Scientific Research Publishing

Non-Regular Example of Confidence-Interval Construction

Jan Vrbik

Department of Mathematics and Statistics, Brock University, Catharines, Canada
Email: jvrbik@brocku.ca

Abstract

When dealing with a regular (fixed-support) one-parameter distribution, the corresponding maximum-likelihood estimator (MLE) is, to a good approximation, Normally distributed. But, when the support boundaries are functions of the parameter, finding good approximation for the sampling distribution of MLE (needed to construct an accurate confidence interval for the parameter's true value) may get very challenging. We demonstrate the nature of this problem, and show how to deal with it, by a detailed study of a specific situation. We also indicate several possible ways to bypass MLE by proposing alternate estimators; these, having relatively simple sampling distributions, then make constructing a confidence interval rather routine.

1. Introduction

Extensive literature exist describing how to find an approximate sampling distribution of one or more maximum-likelihood estimators of parameters of a *regular* distribution. The corresponding procedure is based on Rao-Cramer variance (in one-parameter case) or Fisher-information matrix (several parameters), resulting in (either univariate or multivariate, respectively) Normal distribution; in the former case, this can be further improved by including a few extra terms of a more accurate Edgeworth expansion (see, for example, [1] and [2]). Constructing such approximation is quite routine, even though use of a computer may be necessary to find the required first four moments.

The purpose of this article is to show how much more difficult the same task becomes when the sampled distribution is *not* regular; we do this using the fol-

lowing one-parameter example: sampling a distribution whose probability density function is given by

$$f(x) = \frac{3}{4\theta_0}\left(1 - \frac{x^2}{\theta_0^2}\right) \quad \text{when } -\theta_0 < x < \theta_0 \tag{1}$$

(zero otherwise), and finding an approximation for the probability density function (PDF) of the maximum-likelihood estimator (MLE, denoted $\hat{\theta}$) of the θ_0 (necessarily positive) parameter. We will also discuss several possibilities for constructing a confidence interval (CI) for the unknown value of θ_0, aiming for good accuracy in terms of both the CI's *length* (which needs to be small), and its level of confidence (which can be often only approximated).

Note that our example has two distinctive properties (beyond being a case of non-regular estimation), namely

- it involves what is called a *scaling* parameter; this simplifies the search for a good estimator,
- the corresponding MLE is at the local maximum of the likelihood function, *not* at its boundary (the latter case is easy to deal with—we leave it for another publication).

In what follows, we use the following terminology: **analytic** describes results expressed in formula form, while **numeric** implies using a computational algorithm; similarly, regarding accuracy, we use: **exact** (self-explanatory; only analytic results qualify), while **practically exact** answers can be computed numerically to arbitrary precision; **accurate** implies relatively small errors, and **adequate** restricts errors to less than one percent.

2. Maximum Likelihood Estimator

To find the MLE of θ_0, we differentiate the natural logarithm of the corresponding likelihood function, namely

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \ln\left(1 - \frac{X_i^2}{\theta^2}\right) - n\ln\theta + n\ln\left(\frac{3}{4}\right) \tag{2}$$

with respect to θ, thus getting

$$\frac{2}{\theta^3}\sum_{i=1}^{n}\frac{X_i^2}{1 - \frac{X_i^2}{\theta^2}} - \frac{n}{\theta} \tag{3}$$

Making the last expression equal to 0 results in

$$\frac{\overline{\frac{X^2}{\theta^2}}}{1 - \frac{X^2}{\theta^2}} = \frac{1}{2} \tag{4}$$

where the bar indicates the sample mean of the full expression. Solving the last equation for θ, subject to the $\theta > \max_{1 \leq i \leq n}|X_i|$ constraint (this can be done only numerically) yields the resulting ML *estimate,* denoted $\hat{\theta}$. Note that the distri-

bution of $\Theta := \hat{\theta}/\theta_0$ is parameter free, which simplifies our task; we then need to find the sampling distribution of Θ only (or an adequate approximation to it); the corresponding transformation back to $\hat{\theta}$ is then quite routine.

3. Sampling Distribution of Θ

Let us then investigate the distribution of Θ, a random variable defined as the unique (when $\theta > \max\limits_{1 \le i \le n} |Y_i|$) solution to

$$\frac{\overline{\dfrac{Y^2}{\theta^2}}}{1 - \dfrac{Y^2}{\theta^2}} = \frac{1}{2} \tag{5}$$

where Y stands for a sample of size n from a distribution whose PDF is

$$f(y) = \frac{3}{4}\left(1 - y^2\right) \quad \text{where } -1 < y < 1 \tag{6}$$

The usual (regular-case) approach would be to expand the LHS of (5) at $\theta = 1$ (the asymptotic mean of Θ), thus getting

$$\overline{\frac{Y^2}{1 - Y^2}} - 2 \cdot \overline{\frac{Y^2}{\left(1 - Y^2\right)^2}} \cdot (\theta - 1) + \cdots = \frac{1}{2} \tag{7}$$

and then solving for $\theta - 1$; this yields the following approximate value of $\Theta - 1$:

$$\Theta - 1 \simeq \frac{\overline{\dfrac{Y^2}{1 - Y^2}} - \dfrac{1}{2}}{2 \cdot \overline{\dfrac{Y^2}{\left(1 - Y^2\right)^2}}} \tag{8}$$

When both the numerator and denominator have finite moments, this then leads, without much difficulty, to the usual (Normal or Edgeworth) asymptotic result.

Unfortunately, $U := \dfrac{Y^2}{1 - Y^2}$ has $\dfrac{1}{2}$ as its expected value but its variance is infinite, while $V := \dfrac{Y^2}{\left(1 - Y^2\right)^2}$ has both of these moments infinite. The best we can do at this point is to investigate the marginal distribution of the numerator and then, separately, of the denominator (deriving the *bivariate* distribution of U and V and proceeding in the manner of [3] or [4] proves too difficult).

3.1. The Numerator

Using (6), it is easy to derive the corresponding PDF of U, namely

$$f_n(u) = \frac{3}{4} \cdot \frac{1}{(u + 1)^{5/2} \sqrt{u}} \quad \text{where } u > 0 \tag{9}$$

This can be routinely converted into the corresponding characteristic function

(CF) of $U - \dfrac{1}{2}$, namely

$$\chi_{\mathrm{n}}(t) := \frac{t}{2} \exp(-\mathrm{i}t) \cdot \left(K_0\left(-\frac{\mathrm{i}t}{2}\right) + (t - \mathrm{i}) K_1\left(-\frac{\mathrm{i}t}{2}\right) \right) \tag{10}$$

where K_0 and K_1 are second-kind Bessel functions. When expanded in t, the ln of the last expression yields

$$\ln \chi_{\mathrm{n}}(t) \simeq 0.2591 t^2 - 0.5890\mathrm{i}t\,|t| + 0.375 t^2 \ln|t| + \cdots \tag{11}$$

implying that the ln of the characteristic function of

$$\frac{\sum_{i=1}^{n}\left(\dfrac{Y_i^2}{1 - Y_i^2} - \dfrac{1}{2} \right)}{\sqrt{n \ln n}} \tag{12}$$

is (now expanded in terms which decrease with n)

$$-\frac{3t^2}{16}\left(1 + \frac{\ln\ln n}{\ln n} - \frac{1.382}{\ln n} \right) + \frac{3t^2 \ln|t|}{8 \ln n} - 0.589\mathrm{i}\frac{t\,|t|}{\ln n} + \cdots \tag{13}$$

The $n \to \infty$ limit of the last expression is $-\dfrac{3}{16}t^2$, which means that (12) has, in the same limit, the Normal distribution with the mean of 0 and variance equal to $\dfrac{3}{8}$, in agreement with [5] (yet, the actual variance of (12) remains infinite at any n). Unfortunately, the convergence to this limit is so slow that using it as an approximation is virtually useless; even when employing the full expression (9), adequate approximation is achieved only when $n \geq 1000$. As an example, see **Figure 1** which displays,

- an *empirical* histogram of one million values of (12), generated by Monte Carlo (MC) simulation (done using Mathematica) of the same number of Y_i samples of size $n = 1000$, and converting each of them to a value of (12),
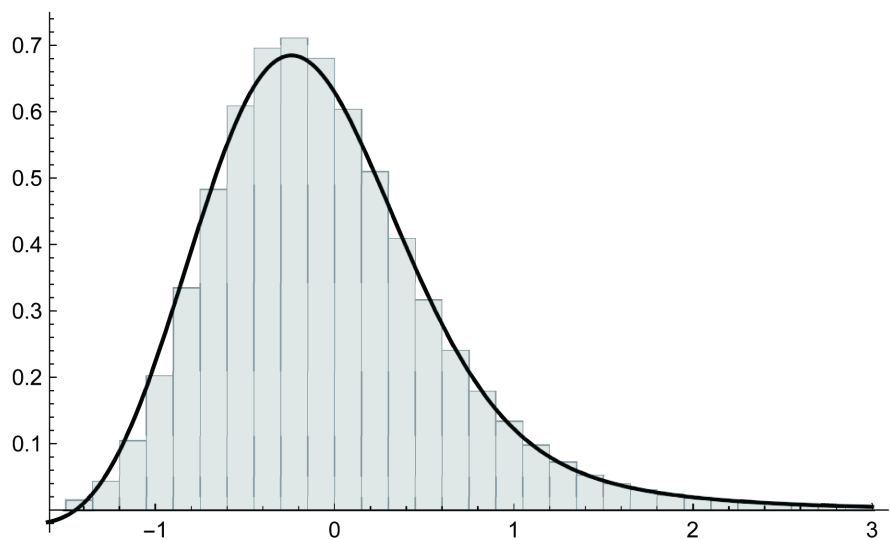

Figure 1. Approximating $\bar{U}$ distribution when $n = 1000$.

- and the corresponding PDF, obtained by raising e to the power of (13) and applying the inverse Fourier transform (also delegated to Mathematica) to the resulting CF; note that $\exp\left(\dfrac{3t^2 \ln|t|}{8 \ln n}\right)$ had to be further expanded to $1 + \dfrac{3t^2 \ln|t|}{8 \ln n}$ to facilitate proper convergence.

The two graphs support our claim of extremely slow convergence to a Normal distribution, which needed to be extended by extra $\dfrac{1}{\ln n}$ and $\dfrac{\ln \ln n}{\ln n}$ -proportional terms of (13) to reach (still less than adequate) accuracy at n as large as 1000. Nevertheless, even this information (the normalizing constant of (12) in particular) is essential in our subsequent attempt to find a useful approximation for the PDF of MLE's distribution.

3.2. The Denominator

By a similarly routine exercise, the PDF of V is

$$f_d(v) = \frac{3}{8\sqrt{2}} \frac{\sqrt{1 + 2v - \sqrt{1 + 4v}}}{v^{5/2}} \left(1 - \frac{1}{\sqrt{1 + 4v}}\right) \tag{14}$$

when $v > 0$ (zero otherwise). This time, there is no simple analytic expression for the corresponding CF; nevertheless, we can still find its t expansion by dividing the required integral into two parts, thus:

$$\int_0^\infty \exp(ivt) f(v) \, dv = \int_0^1 \exp(ivt) f(v) \, dv + \int_1^\infty \exp(ivt) f(v) \, dv \tag{15}$$

To evaluate the first of these, we expand $\exp(ivt)$ at $t = 0$, which lets us proceed by numerical integration; in the second part, we expand the $f(v)$ at $v = \infty$, which enables us to evaluate the resulting integral (term by term) analytically. This yields, for the final sum

$$\chi_d(t) = 1 - 0.4334|t| - it\left(0.4574 + 0.357 \ln|t|\right) + \cdots \tag{16}$$

implying that the ln of the CF of

$$\frac{\sum_{i=1}^n \dfrac{Y_i^2}{\left(1 - Y_i^2\right)^2}}{n \ln n} \tag{17}$$

is (when expanded as a function of n rather than t)

$$\left(0.375 - \frac{0.3018}{\ln n} + \frac{0.375 \ln \ln n}{\ln n}\right) it - \frac{0.589}{\ln n}|t| - 0.375i \frac{t \ln|t|}{\ln n} + \cdots \tag{18}$$

This readily implies that, in the $n \to \infty$ limit, (17) converges to the constant $\dfrac{3}{8}$ (yet, its expected value and variance remain infinite at any n); but again: this asymptotic (degenerate) distribution is modified (at any finite n) by additional terms, which make the actual convergence extremely slow. Similarly to results of the previous section, the empirical distribution of (17) agrees with the PDF con-

structed (via inverse Fourier transform) from (18) only at fairly large values of n; **Figure 2** illustrates that, even when $n = 1000$, the error of this approximation is less than adequate.
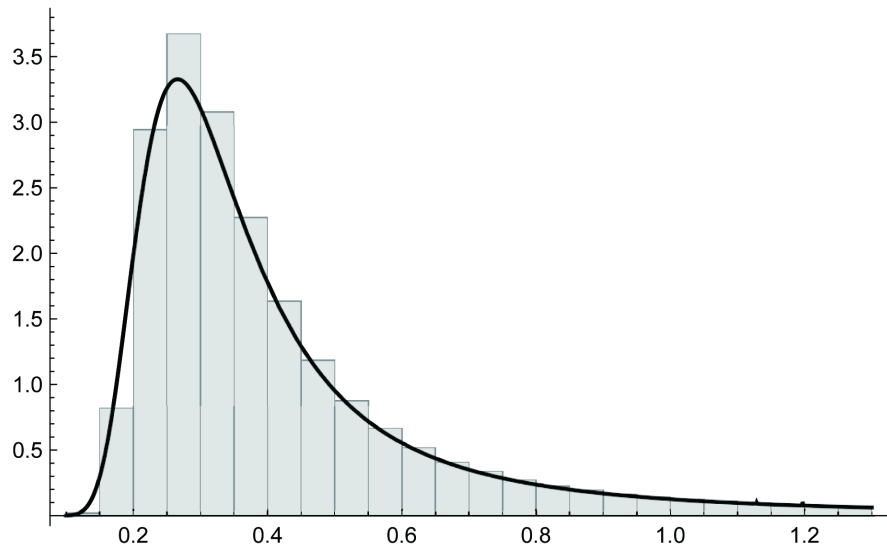

Figure 2. Approximating the $\bar{V}$ distribution when n = 1000.

3.3. And Their Ratio

Having such (rather imperfect) approximations for the distributions of both the numerator and denominator of (8) does not readily translate into a similar approximation for the ratio itself; attempting to find one would be extremely difficult and (considering the very limited accuracy achieved by such approach so far) ultimately of a very limited practical value. There is empirical evidence that the RHS of (8) cannot adequately approximate $\Theta - 1$ unless additional terms of the (7) expansion are included, thus further complicating the matter.

Nevertheless, the previous two sections have given us a good indication that $\Theta - 1$ needs to be multiplied by $\sqrt{n \ln n}$ to converge to an asymptotic distribution whose main part is Normal (with zero mean and variance equal to $\frac{2}{3}$), but which needs additional, $\frac{1}{\ln n}$ and $\frac{\ln \ln n}{\ln n}$ -proportional extra terms to reach adequate accuracy. But, to construct such an approximation, we can now rely only on MC simulation of a large number of random independent samples (RIS) from (6), each yielding one random value of Θ, as described earlier. Doing this, we discover that, unlike $\bar{U}$ and $\bar{V}$, the moments of the $\Theta - 1$ distribution are finite; this enables us to use Edgeworth approximation (described shortly) for the corresponding PDF.

But first we go over some details of the MC simulation.

4. Monte-Carlo Simulation

Existing software (such as Mathematica) can easily generate random values from

any of the commonly used distributions. Since $Y =: 2T - 1$, where T is a random variable having the Beta (2, 2) distribution, it is quite easy to generate a RIS of a specific size n from this distribution. Solving (5) for θ is more difficult, as the procedure needs to be fast, accurate and reliable. The way we do it starts with

$$\theta = \max_{1 \leq i \leq n} |Y_i| + \frac{0.2}{\sqrt{n}} \tag{19}$$

and continues by performing four iteration of the basic Newton technique, where a single iteration replaces the current value of θ by

$$\theta \cdot \left(1 + \frac{\overline{\dfrac{\left(\dfrac{Y}{\theta}\right)^2}{1 - \left(\dfrac{Y}{\theta}\right)^2}} - \dfrac{1}{2}}{2 \cdot \dfrac{\overline{\left(\dfrac{Y}{\theta}\right)^2}}{\left(1 - \left(\dfrac{Y}{\theta}\right)^2\right)^2}} \right) \tag{20}$$

This results in a single, random value of Θ; to get the corresponding empirical distribution of Θ, this is repeated as many times as feasible. The resulting **mega-sample** of the values of

$$W := (\Theta - 1)\sqrt{n \ln n} \tag{21}$$

can then be used to

- display the corresponding histogram;
- convert it into an empirical PDF function (we use Mathematica's "Smooth-Kernel Distribution" for this purpose; for the details of the corresponding algorithm, see [6]);
- find accurate estimates of the first few moments of the W distribution, of selected percentiles, etc.

Note that all our simulations use such large mega-samples (of one *million* RISs, to be specific) that their results can be considered practically exact.

4.1. Example

As an example, we present the results of this procedure when $n = 30$, displaying the histogram of the W values together with the corresponding empirical PDF in **Figure 3**; note that the latter can be evaluated only numerically (it does not have an analytic form). Also note both graphs have been constructed based on the same set of data; this explains their perfect agreement.

The simulation has also yielded the following (practically exact) values of the mean, variance, skewness and excess kurtosis of the W distribution: -0.331, $\frac{2}{3} - 0.120$, -0.571 and 0.391 respectively, while -2.598, -2.342, -1.978, -1.670. 0.733, 0.893, 1.078 and 1.203 are its empirical percentiles (also known as **critical**
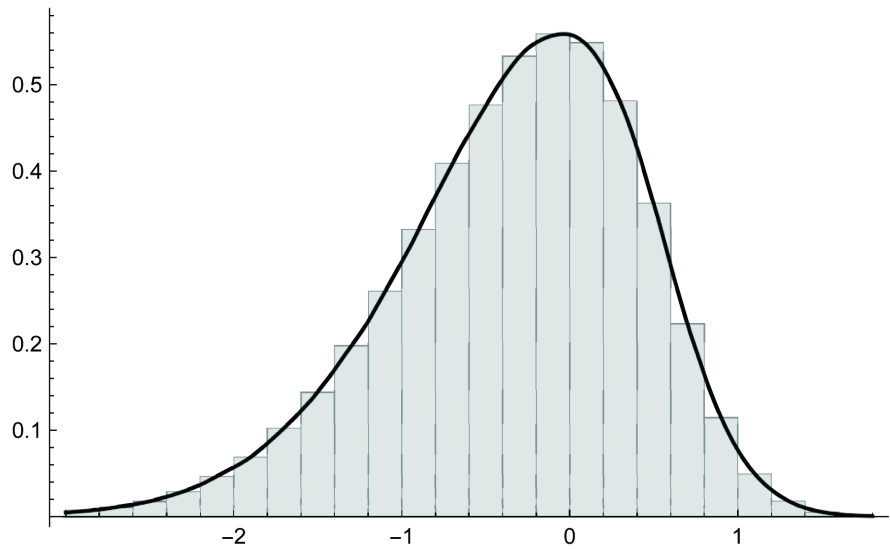
Figure 3. Empirical PDF of *W* when $n = 30$.

values), corresponding to 0.5%, 1%, 2.5%, 5%, 95%, 97.5%, 99% and 99.5% respectively (these enable us to construct CIs for θ_0 when $n = 30$, as explained below).

4.2. Edgeworth Approximation

When a reasonably accurate *analytic* formula is desired (for the PDF of *W*, at a specific value of *n*), it is natural to use the usual Edgeworth-series expansion, namely

$$f(w) = \frac{\exp\left(-\dfrac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 + \frac{z^3 - 3z}{6} \cdot \alpha_3 + \frac{z^6 - 15z^4 + 45z^2 - 15}{72} \cdot \alpha_3^2 \right.$$
$$\left. + \frac{z^4 - 6z^2 + 3}{24} \cdot \alpha_4 \right) \tag{22}$$

where

$$z = \frac{w - \mu}{\sqrt{V}} \tag{23}$$

and μ, *V*, α_3 and α_4, are the mean, variance, skewness and excess kurtosis taken from a MC simulation. This approximation will not be as accurate as the empirical PDF of **Figure 3**—the actual difference (when $n = 30$) is displayed in **Figure 4**—but it can be utilized by people lacking the sophistication and computer resources to do their own simulation. Note that the maximum error of this approximation is never bigger than 1% (the shaded area of **Figure 4**), and is expected (as has been confirmed) to decrease when the sample size gets larger.

By extending the simulation (this will take several hours of CPU) to get the same set of results for $n = 30, 30 \cdot 2, 30 \cdot 2^2, \cdots, 30 \cdot 2^9$, we have computed and are displaying (in **Figure 5**) the values of μ (red), $V - \dfrac{2}{3}$ (blue), α_3 (green) and
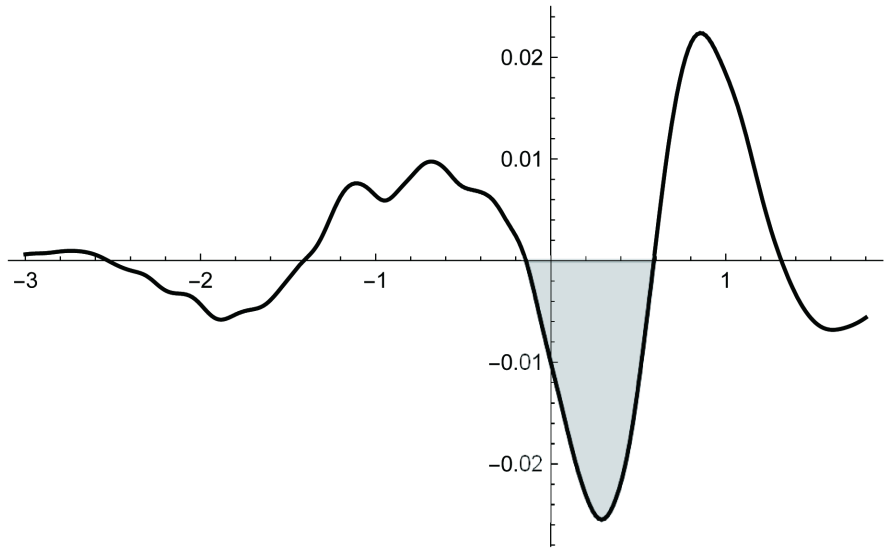
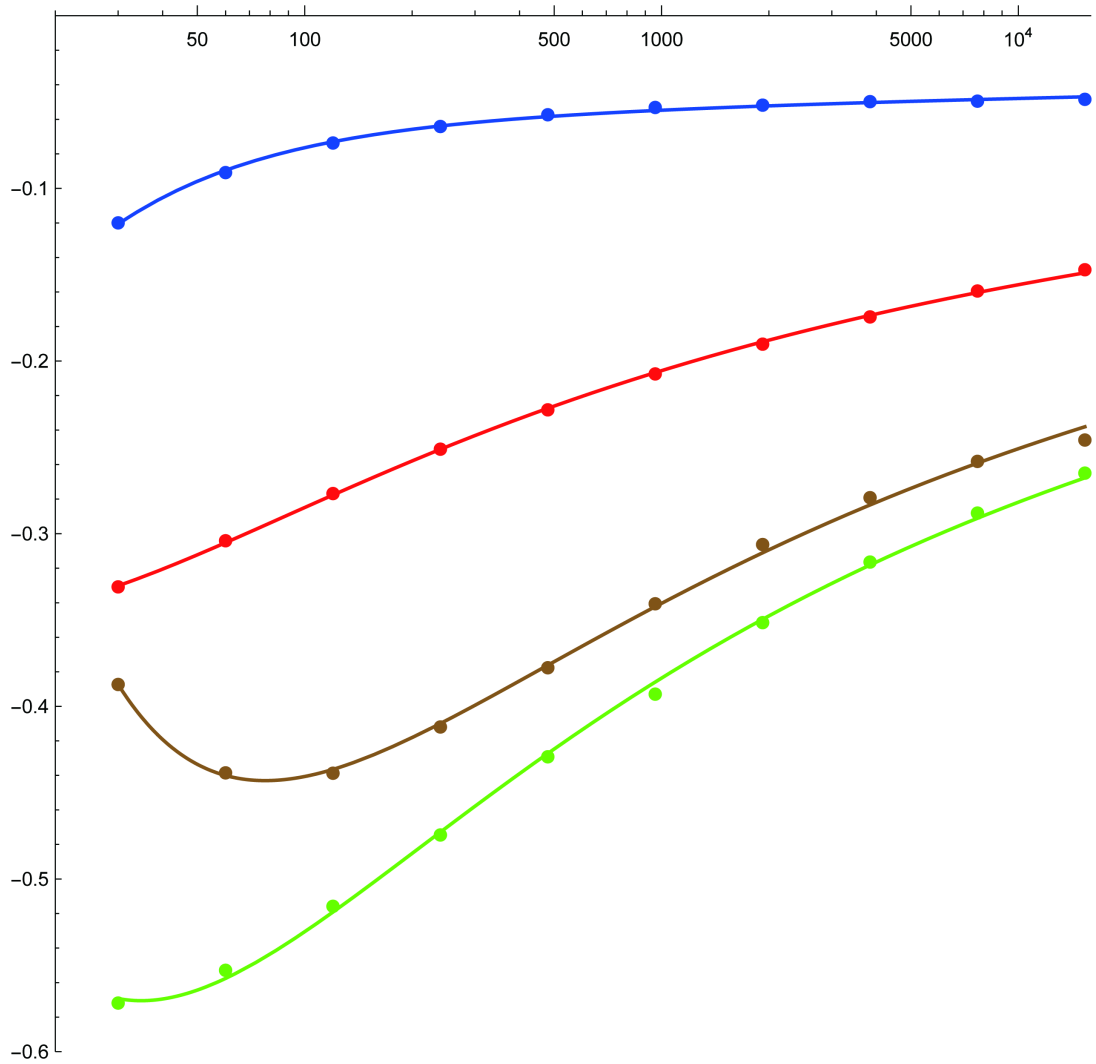**Figure 4.** Error of Edgeworth approximation when $n = 30$.

Figure 5. Four characteristics of W distribution.

$-\alpha_4$ (brown) of the W distribution, for any practical sample size n (on the horizontal, $\log_{10}$ scale); the small dots represent the simulated values, each of the four curves has the (sufficiently flexible, by numerical exploration) form of

$$\frac{a}{\ln n} + \frac{b \cdot \ln\ln n}{\ln n} + \frac{c}{n} \tag{24}$$

where a, b, c have been found by a least-square fit. The accuracy of the resulting (22)-based approximation is similar to what we have observed for $n = 30$ (improving, rather slowly, with increasing n).

Note that the mean (also known as the **expected value**) of W has such a substantial bias (the red line) that (unlike in a regular case) it cannot be ignored even when n gets extremely large (this goes for the other three quantities as well).

5. Confidence Intervals

We will now turn our attention to constructing CIs for θ_0, with the aim of making their expected length short and their claimed level of confidence as accurate as possible.

The CI construction critically depends on the sample statistics used for the point estimator of θ_0; we will consider several possibilities, staring with the obvious choice of MLE (the only estimator discussed so far).

5.1. MLE-Based CI

Most of what is needed to construct a CI for the true value of θ_0 has already been done in the previous section. Now, we just need to

- choose the level of confidence (denoted $1 - \alpha$), with α typically between 1% and 10% (5% being most common);
- find the corresponding critical values (denoted $C_{\alpha/2}$ and $C_{1-\alpha/2}$) of the W distribution (these are obtained either as a by-product of MC simulation or, somehow less accurately, computed from the Edgeworth's approximation to the PDF of W);
- find $\hat{\theta}$, based on a real (not simulated) RIS of size n, and solve

$$\left(\frac{\hat{\theta}}{\theta_0} - 1\right)\sqrt{n \ln n} = C_{1-\alpha/2} \tag{25}$$

for θ_0 to get the CI's lower limit, and repeat with $C_{\alpha/2}$ in place of $C_{1-\alpha/2}$ to get the upper limit.

Note that the expected length of the resulting CI is, to a sufficient approximation

$$\frac{C_{1-\alpha/2} - C_{\alpha/2}}{\sqrt{n \ln n}} \cdot \mathbb{E}\left(\hat{\theta}\right) \tag{26}$$

When $\alpha = 0.05$ and $n = 30$, this yields $0.293 \cdot \theta_0$ (implying that we can be easily off the true value of θ_0 by 30%); increasing n to 30720, the expected CI length is reduced to an impressive $0.00598 \cdot \theta_0$.

We still need to explain how to get (practically exact) critical value C_β for a
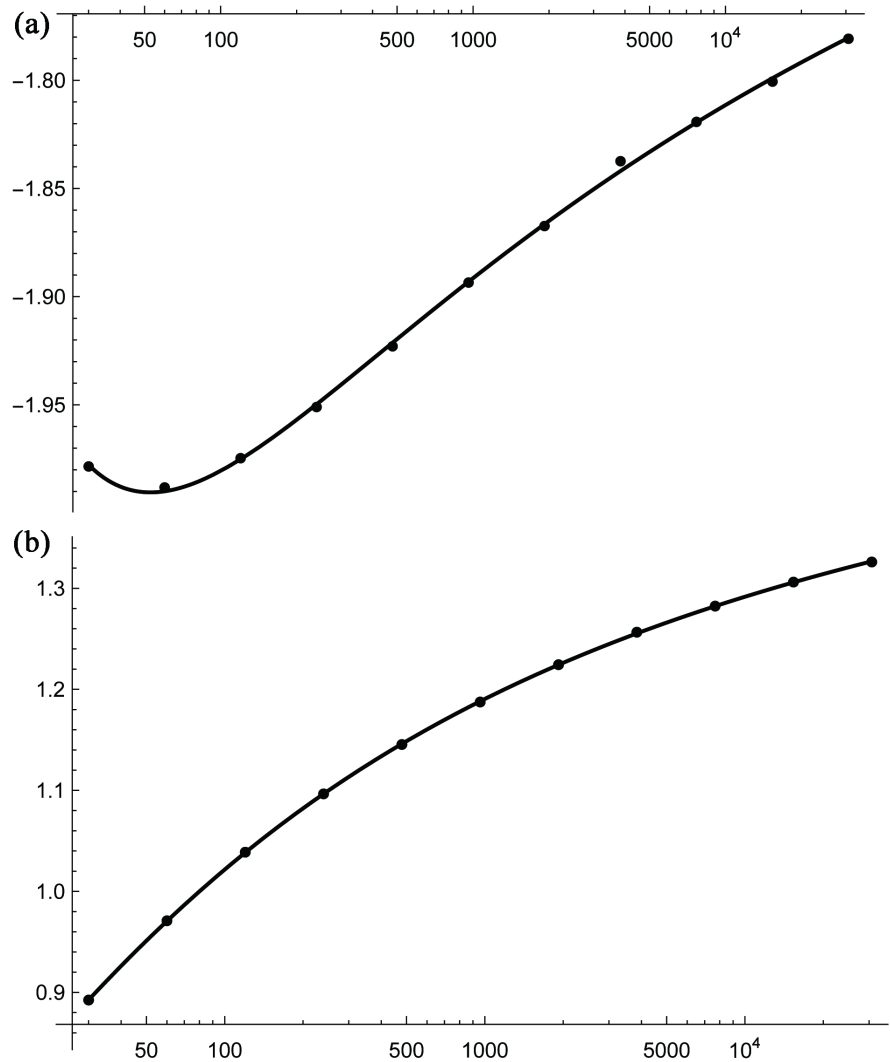
Figure 6. Critical values $C_{0.975}$ (top) and $C_{0.05}$ (bottom).

given sample size n, based on a mega-sample of a million $\hat{\theta}$ estimates: all we need to do is to take the kth smallest of these estimates, where k is the nearest integer to $\beta \cdot 10^6$ (something we have already done for several values of β and $n = 30$). By an extensive simulation (discussed previously) we have extended these results up to $n = 30720$; the values of $C_{0.025}$ and $C_{0.975}$ are displayed in **Figure 6**. This time, the curves were fitted using (24) with an additional constant term.

An alternate (but more elaborate) way of constructing an MLE-based CI is to minimize the distance between two percentiles (denoted C_{β} and $C_{1-\alpha+\beta}$, where $0 < \beta < \alpha$) while keeping the probability of the corresponding interval equal to $1-\alpha$; this is achieved (based on the same mega-sample) by usual minimization of this distance, after displaying it as a function of β (this can be done graphically). In the case of $n = 30$ and $\alpha = 5\%$. this yields -1.855 (instead of -1.978) and 0.977 (instead of 1.078) when $\beta = 0.0333$; the distance of the corresponding critical values has thus been shortened to 2.832 (from the old

3.056, *i.e.* by about 8%), thus improving the CI's predictive power.

The main disadvantage of using the technique of this section rests in our inability to find, without extensive simulation, an analytic and sufficiently accurate expression for the PDF of W (even computing the estimate itself may prove difficult for some). Yet, it needs to be acknowledged that MLE is more efficient than any other potential estimator of θ_0, especially at large values of n (due to the extra $\dfrac{1}{\ln n}$ factor in its variance; most estimators have a variance decreasing with $\dfrac{1}{n}$ only). Therefore, in our subsequent search for a simpler, formula-based point estimator of θ_0, we must expect some reduction in the estimator's relative efficiency, and thus in the predictive power of the resulting CI.

The question now is: when abandoning MLE, what other potential candidates are there for a good (*i.e.* relatively efficient and also unbiased, at least asymptotically) estimator? Furthermore, we would like to be able find, analytically, the asymptotic form of its distribution, and have the exact distribution *rapidly* converge to this $n \to \infty$ limit.

5.2. Method of Moments

The next traditional choice of an estimator (called MME) uses the method of moments. It works by solving the $\mathbb{E}(X) = \bar{X}$ equation for θ_0, where $\mathbb{E}(X)$ stands for the expected value of a single X, and $\bar{X}$ is the usual sample mean of n such values. Since, in our case, $\mathbb{E}(X) = 0$ (thus, not a function of θ_0), the method needs to be modified by using a *function* of X instead of X itself; in our case, a convenient choice is $|X|$, whose PDF, namely

$$\frac{3}{2\theta_0}\left(1 - \frac{x^2}{\theta_0^2}\right) \quad \text{when } 0 < x < \theta_0 \tag{27}$$

leads to $\mathbb{E}(|X|) = \dfrac{3}{8}\theta_0$. The new (fully unbiased) estimator of θ_0 is thus

$$\hat{\theta} := \frac{8}{3}\frac{\sum_{i=1}^{n}|X_i|}{n} \tag{28}$$

(we keep the original notation for the new estimator).

We can now (rather routinely) compute the mean, variance, skewness and excess kurtosis of this estimator to be $\mu = \theta_0$, $V = \dfrac{19}{45}\cdot\dfrac{\theta_0^2}{n}$, $\alpha_3 = 14\sqrt{\dfrac{5}{19^3 n}}$ and $\alpha_4 = -\dfrac{2106}{2527n}$ respectively. The Central Limit Theorem (CLT) tells us that the sampling distribution of $\hat{\theta}$ is approximately Normal, having the above mean and variance. Unlike the previous case of MLE, this approximation is adequate even at relatively small values of n. Furthermore, we can substantially improve its accuracy by incorporating corrections indicated by (22); **Figure 7** displays (using the z scale) the corresponding error of both approximations
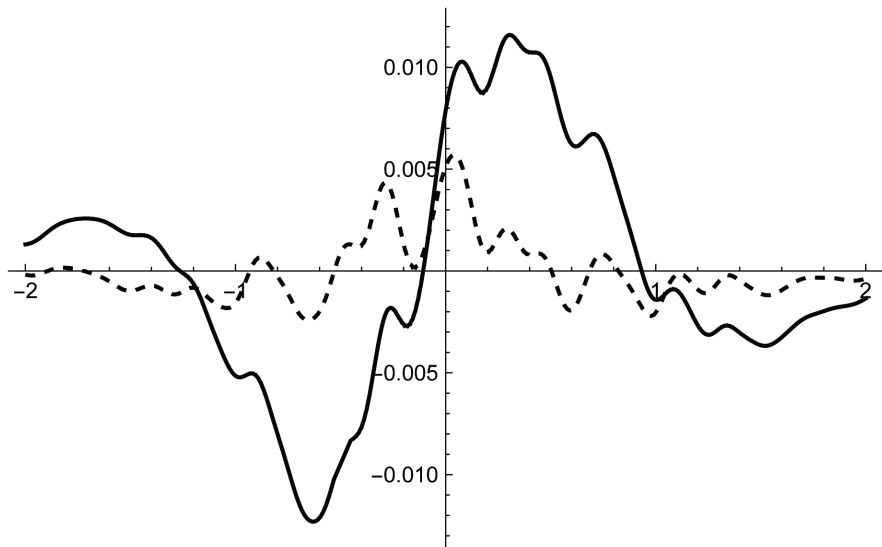
Figure 7. Error of CLT and Edgeworth (dashed) PDFs when n = 30.

relative to (MC generated but practically exact) PDF of

$$W := \left(\frac{\hat{\theta}}{\theta_0} - 1 \right) \sqrt{n} \qquad (29)$$

Note that the simulation was required only for this comparison; it is not a part of the technique itself.

The largest possible error of the CLT approximation at $n = 30$ (the worst-case scenario) is about 1% (adequate), while the Edgeworth approximation is distinctively more accurate (indicated by the area under the curve, not by its height). Note that the CTL's error will decrease with $\frac{1}{\sqrt{n}}$, while that of Edgeworth's approximation does it with $\frac{1}{n^{3/2}}$ (substantially faster).

The advantage of these approximations is that both have a simple analytic form; this enables us to directly compute the critical values required for construction of CIs. Furthermore, these critical values (using the W scale) are, when using the CLT approximation, independent of n; thus, for example $C_{0.025} = -1.960 \cdot \sqrt{\frac{19}{25}} = -1.274$ and similarly $C_{0.975} = 1.274$. Using MC simulation and $n = 30$, we have established the true confidence level of the corresponding CI to be 0.9508; the approximation is thus quite accurate even for such a small samples (better accuracy yet can be achieved via the Edgeworth approximation; something we leave for the reader to try).

Using the appropriate analog of (26), namely

$$\frac{C_{1-\alpha/2} - C_{\alpha/2}}{\sqrt{n}} \cdot \mathbb{E}\left(\hat{\theta}\right) \qquad (30)$$

the expected length of a CI when $\alpha = 0.05$ and $n = 30$ is $0.465 \cdot \theta_0$ (59% longer than MLE-based CI); when n increases to 30,720, the length becomes 0.0145

(2.43 times longer, which makes the technique rather non-competitive at large n).

In summary: using (28) as a basis of CI construction is computationally quite simple, as its distribution is approximately Normal yet reasonably accurate; unfortunately, the predictive power of the resulting CI (measured by its length) is less than impressive.

5.3. Utilizing Order Statistic

In this section, we explore yet another estimator of θ_0, namely the nth-order statistic of a sample of n values of $|X|$, *i.e.* $\hat{\theta} := \max_{1 \leq i \leq n} |X_i|$; note that its *exact* sampling distribution and the corresponding $n \to \infty$ limit are now both readily available in an analytic form. The idea of using this estimator rests on a hint we get from our MLE (which was always only slightly bigger than this maximum).

Transforming $\hat{\theta}$ to

$$W := \left(1 - \frac{\max_{1 \leq i \leq n} |X_i|}{\theta_0} \right) \cdot \sqrt{n} \tag{31}$$

We can easily compute

$$\Pr(W > w) = \Pr\left(1 - \frac{|X|}{\theta_0} > \frac{w}{\sqrt{n}} \right)^n = \left(1 - \frac{3}{2n} w^2 + \frac{w^3}{2n^{3/2}} \right)^n \quad \text{when } 0 < w < \sqrt{n} \tag{32}$$

based on

$$\Pr\left(\frac{|X|}{\theta_0} < 1 - \frac{w}{\sqrt{n}} \right) = \frac{3}{2} \left(1 - \frac{w}{\sqrt{n}} - \frac{\left(1 - \frac{w}{\sqrt{n}} \right)^3}{3} \right) = 1 - \frac{3}{2n} w^2 + \frac{w^3}{2n^{3/2}} \tag{33}$$

which follows from (27), and the corresponding

$$\Pr\left(\frac{|X|}{\theta_0} < x \right) = \frac{3}{2} \left(t - \frac{x^3}{3} \right) \quad \text{when } 0 < x < 1 \tag{34}$$

Note that, when n is sufficiently large, we can approximate (32) by its $n \to \infty$ limit, namely

$$\Pr(W > w) \simeq \exp\left(-\frac{3}{2} w^2 \right) \quad \text{when } w > 0 \tag{35}$$

Both formulas (exact and approximate) indicate that $\hat{\theta}$ is unbiased only asymptotically; the exact bias needs to be computed from (32); for sufficiently large n we get the asymptotic result of

$$\mathbb{E}\left(\hat{\theta} \right) = \left(1 - \sqrt{\frac{\pi}{6n}} \right) \cdot \theta_0 \tag{36}$$

obtained from (35). Similarly, the asymptotic variance of $\hat{\theta}$ is $\frac{4 - \pi}{6n} \cdot \theta_0^2$.

To construct a CI for θ_0, one needs to find $C_{\alpha/2}$ ($C_{1-\alpha/2}$) by solving

$\Pr(W > w) = 1 - \dfrac{\alpha}{2}$ $\left(= \dfrac{\alpha}{2}\right)$ for w, and then solve $W = C_{\alpha/2}$ for θ_0 (getting the lower boundary) and $W = C_{1\alpha/2}$ (upper boundary). In **Figure 8**, we plot the two critical values (using $\alpha = 5\%$) as functions of n; note that there are only minute differences between these and their $n \to \infty$ limits, namely to $C_{0.025} = 0.1299$ and $C_{0.975} = 1.5682$.

This implies that the asymptotic values can be used at practically any n; even when $n = 30$, the true confidence level thus achieved is 94.6% instead of 95% (a tolerable error).

To assess the predictive power of the new technique: the expected length of a CI (when $\alpha = 0.05$ and $n = 30$) is, based on (30), $0.305 \cdot \theta_0$ (only 4% higher than when using MLE); increasing n to 30720 reduces it to $0.00829 \cdot \theta_0$ (about 39% higher—still quite tolerable).

Furthermore, we can make any such CI shorter by optimizing its length while keeping the confidence level at 95%, as done in the previous section; to demonstrate
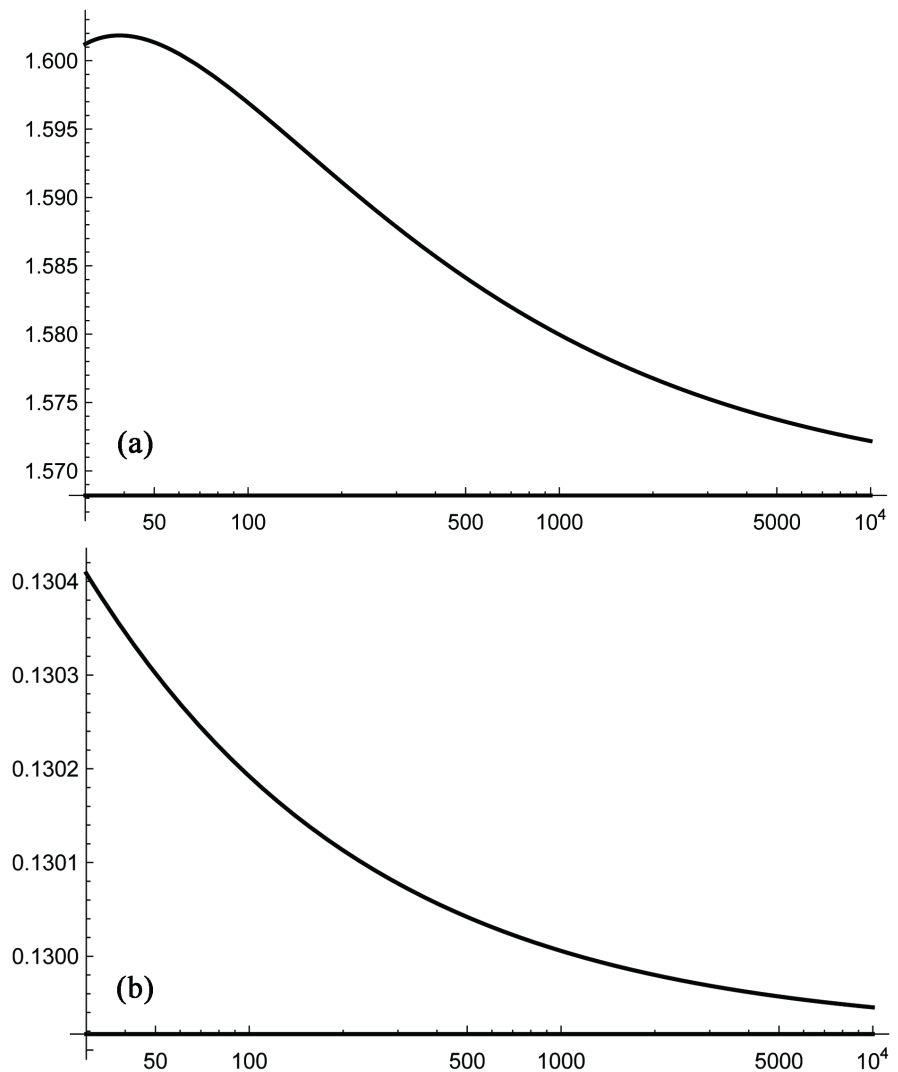

Figure 8. $C_{0.975}$ (top) and $C_{0.025}$ (bottom) of (31) distribution.

this, we use $n = 30720$ and the asymptotic formulas (practically exact at such large n). Skipping routine details, this yields $C_{0.0439} = 0.5375$ and $C_{0.9939} = 1.4435$, resulting in a new length of $0.00787 \cdot \theta_0$ (only 32% higher than MLE-based CI).

To summarize: with the last estimator, we have achieved a combination of simplicity, ease of computation and a fast convergence to asymptotic formulas, without much loss in the predictive power of the corresponding CI.

6. Conclusions and Future Research

We have shown that finding an accurate and efficient CI for a single parameter is, in a case of non-regular estimation, a surprisingly difficult task; as a convincing simple example, we have presented a case study of estimating a scaling parameter of a distribution closely related to Beta (2, 2). We have delineated several possibilities for dealing with this problem by using the MLE, MME, and the largest-order statistics as point estimators, and investigating their sampling distributions. We have discovered that this distribution converges to Normal for both MLE and MME, but that in the former case the convergence is too slow to be of any practical use (the CI construction needs to rely on CPU-intensive numerical approach). Using MME and $\max|X|$ resulted in simple, analytic formulas for each of their sampling distributions and their asymptotic form (reached relatively quickly as n increases). Unfortunately, the variance of MME is too large to make the resulting CI competitive, but using $\max|X|$ proved to be both simple and relatively efficient.

Empirically (and rather surprisingly) we have also discovered that, similarly to *regular* cases, the distribution of

$$2\mathcal{L}(\hat{\theta}) - 2\mathcal{L}(\theta_0) \tag{37}$$

(where $\mathcal{L}(\theta)$ was defined in (2)) is approximately χ_1^2. This is demonstrated in **Figure 9** using $n = 10$ (at such small n, most approximations become visibly inaccurate; the practically perfect fit we see here is truly amazing). To construct a
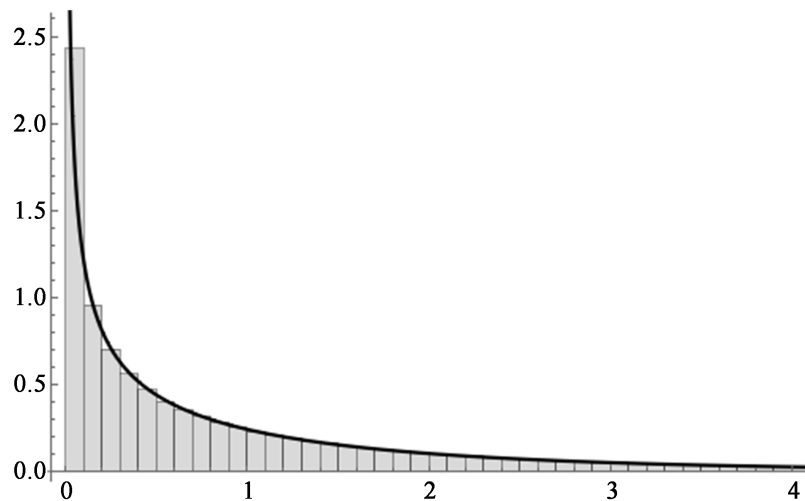

Figure 9. Histogram of (37)-values and PDF of χ_1^2.

95% CI for the value of θ_0, all we have to do is to compute $\hat{\theta}$ (based on a specific RIS), make (37) equal to $C_{0.95}$ of the χ_1^2 distribution (*i.e.* 3.8415), and solve for θ_0 (there will always be two solutions, providing the CI's boundaries). This produces CI's similar to those constructed in the last paragraph of the previous section, including the $1/\sqrt{n}$ -proportional decrease of the CI's length. Theoretical justification of these statements, and finding out whether they are true in general (after excluding MLEs involving max or min which have different, but easy-to-find, asymptotic distributions) will require further investigation.

We acknowledge that every case of non-regular estimation is unique and may require exploring other possibilities than those presented here; nevertheless, the results of our article provide guidelines for such an exploration.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Edgeworth, F.Y. (1905) The Law of Error. *Transactions of the Cambridge Philosophical Society,* **20**, 36-66, 113-141.

[2] Vrbik, J. (2022) Fisher Transformation via Edgeworth Expansion. arXiv: 2208.05070v1.

[3] Logan, B.F., Mallows, C.L., Rice, S.O. and Shepp, A.L. (1973) Limit Distributions of Self-Normalized Sums. *The Annals of Probability*, **1**, 788-809.
https://doi.org/10.1214/aop/1176996846

[4] Spataru, A. (2014) Convergence and Precise Asymptotics for Series Involving Self-Normalized Sums. *Journal of Theoretical Probability*, **29**, 267-276.
https://doi.org/10.1007/s10959-014-0560-1sss

[5] Gnedenko, B.V. and Kolmogorov, A.N. (1954) Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Cambridge.

[6] https://reference.wolfram.com/language/ref/SmoothKernelDistribution.html