

# High-Accuracy Confidence Regions for Distribution Parameters

Jan Vrbik

Department of Mathematics and Statistics, Brock University, St. Catharines, Canada

Email: [jvr bik@brocku.ca](mailto:jvr bik@brocku.ca)

**How to cite this paper:** Vrbik, J. (2022) High-Accuracy Confidence Regions for Distribution Parameters. *Applied Mathematics*, 13, 488-501.

<https://doi.org/10.4236/am.2022.136031>

**Received:** April 8, 2022

**Accepted:** June 14, 2022

**Published:** June 17, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

With the help of today's computers, it is always relatively easy to find maximum-likelihood estimators of one or more parameters of any specific statistical distribution, and use these to construct the corresponding *approximate* confidence interval/region, facilitated by the well-known asymptotic properties of the likelihood function. The purpose of this article is to make this approximation substantially more accurate by extending the Taylor expansion of the corresponding probability density function to include quadratic and cubic terms in several centralized sample means, and thus finding the corresponding  $\frac{1}{n}$ -proportional correction to the original algorithm. We then demonstrate the new procedure's usage, both for constructing confidence regions and for testing hypotheses, emphasizing that incorporating this correction carries minimal computational and programming cost. In our final chapter, we present two examples to indicate how significantly the new approximation improves the procedure's accuracy.

## Keywords

High-Accuracy Confidence Regions, Distribution Parameters

## 1. Introduction

Using the likelihood function to find an approximate confidence region for several parameters (often only one, in which case we speak of a confidence *interval*) of a specific distribution goes back to the classic publication of Kendall and Stuart [1]. To understand the rest of our article, it is necessary to briefly review their main result, summarized by

**Theorem 1.** *Assuming a case of **regular** (meaning the distribution's support is **not** a function of any of the distribution's parameters) estimation, the follow-*

ing random variable

$$2 \ln L(\mathbf{X}; \hat{\boldsymbol{\theta}}) - 2 \ln L(\mathbf{X}; \boldsymbol{\theta}_0) \tag{1}$$

has approximately the chi-square distribution with  $K$  degrees of freedom, where  $K$  is the number of parameters to be estimated,  $\mathbf{X}$  is the set of  $n$  observations (allowing for the possibility of a multivariate distribution),  $L(\mathbf{X}; \boldsymbol{\theta})$  denotes the corresponding likelihood function,  $\hat{\boldsymbol{\theta}}$  is the vector of the resulting ML estimates, and  $\boldsymbol{\theta}_0$  represents the true (but unknown) values of the parameters.

**Proof.** Realizing that

$$\ln L(\mathbf{X}; \boldsymbol{\theta}) := \sum_{i=1}^n \ln f(\mathbf{x}_i; \boldsymbol{\theta}) \tag{2}$$

where  $\mathbf{x}_i$  are the individual observations, then differentiating the RHS with respect to  $\boldsymbol{\theta}$  and setting each component of the answer to 0 yields

$$\sum_{i=1}^n \frac{\partial \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{3}$$

where  $\frac{\partial}{\partial \boldsymbol{\theta}}$  implies differentiating  $\ln f(\mathbf{x}_i; \boldsymbol{\theta})$  with respect to *each* parameter; the result is thus a vector with  $K$  components.

To solve the corresponding set of equations, we expand the LHS of (3) at  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ , thus getting

$$\sum_{i=1}^n \frac{\partial \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \sum_{i=1}^n \frac{\partial^2 \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \dots = \mathbf{0} \tag{4}$$

where the second derivative of  $\ln f(\mathbf{x}_i; \boldsymbol{\theta})$  is a symmetric  $K$  by  $K$  matrix. Dividing each side of the last equation by  $n$ , we solve it for  $\boldsymbol{\theta} - \boldsymbol{\theta}_0$  getting, to the *first-order* accuracy

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \simeq -\mu_2^{-1} \mathbf{Y} \tag{5}$$

where

$$\mathbf{Y} := \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{6}$$

$$\mu_2 := \mathbb{E} \left( \frac{\partial^2 \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \tag{7}$$

Note that the expected value of  $\mathbf{Y}$  is  $\mathbf{0}$ , since

$$\int_{\text{All } \mathbf{x}} \frac{\partial \ln f(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = \int_{\text{All } \mathbf{x}} \frac{\partial f(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\mathbf{x} = \frac{\partial}{\partial \boldsymbol{\theta}} (1) = \mathbf{0} \tag{8}$$

and its variance-covariance matrix equals to  $-\mu_2/n$ , due to

$$\frac{d}{d\boldsymbol{\theta}} \int_{\text{All } \mathbf{x}} \frac{\partial \ln f(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = \mu_2 + \mu_{11} = \mathbf{0} \tag{9}$$

where

$$\mu_{11} := \mathbb{E} \left( \left. \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \circ \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right) \tag{10}$$

is the variance-covariance matrix of  $\left. \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}$  (the small circle implies *direct* product of the two vectors), and  $\mathbf{0}$  is the zero matrix.

Similarly expanding (1) and utilizing (5) we get, to the same level of accuracy

$$2n\mathbf{Y}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mu_2 (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -n\mathbf{Y}^T \mu_2^{-1} \mathbf{Y} \tag{11}$$

resulting in the following simple approximation to (1)

$$-\sqrt{n}\mathbf{Y}^T \mu_2^{-1} \mathbf{Y} \sqrt{n} \tag{12}$$

where  $\mathbf{Y}\sqrt{n}$  has (due to Central Limit Theorem) a  $K$ -variate Normal distribution with the mean of  $\mathbf{0}$  and the variance-covariance matrix of  $-\mu_2$ . Introducing  $\mathbf{Z} = \mathbf{Y}\sqrt{n}$ , the moment generating function of (12) is then computed by

$$\begin{aligned} & \frac{\int_{-\infty}^{\infty} \exp \left( -\frac{\mathbf{z}^T (-\mu_2)^{-1} \mathbf{z}}{2} + t \mathbf{z}^T (-\mu_2)^{-1} \mathbf{z} \right) d\mathbf{z}}{(2\pi)^{K/2} \sqrt{\det(-\mu_2)}} \\ &= \sqrt{\det \left( \frac{-\mu_2}{1-2t} (-\mu_2)^{-1} \right)} = (1-2t)^{-K/2} \end{aligned} \tag{13}$$

easily identified as the MGF of  $\chi_K^2$  distribution. ■

The result then leads to a simple (we call it *basic*) algorithm of constructing the corresponding confidence region based on a random independent sample of  $n$  observations.

### Examples

From the last theorem it follows that an approximate confidence region is found by following these steps:

- Using sample values drawn from the distribution, maximize its likelihood function, thus getting the ML estimates of all parameters;
- Choose a confidence level (denoted  $\tau$ ) and make (1) (where  $\hat{\boldsymbol{\theta}}$  represents the ML estimates, but  $\boldsymbol{\theta}_0$  are now *variables* to be solved for) equal to the corresponding critical value of the  $\chi_K^2$  distribution;
- Solve for  $\boldsymbol{\theta}_0$ , resulting in two limits of the corresponding confidence interval when  $K = 1$ , a closed curve when  $K = 2$ , a closed surface when  $K = 3$ , etc.; this means that, beyond a one-parameter case, the resulting region can be displayed only graphically (a trivial task for today's computers).

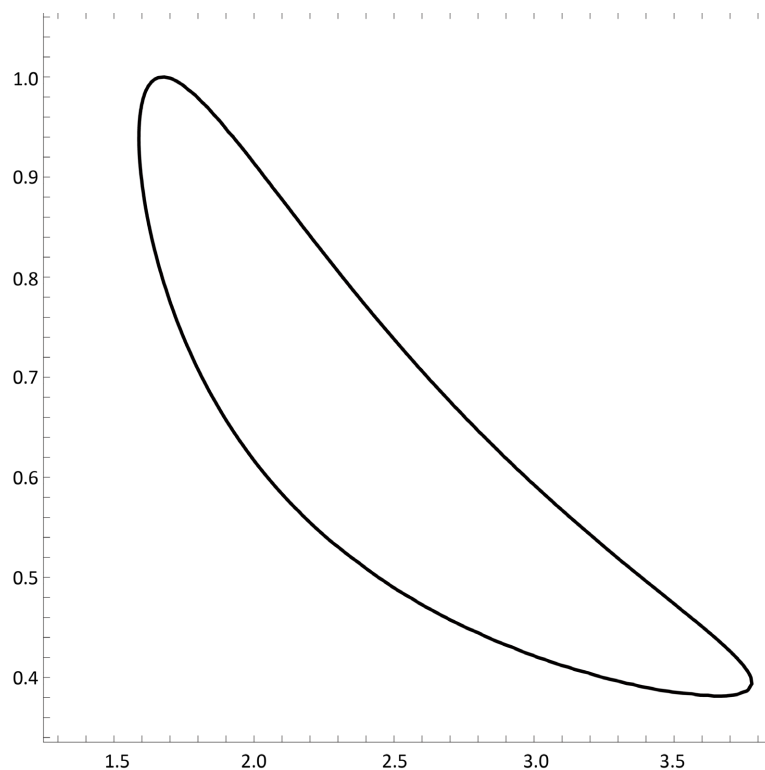
The following Mathematica program shows how it is done, first using a Geometric distribution with a (computer generated) sample of 40 and confidence level of 90%, then a Gamma distribution with a sample of 60 and confidence level of 95%.

```
(* Geometric-distribution example *)
X = RandomVariate[GeometricDistribution[0.37], 40]; (* RIS of size 40 *)
n = Length[X]; (* sample size *)
Inf = Log[p] + x Log[1 - p]; (* logarithm of Geometric PMF *)
lnLF = Apply[Plus, Inf /. x -> X]; (* logarithm of LF *)
max = NMaximize[{lnLF, 0 < p < 1}, p] (* yielding point estimate: *)
{-72.5525, {p -> 0.36036}}
CL = InverseCDF[ChiSquareDistribution[1], 0.90]; (* confidence level *)
NSolve[2max[[1]] - 2lnLF == CL, p, Reals] (* boundaries of CI: *)
{{p -> 0.274971}, {p -> 0.452126}}

(* Gamma-distribution example *)
X = RandomVariate[GammaDistribution[2.13, 0.77], 60]; (* RIS of size 60 *)
n = Length[X]; (* sample size *)
Inf = ( $\alpha - 1$ ) Log[x] - x/ $\beta$  - LogGamma[ $\alpha$ ] -  $\alpha$  Log[ $\beta$ ]; (* log of PDF *)
lnLF = Apply[Plus, Inf /. x -> X]; (* log of LF *)
max = NMaximize[{lnLF, 0 <  $\alpha$ , 0 <  $\beta$ }, { $\alpha$ ,  $\beta$ }] (* point estimates: *)
{-76.3629, { $\alpha$  -> 2.30683,  $\beta$  -> 0.668023}}
CL = InverseCDF[ChiSquareDistribution[2], 0.95]; (* confidence level *)
ContourPlot[2 max[[1]] - 2 lnLF == CL, { $\alpha$ , 1, 3.5}, { $\beta$ , 0.3, 1.2}]
```

The last line of the program has produced the following 95% confidence region (see **Figure 1**) for  $\alpha$  (horizontal scale) and  $\beta$  (vertical scale).

A three-parameter estimation can be done in a similar manner, only the call to “ContourPlot” needs to be replaced by “ContourPlot3D”. Visualizing a confidence region for more than three parameters would require slicing it into a sequence of cross-sections.



**Figure 1.** Confidence region for  $\alpha$  and  $\beta$ .

The error of this procedure depends on the sample size, decreasing with  $n$  in the  $O\left(\frac{1}{n}\right)$  manner; it can reach several percent for small samples. The purpose of the next section is to show how to substantially reduce this error, making it rather insignificant in all practical situations.

## 2. High-Accuracy Extension

There have been many attempts at improving the accuracy of this approximation, starting with Bartlett [2], followed by [3] and many others, all concentrating on higher moments of the approximately Normal distribution of ML estimators. Our approach is different: we aim at finding an appropriate correction to the  $\chi^2$  distribution of the Likelihood function. The technique we use is not unlike that of [4], in spite of diverging objectives.

It has been shown in previous publications [5] and [6] that the distribution of (1) is more accurately described by the following probability density function

$$\chi_K^2(u) \left( 1 + \frac{A}{n} \cdot \left( \frac{u}{K} - 1 \right) \right) \tag{14}$$

where  $\chi_K^2(u)$  is the chi-square PDF of the original theorem,  $A$  is a quantity (often a simple constant) specific to the sampled distribution, and  $K$  is the number of parameters to be estimated. In terms of the corresponding CDF, this becomes

$$1 - \frac{\Gamma\left(\frac{K}{2}, \frac{u}{2}\right) + \frac{2A}{nK} \left(\frac{u}{2}\right)^{K/2} \exp\left(-\frac{u}{2}\right)}{\Gamma\left(\frac{K}{2}\right)} \tag{15}$$

(the two-argument  $\Gamma$  denotes the incomplete Gamma function) or, more explicitly

$$\operatorname{erf}\left(\sqrt{\frac{u}{2}}\right) - \frac{A}{n} \sqrt{\frac{2u}{\pi}} \exp\left(-\frac{u}{2}\right)$$

when  $K = 1$ , and

$$1 - \left( 1 + \frac{Au}{2n} \right) \exp\left(-\frac{u}{2}\right) \tag{16}$$

when  $K = 2$  (the two most important cases).

Note that the expected value of distribution (14) is  $K + \frac{2A}{n}$ ; similarly expanding (1) then yields the formula for computing  $A$ , given the PDF of the sampled distribution.

Leaving out less important details, we now indicate the individual steps of such derivation.

- Extend the LHS of (4) by two more terms of the corresponding expansion, namely by

$$+ \frac{1}{2} \sum_{i=1}^n \left. \frac{\partial^3 \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^3} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^2 + \frac{1}{6} \sum_{i=1}^n \left. \frac{\partial^4 \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^4} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^3 + \dots \quad (17)$$

Note that the third and fourth derivatives of  $\ln f(\mathbf{x}_i; \boldsymbol{\theta})$  constitute fully symmetric tensors of ranks 3 and 4 respectively; each is further multiplied by the corresponding power (one less than the derivative's order) of  $\boldsymbol{\theta} - \boldsymbol{\theta}_0$ . The multiplication is carried out by taking the dot product along one of the tensor's indices with vector  $\boldsymbol{\theta} - \boldsymbol{\theta}_0$ , and repeating as many times as the power of  $\boldsymbol{\theta} - \boldsymbol{\theta}_0$  indicates; the result is thus always a vector with one remaining index.

- Solve the resulting equation (iteratively) for  $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ , to the same order of accuracy. This extends our existing solution (5), which we denote  $\boldsymbol{\Theta}_1$ , by the following quadratic term

$$\boldsymbol{\Theta}_2 = -\mu_2^{-1} \left( \mathbf{Y}_2 - \mu_2 + \frac{1}{2} \mu_3 \boldsymbol{\Theta}_1 \right) \boldsymbol{\Theta}_1 \quad (18)$$

and a cubic term, given by

$$\boldsymbol{\Theta}_3 = -\mu_2^{-1} \left( \frac{1}{2} (\mathbf{Y}_3 - \mu_3) \boldsymbol{\Theta}_1^2 + \mu_3 \boldsymbol{\Theta}_2 \boldsymbol{\Theta}_1 + \frac{1}{6} \mu_4 \boldsymbol{\Theta}_1^3 + (\mathbf{Y}_2 - \mu_2) \boldsymbol{\Theta}_2 \right) \quad (19)$$

where

$$\mathbf{Y}_\ell := \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial^\ell \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\ell} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad (20)$$

$$\mu_\ell := \mathbb{E} \left( \left. \frac{\partial^\ell \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\ell} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \quad (21)$$

(we have already explained how to multiply a symmetric tensor by a power or a product of vectors).

- Similarly expand (1); further divided by the sample size  $n$ , this results in the following *scalar* expression

$$\begin{aligned} & \mathbf{Y} \boldsymbol{\Theta}^{(1)} + \mathbf{Y} \boldsymbol{\Theta}_2 - \frac{1}{6} \mu_3 \boldsymbol{\Theta}_1^3 + \mathbf{Y} \boldsymbol{\Theta}_3 - \frac{1}{2} \mu_3 \boldsymbol{\Theta}_2 \boldsymbol{\Theta}_1^2 - \frac{1}{6} (\mathbf{Y}_3 - \mu_3) \boldsymbol{\Theta}_1^3 - \frac{1}{12} \mu_4 \boldsymbol{\Theta}_1^4 + \dots \\ &= \mathbf{Y} \boldsymbol{\Theta}^{(1)} + (\mathbf{Y}_2 - \mu_2) \boldsymbol{\Theta}_1^2 + \frac{1}{3} \mu_3 \boldsymbol{\Theta}_1^3 + \frac{1}{3} (\mathbf{Y}_3 - \mu_3) \boldsymbol{\Theta}_1^3 + \frac{1}{4} (\mu_3 \boldsymbol{\Theta}_1^2) \times (\mu_3 \boldsymbol{\Theta}_1^2) \quad (22) \\ & \quad + \frac{1}{12} \mu_4 \boldsymbol{\Theta}_1^4 + (\mu_3 \boldsymbol{\Theta}_1^2 + (\mathbf{Y}_2 - \mu_2) \boldsymbol{\Theta}_1) \times ((\mathbf{Y}_2 - \mu_2) \boldsymbol{\Theta}_1) + \dots \end{aligned}$$

where the  $\times$  symbol between two vectors (not used beyond this formula) implies an operation we call *contraction* of the two vectors, carried out by

$$\mathbf{V}_1 \times \mathbf{V}_2 := -\mathbf{V}_1 \mu_2^{-1} \mathbf{V}_2 \quad (23)$$

This is also how we perform a contraction of two specific indices of a tensor (required shortly); e.g. contracting the first two indices of  $\mu_3$  would be done (and denoted) by

$$\boldsymbol{\mu}_3 := \sum_{i,j=1}^K (\mu_3)_{i,j,n} (-\mu_2^{-1})_{i,j} \quad (24)$$

resulting in a vector.

- Finally, take the expected value of (22) to get  $\frac{K}{n} + \frac{2A}{n^2} + \dots$ .

This yields (term by term)

$$\begin{aligned} & \frac{K}{n} + \frac{\mu_{211} + K}{n^2} + \frac{\mu_3 \mu_{111}}{3n^2} + \frac{\mu_{31}}{n^2} + \frac{2\mu_3 \mu_3 + \mu_3 \mu_3}{4n^2} + \frac{\mu_4}{4n^2} \\ & + \frac{2\mu_3 \mu_{21} + \mu_3 \mu_{21} + \mu_{21} \mu_{21} + \mu_{21} \mu_{21} + \mu_{22} - K}{n^2} \end{aligned} \tag{25}$$

where

$$\mu_{21} := \mathbb{E} \left( \frac{\partial^2 \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \circ \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right) \tag{26}$$

with  $\mu_{211}, \mu_{111}, \mu_{22}$  and  $\mu_{31}$  defined analogously. Bold-face notation indicates that each term of (25) has been contracted in *all* of its indices, resulting in a scalar. For some of these terms, there are several possibilities to carry out such a contraction (e.g. the six indices of  $\mu_{21}\mu_{21}$  can be contracted in altogether five non-equivalent ways, only two of which we actually need); it is thus very important to indicate (in the ambiguous cases *only*; note that all pair-wise contractions of  $\mu_4$ , and also of  $\mu_{31}$ , yield the same answer) the proper way of doing it.

A dot implies contraction of two of the indicated indices ( $\mu_3$  thus making the result into a vector, while  $\mu_{211}$  implies, by the *absence* of a dot over 2, that the first two indices have been contracted with the third and fourth index, rather than with each other); similarly, two circles indicate that the corresponding two indices are to be contracted:  $\mu_{21}\mu_{21}$  thus makes the first index of  $\mu_{21}$  contract with its last index, returning a vector yet to be contracted with the *first* index of  $\mu_{21}$  (due to the lack of a dot over 2); similarly  $\mu_{21}\mu_{21}$  indicates contracting the last index of  $\mu_{21}$  with the first index of  $\mu_{21}$ , while the remaining two indices of  $\mu_{21}$  are contracted with the remaining two indices of  $\mu_{21}$ .

Since  $\mu_3 \mu_{111} = -\mu_3 \mu_3 - 2\mu_3 \mu_{21}$ , which can be verified in a manner of (9), the final formula for  $2A$  then reads

$$\mu_{211} + \mu_{31} + \mu_{22} + \frac{1}{4} \mu_4 + \frac{1}{4} \mu_3 \mu_3 + \frac{1}{6} \mu_3 \mu_3 + \mu_3 \mu_{21} + \mu_3 \mu_{21} + \mu_{21} \mu_{21} + \mu_{21} \mu_{21} \tag{27}$$

when  $K = 1$ , the formula simplifies to

$$A = \frac{\mu_{211} + \mu_{31} + \mu_{22} + \frac{1}{4} \mu_4}{2(-\mu_2)^2} + \frac{\frac{5}{12} \mu_3^2 + 2\mu_3 \mu_{21} + 2\mu_{21}^2}{2(-\mu_2)^3} \tag{28}$$

We should mention that, in principle, the  $\frac{1}{n}$ -proportional correction in (14) has two more terms of increasing complexity (see [5]), namely

$$\begin{aligned} & \chi_k^2(u) \left( 1 + \frac{A}{n} \cdot \left( \frac{u}{K} - 1 \right) + \frac{B}{n} \cdot \left( \frac{u^2}{K(K+2)} - \frac{2u}{K} + 1 \right) \right. \\ & \left. + \frac{C}{n} \cdot \left( \frac{u^3}{K(K+2)(K+4)} - \frac{3u^2}{K(K+2)} + \frac{3u}{K} - 1 \right) \right) \end{aligned} \tag{29}$$

where the  $B$  and  $C$  coefficients can be found in a manner very similar to how we have just established the value of  $A$ ; most surprisingly, both  $B$  and  $C$  then turn out to be equal to zero.

### 2.1. Location and Scaling Parameters

An important feature of the last two formulas is the fact that  $A$  may be a function of *shape* parameters (if any) of the sampled distribution, but not of a *location* or a *scaling* parameter, as these always cancel out from the resulting expression. To understand why, consider

$$f(x; \mu, \sigma) = \frac{f_0\left(\frac{x-\mu}{\sigma}\right)}{\sigma} := \frac{f_0(y)}{\sigma} \tag{30}$$

This enables us to simplify the two derivatives of  $\ln f(x; \mu, \sigma)$  thus

$$\frac{\partial(\ln f_0(y) - \ln \sigma)}{\partial \mu} = -\frac{f'_0(y)}{\sigma f_0(y)} \tag{31}$$

$$\frac{\partial(\ln f_0(y) - \ln \sigma)}{\partial \sigma} = -\frac{yf'_0(y)}{\sigma f_0(y)} - \frac{1}{\sigma} \tag{32}$$

from which it is clear that each further differentiation (whether with respect to  $\mu$  or  $\sigma$ ) yields a function of  $y$ , divided by an extra power of  $\sigma$ . Taking the expected value of any product of such derivatives amounts to further multiplying by  $f_0(y)dy$  and integrating over all possible values of  $y$ , leaving us with a number divided by power (equal to the derivative's total order) of  $\sigma$ . These powers of  $\sigma$  then always cancel out from (27).

This implies that, for distributions having only location and/or scaling parameters,  $A$  will turn out to be a *constant*.

### 2.2. Examples

Values of  $A$  are summarized in **Table 1** for some of the most common distributions.

For Gamma and Beta distributions,  $A$  is a complicated function of its first (Gamma) and both (Beta) parameter(s); yet, under the indicated constrains (see the above table), the value of  $A$  remains fairly constant.

For Gamma distribution (to illustrate the complexity of  $A$  as a function of a shape parameter), the full formula reads

$$A = \frac{\left[12\psi_1 + \alpha(16\psi_2 - 9\psi_1^2) + \alpha^2(2\psi_1^3 - 6\psi_1\psi_2 + 3\psi_3) + \alpha^3(5\psi_2^2 - 3\psi_1\psi_3)\right]^2}{288(\alpha\psi_1 - 1)^6} \tag{33}$$

where  $\psi_i$  is the  $i+1^{\text{st}}$  derivative of the natural logarithm of the  $\Gamma(\alpha)$  function, evaluated at  $\alpha$ . Yet, by plotting the last expression as a function of  $\alpha$ , we can confirm that, with the exception of only the smallest (say  $< 0.5$ ) values of  $\alpha$ ,  $A$  is nearly a constant, quickly approaching its  $\alpha \rightarrow \infty$  limit of  $\frac{121}{72}$ . We can



**Table 1.** Value of  $A$  for various distributions.

PDF/PMF	Distribution	$A$
$\frac{\exp\left(-\frac{x}{\beta}\right)}{\beta}$	Exponential	$\frac{1}{12}$
$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$	Normal	$\frac{11}{12}$
$\frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right)^2}$	Logistic	0.75866
$\frac{\sigma}{\pi\left((x-\mu)^2+\sigma^2\right)}$	Cauchy	1
$\frac{\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)-\frac{x-\mu}{\sigma}\right)}{\sigma}$	Gumbel	0.98915
$\frac{x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)}{\Gamma(\alpha)\beta^\alpha}$	Gamma	$\frac{121}{72} (\alpha > 0.5)$
$\frac{\Gamma(\alpha+\beta)\cdot x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}$	Beta	$\frac{11}{12} (\alpha, \beta > 0.05)$
$\frac{x\exp\left(-\frac{x^2}{2\theta}\right)}{\theta}$	Rayleigh	$\frac{1}{12}$
$\frac{\exp\left(-\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)}{2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$	Bi-normal	$\frac{37}{12}$
$p(1-p)^x$	Geometric	$\frac{1}{12(1-p)} - \frac{p}{12}$

make similar conclusion in the case of Beta distribution, as mentioned already. Note that, in the case of bivariate Normal distribution,  $A$  is a constant, free not only of the two location and the two scaling parameters, but also of the shape parameter  $\rho$ .

Incorporating the  $A$ -related correction requires only a trivial modification of the previous program; in the Geometric example, we need to replace the “CL = ...” line by the following *two* lines:

```
A = 1/6/(1 - p) - p/6 /. max[[2]];
CL = u /. FindRoot[Erf[\sqrt{u/2}] - A/n\sqrt{2u/\pi}Exp[-u/2]==0.9, {u, 1}];
```

Note that, to compute  $A$ , instead of using the true but unknown value of  $p$ , we

had to use its point estimate instead.

In the Gamma example (since the  $\alpha$  estimate was bigger than 0.5), we can use the large- $\alpha$  limit of  $A$ ; the “CI = ...” line now has to be replaced by

$$CL = u /. \text{FindRoot}[1 - \text{Exp}[-u/2] - 121/36 u/2/n \text{Exp}[-u/2] == 0.95, \{u, 1\}];$$

### 3. Hypotheses Testing

The same formulas can be utilized for high-accuracy testing of a null hypothesis, claiming that parameters of distribution have specific values. Thus, for example, when the distribution has two parameters, we first need to find their ML estimates (based on a random independent sample of size  $n$ ); using these we then evaluate (1), where  $\theta_0$  now stands for the parameter values as specified by the null hypothesis, and substitute the result for  $u$  in (16); this yields the confidence level of rejecting the null hypothesis—subtracting this from 1 then returns the corresponding  $P$ -value.

The following Mathematica program illustrates how it is done, using the Cauchy distribution and testing whether its two parameters ( $m$  and  $s$ , say) have true values equal to 0 and 1 respectively (our random sample of 40 independent observations has been generated using  $m = 0.3$  and  $s = 0.8$ ).

```
(* Cauchy-distribution example *)
X = RandomVariate[CauchyDistribution[0.3, 0.8], 40];
n = Length[X];
f = s/π/((x - m)2 + s2);
A = 1;
L = Log[f] // PowerExpand;
lnLF = Apply[Plus, L /. x -> X];
res = NMaximize[{lnLF, s > 0}, {m, s}];
u = 2 (res[[1]] - lnLF /. {m -> 0, s -> 1});
P = Exp[-u/2](1 + A u/2/n)
0.00733479
```

Based on the resulting  $P$  value, the null hypothesis would be rejected at 1% level of significance.

In our second example, we test whether parameters of a bivariate Normal distribution have the following values:  $\mu_1 = \mu_2 = 0$ ,  $\sigma_1 = \sigma_2 = 1$  and  $\rho = 0.5$  (these are also the values we used to generate the sample of 100 bivariate observations).

```
(* Bi-Normal-distribution example *)
n = Length[X];
L[x_] := -((x[[1]] - μ1)2/σ12 + (x[[2]] - μ2)2/σ22 - 2ρ(x[[1]] - μ1)(x[[2]] - μ2)/σ1σ2) / (2/(1 - ρ2) - Log[σ1σ2√(1 - ρ2)]);
A = 37/12;
lnLF = Apply[Plus, Map[L, X]];
res = NMaximize[{lnLF, σ1 > 0, σ2 > 0, -1 < ρ < 1}, {μ1, μ2, σ1, σ2, ρ}];
u = 2 (res[[1]] - lnLF /. {μ1 -> 0, μ2 -> 0, σ1 -> 1, σ2 -> 1, ρ -> 0.5});
P = (Gamma[5/2, u/2] + 2/5 A/n (u/2)5/2 Exp[-u/2])/Gamma[5/2]
0.240079
```

The resulting  $P$  value would lead to accepting the null hypothesis at any practical level of significance.

### 4. Accuracy Improvement

To investigate the accuracy of the new approximation, we need to be able to compute the level of confidence of the resulting region *exactly*. This can be done only in a few situations, such as the Exponential and Normal distributions. Exploring these two cases will give us a good indication of the improvement achieved by incorporating the  $A$  correction in general. The details of computing the exact confidence level may be of only passing interest to most readers; it is the resulting error tables which are important.

#### 4.1. Exponential Case

We know that the ML estimator of the  $\beta$  parameter is given by the sample mean  $\bar{X}$  whose exact distribution is  $\text{Gamma}(n, \beta/n)$ . Since

$$L(\mathbf{X}; \beta) = -\frac{\sum_{i=1}^n X_i}{\beta} - n \ln \beta \tag{34}$$

then

$$2L(\mathbf{X}; \bar{X}) - 2L(\mathbf{X}; \beta) = -2n \left( 1 + \ln \frac{\bar{X}}{\beta} - \frac{\bar{X}}{\beta} \right) \tag{35}$$

(note that this expression is always non-negative, reaching a minimum value of zero when  $\bar{X} = \beta$ ). The limits of the confidence interval of the basic technique are found by making (35) equal to a critical value of the  $\chi^2_1$  distribution (let us call it  $C_\tau$ , where  $\tau$  is the corresponding level of confidence) and solving for  $\bar{X}/\beta$  (and then, trivially, for  $\beta$ ).

One can show that the two real solutions are given by

$$-W_0 \left( -\exp \left( -1 - \frac{C_\tau}{2n} \right) \right) \text{ and } -W_{-1} \left( -\exp \left( -1 - \frac{C_\tau}{2n} \right) \right) \tag{36}$$

where  $W$  is the Lambert (also called Omega, or “ProductLog” by Mathematica) function and the subscripts refer to its two real branches. To find the true confidence level, we need to integrate the PDF of  $\bar{X}/\beta$  between these two limits (our code utilizes the corresponding CDF instead); subtracting the result from  $\tau$  establishes the error of the basic algorithm. Further subtracting

$$\frac{\sqrt{C_\tau} \exp \left( -\frac{C_\tau}{2} \right)}{6n\sqrt{2\pi}} \tag{37}$$

then converts it into the error of our new formula (15). The following Mathematica program does that in four simple lines.

```
(* CI errors; Exponential distribution *)
n=3;
c=inverseCDF[ChiSquareDistribution[1],0.95];
```

$$\begin{aligned}
 & 0.95 - (\text{Gamma}[n, -\text{ProductLog}[0, -\text{Exp}[1-c/2/n]]n] \\
 & \quad - \text{Gamma}[n, -\text{ProductLog}[-1, -\text{Exp}[1-c/2/n]], n]) / (n-1)! \\
 & 0.00631436 \\
 & \quad \% - \sqrt{c} \text{Exp}[-c/2] / 6/n / \sqrt{2\pi} \\
 & -0.0000495375
 \end{aligned}$$

Both errors (using various sample sizes) are displayed in **Table 2**, using  $\tau = 95\%$ .

The table demonstrates that, as  $n$  increases, the error decreases with the first power of  $\frac{1}{n}$  for the basic approximation, and with the second power of  $\frac{1}{n}$  for our new algorithm.

### 4.2. Normal Case

Similarly, it is well known that the ML estimators of  $\mu$  and  $\sigma^2$  of the Normal distribution are  $\bar{X}$  and  $s^2 := \bar{X}^2 - \bar{X}^2$  respectively, that they are independent of each other, and that their respective distributions are  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$  and  $\text{Gamma}\left(\frac{n-1}{2}, 2\sigma^2\right)$ . Since

$$L(\mathbf{X}; \mu, \sigma^2) = -\frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2} - \frac{n}{2} \ln \sigma^2 - n \ln \sqrt{2\pi} \tag{38}$$

(1) then becomes

$$2L(\mathbf{X}; \bar{X}, s^2) - 2L(\mathbf{X}; \mu, \sigma^2) = -n \left( 1 + \ln \frac{s^2}{\sigma^2} - \frac{s^2}{\sigma^2} - \frac{(\bar{X} - \mu)^2}{\sigma^2} \right) \tag{39}$$

To get the exact confidence level of the basic algorithm, we need to make this expression less than the critical value (denoted  $C_\tau$ , as in the previous example) of the  $\chi^2_2$  distribution, and integrate the joint PDF of  $s^2/\sigma^2$  and  $\sqrt{n}(\bar{X} - \mu)/\sigma$  over the resulting region. Solving the equation of its boundary for  $s^2/\sigma^2$  in the manner of (35) now yields

$$-W_0\left(-\exp\left(-1 - \frac{C_\tau - Z^2}{n}\right)\right) \text{ and } -W_{-1}\left(-\exp\left(-1 - \frac{C_\tau - Z^2}{n}\right)\right) \tag{40}$$

where  $Z$  stands for  $\sqrt{n}(\bar{X} - \mu)/\sigma$ , having the standard Normal distribution.

**Table 2.** Approximation errors (Exponential case).

$n$	basic approximation	with $A$ correction
3	0.63%	$-4.9 \times 10^{-5}$
10	$1.9 \times 10^{-3}$	$1.5 \times 10^{-7}$
30	$6.4 \times 10^{-4}$	$1.7 \times 10^{-7}$
100	$4.9 \times 10^{-5}$	$2.0 \times 10^{-8}$
300	$6.4 \times 10^{-5}$	$2.4 \times 10^{-9}$

We then integrate the PDF of  $s^2/\sigma^2$  (having the  $\chi^2_{n-1}$  distribution) using these two limits (again, the program utilizes the corresponding CDF instead); finally, we multiply the answer by the PDF of  $Z$  and integrate over all marginal values of  $Z$  (i.e. from 0 to  $\sqrt{C_\tau}$ )—the last step is done numerically. This yields the *true* confidence level of the resulting confidence region; subtracting it from  $\tau$  then provides the corresponding error of the basic algorithm; further subtracting

$$\frac{C_\tau \exp\left(-\frac{C_\tau}{2}\right)}{24n} \tag{41}$$

converts that into the error of the new algorithm. This is how it is done using Mathematica:

```
(* Confidence-region errors, Normal distribution *)
n = 300;
c = InverseCDF[ChiSquareDistribution[2], .95];
int = (Gamma[(n-1)/2, -n/2 ProductLog[-Exp[-(c+n-q)/n]]] -
Gamma[(n-1)/2, -n/2 ProductLog[-1, -Exp[-(c+n-q)/n]]])/Gamma[(n-1)/2];
0.95 - (true = NIntegrate[2 Exp[-z^2/2]/sqrt(2pi) (int /. q -> z^2), {z, 0, sqrt(c]}])
0.000460117
NIntegrate[Exp[-u/2]/2 (1 + (u/2 - 1) 11/12/n), {u, 0, c}] - true
2.43546x10^-6
```

**Table 3** summarizes the results.

**Table 3.** Approximation errors (Normal case).

$n$	basic approximation	with $A$ correction
3	8.1%	3.6%
10	1.6%	0.24%
30	$4.8 \times 10^{-3}$	$2.5 \times 10^{-4}$
100	$1.4 \times 10^{-3}$	$2.2 \times 10^{-5}$
300	$4.6 \times 10^{-4}$	$2.4 \times 10^{-6}$

Similarly to the previous example, one can clearly discern the  $\frac{1}{n}$ -proportional decrease of the basic error, and the more rapid  $\frac{1}{n^2}$ -proportional decrease achieved by our correction.

### 5. Conclusion

In this article, we have modified the traditional way (based on the corresponding likelihood function and a random independent sample of size  $n$ ) of constructing a confidence region for distribution parameters, aiming to reduce its error. This has been achieved by adding, to the usual  $\chi^2$  distribution of the original tech-

nique, a simple,  $\frac{1}{n}$ -proportional correction whose coefficient is computed, for each specific distribution, using a rather complex formula; we have done the latter for many common distributions, summarizing the results in **Table 1**. We have also provided several examples of a Mathematica code, demonstrating the new algorithm and its simplicity (each computer program takes only a few lines); a practically identical approach can then be applied to the task of hypotheses testing. Finally, we have numerically investigated the improvement in accuracy thereby achieved, finding it rather substantial. It is worth emphasizing that, once the value of  $A$  has been established, incorporating the corresponding correction is extremely easy and done with a negligible computational cost.

Future research will hopefully explain why are the  $B$  and  $C$  coefficients of (29) both equal to zero; so far we have been able to prove that only by a brute-force computation, but we believe that there is some more fundamental reason behind this result, which we have not been able to uncover.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Kendall, M.G. and Stuart, A. (1969) *The Advanced Theory of Statistics*, Vol. 1, Chapter 3. Hafner Publishing Company, New York.
- [2] Bartlett, M.S. (1953) Approximate Confidence Intervals. *Biometrika*, **40**, 12-19. <https://doi.org/10.1093/biomet/40.1-2.12>
- [3] Cysneiros, F.J.deA., dos Santos, S.J.P. and Cordeiro, G.M. (2001) Skewness and Kurtosis for Maximum Likelihood Estimator in One-Parameter Exponential Family Models. *Brazilian Journal of Statistics and Probability*, **15**, 85-108.
- [4] Bowman, K.O. and Shenton, L.R. (1998) Aymptotic Skewness and the Distribution of Maximum Likelihood Estimators. *Communications in Statistics—Theory and Methods*, **27**, 2743-2760. <https://doi.org/10.1080/03610929808832252>
- [5] Vrbik, J. (2013) Accurate Confidence Regions Based on MLEs. *Advances and Applications in Statistics*, **32**, 33-56.
- [6] Vrbik, J. (2013) Addendum and Errata to “Accurate Confidence Regions Based on MLEs”. *Advances and Applications in Statistics*, **35**, 161-163.