Austrian Journal of Statistics
January 2023, Volume 52, 104–115.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i2.1406

On Fixed-Accuracy Confidence Intervals for the Parameters of Lindley Distribution and Its Extensions

Sudeep R. Bapat
Indian Institute
of Management
Indore, India

Neeraj Joshi
Indian Institute
of Technology Kanpur
Uttar Pradesh, India

Ashish Kumar Shukla
Ramanujan College
University of Delhi
Delhi, India

Abstract

The purpose of the present paper is to deal with sequential estimation of the parameter θ in a Lindley distribution. A fixed-accuracy confidence interval for θ with a preassigned confidence coefficient is developed. It is established that, no fixed sample size procedure can solve the estimation problem and hence a purely sequential methodology is proposed to deal with the situation. The first-order asymptotic efficiency and consistency properties associated with our purely sequential strategy are derived. Similar estimation strategies are also outlined for a few other extensions of the Lindley distribution. Extensive simulation analysis is carried out to validate the theoretical findings. We also provide a real data example, where we estimate the parameter related to the "initial mass function" for a particular cluster of stars.

Keywords: asymptotic consistency, asymptotic efficiency, confidence interval, fixed - accuracy, initial mass function, interval estimation, Lindley distribution, sequential estimation, simulation.

1. Introduction

The Lindley distribution was first introduced by Lindley (1958) in order to study fiducial distribution and Bayes theorem. In recent years, it has emerged as an excellent model for lifetime data analysis. Ghitany, Atieh, and Nadarajah (2008) have studied the properties and applications of Lindley distribution in detail and established its superiority over the exponential distribution in many ways. The exponential distribution has a constant hazard rate and mean residual life while the Lindley distribution has increasing hazard rate and decreasing mean residual life. Moreover, since Lindley distribution has only one parameter, it is easily tractable. These are some reasons behind the rapidly growing popularity of this distribution.
Numerous authors have worked on the Lindley distribution, its generalizations and the associated statistical inference. For instance, Nadarajah, Bakouch, and Tahmasbi (2011) have proposed a generalized Lindley distribution to model the lifetime data. They established that, this distribution has better hazard rate properties in comparison to some existing lifetime models. Bakouch, Al-Zahrani, Al-Shomrani, Marchi, and Louzada (2012) have introduced an extension of the Lindley distribution which offers a more flexible model for lifetime data. Al-Mutairi, Ghitany, and Kundu (2013) have developed inferential procedures for estimating the stress-strength reliability parameter when both stress

and strength are independent Lindley random variables with different shape parameters. Gomez-Deniz, Sordo, and Calderm-Ojeda (2014) have proposed a Log-Lindley distribution as an alternative to the beta regression model. Bhati, Sastry, and Maha-Qadri (2015) have proposed a generalized Poisson-Lindley distribution and studied its applications and properties in detail. Asgharzadeh, Ng, Valiollahi, and Azizpour (2017) have discussed inferential procedures for the Lindley distribution based on type II censoring scheme. Asgharzadeh, Fallah, Raqab, and Valiollahi (2018) have provided statistical inference for the Lindley distribution based on record data. Maiti and Mukherjee (2018) provided estimation methods for the probability density function (*p.d.f.*) and cumulative distribution function (*c.d.f.*) of the Lindley distribution. Joshi and Jose (2018) came up with the wrapped Lindley distribution which is an excellent model for the analysis of circular data. Abdi, Asgharzadeh, Bakouch, and Alipour (2019) have studied the properties of a compound Lindley distribution with application to failure data. Mazucheli, Bapat, and Menezes (2020) have proposed a new one parameter Unit-Lindley distribution for analyzing data in the unit interval and discussed its properties in detail. Pandey, Kaushik, Singh, and Singh (2021) have provided a detailed statistical analysis for the generalized progressive hybrid censored data from the Lindley distribution under step-stress partially accelerated life test model.

In addition to the above, a bunch of other prolific work has also been done for the Lindley distribution. We are not mentioning many such contributions here for the sake of brevity. However, the aforementioned work and other existing literature surrounding the Lindley distribution is based on the fixed sample size scenarios and as of now, no work has been done for the case when sample size is not fixed in advance. In the present paper, our motive is to fill this gap by using a sequential sampling scheme. We consider a Lindley distribution having the *p.d.f.*:

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

where θ is an unknown parameter. The corresponding *c.d.f.* is given by:

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

Our focus is to sequentially estimate the parameter θ. Of course, one can further easily estimate the mean also, as it is a one-one function of θ. We will develop a fixed-accuracy confidence interval for θ and then propose a purely sequential methodology to handle the estimation problem. It is worth mentioning here that, sequential fixed-accuracy confidence interval estimation approach has been considered widely in the literature of sequential analysis. For a general overview, one may refer to Govindarajulu (1967), Ghosh, Mukhopadhyay, and Sen (1997) and Cha, Lee, and Jeon (2006). Recently, Mukhopadhyay and Banerjee (2014, 2015), Banerjee and Mukhopadhyay (2016) developed a class of fixed-accuracy confidence interval estimation methodologies with a preassigned confidence coefficient for the mean of a negative binomial distribution. Mukhopadhyay and Zhuang (2016) considered a fixed-accuracy interval in the Fisher's Nile example. Bapat (2018) developed a purely sequential fixed-accuracy confidence interval approach for the stress-strength reliability parameter under bivariate exponential models. Khalifeh, Mahmoudi, and Chaturvedi (2020) have worked on sequential fixed-accuracy confidence intervals for the stress-strength reliability parameter under exponential distribution. Zhuang, Hu, and Zou (2020) have proposed purely sequential strategies for estimating the reliability function of a two-parameter gamma distribution under a fixed-accuracy confidence interval estimation set-up. Most recently, Hu, Zhuang, and Goldiner (2021) have considered the problem of fixed-accuracy confidence interval estimation of the stress-strength reliability parameter under a geometric-exponential model, where they developed two-stage and modified two-stage sampling procedures and derived the interesting asymptotic properties. For a comprehensive overview of sequential methodologies, one may refer to Mukhopadhyay and de Silva (2009).

The rest of the paper is organized as follows: In Section 2, we develop a fixed-accuracy confidence interval with preassigned accuracy for the unknown parameter θ of the Lindley distribution (1). We discuss some key points regarding the formulation of this type of confidence interval. Then we propose a purely sequential methodology to handle our interval estimation problem in Section 3. We also provide some interesting first-order efficiency properties of our proposed sequential methodology. In

Section 4 we develop similar sequential estimation strategies for a two-parameter Lindley distribution, which assumes a bounded parameter. Section 5 includes an extensive simulation analysis of our proposed sequential methodology. We present results under various configurations and choices of the concerned parameters. In Section 6, we provide real life data analysis to show the practical utility of our proposed sequential methodology. We end with brief conclusions in Section 7.

2. Fixed-accuracy confidence interval for the Lindley parameter

Let $X_1, X_2, \ldots, X_n$ be n observations from the Lindley distribution (1). Let $\widehat{\theta}_n$ be the maximum likelihood estimator (MLE) of the parameter θ, where $\widehat{\theta}_n$ [see Ghitany *et al.* (2008)] is given by,

$$\widehat{\theta}_n = \frac{-(\overline{X}-1)+\sqrt{(\overline{X}-1)^2+8\overline{X}}}{2\overline{X}}, \ \overline{X} > 0. \tag{3}$$

Ghitany *et al.* (2008) have also proved that, $\widehat{\theta}_n$ is a positively biased estimator of θ. Moreover, $\widehat{\theta}_n$ is consistent and asymptotically normal and satisfies the following asymptotic normality result:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{D} N\left(0, \frac{\theta^2(\theta+1)^2}{\theta^2+4\theta+2}\right), \tag{4}$$

where $\xrightarrow{D}$ stands for convergence in distribution.
For a given $d > 0$, a fixed-width confidence interval I_n for θ with length $2d$ based on $\widehat{\theta}_n$ is given by,

$$I_n = \left\{\theta : \ \theta \in \left[\widehat{\theta}_n - d, \ \widehat{\theta}_n + d\right]\right\}, \tag{5}$$

The confidence interval I_n of the form (5) aims at controlling the width of the interval $2d$ and the confidence limit $1-\alpha$ $(0 < \alpha < 1)$ simultaneously and it is symmetric around the parameter θ. This type of confidence interval may not be reasonable since lower bound of this confidence interval has a positive probability of being negative, even though the parameter is entirely positive. Thus, we adopt a different approach and consider the confidence interval of the form,

$$J_n = \left[\frac{\widehat{\theta}_n}{d}, \ \widehat{\theta}_n d\right], \tag{6}$$

with pre-fixed accuracy $d > 1$. This is known as a fixed-accuracy confidence interval and it is symmetric around $\log \widehat{\theta}_n$ for the unknown parameter $\log \theta$. It thus takes into account the positivity of the parameter θ. Even in this case it may happen that, if the parameter space is bounded (say from above by U), a fixed-accuracy confidence interval may contain bounds which cross U. One may fix this problem by constructing a bounded length fixed-accuracy interval. One may refer to Mukhopadhyay and Banerjee (2014) or Banerjee and Mukhopadhyay (2016) for a general framework.
In the following section, we develop a purely sequential strategy to provide a fixed-accuracy confidence interval for the unknown parameter θ.

3. Purely sequential methodology

We want to construct a confidence interval of the form (6) for the unknown parameter θ with the coverage probability $1-\alpha$ $(0 < \alpha < 1)$ and the pre-fixed accuracy d. There does not exist any fixed-sample size procedure that solves such a problem. Thus, we have to resort to a sequential sampling methodology. It is worth mentioning here that, a sequential estimation strategy involves a stopping rule, which determines the optimal sample size to be used in the study. This strategy can reduce the number of observations needed for the inference part to a great extent, which proves to be beneficial as it reduces both time and cost substantially. For the present problem, the entire set-up is being provided

here.
For some pre-fixed accuracy $d(>1)$, the confidence interval J_n from (6) can be rewritten as,

$$J_n = \left\{ \theta : \ \theta \in \left[\frac{\widehat{\theta}_n}{d}, \ \widehat{\theta}_n d \right] \right\}. \tag{7}$$

Using (4) and delta method, we obtain,

$$\sqrt{n}(\log \widehat{\theta}_n - \log \theta) \xrightarrow{D} N\left(0, \frac{(\theta+1)^2}{\theta^2+4\theta+2}\right), \ as \ n \to \infty. \tag{8}$$

Now for J_n to include θ with the pre-fixed coverage probability $1-\alpha$, the required fixed-sample size can be obtained as follows:

$$P\left(d^{-1}\widehat{\theta}_n \le \theta \le d\widehat{\theta}_n\right) = 1-\alpha,$$

$$\Rightarrow n^* \equiv n_d^* = \left(\frac{z_{\alpha/2}}{\log d}\right)^2 \frac{(\theta+1)^2}{\theta^2+4\theta+2}, \tag{9}$$

where $z_{\alpha/2}$ is the upper $100(\alpha/2)\%$ point of a standard normal distribution.
Now since θ is unknown, there does not exist any fixed sample size procedure that will attain the pre-fixed coverage probability $1-\alpha$. Hence, we propose the following purely sequential methodology: We first fix an integer $m(>1)$ and obtain the pilot sample $X_1, X_2, ..., X_m$ from the density given in (1). We then proceed to collect one additional observation as needed until the sampling is terminated according to the rule,

$$N \equiv N_d = \inf\left\{ n \ge m : \ n \ge \left(\frac{z_{\alpha/2}}{\log d}\right)^2 \frac{(\widehat{\theta}_n+1)^2}{\widehat{\theta}_n^2+4\widehat{\theta}_n+2} \right\}. \tag{10}$$

Now we have a final data set as $X_1, X_2, ..., X_N$ and we will now estimate θ using the interval

$$J_N = \left[d^{-1}\widehat{\theta}_N, \ \widehat{\theta}_N d\right].$$

In the following theorems, we provide some interesting properties associated with our stopping rule (10).

Theorem 1. *For the stopping rule N defined in (10),* $P_\theta(N < \infty) = 1$.

Proof. From the stopping rule (10), we have,

$$\begin{aligned} P(N=\infty) &= \lim_{n\to\infty} P(N>n) \\ &\le \lim_{n\to\infty} P\left[n \le \left(\frac{z_{\alpha/2}}{\log d}\right)^2 \frac{(\widehat{\theta}_n+1)^2}{\widehat{\theta}_n^2+4\widehat{\theta}_n+2} \right] \\ &= 0, \end{aligned} \tag{11}$$

since $\widehat{\theta}_n \to \theta$ almost surely as $n \to \infty$, and Theorem 1 follows. □

Theorem 2. *For the stopping rule* $(N, \widehat{\theta}_N)$ *with N defined in (10), for each fixed value of* $\theta > 0$*, we have as* $d \to 1$*:*

1. $N/n^* \to 1$ *w.p. 1;*
2. $E_\theta[N/n^*] \to 1$ *(asymptotic first-order efficiency);*
3. $P_\theta(\theta \in J_N) \to 1-\alpha$ *(asymptotic consistency);*

where n^* *comes from (9) and* $0 < \alpha < 1$.

Proof. 1. From (10), we have,

$$\left(\frac{z_{\alpha/2}}{\log d}\right)^2 \frac{(\widehat{\theta}_N+1)^2}{\widehat{\theta}_N+4\widehat{\theta}_N+2} \le N \le \left(\frac{z_{\alpha/2}}{\log d}\right)^2 \frac{(\widehat{\theta}_{N-1}+1)^2}{\widehat{\theta}_{N-1}+4\widehat{\theta}_{N-1}+2} + m. \tag{12}$$

On dividing throughout by n^*, we get,

$$\frac{\theta^2+4\theta+2}{\widehat{\theta}_N+4\widehat{\theta}_N+2}\left(\frac{\widehat{\theta}_N+1}{\theta+1}\right)^2 \le \frac{N}{n^*} \le \frac{\theta^2+4\theta+2}{\widehat{\theta}_{N-1}+4\widehat{\theta}_{N-1}+2}\left(\frac{\widehat{\theta}_{N-1}+1}{\theta+1}\right)^2 + \frac{m}{n^*}. \tag{13}$$

Now the proof of part 1 follows by taking limits throughout and noting that $N \to \infty$ w.p. (with probability) 1, $\widehat{\theta}_N \to \theta$ w.p. 1 and $m/n^* \to 0$ as $d \to 1$.

2. To prove part 2, let $\lambda_1 = (1-\varepsilon)n^*$ and $\lambda_2 = (1+\varepsilon)n^*$, where ε is an arbitrary real number such that $0 < \varepsilon < 1$. We note that,

$$\begin{aligned} E(N) &= \sum_{n=m}^{\infty} nP(N=n) \\ &= \sum_{n=m}^{\lambda_1} nP(N=n) + \sum_{n>\lambda_1} nP(N=n) \\ &\ge \lambda_1 P(m \le N \le \lambda_1) \\ \Rightarrow\ E\left(\frac{N}{n^*}\right) &\ge (1-\varepsilon)P(m \le N \le \lambda_1). \end{aligned} \tag{14}$$

Now using the fact that $P(N<\infty)=1$ (Theorem 1) and ε is any arbitrary value, we get,

$$\liminf_{d\to 1} E\left(\frac{N}{n^*}\right) \ge 1. \tag{15}$$

On the other hand,

$$\begin{aligned} E(N) &= \sum_{n=m}^{\infty} nP(N=n) \\ &= \sum_{n=m}^{\lambda_2} nP(N=n) + \sum_{n\ge\lambda_2+1} nP(N=n) \\ &\le \lambda_2 P(N \le \lambda_2) + \sum_{n\ge\lambda_2+1} nP(N=n) \\ \Rightarrow\ E\left(\frac{N}{n^*}\right) &\le (1+\varepsilon) + \frac{1}{n^*}\sum_{n\ge\lambda_2} (n+1)P(N=n+1). \end{aligned} \tag{16}$$

It follows from (16) that,

$$\limsup_{d\to 1} E\left(\frac{N}{n^*}\right) \le 1. \tag{17}$$

Now part 2 follows by combining (15) and (17).

3. The confidence interval for θ is $J_N = [d^{-1}\widehat{\theta}_N,\ d\widehat{\theta}_N]$. We can write,

$$P_\theta(\theta \in J_N) = P_\theta(d^{-1}\widehat{\theta}_N \le \theta \le d\widehat{\theta}_N) \tag{18}$$

$$= P_\theta(\log\widehat{\theta}_N - \log d \le \log\theta \le \log\widehat{\theta}_N + \log d) \tag{19}$$

$$= P_\theta(|\log\widehat{\theta}_N - \log\theta| \le \log d). \tag{20}$$

Now using the random central limit theorem of Anscombe (1952), we can write,

$$\frac{\sqrt{N}(\widehat{\theta}_N-\theta)}{\sqrt{Var(\widehat{\theta}_N)}} \xrightarrow{D} N(0,1)\ \ as\ \ d\to 1. \tag{21}$$

Using (20) and the Mann-Wald theorem, we define W_N, where,

$$W_N=\frac{z_{\alpha/2}(\log\widehat{\theta}_N-\log\theta)}{\log d}\xrightarrow{D} N(0,1)\ \ as\ \ d\to 1. \tag{22}$$

Now using (19), we have,

$$P_\theta(d^{-1}\widehat{\theta}_N\le\theta\le d\widehat{\theta}_N)=P_\theta(\mid W_N\mid\le z_{\alpha/2})\to 1-\alpha.$$

The proof of part 3 is now complete.

□

4. An extension to a two-parameter Lindley distribution

The above strategy can be extended to other distributions, in cases where one/multiple parameters are bounded by a constant. We will now present an analogous confidence interval for one of the parameters of a two parameter Lindley distribution (TLD), to showcase a slightly different strategy. TLD was introduced by Shanker, Sharma, and Shanker (2013) and is seen useful to model waiting and survival times. Its moments, failure rate function, mean residual life function and stochastic orderings appear to be significantly better than a one-parameter Lindley distribution. The density function and the distribution function of a TLD with parameters α and θ are given respectively by,

$$f(x;\alpha,\theta)=\frac{\theta^2}{\theta+\alpha}(1+\alpha x)e^{-\theta x}, \tag{23}$$

$$F(x)=1-\frac{\theta+\alpha+\alpha\theta x}{\theta+\alpha}e^{-\theta x}, \tag{24}$$

where $x>0,\theta>0$ and $\alpha>-\theta$. Now let $X_1,X_2,...,X_n$ be n observations from a TLD as given in (23). We further assume α to be known, either from prior information or data. Let θ_n^* be a method of moments estimator (MME) of θ, which can be obtained as follows:
Let $k=\mu_2'/\mu_1'^2$, where μ_1' and μ_2' are the first two moments about origin for a TLD. Now if we assume $\theta=b\alpha$, we get,

$$k=\frac{2b^2+8b+6}{b^2+4b+4}.$$

On the other hand, since the ratio $k=\mu_2'/\mu_1'^2$ is unknown, one can estimate it using the ratio of the corresponding sample estimates namely, $\bar{X}$ and m_2'. Now using this estimated value of k and the above equation, one can obtain the value of b. Finally after simplification one can evaluate the MME as follows,

$$\theta_n^*=\left(\frac{b+2}{b+1}\right)\frac{1}{\overline{X}}. \tag{25}$$

Now let $\theta_n^*\equiv g(\overline{X},m_2')$. This function g is clearly differentiable and hence one can invoke the Taylor series expansion around the true moment vector (μ_1',μ_2'). Specifically,

$$g(\theta)=\left(\frac{b+2}{b+1}\right)\frac{1}{\theta}\ \text{ and }\ g'(\theta)=-\left(\frac{b+2}{b+1}\right)\frac{1}{\theta^2}. \tag{26}$$

Now on using the Mann-Wald theorem, we have,

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{D} N\left(0, \left(\frac{b+2}{b+1}\right)^2 \frac{\sigma^2}{\theta^4}\right), \tag{27}$$

where σ^2 denotes the variance of TLD and is given by $(\theta^2+4\theta\alpha+2\alpha^2)/(\theta^2(\theta+\alpha)^2)$. Now in order to construct an appropriate fixed-accuracy confidence interval, we make a minor adjustment as follows:
Since $\alpha > -\theta$, $\alpha+\theta > 0$, hence, similar to (6), one can write a fixed-accuracy confidence interval for $\alpha+\theta$ as follows:

$$J_n^* = \left[\frac{(\alpha+\theta_n^*)}{d^*}, (\alpha+\theta_n^*)d^*\right], \tag{28}$$

with a pre-fixed accuracy $d^*(>1)$. We can then diligently follow steps outlined in Section 3, first to come up with an appropriate optimum sample size (say n^{**}) and then to apply a purely sequential strategy (similar to (10)) to obtain a final stopping variable (say N^*). We then have a final dataset $X_1, X_2, ..., X_{N^*}$ and can estimate $\alpha+\theta$ using the interval

$$J_{N^*}^* = [d^{*^{-1}}(\alpha+\theta_{N^*}^*),\ d^*(\alpha+\theta_{N^*}^*)]. \tag{29}$$

One can then easily adjust the above interval to obtain the required fixed-accuracy interval for θ, by subtracting α from both endpoints. Similar theorems can be proved easily in this case as well, but we leave out the details for brevity.

Apart from the above extension and methodology, a similar strategy can be applied to other families of Lindley distribution. To quickly outline a few:

1. The “Quasi Lindley distribution" with parameters α and θ was introduced by Shanker and Mishra (2013), and has the following density function,

$$f(x;\alpha,\theta) = \frac{\theta(\alpha+x\theta)}{\alpha+1}e^{-\theta x}, \quad x>0, \theta>0, \alpha>-1.$$

   Now if the problem is to find a fixed-accuracy confidence interval for α, assuming θ is known, one can adopt the following strategy: since $\alpha > -1, \alpha+1 > 0$, and one can design an interval similar to (6), first for $\alpha+1$ and then suitably for α.

2. The “Pseudo-Lindley distribution" with parameters β and θ was introduced by Zeghdoudi and Nedjar (2016), and has the following density function,

$$f(x;\beta,\theta) = \frac{\theta(\beta-1+\theta x)}{\beta}e^{-\theta x}, \quad x,\theta,\beta > 1.$$

   Now if the problem is to find a fixed-accuracy confidence interval for β, assuming θ is known, one can adopt the following strategy: since $\beta > 1, \beta-1 > 0$, and one can design an interval similar to (6), first for $\beta-1$ and then suitably for β.

5. Analysis from simulations

We now present simulation analyses associated with our proposed purely sequential procedure (10) which gives a fixed-accuracy confidence interval for the unknown parameter θ. These are tabulated in Tables 1 and 2. We first generate a set of pseudorandom observations at-a-time from a Lindley distribution (1) with two different choices of θ being 2 and 0.5. Further, each row in Tables 1 and 2 corresponds to analysis from 10,000 replications for a range of values for d with n^* coming from (9). We also fix the pilot sample size $m = 5$ under each case. The values of α are fixed to be 0.05 and 0.01 in Tables 1 and 2 respectively.

Each block in Tables 1 and 2 relate to different values of θ and shows the value of d (column 1), n^* (column 2), estimated values $\overline{n}$ along with their estimated standard errors $s_{\overline{n}}$ (columns 3 and 4), and the ratio $\overline{n}/n^*$ (column 5). We will denote p to be an indicator variable taking a value 1(or 0) if a constructed confidence interval from a run includes (or does not include) the true value of θ. $\overline{p}$ is hence the average p obtained from 10,000 runs along with its standard errors shown in columns 6 and 7. This gives us a sense of the achieved coverage probability. We fixed the values of α as 0.05 and 0.01 and we thus expect the $\overline{p}$ values to be close to 0.95 and 0.99 in Tables 1 and 2 respectively. Finally just for comparison, column 8 contains the average of the obtained confidence intervals based on N observations. We will denote this as $\bar{J_n}$. This gives an idea about how the actual parameter θ is contained in the interval.

One can note that both Tables 1 and 2 indicate that the values of $\overline{n}$ estimate n^* very accurately across the rows, which become closer for larger values of n^* or smaller values of d. This is consistent with the first-order efficiency assumption. The values of $\overline{p}$ are also seen to be close to 0.95 or 0.99 as expected. Further, a general observation is that the values of $\overline{n}$ seem closer to the corresponding values of n^* when $\alpha = 0.01$, as compared to when $\alpha = 0.05$. The intervals $\bar{J_n}$ also capture the actual values of the corresponding θ in all the cases. Figure 1 depicts the varying optimal sample sizes over a range of d values for fixed values of α. Figure 1(a) assumes $\theta = 2$ whereas Figure 1(b) assumes $\theta = 0.5$. As one can see, and intuitively, the required optimum sample size decreases, as the fixed-accuracy $(= d)$ increases, in both the cases.

Table 1: Simulation results from 10,000 replications for the purely sequential methodology (10) with $m = 5, \alpha = 0.05$

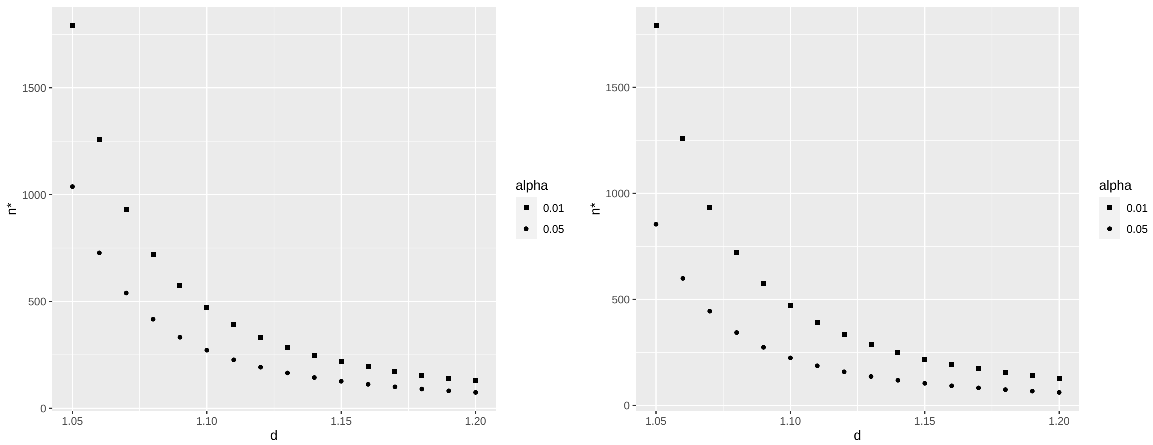
d	n^*	$\bar{n}$	$s_{\bar{n}}$	$\bar{n}/n^*$	$\bar{p}$	$s_{\bar{p}}$	$\bar{J_n}$
				$\theta = 2$			
1.2	74.29	74.82	0.0044	1.0072	0.9599	0.0001	(1.66, 2.40)
1.15	126.42	126.93	0.0067	1.0040	0.9595	0.0001	(1.74, 2.30)
1.09	332.52	333.06	0.0162	1.0016	0.9591	0.0001	(1.83, 2.18)
1.08	416.93	417.46	0.0202	1.0012	0.9564	0.0003	(1.85, 2.16)
1.06	727.33	727.86	0.0352	1.0007	0.9521	0.0014	(1.88, 2.12)
				$\theta = 0.5$			
1.2	61.18	61.95	0.0021	1.0126	0.9595	0.0001	(0.41, 0.60)
1.15	104.11	104.73	0.0044	1.0059	0.9576	0.0001	(0.43, 0.57)
1.09	273.84	274.35	0.0057	1.0018	0.9541	0.0002	(0.45, 0.54)
1.08	343.35	343.87	0.0069	1.0015	0.9522	0.0005	(0.46, 0.54)
1.06	598.98	599.92	0.0111	1.0009	0.9513	0.0009	(0.47, 0.53)

6. Analysis from real data

We now illustrate the applicability of our proposed purely sequential procedure (10) using a real data set from astrophysics. The "initial mass function", denoted as IMF, describes the initial distribution of masses for a population of stars. The IMF is usually fitted using either three or four power laws. One may refer to Hopkins (2018) for more details. However, we will apply our sequential strategies on a low-mass IMF, which happens to follow a Lindley distribution adequately. This low-mass IMF in question, is that of a massive young cluster (2-3 Myr) named as NGC 6611, which ionizes the Eagle Nebula. This cluster contains masses from $1.5M_\odot \geq M \geq 0.02M_\odot$ and hence the brown dwarfs region ($\approx 0.2M_\odot$) is covered. This data set was reported by Oliveira, Jeffries, and van Loon (2009), where they presented very deep photometric observations of the central region of NGC 6611 cluster, using a *Hubble Space Telescope*. One may additionally refer to the corresponding CDS catalog J/MNRAS/392/1034 or Zaninetti (2020) for more details about the dataset. An advantage of adopting a sequential rule in this case, rather than a fixed-sample strategy is the huge reduction in sampling cost and time, since collecting observations using a telescope can quickly get expensive.

Table 2: Simulation results from 10,000 replications for the purely sequential methodology (10) with $m = 5, \alpha = 0.01$

d	n^*	$\bar{n}$	$s_{\bar{n}}$	$\bar{n}/n^*$	$\bar{p}$	$s_{\bar{p}}$	$\bar{J}_n$
				$\theta = 2$			
1.2	128.31	128.81	0.0067	1.0039	0.9995	0.0001	$(1.66, 2.40)$
1.15	218.35	218.86	0.0108	1.0023	0.9994	0.0002	$(1.74, 2.30)$
1.09	574.32	574.86	0.0278	1.0009	0.9992	0.0003	$(1.83, 2.18)$
1.08	720.12	720.66	0.0345	1.0007	0.9978	0.0004	$(1.85, 2.16)$
1.07	931.75	932.32	0.0453	1.0006	0.9913	0.0009	$(1.87, 2.14)$
				$\theta = 0.5$			
1.2	105.67	106.05	0.0022	1.0036	0.9999	0.0001	$(0.41, 0.60)$
1.15	179.82	180.29	0.0046	1.0026	0.9998	0.0001	$(0.43, 0.57)$
1.09	472.97	473.49	0.0090	1.0011	0.9998	0.0002	$(0.45, 0.54)$
1.08	593.04	593.55	0.0110	1.0008	0.9988	0.0003	$(0.46, 0.54)$
1.07	767.32	767.85	0.0141	1.0006	0.9967	0.0005	$(0.46, 0.53)$



(a) $\theta = 2$ (b) $\theta = 0.5$

Figure 1: Varying optimal sample size n^* over d for fixed α

The full data consists of 1023 stars in the cluster. A Lindley distribution fits the data well with a p-value of 0.161, as seen from a suitable chi-square test. One may refer to Zaninetti (2020) for an affirmation of the claim. Treating these data as the universe, the MLE of θ was $\widehat{\theta} = 1.51$. We then implemented our purely sequential estimation procedure (10) to draw observations from the full set of data as required.

Table 3 summarizes the corresponding results. We have considered several fixed values of d ranging from 1.05 to 1.09 along with a couple of extreme values such as 1.15 and 1.20 just as a comparison. We fixed the pilot sample size $m = 5$ and the significance level $\alpha = 0.05$. After implementing the sequential rule (10) with a particular choice of d, we obtained the confidence interval $(d^{-1}\widehat{\theta}_N, d\widehat{\theta}_N)$ for the required parameter θ. As one can note, the confidence intervals contain the value of $\widehat{\theta}(= 1.51)$ which is taken to be as the true θ for illustrative purposes. The ratios N/n^* also appear close to 1 in each case which supports the first-order efficiency. However, it is indicated again that all of these results are outcomes of a single run.

7. Concluding remarks

A purely sequential estimation methodology was developed to construct a fixed-accuracy confidence

Table 3: Analysis of the IMF data using purely sequential methodology (10) with $m = 5$, $\alpha = 0.05$

d	n^*	N	N/n^*	$(d^{-1}\widehat{\theta}_N, d\widehat{\theta}_N)$
1.05	985.13	987	1.001	(1.44, 1.59)
1.06	690.69	692	1.002	(1.42, 1.59)
1.07	512.28	514	1.003	(1.43, 1.62)
1.08	395.92	396	1.001	(1.42, 1.66)
1.09	315.76	316	1.001	(1.39, 1.66)
1.15	120.05	120	0.999	(1.34, 1.78)
1.20	70.54	71	1.006	(1.28, 1.84)

interval for the parameter of a Lindley distribution. Our proposed sequential methodology enjoys asymptotic consistency and efficiency properties as well. Further, suitable strategies were designed under a few extensions of the Lindley distribution, by tweaking the original method slightly. Extensive simulation analysis is carried out to support our claims and a real data example from astrophysics is also presented for the practical utility of our proposed methodology. One may also look to develop multi-stage methodologies for the present problem, which may appear in a sequel.

Acknowledgments

We are grateful to the Editor and anonymous reviewer(s) for their constructive comments and suggestions on an earlier version of this paper leading to an improved presentation. The corresponding author, Neeraj Joshi is also indebted to the Department of Science and Technology, Government of India for providing financial support under the INSPIRE fellowship scheme (Award No. - IF170889).

References

Abdi M, Asgharzadeh A, Bakouch H, Alipour Z (2019). "A New Compound Gamma and Lindley Distribution with Application to Failure Data." *Austrian Journal of Statistics*, **48**, 54–75.

Al-Mutairi D, Ghitany M, Kundu D (2013). "Inferences on Stress-Strength Reliability from Lindley Distributions." *Communications in Statistics - Theory and Methods*, **42**, 1443–1463.

Anscombe F (1952). "Large Sample Theory of Sequential Estimation." *Proceedings of the Cambridge Philosophical Society*, **48**, 600–607.

Asgharzadeh A, Fallah A, Raqab M, Valiollahi R (2018). "Statistical Inference Based on Lindley Record Data." *Statistical Papers*, **59**, 759–779.

Asgharzadeh A, Ng H, Valiollahi R, Azizpour M (2017). "Statistical Inference for Lindley Model Based on Type II Censored Data." *Journal of Statistical Theory and Applications*, **16**, 178–197.

Bakouch H, Al-Zahrani B, Al-Shomrani A, Marchi V, Louzada F (2012). "An Extended Lindley Distribution." *Journal of the Korean Statistical Society*, **41**, 75–85.

Banerjee S, Mukhopadhyay N (2016). "A General Sequential Fixed-Accuracy Confidence Interval Estimation Methodology for a Positive Parameter: Illustrations Using Health and Safety Data." *Annals of the Institute of Statistical Mathematics*, **68**, 541–570.

Bapat S (2018). "Purely Sequential Fixed Accuracy Confidence Intervals for $P(X > Y)$ Under Bivariate Exponential Models." *American Journal of Mathematical and Management Sciences*, **37**, 386–400.

Bhati D, Sastry D, Maha-Qadri P (2015). "A New Generalized Poisson-Lindley Distribution: Applications and Properties." *Austrian Journal of Statistics*, **44**, 35–51.

Cha J, Lee S, Jeon J (2006). "Sequential Confidence Interval Estimation for System Availability." *Quality and Reliability Engineering International*, **22**, 165–176.

Ghitany M, Atieh B, Nadarajah S (2008). "Lindley Distribution and Its Application." *Mathematics and Computers in Simulation*, **78**, 493–506.

Ghosh M, Mukhopadhyay N, Sen P (1997). *Sequential Estimation*. Wiley.

Gomez-Deniz E, Sordo M, Calderm-Ojeda E (2014). "The Log-Lindley Distribution as an Alternative to the Beta Regression Model with Applications in Insurance." *Insurance: Mathematics and Economics*, **54**, 49–57.

Govindarajulu Z (1967). "Two Sided Confidence Limits for $P(X < Y)$ Based on Normal Samples of X and Y." *Sankhya*, **29**, 35–40.

Hopkins A (2018). "The Dawes Review 8: Measuring the Stellar Initial Mass Function." *Publications of the Astronomical Society of Australia*, **35**, 1–39.

Hu J, Zhuang Y, Goldiner C (2021). "Fixed-Accuracy Confidence Interval Estimation of $P(X < Y)$ under a Geometric-Exponential Model." *Japanese Journal of Statistics and Data Science*, **4**, 1079–1104.

Joshi S, Jose K (2018). "Wrapped Lindley Distribution." *Communications in Statistics - Theory and Methods*, **47**, 1013–1021.

Khalifeh A, Mahmoudi E, Chaturvedi A (2020). "Sequential Fixed-Accuracy Confidence Intervals for the Stress-Strength Reliability Parameter for the Exponential Distribution: Two-Stage Sampling Procedure." *Computational Statistics*, **35**, 1553–1575.

Lindley D (1958). "Fiducial Distributions and Bayes Theorem." *Journal of the Royal Statistical Society: Series B*, **20**, 102–107.

Maiti S, Mukherjee I (2018). "On Estimation of the PDF and CDF of the Lindley Distribution." *Communications in Statistics - Simulation and Computation*, **47**, 1370–1381.

Mazucheli J, Bapat S, Menezes A (2020). "A New One Parameter Unit-Lindley Distribution." *Chilean Journal of Statistics*, **11**, 53–66.

Mukhopadhyay N, Banerjee S (2014). "Purely Sequential and Two Stage Fixed-Accuracy Confidence Interval Estimation Methods for Count Data for Negative Binomial Distributions in Statistical Ecology: One-Sample and Two-Sample Problems." *Sequential Analysis*, **33**, 251–285.

Mukhopadhyay N, Banerjee S (2015). "Purely Sequential Negative Binomial Problems and Statistical Ecology: A Selected Review with New Directions." *Statistical Methodology*, **26**, 34–60.

Mukhopadhyay N, de Silva B (2009). *Sequential Methods and Their Applications*. Boca Raton: CRC.

Mukhopadhyay N, Zhuang Y (2016). "On Fixed-Accuracy and Bounded Accuracy Confidence Interval Estimation Problems in Fisher's Nile Example." *Sequential Analysis*, **35**, 516–535.

Nadarajah S, Bakouch H, Tahmasbi R (2011). "A Generalized Lindley Distribution." *Sankhya B*, **73**, 331–359.

Oliveira J, Jeffries R, van Loon J (2009). "The Low-Mass Initial Mass Function in the Younger Cluster NGC 6611." *Monthly Notices of the Royal Astronomical Society*, **392**, 1034–1050.

Pandey A, Kaushik A, Singh S, Singh U (2021). "Statistical Analysis for Generalized Progressive Hybrid Censored Data from Lindley Distribution under Step-Stress Partially Accelerated Life Test Model." *Austrian Journal of Statistics*, **50**, 105–120.

Shanker R, Mishra A (2013). "A Quasi Lindley Distribution." *African Journal of Mathematics and Computer Science Research*, **6**, 64–71.

Shanker R, Sharma S, Shanker R (2013). "A Two-Parameter Lindley Distribution for Modeling Waiting and Survival Times Data." *Applied Mathematics*, **4**, 363–368.

Zaninetti L (2020). "New Probability Distributions in Astrophysics: II. The Generalized and Double Truncated Lindley." *International Journal of Astronomy and Astrophysics*, **10**, 39–55.

Zeghdoudi H, Nedjar S (2016). "A Pseudo Lindley Distribution and Its Application." *Afrika Statistika*, **11**, 923–932.

Zhuang Y, Hu J, Zou Y (2020). "Fixed-Accuracy Confidence Interval Estimation of $P(X > c)$ for a Two-Parameter Gamma Population." *Communications for Statistical Applications and Methods*, **27**, 625–639.

Affiliation:

Sudeep R. Bapat
Department of Operations Management & Quantitative Techniques
Indian Institute of Management, Indore
Indore-453556, M.P., India
E-mail: sudeepb@iimidr.ac.in

Neeraj Joshi
Department of Industrial and Management Engineering
Indian Institute of Technology Kanpur
Kanpur-208016, Uttar Pradesh, India
E-mail: stats.joshi@gmail.com

Ashish Kumar Shukla
Department of Statistics, Ramanujan College
University of Delhi
Delhi-110019, Delhi, India
E-mail: ashish.iasqm@gmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2021-08-20
January 2023 *Accepted:* 2021-12-17