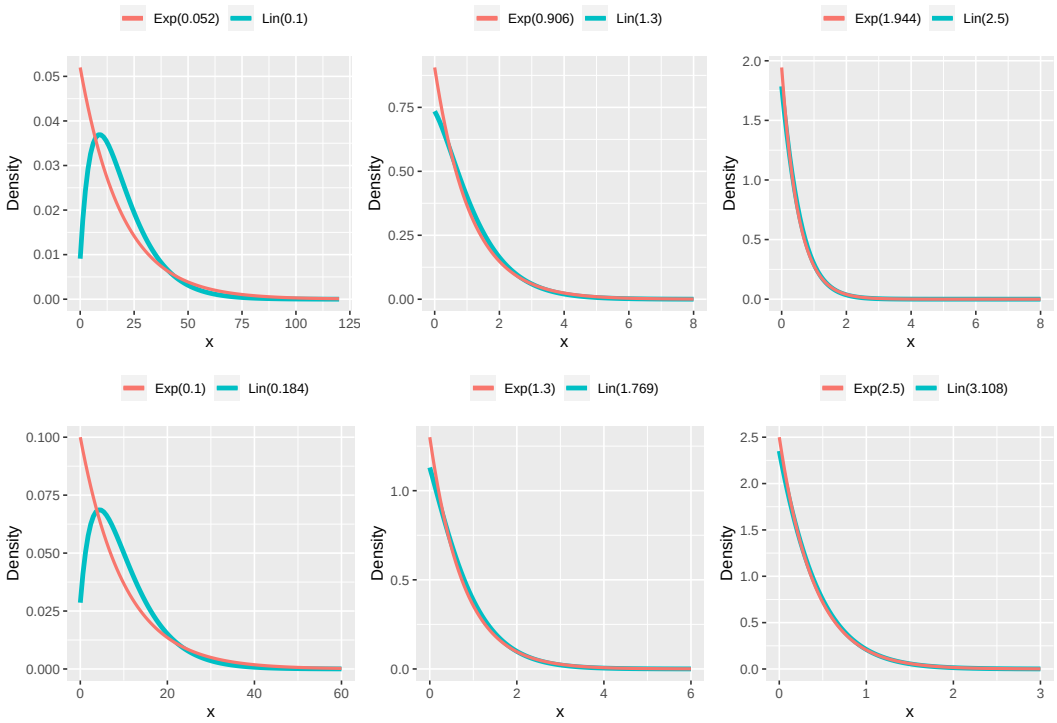

Figure 3: PDF of Lindley (Lin) and exponential (Exp) distributions for different parameters.

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.

The problem of selecting the correct distribution is not new in the statistics literature. The problem of discriminating between two non-nested models for a complete data set was first considered by Cox [5, 6] and was later studied by other scholars, including Bain and Engelhardt [2], and Fearn and Nebenzahl [9]. 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 ([19], [14]), the Gamma and Log-Normal distributions ([20]), and the exponential and Lindley distributions ([7], [36]). 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 Type-I censored samples ([29]), and Weibull and lognormal distributions for Type II censored samples ([3], [15], [8]).

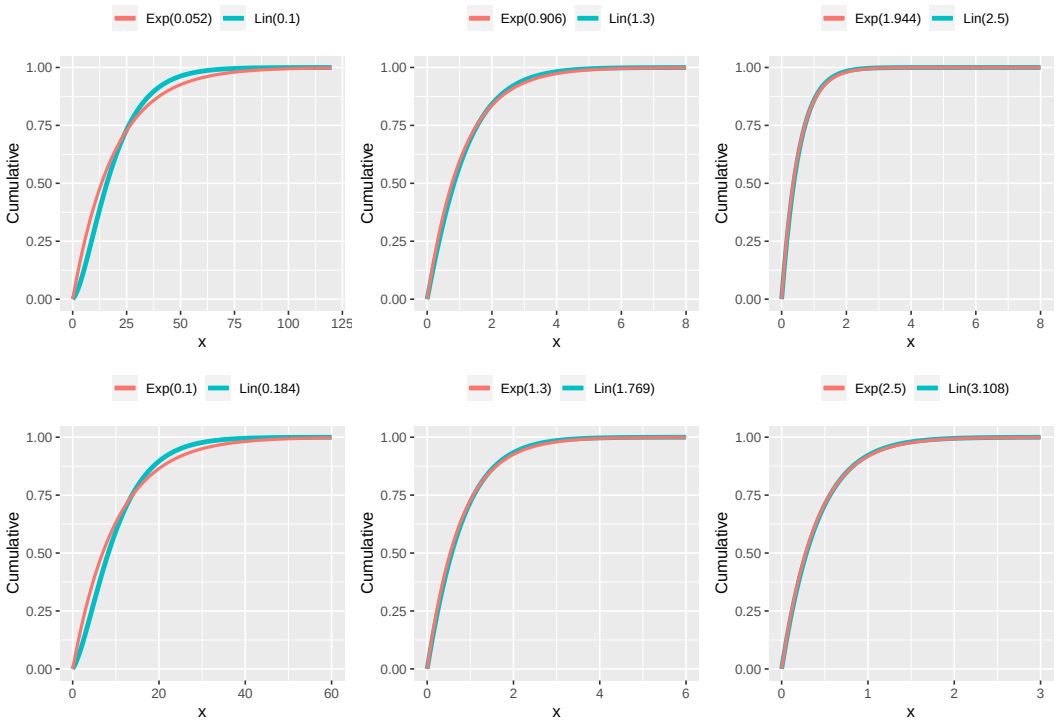

Figure 4: CDF of Lindley (Lin) and Exponential (Exp) distributions for different parameters.

5 Discrimination Procedure

Data analyzed in biological 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 [1, 23]. Two common types of censoring discussed in statistics are Type I, where a study ends at a predetermined time, and Type II, where it ends after a set number of events occur. While these specific censoring schemes are uncommon in typical cross-sectional biological psychology experiments, the underlying issue of incomplete observation can arise. For example, a scenario resembling Type I censoring occurs when a researcher imposes a maximum allowable response value; any true response exceeding this limit is only recorded as the maximum, effectively censoring the actual value. An analogue to Type II censoring might emerge if a researcher stopped a study once a specific count of subjects yielded a response within a defined range, thereby censoring the potential responses of other participants who had not yet met that criterion.

To further clarify how censoring can occur in a biological psychology context, consider

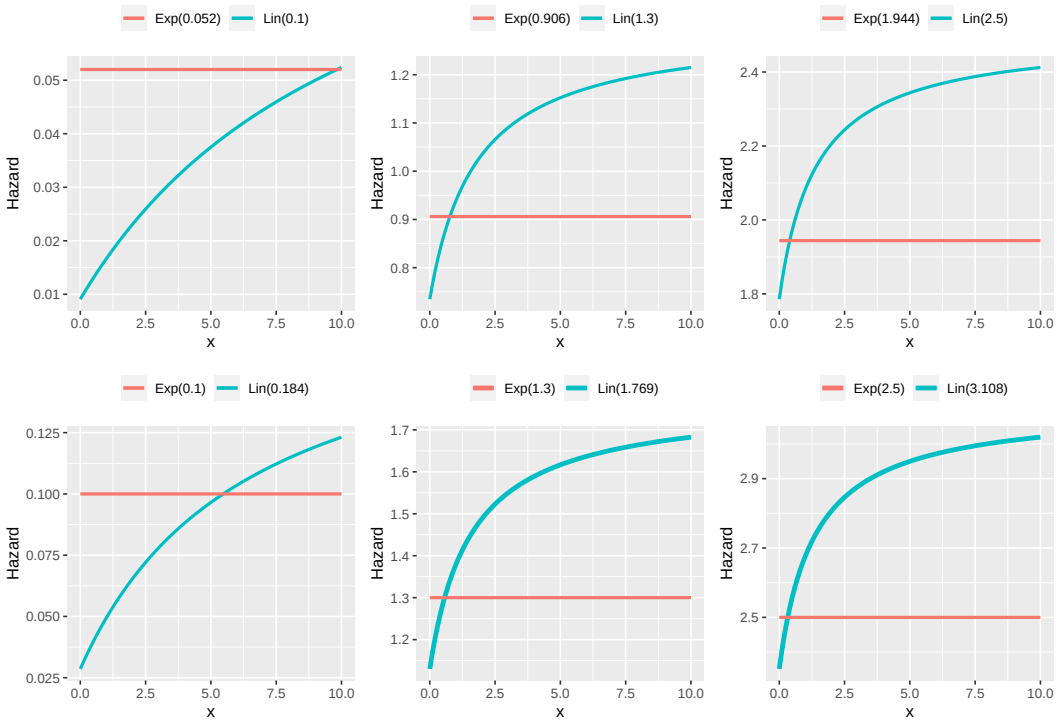

Figure 5: HRF of Lindley (Lin) and exponential (Exp) distributions for different parameters.

a hypothetical longitudinal cognitive neuroscience study. Suppose researchers are following 10 patients recovering from traumatic brain injury undergoing a novel cognitive rehabilitation program, periodically assessing their cognitive function using standardized cognitive assessments. The event of interest for analysis could be achieving a specific milestone of cognitive recovery, defined by reaching a pre-defined performance criterion on these assessments. However, not all patients may reach this milestone during the study period, or some might drop out for unrelated reasons, leading to incomplete observation of their event time – a situation known as censoring. Two ways this censoring might be structured, following standard definitions, are:

- *Type I Censoring:* The study has a fixed duration, for example, 12 months. For any patient who has not reached the cognitive recovery milestone by the end of 12 months, their event time is censored at 12 months. We know they remained in the 'at-risk' state for at least this period.

- *Type II Censoring:* The study continues until a fixed number of patients achieve the milestone, for instance, the first 7 patients. For the 3 patients who have not yet reached the milestone when the 7th patient does, their event time is censored. Data collection stops for everyone once the 7th event occurs.

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

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

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 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

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

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

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

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:

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

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.

$\theta \rightarrow$	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.

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 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 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: First, we determined a value t_0 corresponding to a percentile to ensure that a certain percentage (10%

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 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.

$\lambda\downarrow\ n\rightarrow$	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.

$\theta\downarrow\ n\rightarrow$	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.

$\lambda\downarrow\ n\rightarrow$	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.

7 Application to Biological Psychology

Posada-Quintero and Bolkhovsky [28] 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 [28] 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 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

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

$PCS = 0.633$ 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

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, in this case, the PCS is at least $min(0.594, 0.596) = 0.594$. 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, [28] 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 [16] would have been helpful to investigate the potential effects of

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 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. 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 biological psychology as well as in other fields where censored data are 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 [27]). We believe that while GAMLSS offers a powerful tool for selecting optimal distributions, our proposed method pro-

vides 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] 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.

[2] 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.

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

[4] Changbo, Z., Zhou, K., Tang, F., Tang, Y., and Si, B. (in press). A hierarchical Bayesian inference model for volatile multivariate exponentially distributed signals. *Frontiers in Computational Neuroscience*

[5] 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.

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

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

[8] 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.

[9] 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.

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

[11] 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.

[12] Jiménez-Figueroa, G., Ardila-Duarte, C., Pineda, D.A., Acosta-López, J. E., Cervantes-Henríquez, M. L., Pineda-Alhucema, W., Cervantes-Gutiérrez, J., Quintero-Ibarra, M., Sánchez-Rojas, M., Vélez, J., Puentes-Rozo, P. J. (2017). Prepotent response inhibition and reaction times in children with attention deficit/hyperactivity disorder from a Caribbean community. *ADHD Attention Deficit and Hyperactivity Disorders*, **9 (4)**, 199-211.

[13] Kang, H. A. (2017). Penalized partial likelihood inference of proportional hazards latent trait models. *British Journal of Mathematical and Statistical Psychology*, **70**: 187-208.

[14] 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.

[15] 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.

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

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

[18] 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

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

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

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

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

[23] 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

[24] Loeys, T., Legrand, C., Schettino, A. and Pourtois, G. (2014). Semi-parametric proportional hazards models with crossed random effects for psychometric response times. *British Journal of Mathematical and Statistical Psychology*, **67**: 304-327.

[25] Marmolejo-Ramos, F., Barrera-Causil, C., Kuang, S., Fazlali, Z., Wegener, D., Kneib, T., De Bastiani, F., and Martinez-Flórez, G. (2023). Generalised exponential-Gaussian distribution: A method for neural reaction time analysis. *Cognitive Neurodynamics*, **17(1)**, 221–237. https://doi.org/10.1007/s11571-022-09813-2

[26] 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.

[27] 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)**.

[28] 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.

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

[30] Roberts, W. M., Augustine, S. B., Lawton, K. J., Lindsay, T. H., Thiele, T. R., Izquierdo, E. J., Faumont, S., Lindsay, R. A., Britton, M. C., Pokala, N., Bargmann, C. I., and Lockery, S. R. (2016). A stochastic neuronal model predicts random search behaviors at multiple spatial scales in C. elegans. *eLife*, **5, e12572**. https://doi.org/10.7554/eLife.12572

[31] 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.

[32] Tomitaka, S., Kawasaki, Y., Ide, K. et al. (2017). Exponential distribution of total depressive symptom scores in relation to exponential latent trait and item threshold distributions: a simulation study. *BMC Reseach Notes*, **10 (614)**. https://doi.org/10.1186/s13104-017-2937-6

[33] Tomitaka S, Kawasaki Y, Ide K, Akutagawa M, Yamada H, Yutaka O, Furukawa TA. (2017). Pattern analysis of total item score and item response of the Kessler Screening Scale for Psychological Distress (K6) in a nationally representative sample of US adults. *PeerJ*. **5:e2987**. doi: 10.7717/peerj.2987.

[34] 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.

[35] 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

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

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