PURELY SEQUENTIAL AND TWO-STAGE
BOUND-LIGHT LENGTH CONFIDENCE INTERVAL
ESTIMATION PROBLEMS IN FISHER’S
“NILE” EXAMPLE

Nitis Mukhopadhyay* and Yan Zhuang*

Fisher’s “Nile” example is a classic which involves a bivariate random variable
\((X, Y)\) having a joint probability density function given by
\[ f(x, y; \theta) = \exp(-\theta x - \theta^{-1} y), \quad 0 < x, y < \infty, \quad \text{where } \theta > 0 \]
where \(\theta > 0\) is a single unknown parameter. We develop
bounded-length confidence interval estimations for \(P_\theta(X > a)\) with a preassigned
confidence coefficient using both purely sequential and two-stage methodologies. We
show: (i) Both methodologies enjoy asymptotic first-order efficiency and asymptotic
consistency properties; (ii) Both methodologies enjoy second-order efficiency prop-
erties. After presenting substantial theoretical investigations, we have also imple-
mented extensive sets of computer simulations to empirically validate the theoretical
properties.

Key words and phrases: Asymptotic consistency, bounded-length, confidence inter-
vals, first-order asymptotic efficiency, maximum likelihood estimator, “Nile” example,
purely sequential sampling, second-order asymptotic efficiency, two-stage sampling.

1. Introduction

Fisher (1973) revisited an interesting bivariate distribution from his previous
discourses (Fisher (1934, 1956)) on classical “Nile” example. That joint distri-
bution was given by the following probability density function (p.d.f.):

\[
(1.1) \quad f(x, y; \theta) = \begin{cases} 
\exp(-\theta x - \theta^{-1} y) & \text{if } x > 0, y > 0 \\
0 & \text{otherwise} 
\end{cases}
\]

where \(\theta(> 0)\) is an unknown parameter. What is special about this distribution
may be briefly summarized as follows: The maximum likelihood estimator (MLE)
for \(\theta\), namely \(\hat{\theta}_{\text{MLE}}\), is not sufficient for \(\theta\), but there exists an ancillary comple-
ment \(U\). In other words, while \(U\) is an ancillary statistic, the statistic \((\hat{\theta}_{\text{MLE}}, U)\)
is a jointly sufficient statistic for \(\theta\).

The parameter \(\theta\) may indicate the depth of a river (for example, Nile) in
some location within its path where the river is prone to flooding during heavy
rainy season. In order to estimate \(\theta\), one may record two measurements at-a-
time: \(X\) may be the speed at which water flows whereas \(Y\) may be the volume of
water that flows by. If \(\theta\) is large (small), one may expect \(X\) to be small (large),
but $Y$ would be accordingly large (small). Thus, estimation of $\theta$ is directly linked to chances of potential flooding of the banks of the river.

This distribution (1.1) led Fisher to propose generally that inferences based on some non-sufficient MLE $\hat{\theta}_{\text{MLE}}$ ought to be conditioned on the observed value of an ancillary complement $U$ in order to recoup or make up for lost information due to one’s use of $\hat{\theta}_{\text{MLE}}$. This path-breaking direction from Fisher (1934, 1956, 1973) led him to create the foundation of conditional inference in a solid footing within statistical science.

1.1. Brief literature review

Over the years, many researchers have returned to Fisher’s example (1.1). One may refer to Basu (1964), Rao (1973), Cox and Hinkley (1974), Lehmann and Casella (1998), Ghosh et al. (2010), Kagan and Malinovsky (2013, 2016), Mukhopadhyay (2000, 2014), Mukhopadhyay and Zhuang (2016, 2017) and other sources. Mukhopadhyay and Zhuang (2017) constructed a number of Nile-like illustrations where the MLE of $\theta$ was non-sufficient, had less than full information, but its ancillary complement helped in recovering the full information.

Joshi and Shah (1999) handled unbiased point estimation problem for the parametric function $\tau(\theta) \equiv P_\theta(Y < X) = (1 + \theta^2)^{-1}$ based on sufficient statistics of $\theta$. They derived a number of crucial expressions of the moments of the MLE for $\tau(\theta)$ including the bias and mean squared error of the MLE.

Mukhopadhyay and Banerjee (2014, 2015a) introduced fixed-accuracy confidence interval estimations for the mean parameter in a negative binomial distribution. Banerjee and Mukhopadhyay (2016) developed a general structure for fixed-accuracy confidence interval estimation methodologies for a positive parameter of an arbitrary distribution, which would enjoy asymptotic consistency and asymptotic first-order efficiency properties. They handled precisely both implementation and validity of their proposed methodologies with the help of illustrations involving odds-ratio estimation in a Bernoulli($\theta$) distribution and mean estimations in the case of Poisson($\theta$) and Normal($\theta$, $\theta$) distributions. Mukhopadhyay and Banerjee (2015b) proposed two-stage and purely sequential bounded-length confidence intervals for $\theta$ in a Bernoulli($\theta$) distribution, where $0 < \theta < 1$ is an unknown parameter.

More specifically, Mukhopadhyay and Zhuang (2016) developed both fixed-width and fixed-accuracy confidence intervals for $\theta$ in Fisher’s example (1.1). Their fixed-accuracy confidence interval estimation methodology for $\theta$ turned out to have some major advantages over the fixed-width confidence interval estimation methodology for $\theta$. An appropriate fixed-sample-size estimation strategy was developed with both exact and approximate properties which could be guaranteed to produce fixed-accuracy confidence intervals for $\theta$. Also, a bounded-accuracy confidence interval for $P_\theta(Y < X)$ associated with (1.1), and the requisite fixed-sample-size methodology were introduced.

In this paper, we aim at estimating the parametric function, $P_\theta(X > a)$, $a > 0$, with the help of bounded-length confidence intervals. Here, the parametric function $P_\theta(X > a)$ lies between 0 and 1, and so Banerjee and Mukhopadhyay’s
general methodology does not immediately apply. Our goal is to appropriately improvise upon Banerjee and Mukhopadhyay’s (2016) general methodology in order to make it fit with requirements that we must face.

The parametric function \( P_\theta(X > a) \) may be directly linked to calibrate chances of flooding the banks of the (Nile) river. Its accurate estimation may lead to a good warning system for thousands of people whose livelihood relies upon the status of the river and its many estuaries. Early warning of possible severe flooding will save thousands of lives, hopefully allowing families time to move to higher grounds and stay away from harm’s way.

Now, if one happens to decide to propose sampling strategies based on utilizing the \( X \)-data (or \( Y \)-data) alone, many technical details arising from the associated estimation problems may become rather simple in nature, but we do not advocate such a route at all. We emphasize that our data will consist of paired observations on \((X, Y)\) and then utilizing the \( X \)-data or \( Y \)-data alone would amount to significant loss of information about \( \theta \). This requires us to develop new and innovative ideas as well as interesting and productive new formulations, theoretical challenges, and methodologies. Next, we provide a layout of our presentation.

1.2. Layout of the paper

We begin Section 2 describing the foundation, formulation and motivation. First, we may transform \( p_\theta \) from (2.1) to another suitable parametric function \( q_\theta \) defined on the space \((0, \infty)\) where \( q_\theta \) is a one-to-one function of \( p_\theta \). Then, in the spirits of Mukhopadhyay and Banerjee (2015a, b) and Banerjee and Mukhopadhyay (2016), we proceed to build a fixed-accuracy confidence interval \( J_n \subset (0, \infty) \) for \( q_\theta \) as in (2.2) with confidence coefficient at least (or approximately) \( 1 - \alpha \) where \( 0 < \alpha < 1 \) is preassigned. This will lead to a bounded-length confidence interval \( K_n \) for \( p_\theta \) as illustrated in (2.6). Theorem 2.1 shows that the length of the finally proposed confidence interval \( K_n \) for \( p_\theta \) is bounded from above by a preassigned positive number.

In Section 3, we supply more details surrounding the bounded-length confidence interval estimation problem for \( p_\theta \) and obtain an expression of \( n^*_d \equiv n^*_d(\theta) \) given by (3.4), the required optimal fixed sample size. Lemma 3.1, which may be of some independent interest, helps us to show the existence of a natural positive lower bound \( n_{0d} \) for \( n^*_d(\theta) \) in (3.8) where \( n_{0d} \) is known and \( n_{0d}, n^*_d \) have the same order as \( d \to 1+ \).

Section 4 develops the bounded-length confidence interval estimation problem for \( p_\theta \) using a properly designed purely sequential sampling strategy along with nonlinear renewal theoretic representations (Subsection 4.1). We prove that the purely sequential estimation methodology enjoys attractive properties (Theorems 4.1–4.3) including asymptotic first-order efficiency and asymptotic consistency (Theorem 4.1). Even though the boundary condition in our purely sequential stopping time (4.1) admittedly appears complicated, we are able to obtain asymptotic second-order properties (Theorem 4.4). In doing so, we have been mindful in addressing the role of second-order approximate expression \( \kappa(\theta) \)
from (4.19) numerically (Table 1). Subsection 4.4 summarizes truly encouraging findings obtained from computer simulations.

In Section 5, we address the bounded-length confidence interval estimation problem for $p_\theta$ under a properly designed two-stage sequential sampling strategy. After developing some requisite preliminaries (Subsection 5.1), we move to verify that the two-stage estimation methodology (5.1)–(5.2) enjoys attractive properties (Theorem 5.1) including asymptotic first-order efficiency and asymptotic consistency. We have also strengthened our assertion considerably by obtaining asymptotic second-order properties (Theorems 5.2–5.3) including asymptotic second-order efficiency (Theorem 5.3). In doing so, we have again been especially mindful in addressing the role of second-order approximate expressions numerically (Subsections 5.3.1–5.3.2). Subsection 5.4 summarizes truly encouraging findings obtained from computer simulations.

Based on the summaries of data analyses, we find that both proposed purely sequential as well as the two-stage estimation methodologies perform remarkably well across the board. This sentiment is clearly validated under all circumstances whether or not the sample size happens to be small, moderate, or large. Section 6 lays down a number of concluding thoughts.

2. Foundation-formulation-motivation

We believe that it is important to estimate $p_\theta \equiv P_\theta(X > a)$, $a > 0$ by means of fixed-accuracy confidence intervals. In the same spirit, however, one may argue and want to estimate $P_\theta(Y > b)$, $b > 0$ instead of $p_\theta$. For brevity, we focus on estimating $p_\theta$. Obviously,

$$p_\theta \equiv P_\theta(X > a) = e^{-a\theta}, \quad (2.1)$$

where $a(> 0)$ is a fixed constant.

Obviously, $p_\theta \in (0, 1)$ and we transform $p_\theta$ to $q_\theta$ where

$$q_\theta \equiv \frac{p_\theta}{1 - p_\theta} = \frac{e^{-a\theta}}{1 - e^{-a\theta}}. \quad (2.2)$$

Clearly, the space for $q_\theta$ is $(0, \infty)$ as Banerjee and Mukhopadhyay (2016) would have required and $q_\theta$ is also a one-to-one function of $p_\theta$. Thus, we will accordingly tie a fixed accuracy confidence interval for $q_\theta$ with an associated bounded-length confidence interval for $p_\theta$. We will explain more as we move forward.

In the light of Mukhopadhyay and Banerjee (2015a, b), Banerjee and Mukhopadhyay (2016), Mukhopadhyay and Zhuang (2016), we begin with independent and identically distributed (i.i.d.) random samples $(X_i, Y_i)$, $i = 1, 2, \ldots, n$, each having the common p.d.f. (1.1). The MLE of $\theta$ is given by:

$$T_n \equiv \hat{\theta}_{n,MLE} = (\sum_{i=1}^{n} Y_i / \sum_{i=1}^{n} X_i)^{1/2}. \quad (2.3)$$

For fixed $n$, we note that $T_n$ is a biased but consistent estimator of $\theta$. 
Now, according to the invariance property of the MLE, in view of (2.2), the MLE of $q_\theta$ can be expressed as follows:

$$U_n = e^{-aT_n}(1 - e^{-aT_n})^{-1}.$$  \hfill (2.4)

We construct a fixed-accuracy confidence interval $J_n$ for $q_\theta$ as follows:

$$J_n = \{q_\theta : q_\theta \in [d^{-1}U_n, dU_n]\},$$  \hfill (2.5)

involving the MLE $U_n$ for $q_\theta$ from (2.4) where $d(>1)$ is the preassigned fixed-accuracy measure.

Once the fixed-accuracy confidence interval for $q_\theta$ is constructed, the associated confidence interval $K_n$ for $p_\theta$ can be derived as follows in view of (2.2) and (2.4):

$$K_n = \{p_\theta : p_\theta \in [(d + U_n)^{-1}U_n, (dU_n + 1)^{-1}dU_n]\}.$$  \hfill (2.6)

**Theorem 2.1.** For all fixed sample size $n$ and for all fixed but otherwise arbitrary $d(>1)$, the length of the confidence interval $K_n$ from (2.6) proposed for $p_\theta$ defined in (2.1) is bounded from above by the expression $\frac{d-1}{d+1}$ w.p.1 where $d$ stands for the preassigned fixed-accuracy measure associated with $J_n$ from (2.5).

**Proof.** The length of the confidence interval is given by:

$$\text{Length}(K_n) = (dU_n + 1)^{-1}dU_n - (d + U_n)^{-1}U_n = \frac{(d^2 - 1)U_n}{(dU_n + 1)(d + U_n)}.$$  \hfill (2.7)

Now, letting

$$m(x) \equiv \frac{(d^2 - 1)x}{(dx + 1)(d + x)}, \quad x > 0,$$

we can easily check that the function $m(x)$ attains its global maximum $x = 1$. Thus, we can immediately claim that $\text{Length}(K_n) \leq \frac{d-1}{d+1}$ w.p.1. \hfill $\square$

At this point, it is understood that the upper bound $\frac{d-1}{d+1}$ meant for the length of the proposed confidence interval $K_n$ for $p_\theta$ is smaller than 1 and it goes to zero as $d \downarrow 1$. That ought to make good sense because after all the confidence interval $K_n$ is proposed for $p_\theta$ which is a number between zero and one.

In other words, if we alternatively begin with a preassigned number $0 < \delta < 1$ and we require that the proposed confidence interval $K_n$ has its length $\leq \delta$, then we should equate $\frac{d-1}{d+1}$ with $\delta$. We will thereby employ the proposed confidence interval procedure with $d = \frac{1+\delta}{1-\delta}$.

In closing this section, we should record one other important point: We have deliberately not yet tied the proposed confidence interval $K_n$ for $p_\theta$ with a given preassigned confidence level $1 - \alpha$, $0 < \alpha < 1$. Obviously, an expression for an appropriate minimum requisite fixed-sample-size $n$ associated with (2.5), and hence equivalently with (2.6), must be determined. This is explored explicitly in Section 3.
Remark 2.1. Generally speaking, suppose that one is interested in a bounded-length confidence interval of an unknown parametric function $\psi(\xi)$ of an unknown parameter $\xi$ where $c_L < \psi(\xi) < c_U$ with known lower and upper bounds $c_L, c_U$ respectively. Then, clearly, much like $p_\theta$ from (2.1), the parametric function $(c_U - \psi(\xi))/(c_U - c_L)$ will belong to the space $(0, 1)$. In other words, we should consider estimating a one-to-one transform of $\psi(\xi)$, namely $(c_U - \psi(\xi))/(\psi(\xi) - c_L) \in (0, \infty)$, much in the spirit of $q_\theta$ from (2.2). We live out a full-blown general discourse for brevity.

3. Optimal fixed sample size for a fixed preassigned confidence coefficient

The Fisher information about $\theta$ in a single pair of observation $(X, Y)$ is given by:

$$I_{(X,Y)}(\theta) = \frac{2}{\theta^2}.$$  

(3.1)

One may refer to Fisher (1934) and Mukhopadhyay and Zhuang (2016, 2017).

Now, having recorded i.i.d. random samples $(X_i, Y_i)$, $i = 1, \ldots, n$, each following the p.d.f. (1.1), recall $T_n \equiv (\Sigma_{i=1}^n Y_i/\Sigma_{i=1}^n X_i)^{1/2}$ from (2.3), the MLE for $\theta$. Then, we will have:

$$n^{1/2}(T_n - \theta) \xrightarrow{L} N \left(0, \frac{1}{2} \theta^2 \right) \quad \text{as } n \to \infty.$$  

(3.2)

With $q_\theta$ defined in (2.2), we express:

$$b(\theta) \equiv \log q_\theta = -a\theta - \log(1 - e^{-a\theta}),$$

which implies:

$$\frac{\partial b(\theta)}{\partial \theta} = -a - \frac{ae^{-a\theta}}{1 - e^{-a\theta}} = \frac{-a}{1 - e^{-a\theta}}.$$

Next, applying Mann-Wald theorem to (3.2), we can conclude:

$$n^{1/2}(\log U_n - \log q_\theta) \xrightarrow{L} N(0, \sigma^2(\theta)) \quad \text{as } n \to \infty$$  

(3.3)

with $\sigma^2(\theta) \equiv \frac{1}{2} a^2 \theta^2 (1 - e^{-a\theta})^{-2}$.

One may refer to Rao (1973, pp. 385–386), Mukhopadhyay (2000, pp. 261–262) or another source. Some readers may prefer to appeal to delta-method (Sen and Singer (1993, pp. 131–132)) in order to claim (3.3) from (3.2).

Thus, in order to obtain a fixed-accuracy confidence interval $J_n$ defined via (2.5) for the parametric function $q_\theta$ defined in (2.2), we utilize (3.3) to claim that $P_\theta\{q_\theta \in J_n\}$ is approximately $1 - \alpha$ for large $n$. Here, we assume that $0 < \alpha < 1$ is fixed and preassigned.
Hence, we can claim that $P_\theta\{q_\theta \in J_n\} \approx 1 - \alpha$ for large $n$ when the required minimum fixed sample size $n$ satisfies:

$$\text{(3.4)} \quad n \geq n^*_d \equiv n^*_d(\theta) = \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 a^2 \theta^2 (1 - e^{-a\theta})^{-2}.$$ 

Here, $z_{\alpha/2}$ is the upper $100(\alpha/2)$% point of a standard normal distribution. Obviously, the magnitude of $n^*_d$, the optimal fixed sample size, remains unknown since it involves the unknown parameter $\theta(>0)$.

### 3.1. Further examination of the optimal fixed sample size from (3.4)

We observe that $n^*_d$ from (3.4) can be alternatively expressed as follows:

$$\text{(3.5)} \quad n^*_d = \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \{g(a\theta)\}^2,$$

where we define the function $g(\cdot)$ as:

$$\text{(3.6)} \quad g(x) = x(1 - e^{-x})^{-1}, \quad x > 0.$$

Parts of the following lemma may already be known. It will help in arriving at an appropriate lower bound for $n^*_d$ and making it easier in developing our proposed sampling strategies in the sequel. Hence, we state it and sketch its proof for completeness.

**Lemma 3.1.** For all fixed $x > 0$, we have the following result:

$$1 < x(1 - e^{-x})^{-1} < 1 + x.$$

**Proof.** We define $h(x) \equiv x - (1 - e^{-x})$ for $x > 0$. Then, the first derivative, $h'(x) = 1 - e^{-x}$, which is positive for all $x > 0$. That is, $h(x)$ is a monotonically increasing function in $x$ for all $x > 0$. But, since $h(0) = 0$, we have $h(x) > 0$ for all $x > 0$. Thus, the lower bound holds.

Also, for all $x > 0$, we have:

$$e^x > 1 + x \iff 1 - e^{-x} > x(1 + x)^{-1},$$

which shows the upper bound. \(\square\)

Using the lower limit from Lemma 3.1 and then by appealing to (3.5), we obtain the following lower bound:

$$\text{(3.7)} \quad n^*_d \equiv n^*_d(\theta) > \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2,$$

for all $\theta > 0$. Hence, the pilot sample size will be defined as:

$$\text{(3.8)} \quad n_{0d} \equiv \left\lceil \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \right\rceil + 1,$$

where $\langle u \rangle$ denotes the largest integer smaller than $u$ with $u > 0$. We may emphasize that $n_{0d}$ does not involve the unknown parameter $\theta$. 
4. A purely sequential estimation methodology

We begin with the pilot set of data \((X_i, Y_i), i = 1, \ldots, n_0D\) with \(n_0D\) coming from (3.8). After that we continue by recording one additional pair of observation \((X, Y)\) at-a-time as needed by an appropriately defined stopping criterion until it decides to terminate gathering more data. Our stopping time is defined as follows:

\[
N_d \equiv N = \inf \left\{ n \geq n_0D : n \geq \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \left\{ g(aT_n) \right\}^2 \right\}.
\] (4.1)

One sees readily that \(N_d\) estimates the optimal fixed-sample-size \(n^*_d\) defined by (3.5). By referring to Chow and Robbins (1965), we can claim that

\[ P_{\theta} \{ N_d < \infty \} = 1 \]

for every fixed \(\theta > 0, d > 1, \) and \(a > 0.\) Upon termination, we’ll have the fully gathered data

\[ \{(X_i, Y_i), i = 1, 2, \ldots, n_0D, \ldots, N_d\}, \]

and we obtain the randomly stopped version of the MLE from (2.3) for \(\theta:\)

\[
T_{N_d} \equiv \hat{\theta}_{N_d, \text{MLE}} = (\Sigma_{i=1}^{N_d} Y_i / \Sigma_{i=1}^{N_d} X_i)^{1/2},
\] (4.2)

and the randomly stopped version of the MLE from (2.4) for \(q_\theta:\)

\[
U_{N_d} = e^{-aT_{N_d}} (1 - e^{-aT_{N_d}})^{-1}.
\] (4.3)

Then, we propose the associated fixed-accuracy confidence interval for \(q_\theta\) as follows:

\[
J_{N_d} = \{ q_\theta : q_\theta \in [d^{-1} U_{N_d}, d U_{N_d}] \},
\] (4.4)

and the bounded-length confidence interval for \(p_\theta\) as follows:

\[
K_{N_d} = \{ p_\theta : p_\theta \in [(d + U_{N_d})^{-1} U_{N_d}, (d U_{N_d} + 1)^{-1} d U_{N_d}] \},
\] (4.5)
in the spirits of (2.5) and (2.6) respectively. In view of Theorem 2.1, one can easily verify that the length of the confidence interval \(K_{N_d}\) for \(p_\theta\) is bounded above by \(\frac{d-1}{d+1}\).

4.1. Nonlinear renewal theoretic representation of the stopping time from (4.1)

Observe that the stopping time \(N_d\) from (4.1) can be rewritten as follows:

\[
N_d \equiv N = \inf \left\{ n \geq n_0D : n \{ g(aT_n) \}^{-2} \geq \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \right\}.
\] (4.6)

Recall that \(n_0D\) comes from (3.8) and \(g(\cdot)\) was defined in (3.6).

We refer to nonlinear renewal theoretic representation, originally developed by Woodroofe (1977, 1982) and Lai and Siegmund (1977, 1979), in the case of
our stopping variable $N_d$. In the light of Ghosh et al. (1997, Section 2.9), we may define:

\begin{equation}
(4.7) \quad n_0 = \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2, \quad W_i = \left( \frac{X_i}{Y_i} \right), \quad W_n = \left( \frac{X_n}{Y_n} \right),
\end{equation}

which gives:

\begin{align*}
E_{\theta}[W_i] &= \left( \frac{1}{\theta} \right), \quad V_{\theta}[W_i] = \begin{pmatrix} 1/\theta^2 & 0 \\ 0 & 0 \end{pmatrix} = \Sigma, \\
(4.8) h(x, y) &= \frac{1 - \exp(-a\sqrt{y/x})}{a^2y/x}, \\
(x, y) &\in R^+ leading to $h(W_n) \equiv h(X_n, Y_n)$.
\end{align*}

Next, we obtain the first- and second-order partial derivatives of $h(x, y)$:

\begin{align*}
\partial h(x, y)/\partial x &= a^{-2}y^{-1}(1 - e^{-a\sqrt{y/x}})^2 \\
&\quad - a^{-1}x^{-1/2}y^{-1/2}e^{-a\sqrt{y/x}}(1 - e^{-a\sqrt{y/x}}); \\
\partial h(x, y)/\partial y &= -a^{-2}xy^{-2}(1 - e^{-a\sqrt{y/x}})^2 \\
&\quad + a^{-1}x^{1/2}y^{-3/2}e^{-a\sqrt{y/x}}(1 - e^{-a\sqrt{y/x}}); \\
\partial^2 h(x, y)/\partial x^2 &= \frac{1}{2}a^{-1}x^{-3/2}y^{-1/2}e^{-a\sqrt{y/x}} - \frac{1}{2}x^{-2}e^{-a\sqrt{y/x}} \\
&\quad + \frac{1}{2}a^{-1}x^{-3/2}y^{-1/2}e^{-2a\sqrt{y/x}} + x^{-2}e^{-2a\sqrt{y/x}}, \\
\partial^2 h(x, y)/\partial y^2 &= 2a^{-2}xy^{-3}(1 - e^{-a\sqrt{y/x}})^2 \\
&\quad - a^{-1}x^{1/2}y^{-5/2}e^{-a\sqrt{y/x}}(1 - e^{-a\sqrt{y/x}}) \\
&\quad - \frac{3}{2}a^{-1}x^{1/2}y^{-5/2}e^{-a\sqrt{y/x}} - \frac{1}{2}y^{-2}e^{-a\sqrt{y/x}} \\
&\quad + \frac{3}{2}a^{-1}x^{1/2}y^{-5/2}e^{-2a\sqrt{y/x}} + y^{-2}e^{-2a\sqrt{y/x}}.
\end{align*}

The second-order mixed partial derivative of $h(x, y)$ is given by:

\begin{align*}
(4.10) \quad \partial^2 h(x, y)/\partial x \partial y &= -a^{-2}y^{-2}(1 - e^{-a\sqrt{y/x}})^2 \\
&\quad + a^{-1}x^{-1/2}y^{-3/2}e^{-a\sqrt{y/x}}(1 - e^{-a\sqrt{y/x}}) \\
&\quad - x^{-1}y^{-1}e^{-2a\sqrt{y/x}} + \frac{1}{2}a^{-1}x^{-1/2}y^{-3/2}e^{-a\sqrt{y/x}} \\
&\quad + \frac{1}{2}x^{-1}y^{-1}e^{-a\sqrt{y/x}} - \frac{1}{2}a^{-1}x^{-1/2}y^{-3/2}e^{-2a\sqrt{y/x}}.
\end{align*}
Now, we evaluate these derivatives from (4.9)–(4.10) at \((x = 1/\theta, y = \theta)\) and obtain the following expressions:

\[
\begin{align*}
\frac{\partial h(x, y)}{\partial x} &\bigg|_{(x=1/\theta, y=\theta)} = a^{-2}\theta^{-1}(1 - e^{-a\theta})^2 - a^{-1}e^{-a\theta}(1 - e^{-a\theta}) \\
\equiv &\ h_1, \quad \text{say}; \\
\frac{\partial h(x, y)}{\partial y} &\bigg|_{(x=1/\theta, y=\theta)} = -a^{-2}\theta^{-3}(1 - e^{-a\theta})^2 + a^{-1}\theta^{-2}e^{-a\theta}(1 - e^{-a\theta}) \\
\equiv &\ h_2, \quad \text{say}; \\
\frac{\partial^2 h(x, y)}{\partial x^2} &\bigg|_{(x=1/\theta, y=\theta)} = -\frac{1}{2}a^{-1}\theta e^{-a\theta} - \frac{1}{2}\theta^2 e^{-a\theta} + \frac{1}{2}a^{-1}\theta e^{-2a\theta} \\
+ &\ \theta^2 e^{-2a\theta}; \\
\frac{\partial^2 h(x, y)}{\partial y^2} &\bigg|_{(x=1/\theta, y=\theta)} = 2a^{-2}\theta^{-4}(1 - e^{-a\theta})^2 - \frac{5}{2}a^{-1}\theta^{-3}e^{-a\theta} \\
+ &\ \frac{5}{2}a^{-1}\theta^{-3}e^{-2a\theta} - \frac{1}{2}\theta^{-2}e^{-a\theta} + \theta^{-2}e^{-2a\theta}; \\
\frac{\partial^2 h(x, y)}{\partial x\partial y} &\bigg|_{(x=1/\theta, y=\theta)} = -a^{-2}\theta^{-2}(1 - e^{-a\theta})^2 \\
+ &\ a^{-1}\theta^{-1}e^{-a\theta}(1 - e^{-a\theta}) + \frac{1}{2}a^{-1}\theta^{-1}e^{-a\theta} \\
+ &\ \frac{1}{2}e^{-a\theta} - \frac{1}{2}a^{-1}\theta^{-1}e^{-2a\theta} - e^{-2a\theta}.
\end{align*}
\]  

(4.11)

Thus, with \(h_1, h_2\) from (4.11) and \(\Sigma\) from (4.8), we obtain the following expressions:

\[
\begin{align*}
u &\equiv h\left(\frac{1}{\theta}, \theta\right) = a^{-2}\theta^{-2}(1 - e^{-a\theta})^2; \\
\sigma^2 &\equiv (h_1, h_2)\Sigma(h_1, h_2) = 2a^{-4}\theta^{-4}(1 - e^{-a\theta})^4 - 4a^{-3}\theta^{-3}e^{-a\theta}(1 - e^{-a\theta})^3 \\
+ &\ 2a^{-2}\theta^{-2}e^{-2a\theta}(1 - e^{-a\theta})^2; \\
S_n &\equiv \sum_{i=1}^{n}(u + h_1X_i + h_2Y_i) = nu + \frac{1}{2}h_1\theta^{-1}H_n + \frac{1}{2}\theta h_2L_n \quad \text{and} \\
H_n &\equiv 2\theta\sum_{i=1}^{n}X_i, \quad L_n \equiv 2\theta^{-1}\sum_{i=1}^{n}Y_i \quad \text{are i.i.d.} \ \chi_{2n}^2; \\
\xi_n &\equiv Z_n - S_n = \frac{nh(W_n) - S_n};
\end{align*}
\]

(4.12)

We also have the following expression for the trace of a required matrix:

\[
\begin{align*}
\text{tr} \left( \begin{array}{ccc}
\sum & \partial^2 h(x, y)/\partial x^2 & \partial^2 h(x, y)/\partial x\partial y \\
\partial^2 h(x, y)/\partial y\partial x & \partial^2 h(x, y)/\partial y^2 \\
\end{array} \right) &\bigg|_{(x=1/\theta, y=\theta)} \\
= &\ -3a^{-1}\theta^{-1}e^{-a\theta} \\
- &\ e^{-a\theta} + 3a^{-1}\theta^{-1}e^{-2a\theta} + 2e^{-2a\theta} + 2a^{-2}\theta^{-2}(1 - e^{-a\theta})^2 \equiv \text{tr}_{a, \theta}, \quad \text{say}.
\end{align*}
\]

(4.13)

From (4.1), recall that \(N_d \geq n_{0d}\) w.p.1 for fixed \(\alpha\) and \(d\). We may pick \(\varepsilon = 1 - \{g(a\theta)\}^{-2}\) so that we have \(0 < \varepsilon < 1\) and then clearly \((1 - \varepsilon)n_d^* = n_0\). Thus, we have the following property holds:

\[
P_{\theta}\{N_d \leq (1 - \varepsilon)n_d^*\} = 0.
\]

(4.14)
We see immediately that our stopping rule \( N_d \) from (4.1) has the same representation required in Theorem 2.9.6 in Ghosh et al. (1997, p. 64) with the verifiable conditions (A.1)–(A.7) from Ghosh et al. (1997, p. 63) holding. See also Woodroofe (1982, pp. 47–48) and Siegmund (1985, pp. 194–195).

### 4.2. First-order asymptotics

We first set out to introduce a number of desirable interesting properties associated with our proposed bounded-length purely sequential confidence interval estimation strategy \((N_d, K_{N_d})\) defined via (4.1) and (4.5) for the parametric function \( p_\theta \) from (2.1).

**Theorem 4.1.** For the purely sequential sampling strategy \((N_d, K_{N_d})\) defined via (4.1) and (4.5) for the parametric function \( p_\theta \) from (2.1), with \( 0 < \alpha < 1 \) and \( \theta > 0 \) fixed but otherwise arbitrary, we have the following asymptotic results as \( d \to 1^+ \):

1. \( N_d/n_d^* \to 1 \text{ w.p.1} (P_\theta) \);
2. \( \mathbb{E}_{\theta}[N_d/n_d^*] \to 1 \text{ [Asymptotic First-Order Efficiency]} \);
3. \( P_\theta \{ p_\theta \in K_{N_d} \} \to 1 - \alpha \text{ [Asymptotic Consistency]} \);

where \( n_d^* \) comes from (3.5) and \( a(> 0) \) is known.

**Proof.** In what follows, we sketch an outline of the proof.

Part (i). For the purely sequential stopping time \( N_d \) defined in (4.1), it is obvious that \( N_d \to \infty \) w.p.1 \((P_\theta)\) and both \( T_{N_d}, T_{N_d-1} \) converge to \( \theta \) w.p.1 \((P_\theta)\) as \( d \to 1^+ \). Then, using (4.1), we can claim the following basic inequality w.p.1 \((P_\theta)\):

\[
\frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \{g(aT_{N_d})\}^2 \leq N_d \leq 1 + \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \{g(aT_{N_d-1})\}^2,
\]

(4.15)

since \( N_d > n_{0d} \) w.p.1 \((P_\theta)\). Then, dividing throughout (4.15) by \( n_d^* \) we can claim that \( N_d/n_d^{* - 1} \to 1 \) w.p.1 \((P_\theta)\) as \( d \to 1^+ \).

Part (ii). Given the discussions in Subsection 4.1, this claim follows from Theorem 2.9.3 in Ghosh et al. (1997).

Part (iii). By combining Anscombe’s (1952) random central limit theorem (CLT) for the MLE with Slusky’s theorem, we conclude:

\[
n_d^{* 1/2} (T_{N_d} - \theta) \overset{\mathcal{L}}{\to} N \left( 0, \frac{1}{2} \theta^2 \right) \quad \text{as} \quad d \to 1^+.
\]

(4.16)

with \( n_d^* \) is defined in (3.5).

Then, applying Mann-Wald theorem to (4.16), we can alternatively express (4.16) as:

\[
n_d^{* 1/2} (\log U_{N_d} - \log q_\theta) \overset{\mathcal{L}}{\to} N \left( 0, \frac{1}{2} a^2 \theta^2 (1 - e^{-\theta a})^{-2} \right) \quad \text{as} \quad d \to 1^+,
\]

(4.17)

where \( U_{N_d} \) is the randomly stopped version of the MLE of \( q_\theta \) obtained from \( \{(X_i, Y_i), i = 1, 2, \ldots, N_d\} \). Recall (3.3). See also Gut (2012) and Mukhopadhyay and Chattopadhyay (2012).
Now, in view of (4.17) let us note that

\[ W_{N_d} \equiv \frac{z_{\alpha/2}}{\log d} \left( \log U_{N_d} - \log q_\theta \right) \xrightarrow{p} N(0,1) \quad \text{as} \quad d \to 1. \tag{4.18} \]

Then, in view of (4.18), the following will hold:

\[ P_\theta \{ q_\theta \in J_{N_d} \} = P_\theta \{ |\log U_{N_d} - \log q_\theta| < \log d \} = P_\theta \{|W_N| < z_{\alpha/2}\}, \]

which converges to \(1 - \alpha\) as \(d \to 1^+\).

Now, the proof is complete. \(\square\)

The following result shows convergence of the negative moments of \(N_d\).

**Theorem 4.2.** For the purely sequential sampling strategy \((N_d, K_{N_d})\) defined via (4.1) and (4.5) for the parametric function \(p_\theta\) from (2.1), with \(0 < \alpha < 1\) and \(\theta > 0\) fixed but otherwise arbitrary, we have the following asymptotic results as \(d \to 1^+\):

\[ E_\theta[(n_d^*/N_d)^\omega] \to 1 \quad \text{for all fixed} \quad \omega > 0. \]

where \(n_d^*\) comes from (3.5).

**Proof.** In what follows, we sketch an outline of the proof. We can write w.p.1\((P_\theta)\):

\[ 0 < (n_d^*/N_d)^\omega \leq (n_d^*/n_0d)^\omega = \{g(a\theta)\}^{2\omega}, \]

which shows that \((n_d^*/N_d)^\omega\) remains bounded for all \(d > 1\). Hence, \((n_d^*/N_d)^\omega\) is uniformly integrable. Next, for the purely sequential stopping variable \(N_d\), we can obviously claim that \((n_d^*/N_d)^\omega \to 1\) w.p.1\((P_\theta)\) for all fixed \(\omega(>0)\) as \(d \to 1^+\). Hence, the result follows for all fixed \(\omega > 0\). \(\square\)

The following result shows moment convergence for the standardized sample means, \(\bar{X}_{N_d}\) and \(\bar{Y}_{N_d}\). This result may be of independent interest.

**Theorem 4.3.** For the purely sequential sampling strategy \((N_d, K_{N_d})\) defined via (4.1) and (4.5) for the parametric function \(p_\theta\) from (2.1), with \(0 < \alpha < 1\) and \(\theta > 0\) fixed but otherwise arbitrary, we have the uniformly integrability of \(|n_d^{*1/2}(\bar{X}_{N_d} - \theta^{-1})|^\omega\) and \(|n_d^{*1/2}(\bar{Y}_{N_d} - \theta)|^\omega\) so that we have as \(d \to 1^+\):

\[ E_\theta[|n_d^{*1/2}(\bar{X}_{N_d} - \theta^{-1})|^\omega] \quad \text{and} \quad E_\theta[|n_d^{*1/2}(\bar{Y}_{N_d} - \theta)|^\omega] \quad \text{are both} \quad O(1), \]

for all fixed \(\omega > 0\), where \(n_d^*\) comes from (3.5).

**Proof.** From Chow et al. (1979), it is known that \(|n_d^{*1/2}(\bar{X}_{N_d} - \theta^{-1})|^\omega\) is uniformly integrable for all fixed \(\omega > 0\) if \((n_d^*/N_d)^\omega\) is uniformly integrable for all fixed \(\omega > 0\). But, our Theorem 4.1 part (i) combined with Theorem 4.2 show that \((n_d^*/N_d)^\omega\) is uniformly integrable for all fixed \(\omega > 0\). Thus, \(|n_d^{*1/2}(\bar{X}_{N_d} - \theta^{-1})|^\omega\) is uniformly integrable for all fixed \(\omega > 0\).
Let $Z$ be a random variable that is distributed as $N(0, \theta^{-2})$. Now, from Anscombe’s (1952) random CLT, we know that
\[
n_d^{*1/2}(\bar{X}_d - \theta^{-1}) \overset{L}{\to} N(0, \theta^{-2}) \quad \text{as} \quad d \to 1^+.
\]
One may also refer to Gut (2012) and Mukhopadhyay and Chattopadhyay (2012). Hence, we have:
\[
\lim_{d \to 1^+} E_\theta[|n_d^{*1/2}(\bar{X}_d - \theta^{-1})|^\omega] = E_\theta[|Z|^\omega] = O(1),
\]
for all fixed $\omega > 0$. The other result follows similarly. □

4.3. Second-order asymptotics

We recall $u$, $\sigma^2$, and $S_n$ from (4.12) as well as $\text{tr}_{a,\theta}$ from (4.13) in order to define two new entities as follows:
\[
\rho \equiv \rho(\theta) = (u^2 + \sigma^2)\{2u - \sum_{n=1}^\infty n^{-1}E_\theta[S_n^-]\}^{-1};
\]
\[
\kappa \equiv \kappa(\theta) = u^{-1}\left\{\rho(\theta) - \frac{1}{2} \text{tr}_{a,\theta}\right\};
\]
(4.19)
where $S_n^- = \min\{0, S_n\}$.

Now, we are in a position to significantly strengthen our previous conclusion in part (ii), Theorem 4.1. The next result shows asymptotic second-order efficiency of our proposed purely sequential estimation methodology (4.1) in the sense of Ghosh and Mukhopadhyay (1981).

**Theorem 4.4.** For the purely sequential sampling strategy $(N_d, K_{N_d})$ defined via (4.1) and (4.5) for the parametric function $p_\theta$ from (2.1), with $0 < \alpha < 1$ and $\theta > 0$ fixed but otherwise arbitrary, we have the following asymptotic result under true $\theta$:
\[
\lim_{d \to 1^+} E_\theta[N_d - n_d^*] = \kappa(\theta),
\]
where $n_d^*$ and $\kappa(\theta)$ come from (3.5) and (4.19) respectively.

**Proof.** Given our discussions in Subsection 4.1, this result follows immediately from Theorem 2.9.6 in Ghosh et al. (1997, p. 64). Thus, with the function $h(x, y)$ defined in (4.8), we claim:
\[
E_\theta[N_d] = \{h(\theta^{-1}, \theta)\}^{-1}\left\{ n_{0d} + \rho(\theta) - \frac{1}{2} \text{tr}_{a,\theta} \right\} + o(1)
\]
\[
= n_d^* + \kappa(\theta) + o(1).
\]
This completes the proof. □
4.3.1. Numerical evaluation of $\kappa(\theta)$ Defined in (4.19)

Computing $u, \sigma^2$ from (4.12) and $\text{tr}_{a,\theta}$ from (4.13) involves straightforward calculations given values of $a$ and $\theta$. But, in order to evaluate $\kappa(\theta)$ from (4.19), we also need to evaluate $\rho(\theta)$ requiring a numerical value of $\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}]$ with $S_n$ coming from (4.12).

Since exact evaluation of $\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}]$ is complicated, we decided to utilize large-scale computer simulations to estimate the expression $\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}]$ fairly accurately under a number of fixed combination of $a$ and $\theta$ values. Now, we explain this computer algorithm.

First, we fixed $n$ and approximated $E_\theta[S_n^{-}]$ for each fixed $n = 1, \ldots, Q$, say. Having fixed $n$, we generated independent pairs of pseudorandom observations $(H_{n,r}, L_{n,r})$, $r = 1, \ldots, R$, say, on the pair of random variables $(H_n, L_n)$ defined in (4.12) where $H_n, L_n$ are i.i.d. $\chi^2_{2n}$. Thus, with $n$ fixed, we calculated $r$ pseudorandom values

$$S_{n,r} = nu + \frac{1}{2} h_1 \theta^{-1} H_{n,r} + \frac{1}{2} \theta h_2 L_{n,r}, \quad r = 1, \ldots, R,$$

on the random variable $S_n$ defined via (4.12).

Then,

$$E_\theta[S_n^{-}] \text{ was estimated by } \hat{E}_\theta[S_n^{-}] \equiv R^{-1} \Sigma_{r=1}^{R} \min\{0, S_{n,r}\},$$

with $n$ fixed, $n = 1, \ldots, Q$. Next, we estimated $\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}]$ as follows:

$$\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}] \equiv \Sigma_{n=1}^{Q} n^{-1} \hat{E}_\theta[S_n^{-}].$$

This led us to obtain:

$$\rho(\theta) \equiv (u^2 + \sigma^2) \{2u - \Sigma_{n=1}^{\infty} n^{-1} \hat{E}_\theta[S_n^{-}]\}^{-1},$$

$$\kappa(\theta) \equiv u^{-1} \left\{ \rho(\theta) - \frac{1}{2} \text{tr}_{a,\theta} \right\},$$

in the spirit of (4.19).

Table 1 shows the estimated values of $\rho$, $\kappa$, and other requisite entities corresponding to number of fixed combination of $a$ and $\theta$ when we took $Q = 5,000$ and $R = 10,000$. Some comments are in order:

- $R = 10,000$ replications are expected to be enough to estimate the true value of $E_\theta[S_n^{-}]$;
- In the case of the given combinations of $a$ and $\theta$ values, it may be fairly reasonable to say that it would be rather nearly impossible to observe negative values of $S_n$ when $n$ is especially large enough ($n > 5000$). That is, we would largely expect $S_n^{-} \equiv \min\{0, S_n\} = 0$. Hence, we pick the value $Q = 5000$ to estimate $\Sigma_{n=1}^{\infty} n^{-1} E_\theta[S_n^{-}]$;
- Estimated standard errors from $R = 10,000$ replications appeared very small ($< 0.001$) for each fixed $n(= 1, \ldots, Q)$ and for each fixed combination of $a$ and $\theta$. Thus, $\rho$ and $\kappa$ could be estimated very accurately.
Table 1. Estimating $\kappa(\theta)$ from (4.19) together with $u, \sigma^2, \Sigma_{n=1}^{\infty} n^{-1} E_{\theta}[S_n]$ from (4.12), $\text{tr}_{a,\theta}$ from (4.13), when $R = 10000, Q = 5000$, and $\rho$ comes from (4.19): $a = (1,2)$ and $\theta = (0.1,1.0,2.0,5.0)$.

|   | $\theta$ |
|---|----------|
| $a$ | terms | 0.1 | 1.0 | 2.0 | 5.0 |
| 1 | $u$ | 0.906 | 0.400 | 0.187 | 0.039 |
|   | $\sigma^2$ | 0.004 | 0.056 | 0.033 | 0.003 |
|   | $\Sigma_{n=1}^{\infty} n^{-1} E_{\theta}[S_n]$ | 0.000 | -0.010 | -0.025 | -0.014 |
|   | $\hat{\rho}$ | 0.455 | 0.267 | 0.170 | 0.049 |
|   | $\text{tr}_{a,\theta}$ | -0.039 | 0.004 | 0.100 | 0.068 |
|   | $\hat{\kappa}$ | **0.524** | **0.663** | **0.642** | **0.385** |
| 2 | $u$ | 0.821 | 0.187 | 0.060 | 0.010 |
|   | $\sigma^2$ | 0.013 | 0.033 | 0.006 | 0.000 |
|   | $\Sigma_{n=1}^{\infty} n^{-1} E_{\theta}[S_n]$ | 0.000 | -0.025 | -0.020 | -0.004 |
|   | $\hat{\rho}$ | 0.418 | 0.170 | 0.069 | 0.004 |
|   | $\text{tr}_{a,\theta}$ | -0.061 | 0.100 | 0.089 | 0.020 |
|   | $\hat{\kappa}$ | **0.546** | **0.642** | **0.408** | **-0.600** |

- In the context of our Table 1, we additionally estimated $\rho$ and $\kappa$ when

$$Q = \frac{1}{3} R = 5000 \quad \text{(a)} \quad Q = R = 10000 \quad \text{(b)} \quad Q = \frac{2}{3} R = 10000 \quad \text{(c)} \quad Q = \frac{1}{2} R = 10000. \quad \text{(d)}$$

We saw no changes up to 3 decimal places in the estimated values of $\rho$ and $\kappa$ in comparisons with those entries shown in Table 1 in the process of implementing $Q, R$ from (4.25).

4.4. Data illustrations using simulations

In this section, we summarize some interesting features obtained from analyzing simulated data for the purely sequential bounded-length confidence interval estimation methodology $(N_d, K_{N_d})$ defined via (4.1) and (4.5) for the parametric function $p_\theta$ from (2.1). Simulations were carried out under these pre-fixed values: $\theta = 1, 2, 5, a = 2$, and $\alpha = 0.10, 0.05, 0.01$.

Also, we fixed the following choices of values of $d$:

$$d = 2.00, 1.65, 1.60, 1.55, 1.50, 1.35, 1.30, 1.20, 1.10, 1.08, 1.05. \quad (4.26)$$

The features and performances highlighted here remain nearly the same for many other choices of $d$ and $(\theta, a, \alpha)$ values, and so we omit those for brevity.

Under each fixed set of values of $\theta$, $a$, $\alpha$, and $d$, we determined $n_{0d}$ from (3.8), the pilot sample size for purely sequential procedure. Also, we determined $n^*_d$ using (3.4), the optimal fixed sample size, but treated $n^*_d$ as unknown. We first generated $n_{0d}$ pseudorandom observations

$$\{(X_i, Y_i), i = 1, \ldots, n_{0d}\}$$
Table 2. Simulated performances of the purely sequential estimation strategy defined via (4.1) and (4.5) with 10,000 replications: \( \theta = 1, a = 2 \) along with \( \hat{\kappa}(\theta) = 0.642 \) defined via (4.19).

| \( \alpha \) | \( d \) | \( n_{0d} \) | \( n_d^* \) | \( \pi \) | \( s_{\pi} \) | \( \bar{\pi} - n_{d}^* \) | \( \bar{\pi}/n_{d}^* \) | Cov \( \hat{b} \) | \( s_{r} \) |
|---|---|---|---|---|---|---|---|---|---|
| 0.10 | 2.00 | 3 | 15.064 | 15.717 | 0.038 | 0.653 | 1.043 | 0.885 | 0.003 |
| 1.65 | 6 | 28.860 | 29.513 | 0.053 | 0.653 | 1.023 | 0.895 | 0.003 |
| 1.60 | 7 | 32.763 | 33.414 | 0.055 | 0.651 | 1.020 | 0.901 | 0.003 |
| 1.55 | 8 | 37.682 | 38.293 | 0.060 | 0.611 | 1.016 | 0.894 | 0.003 |
| 1.50 | 9 | 44.023 | 44.605 | 0.065 | 0.582 | 1.013 | 0.896 | 0.003 |
| 1.35 | 16 | 80.361 | 80.896 | 0.087 | 0.536 | 1.007 | 0.901 | 0.003 |
| 1.30 | 20 | 105.143 | 105.807 | 0.100 | 0.665 | 1.006 | 0.893 | 0.003 |
| 1.20 | 41 | 217.727 | 218.320 | 0.143 | 0.592 | 1.003 | 0.900 | 0.003 |
| 1.10 | 149 | 796.729 | 797.799 | 0.273 | 1.070 | 1.001 | 0.902 | 0.003 |
| 1.08 | 229 | 1221.932 | 1222.492 | 0.340 | 0.560 | 1.000 | 0.901 | 0.003 |
| 1.05 | 569 | 3040.356 | 3040.186 | 0.541 | 0.170 | 1.000 | 0.893 | 0.003 |
| 0.05 | 2.00 | 4 | 21.388 | 21.910 | 0.046 | 0.522 | 1.024 | 0.936 | 0.002 |
| 1.65 | 8 | 40.977 | 41.496 | 0.063 | 0.519 | 1.013 | 0.945 | 0.002 |
| 1.60 | 9 | 46.519 | 47.088 | 0.066 | 0.569 | 1.012 | 0.943 | 0.002 |
| 1.55 | 11 | 53.503 | 53.998 | 0.072 | 0.495 | 1.009 | 0.942 | 0.002 |
| 1.50 | 12 | 62.506 | 63.154 | 0.076 | 0.647 | 1.010 | 0.948 | 0.002 |
| 1.35 | 22 | 114.100 | 114.642 | 0.103 | 0.542 | 1.005 | 0.951 | 0.002 |
| 1.30 | 28 | 149.287 | 149.801 | 0.118 | 0.514 | 1.003 | 0.951 | 0.002 |
| 1.20 | 58 | 309.140 | 309.590 | 0.171 | 0.450 | 1.001 | 0.949 | 0.002 |
| 1.10 | 212 | 1131.233 | 1131.517 | 0.329 | 0.284 | 1.000 | 0.948 | 0.002 |
| 1.08 | 325 | 1734.957 | 1735.445 | 0.411 | 0.488 | 1.000 | 0.944 | 0.002 |
| 1.05 | 807 | 4316.842 | 4316.624 | 0.639 | 0.218 | 1.000 | 0.949 | 0.002 |
| 0.01 | 2.00 | 7 | 36.942 | 37.385 | 0.059 | 0.443 | 1.012 | 0.985 | 0.001 |
| 1.65 | 14 | 70.775 | 71.307 | 0.082 | 0.532 | 1.008 | 0.990 | 0.001 |
| 1.60 | 16 | 80.346 | 80.920 | 0.088 | 0.574 | 1.007 | 0.988 | 0.001 |
| 1.55 | 18 | 92.409 | 92.841 | 0.095 | 0.432 | 1.005 | 0.987 | 0.001 |
| 1.50 | 21 | 107.960 | 108.538 | 0.101 | 0.579 | 1.005 | 0.989 | 0.001 |
| 1.35 | 37 | 197.071 | 197.721 | 0.135 | 0.649 | 1.003 | 0.990 | 0.001 |
| 1.30 | 49 | 257.845 | 258.266 | 0.157 | 0.421 | 1.002 | 0.989 | 0.001 |
| 1.20 | 100 | 533.940 | 534.795 | 0.227 | 0.855 | 1.002 | 0.989 | 0.001 |
| 1.10 | 366 | 1953.845 | 1954.765 | 0.426 | 0.920 | 1.000 | 0.990 | 0.001 |
| 1.08 | 561 | 2996.586 | 2996.685 | 0.530 | 0.099 | 1.000 | 0.989 | 0.001 |

from the p.d.f. (1.1). Then, we generated one new pair of observations \((X, Y)\) at-a-time until termination according to the purely sequential rule (4.1).

Under each configuration, we replicated the purely sequential procedure (4.1) 10,000(= \( B \), say) times. In the \( i \)th replication, suppose that we observed terminal values \( N_d = n_i, b_i = 1 \) (or 0) if \( p_\theta \) belonged (or did not belong) to the constructed interval \( K_{n_i} \) in (4.5), \( i = 1, \ldots, B \). From such data observed across
Table 3. Simulated performances of the purely sequential estimation strategy defined via (4.1) and (4.5) with 10,000 replications: $\theta = 2$, $a = 2$ along with $\hat{\kappa}(\theta) = 0.408$ defined via (4.19).

| $\alpha$ | $d$ | $n_{0d}$ | $n^*_d$ | $\pi$ | $s_{\pi}$ | $\pi - n^*_d$ | $\pi/n^*_d$ | Cov $\hat{b}$ | $s_{\hat{b}}$ |
|----------|-----|----------|--------|------|--------|-------------|-------------|--------------|-----------|
| 0.1      | 2.00 | 3        | 46.747 | 47.020 | 0.090  | 0.273        | 1.006       | 0.891        | 0.003      |
| 1.65     | 6    | 89.560   | 89.677 | 0.125 | 0.117  | 1.001       | 0.894       | 0.003        |
| 1.60     | 7    | 101.671  | 102.175| 0.132 | 0.504  | 1.005       | 0.898       | 0.003        |
| 1.55     | 8    | 116.936  | 117.148| 0.142 | 0.212  | 1.002       | 0.897       | 0.003        |
| 1.50     | 9    | 136.614  | 136.843| 0.154 | 0.229  | 1.002       | 0.895       | 0.003        |
| 1.35     | 16   | 249.377  | 249.642| 0.207 | 0.265  | 1.001       | 0.899       | 0.003        |
| 1.30     | 20   | 326.281  | 326.535| 0.236 | 0.254  | 1.001       | 0.900       | 0.003        |
| 1.20     | 41   | 675.655  | 675.764| 0.341 | 0.109  | 1.000       | 0.897       | 0.003        |
| 1.10     | 149  | 2472.420 | 2473.567| 0.656 | 0.147  | 1.000       | 0.899       | 0.003        |
| 1.08     | 229  | 3791.918 | 3792.201| 0.800 | 0.283  | 1.000       | 0.904       | 0.003        |
| 0.05     | 2.00 | 4        | 66.373 | 66.508 | 0.107  | 0.135       | 1.002       | 0.945        | 0.002      |
| 1.65     | 8    | 127.162  | 127.492| 0.147 | 0.330  | 1.003       | 0.948       | 0.002        |
| 1.60     | 9    | 144.358  | 144.648| 0.158 | 0.291  | 1.002       | 0.946       | 0.002        |
| 1.55     | 11   | 166.031  | 166.145| 0.169 | 0.114  | 1.001       | 0.948       | 0.002        |
| 1.50     | 12   | 193.970  | 193.933| 0.184 | -0.038 | 1.000       | 0.945       | 0.002        |
| 1.35     | 22   | 354.077  | 354.517| 0.247 | 0.441  | 1.001       | 0.949       | 0.002        |
| 1.30     | 28   | 463.269  | 463.880| 0.282 | 0.611  | 1.001       | 0.949       | 0.002        |
| 1.20     | 58   | 959.326  | 960.043| 0.401 | 0.716  | 1.001       | 0.955       | 0.002        |
| 1.10     | 212  | 3510.459 | 3510.788| 0.800 | 0.329  | 1.000       | 0.946       | 0.002        |
| 1.08     | 325  | 5383.944 | 5385.243| 0.957 | 1.299  | 1.000       | 0.951       | 0.002        |
| 0.01     | 2.00 | 7        | 114.638| 114.920| 0.141  | 0.281       | 1.002       | 0.987        | 0.001      |
| 1.65     | 14   | 219.632  | 220.066| 0.195 | 0.434  | 1.002       | 0.989       | 0.001        |
| 1.60     | 16   | 249.332  | 249.670| 0.210 | 0.338  | 1.001       | 0.986       | 0.001        |
| 1.55     | 18   | 286.765  | 286.968| 0.221 | 0.202  | 1.001       | 0.989       | 0.001        |
| 1.50     | 21   | 335.022  | 335.602| 0.242 | 0.580  | 1.002       | 0.988       | 0.001        |
| 1.35     | 37   | 611.554  | 612.062| 0.326 | 0.508  | 1.001       | 0.989       | 0.001        |
| 1.30     | 49   | 800.149  | 800.932| 0.370 | 0.783  | 1.001       | 0.989       | 0.001        |
| 1.20     | 100  | 1656.931 | 1657.671| 0.535 | 0.740  | 1.000       | 0.991       | 0.001        |
| 1.10     | 366  | 6063.199 | 6063.742| 1.016 | 0.543  | 1.000       | 0.990       | 0.001        |

Using the notation from (4.27), Tables 2–4 summarize our findings. All $\bar{n}$ values shown in column 5 are nearly the same as $n^*_d$ across the board whether the sample sizes are small ($n^*_d \leq 100$), moderate (100 < $n^*_d$ < 300) or large ($n^*_d \geq 300$). This is consistent with the notion of asymptotic first-order efficiency property (Theorem 4.1, part (ii)).
Table 4. Simulated performances of the purely sequential estimation strategy defined via (4.1) and (4.5) with 10,000 replications: $\theta = 5$, $a = 2$ along with $\hat{\kappa}(\theta) = -0.600$ defined via (4.19).

| $\alpha$ | $d$ | $n_{0d}$ | $n_d^*$ | $\pi$ | $s_\pi$ | $\pi - n_d^*$ | $\pi/n_d^*$ | Cov $\hat{b}$ | $s_\Sigma$ |
|----------|-----|----------|---------|------|--------|-------------|-------------|-------------|---------|
| 0.10     | 2.00| 3        | 281.587 | 281.945 | 0.238  | 0.357       | 1.001       | 0.897       | 0.003   |
| 1.65     | 6   | 539.484  | 539.352 | 0.327 | -0.131 | 1.000       | 0.901       | 0.003       |
| 1.60     | 7   | 612.437  | 612.264 | 0.355 | -0.173 | 1.000       | 0.894       | 0.003       |
| 1.55     | 8   | 704.385  | 704.315 | 0.373 | -0.070 | 1.000       | 0.903       | 0.003       |
| 1.50     | 9   | 822.919  | 823.468 | 0.406 | 0.549  | 1.001       | 0.901       | 0.003       |
| 1.35     | 16  | 1502.169 | 1503.072| 0.550 | 0.131  | 1.000       | 0.902       | 0.003       |
| 1.30     | 20  | 1965.416 | 1964.865| 0.631 | -0.131 | 1.000       | 0.898       | 0.003       |
| 1.20     | 41  | 4069.939 | 4070.838| 0.899 | 0.899  | 1.000       | 0.897       | 0.003       |
| 0.05     | 2.00| 4        | 399.811 | 399.751 | 0.282  | -0.060      | 1.000       | 0.950       | 0.002   |
| 1.65     | 8   | 765.984  | 765.756 | 0.388 | -0.228 | 1.000       | 0.953       | 0.002       |
| 1.60     | 9   | 869.567  | 869.625 | 0.417 | 0.057  | 1.000       | 0.948       | 0.002       |
| 1.55     | 11  | 1000.120 | 1000.015| 0.448 | -0.105 | 1.000       | 0.947       | 0.002       |
| 1.50     | 12  | 1168.419 | 1168.885| 0.485 | 0.466  | 1.000       | 0.951       | 0.002       |
| 1.35     | 22  | 2132.850 | 2133.827| 0.646 | 0.977  | 1.000       | 0.948       | 0.002       |
| 1.30     | 28  | 2790.591 | 2789.685| 0.750 | -0.906 | 1.000       | 0.948       | 0.002       |
| 1.20     | 58  | 5778.692 | 5780.603| 1.096 | 1.911  | 1.000       | 0.943       | 0.002       |
| 0.01     | 2.00| 7        | 690.546 | 690.183 | 0.370  | -0.363      | 0.999       | 0.991       | 0.001   |
| 1.65     | 14  | 1322.994 | 1322.005| 0.516 | -0.989 | 0.999       | 0.989       | 0.001       |
| 1.60     | 16  | 1501.901 | 1503.627| 0.553 | 1.726  | 1.001       | 0.989       | 0.001       |
| 1.55     | 18  | 1727.388 | 1727.584| 0.593 | 0.196  | 1.000       | 0.989       | 0.001       |
| 1.50     | 21  | 2018.072 | 2018.434| 0.636 | 0.362  | 1.000       | 0.988       | 0.001       |
| 1.35     | 37  | 3683.819 | 3683.795| 0.848 | -0.024 | 1.000       | 0.990       | 0.001       |
| 1.30     | 49  | 4819.858 | 4822.132| 0.982 | 2.274  | 1.000       | 0.990       | 0.001       |

We also note that the $\hat{b}$ values (column 9) are very close to the target coverage (Cov), $1 - \alpha$. This validates the notion of asymptotic consistency property (Theorem 4.1, part (iii)). All estimated standard error values, namely $s_\pi$ and $s_b$, came out small. And for each fixed combination of $\alpha$ and $d$, all 10,000 confidence intervals had their lengths smaller than the corresponding $\frac{d-1}{d+1}$, which validates our conclusion from Theorem 2.1.

From Theorem 4.4, we know that the purely sequential sampling strategy should enjoy asymptotic second-order efficiency property (4.20) for large $n_d^*$, that is as $d \to 1 +$. We may reasonably expect that the differences between $\pi$ and $n_d^*$ should hover around $\kappa(\theta)$ defined via (4.19). From Table 1, we observe: (i) $\hat{\kappa} = 0.642$ when $\theta = 1, a = 2$; (ii) $\hat{\kappa} = 0.408$ when $\theta = 2, a = 2$; and (iii) $\hat{\kappa} = -0.600$ when $\theta = 5, a = 2$.

From column 7 in Tables 2–4, we see that the values of $\pi - n_d^*$ are very close to the corresponding $\kappa$ values in the sense that the corresponding intervals of $(\pi - n_d^*) \pm 2s_\pi$ include these $\hat{\kappa}$ values. These validates asymptotic second-order efficiency property (4.20) in practice.
5. A two-stage estimation methodology

Recall (i) the expression of $n_d^*$ from (3.4), that is the optimal fixed sample size whose magnitude remains unknown, and (ii) the fact that $n_d^* > n_{0d}$ given by (3.7) where $n_{0d}$ is positive and completely known. We had utilized these information in developing our purely sequential bounded-length confidence interval estimation strategy and its interesting properties (Section 4).

It may be more convenient, however, to implement a two-stage bounded-length confidence interval estimation strategy especially since batch sampling in two steps will enjoy significant operational convenience when compared with purely sequential sampling. Thus, we proceed to develop a two-stage estimation methodology and its associated first-order as well as second-order properties, assuming that we may be allowed to gather data in two batches given a practical scenario on hand.

We let the pilot sample size be $n_{0d}$ as in (3.8), gather initial data, and obtain the MLE for $\theta$ based on pilot data summarized as follows:

$$ \text{Pilot Size: } n_{0d} \equiv \left\lfloor \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \right\rfloor + 1; $$

(5.1)

$$ \text{Pilot Data: } \{(X_i, Y_i), i = 1, \ldots, n_{0d}\}; $$

$$ \text{MLE for } \theta : T_{n_{0d}} \equiv \hat{\theta}_{n_{0d}, \text{MLE}} = (\Sigma_{i=1}^{n_{0d}} Y_i / \Sigma_{i=1}^{n_{0d}} X_i)^{1/2}. $$

Next, we let $\langle u \rangle$ denote the largest integer $< u$ with $u > 0$. Recall the function $g(x) = x(1 - e^{-x})^{-1}, x > 0$, from (3.6). Now, based on the information laid out in (5.1), we define our two-stage stopping time:

$$ L_d \equiv L = \left\lfloor \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \left\{ g(aT_{n_{0d}}) \right\}^2 \right\rfloor + 1, $$

(5.2)

to construct a $100(1 - \alpha)$% fixed-accuracy confidence interval for $q_\theta$ along with $g(\cdot)$ function coming from (3.6).

It should be clear that the stopping variable $L_d$ estimates $n_d^*$ in two steps. In the second stage, we record additional data $\{(X_i, Y_i), i = n_{0d} + 1, \ldots, L_d\}$ in a single batch. Then, based on the combined set of data, that is $\{(X_i, Y_i), i = 1, \ldots, L_d\}$, we propose the following bounded-length confidence interval for the parametric function defined via (2.1), namely $p_\theta$:

$$ T_{L_d} \equiv \hat{\theta}_{L_d, \text{MLE}} = (\Sigma_{i=1}^{L_d} Y_i / \Sigma_{i=1}^{L_d} X_i)^{1/2} \text{ in the spirit of (4.2)}; $$

$$ U_{L_d} = e^{-aT_{L_d}}(1 - e^{-aT_{L_d}})^{-1} \text{ in the spirit of (4.3)}; $$

$$ K_{L_d} = \{ p_\theta : p_\theta \in [(d + U_{L_d})^{-1}U_{L_d}, (dU_{L_d} + 1)^{-1}dU_{L_d}] \} \text{ in the spirit of (4.5)}. $$

(5.3)

Again, the length of the proposed confidence interval $K_{L_d}$ for $p_\theta$ is bounded from above by $d^{-1} d + 1$ in view of Theorem 2.1.
5.1. Moments of the MLE and their expansions

In order to develop asymptotic first-order and second-order properties for the two-stage estimation strategy, we will require behaviors of the moments of $T_{n0d}$, the MLE of $\theta$, up to certain specific orders. These results may be of independent interest, but some of these may be well-known. We collect them here for immediate attention and completeness.

We explicitly focus on positive moments of $T_{n0d}$ since the expressions and expansions of the negative moments of $T_{n0d}$ will obviously follow from the positive moments of $T_{n0d}$ once we note that $T_{n0d} \sim \theta F_{2n0d, 2n0d}$ under $P_\theta$. Let us consider small enough $d (> 1)$ so that $n0d$ may be large enough enabling us to talk about the moments of $T_{n0d}$. We begin with the following explicit expressions:

$$E_\theta[T_{n0d} - \theta] = \theta \left[ \frac{\Gamma \left( n0d + \frac{1}{2} \right) \Gamma \left( n0d - \frac{1}{2} \right)}{\{ \Gamma(n0d) \}^2} - 1 \right] ,$$

(5.4)

$$E_\theta[(T_{n0d} - \theta)^2] = \theta^2 \left[ - \frac{n0d}{n0d - 1} - 2 \frac{\Gamma \left( n0d + \frac{1}{2} \right) \Gamma \left( n0d - \frac{1}{2} \right)}{\{ \Gamma(n0d) \}^2} + 1 \right] .$$

Now, according to Property 6.1.47 in Abramowitz and Stegun (1972, p. 257), for large $n$, we know:

$$n^b - a \frac{\Gamma(n + a)}{\Gamma(n + b)} = 1 + \frac{1}{2} (a - b)(a + b - 1)n^{-1} + O(n^{-2}) ,$$

(5.5)

where $a, b$ are any two fixed real numbers.

Thus, for large $n0d$ or small $d (> 1)$, by exploiting (5.4)–(5.5) we have the following results:

$$n0d^{-1/2} \frac{\Gamma \left( n0d + \frac{1}{2} \right)}{\Gamma(n0d)} = 1 - \frac{1}{8} n0d^{-1} + O(n0d^{-2}) , \quad \text{and}$$

(5.6)

$$n0d^{1/2} \frac{\Gamma \left( n0d - \frac{1}{2} \right)}{\Gamma(n0d)} = 1 + \frac{3}{8} n0d^{-1} + O(n0d^{-2}) .$$

Using (5.6), we immediately obtain:

$$\frac{\Gamma \left( n0d + \frac{1}{2} \right) \Gamma \left( n0d - \frac{1}{2} \right)}{\{ \Gamma(n0d) \}^2} = 1 + \frac{1}{4} n0d^{-1} + O(n0d^{-2}) .$$

(5.7)
Combining (5.4) with (5.7), we obtain the following expressions:

\[ E_\theta[T_{n_0d} - \theta] = \frac{1}{4} n_0d^{-1} \theta + O(n_0d^{-2}), \]
\[ E_\theta[(T_{n_0d} - \theta)^2] = \frac{1}{2} n_0d^{-1} \theta^2 + O(n_0d^{-2}). \]

(5.8)

For large enough \( n_0d (> k > 0) \), we can also express:

\[ E_\theta[(T_{n_0d} - \theta)^2k] = \theta^2k \frac{\Gamma(n_0d + k)\Gamma(n_0d - k)}{\{\Gamma(n_0d)\}^2} \Rightarrow |E_\theta[(1 + a(T_{n_0d} + \theta))^k]| < \infty. \]

(5.9)

Using (5.5) and (5.9), we obtain:

\[ E_\theta[|T_{n_0d} - \theta|^4] = \theta^4O(n_0d^{-2}). \]

(5.10)

5.2. First-order asymptotics

We first set out to introduce a number of desirable interesting properties associated with our proposed bounded-length two-stage confidence interval estimation strategy \((L_d, K_{L_d})\) defined via (5.2) and (5.3) for the parametric function \(p_\theta\) from (2.1).

**Theorem 5.1.** For the two-stage sampling strategy \((L_d, K_{L_d})\) defined via (5.2)–(5.3) for the parametric function \(p_\theta\) from (2.1), with \(0 < \alpha < 1\) and \(\theta > 0\) fixed but otherwise arbitrary, we have the following asymptotic results as \(d \to 1^+\):

(i) \(L_d/n_0^* \to 1\) w.p.1(\(P_\theta\));
(ii) \(E_\theta[L_d/n_0^*] \to 1\) [Asymptotic First-Order Efficiency];
(iii) \(P_\theta[p_\theta \in K_{L_d}] \to 1 - \alpha\) [Asymptotic Consistency];

where \(n_0^*\) comes from (3.5) and \(a(> 0)\) is known.

**Proof.** In what follows, we sketch an outline of the proof.

Part (i). Using (5.2), we will have the following inequality w.p.1(\(P_\theta\)):

\[ \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \{g(aT_{n_0d})\}^2 \leq L_d \leq \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \{g(aT_{n_0d})\}^2 + 1. \]

(5.12)
Now, since \( L_d \to \infty \) w.p.1\((P_\theta)\), \( T_{n_0d} \to \theta \) w.p.1\((P_\theta)\), and \( n_d^* \to \infty \) as \( d \to 1+ \), the result follows by dividing throughout (5.12) with \( n_d^* \) and then taking limits as \( d \to 1+ \).

Part (ii). In view of the upper bound from Lemma 3.1, we begin by expressing:

\[
E_\theta[\{g(aT_{n_0d})\}^4] \leq E_\theta[1 + 4aT_{n_0d} + 6a^2T_{n_0d}^2 + 4a^3T_{n_0d}^3 + a^4T_{n_0d}^4].
\]

Then, utilizing (5.9), we can show that the upper bound from (5.13) is finite and it converges to

\[
1 + 4a\theta + 6a^2\theta^2 + 4a^3\theta^3 + a^4\theta^4,
\]
as \( d \to 1+ \). Thus, clearly, \( \{g(aT_{n_0d})\}^2 \) is uniformly integrable. Now, part (ii) follows in view of part (i).

Part (iii). This result follows along the lines of Theorem 4.1, part (iii) once we realize:

\[
L_d^{1/2}(L_d - \theta) \xrightarrow{\mathcal{L}} N \left(0, \frac{1}{2}\theta^2\right) \quad \text{and} \quad n_d^{1/2}(L_d - \theta) \xrightarrow{\mathcal{L}} N \left(0, \frac{1}{2}\theta^2\right) \quad \text{as} \quad d \to 1+.
\]

Now, the proof is complete. \( \square \)

Remark 5.2. The conclusions from Theorems 4.2–4.3 continue to hold when \( N_d \) is replaced by \( L_d \). For brevity, the detailed proofs are omitted.

5.3. Second-order asymptotics

Before we get to asymptotic second-order analysis, we define two new entities:

\[
h(x) \equiv a^2x^2(1 - e^{-ax})^{-2} = \{g(ax)\}^2, \quad x > 0;
\]

\[
\phi(\theta) \equiv \frac{1}{4} \{\theta h'(\theta) + \theta^2 h''(\theta)\};
\]

where the first three derivatives of the \( h \)-function are expressed as:

\[
h'(x) = 2a^2x(1 - e^{-ax})^{-2} - 2a^3x^2e^{-ax}(1 - e^{-ax})^{-3};
\]

\[
h''(x) = 2a^2(1 - e^{-ax})^{-2} + 2a^3xe^{-ax}(ax - 4)(1 - e^{-ax})^{-3}
\]

\[
+ 6a^4x^2e^{-2ax}(1 - e^{-ax})^{-4};
\]

\[
h'''(x) = -2a^3e^{-ax}(a^2x^2 - 6ax + 6)(1 - e^{-ax})^{-3}
\]

\[
+ 18a^4xe^{-2ax}(2 - ax)(1 - e^{-ax})^{-4}
\]

\[
- 24a^5x^2e^{-3ax}(1 - e^{-ax})^{-5}.
\]

Now, we proceed to obtain an expansion for \( E_\theta[h(T_{n_0d}) - h(\theta)] \) up to a desired order when \( n_{0d} \) is large. This result would help in the sequel in securing asymptotic second-order efficiency property in the sense of Ghosh and Mukhopadhyay (1981) for the two-stage sampling strategy \((L_d, K_{L_d})\) from (5.2)–(5.3).
Theorem 5.2. With $0 < \alpha < 1$ and $\theta > 0$ fixed but otherwise arbitrary, as $d \to 1+$, we have the following expansion:

\[(5.16) \quad E_\theta[h(T_{n_0 d})] = h(\theta) + \frac{1}{4} \{\theta h'(\theta) + \theta^2 h''(\theta)\} n_{0d}^{-1} + O(n_{0d}^{-3/2}),\]

where $n_{0d}$ comes from (5.1) and $h(\cdot)$, $h'(\cdot)$ and $h''(\cdot)$ come from (5.14)–(5.15).

Proof. We express $h(x)$ from (5.14) using Taylor expansion around $x = \theta$:

\[(5.17) \quad h(x) = h(\theta) + (x - \theta) h'(\theta) + \frac{1}{2} (x - \theta)^2 h''(\theta) + \frac{1}{6} (x - \theta)^3 h'''(\xi),\]

with some appropriate $\xi$ that lies between $x$ and $\theta$. From (5.17), we obtain:

\[(5.18) \quad E_\theta[h(T_{n_0 d})] = h(\theta) + h'(\theta) E_\theta[T_{n_0 d} - \theta] + \frac{1}{2} h''(\theta) (T_{n_0 d} - \theta)^2 + \frac{1}{6} R_d,\]

where the remainder term is $R_d \equiv E_\theta[h'''(W_d)(T_{n_0 d} - \theta)^3]$, with an appropriate random variable $W_d$ lying between $T_{n_0 d}$ and $\theta$.

Now, using Jensen’s inequality, Lemma 3.1, and the expression of $h'''(x)$ from (5.15), we obtain:

\[(5.19) \quad |R_d| \leq E_\theta[|h'''(W_d)||T_{n_0 d} - \theta|^3]\]

\[\leq 2 E_\theta \left[ |T_{n_0 d} - \theta|^3 \left( 1 + a W_d \right)^3 \left( a^2 W_d^2 + 6 a W_d + 6 \right) \right] \]

\[+ 18 E_\theta \left[ |T_{n_0 d} - \theta|^3 \left( 1 + a W_d \right)^4 \left( 2 + a W_d \right) \right] \]

\[+ 24 E_\theta \left[ |T_{n_0 d} - \theta|^3 \left( 1 + a W_d \right)^5 \right] \]

\[= 2 E_\theta[A_{1,d}] + 18 E_\theta[A_{2,d}] + 24 E_\theta[A_{3,d}], \quad \text{say.}\]

Next, we begin to handle each term $E_\theta[A_{i,d}]$ from (5.19) by splitting it as follows:

\[(5.20) \quad E_\theta[A_{i,d}] = E_\theta \left[ A_{i,d} I \left( W_d > \frac{1}{2} \theta \right) \right] + E_\theta \left[ A_{i,d} I \left( W_d \leq \frac{1}{2} \theta \right) \right] \]

\[= E_\theta[A_{i1,d}] + E_\theta[A_{i2,d}], \quad \text{say,}\]

for $i = 1, 2, 3$. At this point, in view of (5.20), our goal is to show:

\[(5.21) \quad E_\theta[A_{i1,d}] = O(n_{0d}^{-3/2}) \quad \text{and} \quad E_\theta[A_{i2,d}] = O(n_{0d}^{-3/2}),\]

for all $\theta > 0$, $i = 1, 2, 3$.

Let us fix $i = 1$ and let $c \equiv c(a, \theta)$ stand for a generic positive constant. We note that $W_d$ lies between $T_{n_0 d}$ and $\theta$ and on the set $[W_d > \frac{1}{2} \theta]$, we certainly have $\frac{1}{2} \theta < W_d < T_{n_0 d} + \theta$. Thus, using triangular inequality, we look at an upper bound
for $|E_\theta[A_{11,d}]|$ which would consist of a sum of terms such as $cE_\theta[[T_{n_{od}} - \theta]^kT_{n_{od}}^k]$ with $k = 0, \pm 1, \pm 2, \pm 3$. Applying Holder's inequality, (5.9), and (5.11), we can claim that $|E_\theta[A_{11,d}]| = O(n_{od}^{-3/2})$. Similarly, we can argue that $|E_\theta[A_{12,d}]| = O(n_{od}^{-3/2})$.

With analogous analyses, we can verify (5.21) when $i = 2, 3$. We combine these findings with (5.8), (5.18)–(5.20) to complete the proof. □

**Theorem 5.3.** For the two-stage sampling strategy $(L_d, K_{L_d})$ defined via (5.2)–(5.3) for the parametric function $p_\theta$ from (2.1), with $0 < \alpha < 1$ and $\theta > 0$ fixed but otherwise arbitrary, we have the following asymptotic second-order result:

$$
(5.22) \quad \phi \leq \lim_{d \to 1^+} E_\theta[L_d - n_d^*] \leq \phi + 1,
$$

where $n_d^*$ and $\phi \equiv \phi(\theta)$ come from (3.5) and (5.14) respectively as well as $h(\cdot)$, $h'(\cdot)$ and $h''(\cdot)$ come from (5.14)–(5.15).

**Proof.** Let us denote $H_{\alpha,d} \equiv \frac{1}{2}(\frac{2d}{\log d})^2$ and $L_d^* \equiv H_{\alpha,d}\{g(\alpha T_{n_{od}})\}^2 = H_{\alpha,d}h(T_{n_{od}})$ with $d > 1$. We emphasize that w.p.1($P_\theta$), $L_d^*$ is a positive random variable, but it is not a positive integer valued random variable. We recall that $n_{od} \equiv \langle H_{\alpha,d} \rangle + 1$ where $\langle u \rangle$ is the largest integer $< u$, $u > 0$, and $n_d^* = H_{\alpha,d}h(\theta)$.

From Theorem 5.2, we immediately conclude:

$$
(5.23) \quad E_\theta[L_d^* - n_d^*] \equiv E_\theta[H_{\alpha,d}\{h(T_{n_{od}}) - h(\theta)\}] = H_{\alpha,d}\{\phi(\theta)n_{od}^{-1} + O(n_{od}^{-3/2})\}
$$
$$
= \phi(\theta)\{1 + o(1)\} + H_{\alpha,d}O(n_{od}^{-3/2}) = \phi(\theta) + o(1),
$$

since $H_{\alpha,d}/n_{od} = 1 + o(1)$ as $d \to 1+$.

Obviously, w.p.1($P_\theta$), we have:

$$
(5.24) \quad L_d^* \leq L_d \leq L_d^* + 1.
$$

Now, combining (5.23)–(5.24), the theorem follows. □

**5.3.1. Some heuristics to estimate $E_\theta[L_d - L_d^*]$ tightly**

Let us record our heuristic thoughts step-by-step in support of our understanding of possible validity of this conjecture.

**Step 1.** Observe that $h(x)$ is a strictly increasing function in $x > 0$. Thus, the distribution function $P_\theta\{h(T_{n_{od}}) \leq x\}$ for $h(T_{n_{od}})$ can be equivalently expressed as $P_\theta\{T_{n_{od}} \leq x^*\}$ with $x^* = h^{-1}(x)$ for all $x > 0$.

**Step 2.** Conditionally, given $X_1, \ldots, X_{n_{od}}$ which are independent of $Y_1, \ldots, Y_{n_{od}}$, we are able to express $P_\theta\{T_{n_{od}} \leq x^*\}$ in the form of a distribution function of a chi-square random variable with $2n_{od}$ degrees of freedom.

**Step 3.** Then appropriately modified versions of Lemma 2 and Lemma 3 from Aoshima and Yata (2010) would be expected to hold to claim $E_\theta[T_{n_{od}} -$
\( \langle T_{n_0d} \rangle \approx \frac{1}{2} + o(1) \). One could look over other related publications of Aoshima and his collaborators.

**Step 4.** Next, an appropriately improvised version of Lemma 4 from Aoshima and Yata (2010) would be expected to hold thereby leading us to our heuristic claim that \( L_d - L^*_d \) is asymptotically distributed as Uniform(0, 1). This, upon verification, would lead to (5.25).

Our heuristic Steps 1–4 would lead us to propose the following approximation:

\[
E_{\theta} [L_d - L^*_d] \approx \frac{1}{2},
\]

as a suggested guideline for use in practice.

### 5.3.2. Empirical validation of (5.26) via computer simulations

We may look into some empirical characteristics of a random variable defined as follows:

\[
U_d = \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \left\{ g(aT_{n_0d}) \right\}^2 - \left\{ \frac{1}{2} \left( \frac{z_{\alpha/2}}{\log d} \right)^2 \left\{ g(aT_{n_0d}) \right\}^2 \right\},
\]

with the help of computer simulations. In order to facilitate that part of our investigation, we first fixed a set of values \( \alpha, d, \theta, a \) which led to the associated well-defined pilot size, \( n_{0d} \). Then, for such a fixed set of values \( \alpha, d, \theta, a \), we went through the following steps:

**Step 1.** We draw \( n_{0d} \) pseudo random observations on \((X, Y)\) following the distribution (1.1). These would lead to one observed value \( u_{d1} \) of \( U_d \) defined via (5.26). This is only the first iteration out of 100,000 (= \( R \), say) iterations. After \( R \) iterations under a fixed set of values \( \alpha, d, \theta, a \), we would have recorded \( R \) independently observed values \( u_{d1}, \ldots, u_{dR} \) of \( U_d \).

**Step 2.** In Tables 5–7, we summarize a number of customary descriptive statistics: mean \( \bar{u} \) (column 4), estimated standard error \( s_{\bar{u}} \) (column 5), median \( u_{med} \) (column 6), lower quartile \( Q_L \) (column 7), upper quartile \( Q_U \) (column 8). We saw the minimum \( u_{min} \) and maximum \( u_{max} \) coincide with zero and one respectively throughout this exercise. Thus, in Tables 5–7, we do not show the \( u_{min}, u_{max} \) values.

**Step 3.** Additionally, we ran tests to decide whether or not \( u_{d1}, \ldots, u_{dR} \) could be reasonably assumed to arrive from a Uniform(0, 1) universe. We performed both chi-square goodness-of-fit test as well as the Kolmogorov-Smirnov (KS) test. The chi-square goodness-of-fit test showed P-value = 1.0 throughout this exercise and thus we keep these out from Tables 5–7. However, these tables show the P-values (column 9) associated with the KS test.

**Step 4.** We also looked at the histograms and boxplots illustrated in Fig. 1 obtained from our observed dataset \( u_{d1}, \ldots, u_{dR} \). In a very small number of situations, our reported P-values have fallen under 0.05. However, considering all our histograms and boxplots obtained in the contexts of Tables 5–7, we saw no appreciable departure overall from a Uniform(0, 1) distribution’s fit.
Step 5. Then, we successively fixed other sets of values $\alpha, d, \theta, a$, and ran through Steps 1–4.

Our summary findings are highlighted in Tables 5–7 corresponding to several sets of values $\alpha, d, \theta, a$ that are consistent with such choices highlighted in Tables 2–4. We also include a selected set of histograms obtained from $u_{d1}, \ldots, u_{dR}$ with the Uniform$(0, 1)$ p.d.f. with side-by-side boxplots as illustrations in Fig. 1.

Given that we considered a wide variety of choices of $\alpha, d, \theta, a$ (and hence, $n_{0d}$), these numerical analyses seem to validate reasonably well our strong sen-
Table 6. Summary statistics from 100,000 pseudorandom observations on $U_d$ in (5.26) under the same configuration as in Table 3. Column 9 shows P-value associated with the KS test: $\theta = 2$, $\alpha = 2$.

| $\alpha$ | $d$ | $n_{d_{\alpha}}$ | $n_{\alpha}$ | $s_{\alpha}$ | $u_{med}$ | $Q_L$ | $Q_U$ | P-values |
|----------|-----|------------------|--------------|-------------|-----------|-------|-------|----------|
| 0.10     | 2.00| 3                | 0.499        | 0.001       | 0.499     | 0.248 | 0.749 | 0.429    |
|          | 1.65| 6                | 0.501        | 0.001       | 0.5        | 0.249 | 0.752 | 0.302    |
|          | 1.60| 7                | 0.501        | 0.001       | 0.501      | 0.249 | 0.751 | 0.611    |
|          | 1.55| 8                | 0.499        | 0.001       | 0.498      | 0.246 | 0.751 | 0.061    |
|          | 1.50| 9                | 0.498        | 0.001       | 0.498      | 0.247 | 0.750 | 0.051    |
|          | 1.35| 16               | 0.5          | 0.001       | 0.5        | 0.250 | 0.750 | 0.681    |
|          | 1.30| 20               | 0.499        | 0.001       | 0.498      | 0.248 | 0.749 | 0.294    |
|          | 1.20| 41               | 0.499        | 0.001       | 0.496      | 0.250 | 0.749 | 0.055    |
|          | 1.10| 149              | 0.500        | 0.001       | 0.500      | 0.249 | 0.749 | 0.959    |
|          | 1.08| 229              | 0.500        | 0.001       | 0.499      | 0.249 | 0.751 | 0.860    |
| 0.05     | 2.00| 4                | 0.499        | 0.001       | 0.498      | 0.247 | 0.749 | 0.098    |
|          | 1.65| 8                | 0.499        | 0.001       | 0.499      | 0.251 | 0.749 | 0.608    |
|          | 1.60| 9                | 0.499        | 0.001       | 0.498      | 0.248 | 0.749 | 0.438    |
|          | 1.55| 11               | 0.499        | 0.001       | 0.499      | 0.248 | 0.749 | 0.620    |
|          | 1.50| 12               | 0.500        | 0.001       | 0.501      | 0.251 | 0.748 | 0.659    |
|          | 1.35| 22               | 0.500        | 0.001       | 0.500      | 0.250 | 0.750 | 0.895    |
|          | 1.30| 28               | 0.501        | 0.001       | 0.504      | 0.250 | 0.751 | 0.090    |
|          | 1.20| 58               | 0.499        | 0.001       | 0.499      | 0.248 | 0.750 | 0.413    |
|          | 1.10| 212              | 0.501        | 0.001       | 0.501      | 0.251 | 0.750 | 0.852    |
|          | 1.08| 325              | 0.501        | 0.001       | 0.500      | 0.252 | 0.750 | 0.513    |
| 0.01     | 2.00| 7                | 0.500        | 0.001       | 0.499      | 0.249 | 0.750 | 0.333    |
|          | 1.65| 14               | 0.500        | 0.001       | 0.500      | 0.250 | 0.749 | 0.889    |
|          | 1.60| 16               | 0.500        | 0.001       | 0.500      | 0.249 | 0.750 | 0.996    |
|          | 1.55| 18               | 0.499        | 0.001       | 0.498      | 0.249 | 0.748 | 0.479    |
|          | 1.50| 21               | 0.499        | 0.001       | 0.499      | 0.248 | 0.749 | 0.188    |
|          | 1.35| 37               | 0.499        | 0.001       | 0.499      | 0.247 | 0.750 | 0.192    |
|          | 1.30| 49               | 0.501        | 0.001       | 0.500      | 0.252 | 0.751 | 0.447    |
|          | 1.20| 100              | 0.499        | 0.001       | 0.499      | 0.251 | 0.749 | 0.473    |
|          | 1.10| 366              | 0.500        | 0.001       | 0.500      | 0.250 | 0.751 | 0.834    |

The sentiment that the asymptotic distribution of $U_d$ from (5.27) is empirically approximated rather accurately by the Uniform$(0,1)$ distribution. In other words, we put forward the approximation suggested via (5.26) as a practical guideline with reasonable confidence.

On a related note, in the light of our extensive sets of discussions combined from Subsections 5.3.1–5.3.2, for all practical purposes, we modify the conclusion from Theorem 5.3 to propose the following reasonable approximation:

\[
E_{\theta}[L_d - n_{d_s}^*] \approx \phi + \frac{1}{2} + o(1).
\]

Instead of the obvious bounds seen in (5.24), now we feel confident enough to
Table 7. Summary statistics from 100,000 pseudorandom observations on $U_d$ in (5.26) under same configuration as in Table 4. Column 9 shows P-value associated with the KS test: $\theta = 5$, $a = 2$.

| $\alpha$ | $d$ | $n_{0d}$ | $\pi$ | $s_\pi$ | $u_{med}$ | $Q_L$ | $Q_U$ | P-values |
|----------|-----|---------|-------|--------|-----------|-------|-------|---------|
| 0.10     | 2.00| 3       | 0.498 | 0.001  | 0.496     | 0.248 | 0.749 | 0.008   |
| 1.65     | 6   | 5.000   | 0.001 | 0.500  | 0.250     | 0.751 | 0.969 |
| 1.60     | 7   | 0.499   | 0.001 | 0.500  | 0.249     | 0.750 | 0.563 |
| 1.55     | 8   | 0.502   | 0.001 | 0.502  | 0.253     | 0.753 | 0.020 |
| 1.50     | 9   | 0.500   | 0.001 | 0.501  | 0.250     | 0.750 | 0.830 |
| 1.35     | 16  | 0.499   | 0.001 | 0.499  | 0.248     | 0.749 | 0.507 |
| 1.30     | 20  | 0.499   | 0.001 | 0.500  | 0.248     | 0.747 | 0.197 |
| 1.20     | 41  | 0.500   | 0.001 | 0.500  | 0.249     | 0.750 | 0.879 |
| 0.05     | 2.00| 4       | 0.499 | 0.001  | 0.499     | 0.249 | 0.748 | 0.589   |
| 1.65     | 8   | 0.499   | 0.001 | 0.498  | 0.248     | 0.750 | 0.315 |
| 1.60     | 9   | 0.502   | 0.001 | 0.503  | 0.251     | 0.753 | 0.022 |
| 1.55     | 11  | 0.500   | 0.001 | 0.500  | 0.249     | 0.749 | 0.757 |
| 1.50     | 12  | 0.500   | 0.001 | 0.502  | 0.251     | 0.750 | 0.741 |
| 1.35     | 22  | 0.500   | 0.001 | 0.501  | 0.249     | 0.750 | 0.839 |
| 1.30     | 28  | 0.500   | 0.001 | 0.500  | 0.250     | 0.750 | 0.997 |
| 1.20     | 58  | 0.500   | 0.001 | 0.499  | 0.249     | 0.750 | 0.899 |
| 0.01     | 2.00| 7       | 0.501 | 0.001  | 0.499     | 0.252 | 0.750 | 0.219   |
| 1.65     | 14  | 0.501   | 0.001 | 0.501  | 0.252     | 0.751 | 0.802 |
| 1.60     | 16  | 0.499   | 0.001 | 0.499  | 0.249     | 0.748 | 0.625 |
| 1.55     | 18  | 0.500   | 0.001 | 0.501  | 0.248     | 0.750 | 0.387 |
| 1.50     | 21  | 0.498   | 0.001 | 0.499  | 0.247     | 0.749 | 0.012 |
| 1.35     | 37  | 0.499   | 0.001 | 0.498  | 0.249     | 0.751 | 0.586 |
| 1.30     | 49  | 0.500   | 0.001 | 0.500  | 0.251     | 0.751 | 0.850 |

heuristically proceed and conjecture:

(5.28) \[ E_\theta[L_d - n^*_d] = \phi + \frac{1}{2} + o(1). \]

5.4. Data illustrations using simulations

In the spirit of Subsection 4.4, we summarize some interesting features obtained from analyzing simulated data for the two-stage bounded-length confidence interval estimation methodology ($L_d, K_{L_d}$) defined via (5.2)–(5.3) for the parametric function $p_\theta$ from (2.1). Simulations were analogously carried out under these pre-fixed values: $\theta = 1, 2, 5$, $a = 2$, $\alpha = 0.10, 0.05, 0.01$, and $d = 2.00, 1.65, 1.60, 1.55, 1.50, 1.35, 1.30, 1.20, 1.10, 1.08, 1.05$.

The features and performances highlighted here remain nearly the same for many other choices of $d$ and $(\theta, a, \alpha)$ values, and so we omit those for brevity.

Under each fixed set of values of $\theta, a, \alpha$, and $d$, we determined $n_{0d}$ from (3.8), the pilot sample size. Also, we determined $n^*_d$ using (3.4), the optimal fixed
sample size, but treated $n_d^*$ as unknown. We first generated $n_{0d}$ pseudorandom observations

$$\{(X_i, Y_i), i = 1, 2, \ldots, n_{0d}\}$$

from the p.d.f. (1.1). Then, we generated $L_d - n_{0d}$ new pair of observations $(X, Y)$ in a single batch at the second stage.

Under each configuration, we replicated the two-stage procedure (5.2)–(5.3) 10,000 (= $B$, say) times. In the $i$th replication, suppose that we observed terminal values $L_d = l_i$, $b_i = 1$ (or 0) if $p_y$ belonged (or did not belong) to the constructed interval $K_{l_i}$ in (4.5), $i = 1, \ldots, B$. From such data observed across $B$ replications,
Table 8. Simulated performances of the two-stage estimation strategy defined via (5.2) and (5.3) with 10,000 replications: $\theta = 1$, $a = 2$ along with $\phi(\theta) = 3.624$ defined via (5.14).

| $\alpha$ | $d$ | $n_{0d}$ | $n_d^*$ | $\bar{l}$ | $s_\bar{l}$ | $\bar{l} - n_d^*$ | $\bar{l}/n_d^*$ | Cov $\bar{n}$ | $s_\bar{n}$ |
|----------|-----|-----------|---------|-----------|-------------|-----------------|----------------|-------------|-----------|
| 0.10     | 2.00| 3         | 15.064  | 21.172    | 0.229       | 6.108           | 1.405          | 0.884       | 0.003     |
| 1.65     | 6   | 28.860    | 33.238  | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.60     | 7   | 32.763    | 37.175  | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.55     | 8   | 37.682    | 41.873  | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.50     | 9   | 44.023    | 47.963  | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.35     | 16  | 80.361    | 84.592  | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.30     | 20  | 105.143   | 109.408 | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.20     | 41  | 217.727   | 221.723 | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |
| 1.10     | 149 | 796.729   | 801.284 | 4.126     | 4.378       | 1.152           | 0.900          | 0.003       |


we determined the following entities:

\[
\bar{l} = B^{-1}\Sigma_{i=1}^{B}l_i : \text{should estimate } n_d^* \text{ or } E[L_d];
\]

\[
s_\bar{l} = \{(B^2 - B)^{-1}\Sigma_{i=1}^{B}(l_i - \bar{l})^2\}^{1/2} : \text{estimated standard error of } \bar{l};
\]

\[
\bar{b} = B^{-1}\Sigma_{i=1}^{B}b_i : \text{should estimate } P_{\theta}(\theta \in K_{L_d});
\]

\[
s_\bar{b} = \{B^{-1}\bar{b}(1 - \bar{b})\}^{1/2} : \text{estimated standard error of } \bar{b};
\]

in the spirit of (4.26).

Using the notation explained in (5.29), Tables 8–10 summarize our findings.
Table 9. Simulated performances of the two-stage estimation strategy defined via (5.2) and (5.3) with 10,000 replications: $\theta = 2$, $a = 2$ along with $\phi(\theta) = 16.122$ defined via (5.14).

| $\alpha$ | $d$ | $n_{0d}$ | $n_{d}^*$ | $\bar{l}$ | $s_\pi$ | $\bar{l} - n_{d}^*$ | $\frac{\pi}{n_{d}^*}$ | Cov $\bar{b}$ | $s_\pi$ |
|---------|-----|---------|-----------|---------|--------|----------------|----------------|-------------|--------|
| 0.10    | 2.00| 3       | 46.747    | 69.350  | 0.363  | 22.603         | 1.484          | 0.869       | 0.002  |
|         | 1.65| 6       | 89.560    | 107.791 | 0.309  | 18.231         | 1.204          | 0.880       | 0.001  |
|         | 1.60| 7       | 101.671   | 118.883 | 0.306  | 17.212         | 1.169          | 0.886       | 0.001  |
|         | 1.55| 8       | 116.936   | 133.357 | 0.315  | 16.421         | 1.140          | 0.885       | 0.001  |
|         | 1.50| 9       | 136.614   | 153.807 | 0.335  | 17.193         | 1.126          | 0.887       | 0.001  |
|         | 1.35| 16      | 249.377   | 265.963 | 0.415  | 16.586         | 1.067          | 0.892       | 0.001  |
|         | 1.30| 20      | 326.281   | 343.933 | 0.469  | 16.868         | 1.025          | 0.894       | 0.001  |
|         | 1.20| 41      | 675.655   | 692.523 | 0.649  | 16.373         | 1.007          | 0.899       | 0.001  |
|         | 1.10| 149     | 2472.420  | 2488.793| 1.194  | 16.373         | 1.004          | 0.898       | 0.001  |
|         | 1.08| 229     | 3791.918  | 3808.575| 1.474  | 16.657         | 1.004          | 0.898       | 0.001  |

| 0.05    | 2.00| 4       | 66.373    | 88.046  | 0.353  | 21.673         | 1.327          | 0.922       | 0.001  |
|         | 1.65| 8       | 127.162   | 145.463 | 0.342  | 18.301         | 1.144          | 0.934       | 0.001  |
|         | 1.60| 9       | 144.358   | 162.460 | 0.357  | 18.102         | 1.125          | 0.935       | 0.001  |
|         | 1.55| 11      | 166.031   | 182.527 | 0.351  | 16.496         | 1.099          | 0.937       | 0.001  |
|         | 1.50| 12      | 193.970   | 211.582 | 0.387  | 17.612         | 1.091          | 0.940       | 0.001  |
|         | 1.35| 22      | 354.077   | 370.903 | 0.483  | 16.827         | 1.048          | 0.943       | 0.001  |
|         | 1.30| 28      | 463.269   | 479.813 | 0.550  | 16.545         | 1.036          | 0.945       | 0.001  |
|         | 1.20| 58      | 959.326   | 977.251 | 0.763  | 17.924         | 1.019          | 0.947       | 0.001  |
|         | 1.10| 212     | 3510.459  | 3526.178| 1.423  | 15.719         | 1.004          | 0.948       | 0.001  |
|         | 1.08| 325     | 5383.944  | 5401.646| 1.758  | 17.702         | 1.003          | 0.949       | 0.001  |

| 0.01    | 2.00| 7       | 114.638   | 133.621 | 0.342  | 18.982         | 1.166          | 0.978       | 0.001  |
|         | 1.65| 14      | 219.632   | 236.506 | 0.395  | 16.875         | 1.077          | 0.983       | 0.001  |
|         | 1.60| 16      | 249.332   | 265.442 | 0.411  | 16.110         | 1.065          | 0.985       | 0.001  |
|         | 1.55| 18      | 286.765   | 304.224 | 0.442  | 17.458         | 1.061          | 0.985       | 0.001  |
|         | 1.50| 21      | 335.022   | 352.218 | 0.466  | 17.196         | 1.051          | 0.985       | 0.001  |
|         | 1.35| 37      | 611.554   | 628.981 | 0.619  | 17.426         | 1.028          | 0.988       | 0.000  |
|         | 1.30| 49      | 800.149   | 817.234 | 0.696  | 17.085         | 1.021          | 0.988       | 0.000  |
|         | 1.20| 100     | 1656.931  | 1675.262| 0.988  | 18.331         | 1.011          | 0.989       | 0.000  |
|         | 1.10| 366     | 6063.199  | 6081.013| 1.870  | 17.814         | 1.003          | 0.990       | 0.000  |

All $\bar{l}$ values shown in column 5 overestimate $n_{d}^*$ across the board whether the sample sizes are small ($n_{d}^* \leq 100$), moderate ($100 < n_{d}^* < 300$) or large ($n_{d}^* \geq 300$), but the extent of overestimation goes down fast as $n_{d}^*$ grows. This is consistent with the notion of asymptotic first-order efficiency property seen from Theorem 5.1, part (ii).

We also note that the $\bar{b}$ values (column 9) are very close to the target coverage (Cov), $1 - \alpha$. This validates the notion of asymptotic consistency property (Theorem 5.1, part (iii)). All estimated standard error values, namely $s_\pi$ and $s_b$, came out small. And for each fixed combination of $\alpha$ and $d$, all 10,000 confidence intervals had their lengths smaller than the corresponding $\frac{d-1}{d+1}$, which validates our conclusion from Theorem 2.1.

From Theorem 5.3, we know that the two-stage sampling strategy should
Table 10. Simulated performances of the two-stage estimation strategy defined via (5.2) and (5.3) with 10,000 replications: $\theta = 5$, $a = 2$ along with $\phi(\theta) = 100.123$ defined via (5.14).

| $\alpha$ | $d$ | $n_{ad}$ | $n_d^*$ | $l$ | $s_l$ | $l - n_d^*$ | $\bar{n}/n_d^*$ | $\text{Cov} \, b$ | $s_\bar{b}$ |
|----------|-----|----------|---------|-----|------|-----------|----------------|------------|----------|
| 0.10     | 2.00| 3        | 281.587 | 429.335 | 5.468 | 147.748 | 1.525 | 0.855 | 0.004 |
| 1.65     | 6   | 539.484  | 647.041 | 4.359 | 107.558 | 1.199 | 0.875 | 0.003 |
| 1.60     | 7   | 612.437  | 716.128 | 4.540 | 103.691 | 1.134 | 0.878 | 0.003 |
| 1.55     | 8   | 704.385  | 799.077 | 4.461 | 94.691  | 1.134 | 0.878 | 0.003 |
| 1.50     | 9   | 822.919  | 926.902 | 4.730 | 103.983 | 1.126 | 0.885 | 0.003 |
| 1.35     | 16  | 1502.169 | 1604.316 | 5.985 | 102.147 | 1.068 | 0.889 | 0.003 |
| 1.30     | 20  | 1965.416 | 2060.453 | 6.808 | 95.037  | 1.048 | 0.895 | 0.003 |
| 1.20     | 41  | 4069.939 | 4173.376 | 9.381 | 103.437 | 1.025 | 0.898 | 0.003 |
| 0.05     | 2.00| 4        | 399.811 | 527.953 | 4.732 | 128.142 | 1.321 | 0.910 | 0.003 |
| 1.65     | 8   | 765.984  | 873.174 | 4.877 | 107.190 | 1.140 | 0.933 | 0.003 |
| 1.60     | 9   | 869.567  | 983.913 | 5.146 | 114.345 | 1.131 | 0.933 | 0.003 |
| 1.55     | 11  | 1000.120 | 1102.890 | 5.200 | 102.770 | 1.103 | 0.933 | 0.003 |
| 1.50     | 12  | 1168.419 | 1268.631 | 5.473 | 100.212 | 1.086 | 0.934 | 0.002 |
| 1.35     | 22  | 2132.850 | 2246.196 | 7.046 | 113.346 | 1.053 | 0.944 | 0.002 |
| 1.30     | 28  | 2790.591 | 2891.366 | 7.948 | 100.775 | 1.036 | 0.943 | 0.002 |
| 1.20     | 58  | 5778.692 | 5888.524 | 11.091 | 109.832 | 1.019 | 0.949 | 0.002 |
| 0.01     | 2.00| 7        | 690.546 | 811.095 | 4.998 | 120.549 | 1.175 | 0.973 | 0.002 |
| 1.65     | 14  | 1322.994 | 1425.289 | 5.682 | 102.295 | 1.077 | 0.983 | 0.001 |
| 1.60     | 16  | 1501.901 | 1602.945 | 5.953 | 101.044 | 1.067 | 0.985 | 0.001 |
| 1.55     | 18  | 1727.388 | 1842.444 | 6.447 | 115.056 | 1.067 | 0.985 | 0.001 |
| 1.50     | 21  | 2018.072 | 2117.868 | 6.763 | 99.796  | 1.049 | 0.984 | 0.001 |
| 1.35     | 37  | 3683.819 | 3787.650 | 9.006 | 103.831 | 1.028 | 0.987 | 0.001 |
| 1.30     | 49  | 4819.858 | 4916.647 | 10.255 | 96.789  | 1.020 | 0.989 | 0.001 |

enjoy asymptotic second-order efficiency property (5.22) for large $n_d^*$, that is as $d \to 1+$. We may reasonably expect that the differences between $l$ and $n_d^*$ should approximately hover around $\phi(\theta)$ or $\phi(\theta) + \frac{1}{2}$ defined via (5.22) or (5.28) respectively. We may record: (i) $\phi(\theta) = 3.624$ when $\theta = 1$, $a = 2$; (ii) $\phi(\theta) = 16.122$ when $\theta = 2$, $a = 2$; and (iii) $\phi(\theta) = 100.123$ when $\theta = 5$, $a = 2$.

From column 7 in Tables 8–10, we see that the values of $l - n_d^*$ appear reasonably close to corresponding $\phi + \frac{1}{2}$ values in the sense that the corresponding intervals of $(l - n_d^*) \pm 2s_l$ include such values. These empirically validate asymptotic second-order efficiency property seen from (5.22) or (5.28) in practice.

We have made available our own developed R codes in a “dropbox”. These facilitated the data analyses presented in Subsection 4.3.1, Subsection 4.4, Subsection 5.3.2, and Subsection 5.4. We have provided an unrestricted link to access such codes in the list of references. One may additionally refer to R Core Team (2014).

6. Some concluding thoughts

In Section 3, we noted a straightforward and naturally arriving positive and known lower bound $n_{0d}$ in (3.7) for the expression of $n_d^*$ in (3.4) which led to the
specific choice of a pilot sample size $m$ from (3.8) used both in Sections 4 and 5. Such a choice as our pilot size led to asymptotic second-order considerations under the two-stage estimation strategy (5.1)–(5.2).

In a different vein, Mukhopadhyay and Duggan (1997) developed a remarkable two-stage fixed-width confidence interval methodology for the normal mean when it was assumed that the unknown population variance had a known positive lower bound. It was truly remarkable because Mukhopadhyay and Duggan (1997) could develop asymptotic second-order properties for an appropriately modified two-stage estimation methodology. It was a clear vindication of Stein’s (1945, 1949) original two-stage fixed-width confidence interval methodology.

The proliferation of the core ideas from Mukhopadhyay and Duggan (1997) in many directions has been rather widespread and that continues to grow in areas including big data problems as well as small $n$ large $p$ problems. For brevity, we only mention some of the important references in order to connect the dots: Mukhopadhyay and Aoshima (1998), Aoshima and Mukhopadhyay (1998, 1999), Mukhopadhyay (1999a, b), Mukhopadhyay and Duggan (2000, 2001), Aoshima and Takada (2000), and Aoshima and Yata (2010).

We recall that our purely sequential estimation strategy (4.1) and (4.5) as well as our two-stage estimation strategy (5.2)–(5.3) used the same pilot size $n_{0d}$ from (3.8). But, when we compare the columns corresponding to the values of $\pi - n^*_d$ from Tables 2–4 with the values of $\bar{l} - n^*_d$ from Tables 8–10, it becomes apparent that $\pi$ values are much tighter around $n^*_d$ than the $\bar{l}$ values. On the other hand, the two-stage estimation strategy (5.2)–(5.3) is operationally more convenient than the purely sequential estimation strategy (4.1) and (4.5).

So, here is an important issue that we must grapple with: Should one implement the purely sequential estimation strategy or the two-stage estimation strategy? Assume that in a practical situation, one is able to implement either sampling methodology. Then, one should pick the more appropriate methodology by properly balancing the cost due to increased logistics and sampling operations with the intrinsic value of the extent of tightness required between the average sample size and $n^*_d$. A practical problem must take into account all practical considerations as well as restrictions. Nothing less should be acceptable.

Acknowledgements

We received thoughtful and enthusiastic commentaries from two anonymous referees, an associate editor, and the Editor-in-Chief, Professor Makoto Aoshima. We are grateful to them for their most careful reading of the original version, especially for pointing out a number of necessary corrections very graciously. We thank them all.

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, tenth edition, New York, Dover.

Anscombe, F. J. (1952). Large sample theory of sequential estimation, *Proceedings of Cambridge Philosophical Society*, 48, 600–607.
Aoshima, M. and Mukhopadhyay, N. (1998). Fixed-width simultaneous confidence intervals for multinormal means in several intraclass correlation models, *J. Multivar. Anal.*, **66**, 46–63.

Aoshima, M. and Mukhopadhyay, N. (1999). Second-order properties of a two-stage fixed-size confidence region when the covariance matrix has a structure, *Statistical Inference and Data Analysis* (Tokyo, 1997), *Commun. Stat., Theory and Methods*, **28**, 839–855.

Aoshima, M. and Takada, Y. (2000). Second order properties of a two stage procedure for comparing several treatments with a control, *J. Japan Statist. Soc.*, **30**, 27–41.

Aoshima, M. and Yata, K. (2010). Asymptotically second-order consistency for two-stage estimation methodologies and its applications, *Ann. Inst. Stat. Math.*, **62**, 571–600.

Banerjee, S. and Mukhopadhyay, N. (2016). A general sequential fixed-accuracy confidence interval estimation methodology and its applications, *Ann. Inst. Stat. Math.*, **68**, 541–570.

Basu, D. (1964). Recovery of ancillary information, *Sankhyā*, **26**, 3–16.

Chow, Y. S., Hsiung, C. and Lai, T. L. (1979). Extended-renewal theory and moment convergence in anscombe’s theorem, *Ann. Probab.*, **7**, 304–318.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean, *Ann. Math. Stat.*, **36**, 457–462.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, New York, Chapman & Hall. Reprinted 2000, Boca Raton, CRC.

Dropbox Link for R Programs. https://www.dropbox.com/s/0g0schj34aoy2px/R-Programs.pdf?dl=0, August 2017.

Fisher, R. A. (1934). Two new properties of mathematical likelihood, *Proceedings of Royal Society, Series A*, **144**, 285–307.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, Edinburgh and London, Oliver and Boyd.

Fisher, R. A (1973). *Statistical Methods and Scientific Inference*, third edition, London, Macmillan.

Ghosh, M., Fraser, D. A. S. and Reid, N. (2010). Ancillarity statistics: A review, *Statistica Sinica*, **20**, 1309–1322.

Ghosh, M. and Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential procedures, *Sankhyā, Series A*, **43**, 220–227.

Ghosh, M., Mukhopadhyay, N. and Sen, P. K. (1997). *Sequential Estimation*, New York, Wiley.

Gut, A. (2012). Anscombe’s theorem 60 years later, *Sequential Analysis*, **31**, 368–396.

Joshi, S. N. and Shah, M. N. (1999). Estimation of \(P(Y < X)\) in the problem of the Nile, *Statistical Inference and Design of Experiments* (eds. U. J. Dixit and M. R. Satam), pp. 28–35, New Delhi, Narosa.

Kagan, A. and Malinovsky, Y. (2013). On the Nile problem by Sir Ronald Fisher, *Electronic Journal of Statistics*, **7**, 1968–1982.

Kagan, A. and Malinovsky, Y. (2016). On the structure of UMVUEs, *Sankhyā, Series A*, **78**, 124–132.

Kagan, A. and Malinovsky, Y. (2016). On the structure of UMVUEs, *Sankhyā, Series A*, **78**, 124–132.

Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I, *Annals of Statistics*, **5**, 946–954.

Lai, T. L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis II, *Annals of Statistics*, **7**, 60–76.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, second edition, New York, Springer.

Mukhopadhyay, N. (1999a). Second-order properties of a two-stage fixed-size confidence region for the mean vector of a multivariate normal distribution, *J. Multivar. Anal.*, **68**, 250–263.

Mukhopadhyay, N. (1999b). Higher than second-order approximations via two-stage sampling, *Sankhyā, Series A*, **61**, 254–269.

Mukhopadhyay, N. (2000). *Probability and Statistical Inference*, New York, Dekker.

Mukhopadhyay, N. (2014). On rereading D. Basu’s jointly sufficient statistic example made up of two ancillaries and miscellany, *Sankhyā, Series A*, **76**, 280–287.

Mukhopadhyay, N. and Aoshima, M. (1998). Multivariate multistage methodologies for simultaneous all pairwise comparisons, *Metrika*, **47**, 185–201.
Mukhopadhyay, N. and Banerjee, S. (2014). Purely sequential and two-stage fixed-accuracy confidence interval estimation methods for count data from negative binomial distributions in statistical ecology: One-sample and two-sample problems, *Sequential Analysis*, **33**, 251–285.

Mukhopadhyay, N. and Banerjee, S. (2015a). Sequential negative binomial problems and statistical ecology: A selected review with new directions, *Statistical Methodology*, **26**, 34–60.

Mukhopadhyay, N. and Banerjee, S. (2015b). Purely sequential and two-stage bounded-length confidence intervals for the Bernoulli parameter with illustrations from health studies and ecology, *Ordered Data Analysis, Modeling and Health Research Methods*, Honor of H. N. Nagaraja’s 60th Birthday (eds. P. Choudhary, C. Nagaraja, and H. K. T. Ng), pp. 211–234, New York, Springer.

Mukhopadhyay, N. and Chattopadhyay, B. (2012). A tribute to Frank Anscombe and random central limit theorem from 1952, *Sequential Analysis*, **31**, 265–277.

Mukhopadhyay, N. and Duggan, W. T. (1997). Can a two-stage procedure enjoy second order properties?, *Sankhyā, Series A*, **59**, 435–448.

Mukhopadhyay, N. and Duggan, W. T. (2000). On a two-stage procedure having second-order properties with applications, *Ann. Inst. Stat. Math.*, **51**, 621–636.

Mukhopadhyay, N. and Duggan, W. T. (2001). A two-stage point estimation procedure for the mean of an exponential distribution and second-order results, *Statistics & Decisions*, **19**, 155–171.

Mukhopadhyay, N. and Zhuang, Y. (2016). On fixed-accuracy and bounded-accuracy confidence interval estimation problems in Fisher’s “nile” example, *Sequential Analysis*, **35**, 516–535.

Mukhopadhyay, N. and Zhuang, Y. (2017). MLE, information, ancillary complement, and conditional inference with illustrations, *Methodology and Computing in Applied Probability*, **19**, 615–629.

Rao, C. R. (1973). *Linear Statistical Inference*, second edition, New York, Wiley.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*, Vienna, R Foundation for Statistical Computing.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics*, New York, Chapman & Hall.

Siegmund, D. (1985). *Sequential Analysis*, New York, Springer.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Stat.*, **16**, 243–258.

Stein, C. (1949). Some problems in sequential estimation (abstract), *Econometrica*, **17**, 77–78.

Woodroofe, M. (1977). Second-order approximations for sequential point and interval estimation, *Annals of Statistics*, **5**, 984–995.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, CBMS 39, Philadelphia, SIAM.