

Confidence Regions with Nuisance Parameters

Jan Vrbik

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

Email: jvr bik@brocku.ca

How to cite this paper: Vrbik, J. (2022) Confidence Regions with Nuisance Parameters. *Open Journal of Statistics*, 12, 658-675. <https://doi.org/10.4236/ojs.2022.125039>

Received: August 30, 2022

Accepted: October 15, 2022

Published: October 18, 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

Consider a distribution with several parameters whose exact values are unknown and need to be estimated using the maximum-likelihood technique. Under a regular case of estimation, it is fairly routine to construct a confidence region for *all* such parameters, based on the natural logarithm of the corresponding likelihood function. In this article, we investigate the case of doing this for only *some* of these parameters, assuming that the remaining (so called nuisance) parameters are of no interest to us. This is to be done at a chosen level of confidence, maintaining the usual accuracy of this procedure (resulting in about 1% error for samples of size $n = 30$, and further decreasing with $1/n$). We provide a general solution to this problem, demonstrating it by many explicit examples.

Keywords

Confidence Regions, Maximum Likelihood, Nuisance Parameters, Asymptotic Distribution

1. Introduction

There is a basic technique (expounded in detail by M. S. Bartlett—see [1] [2] and [3]—nicely summarized in [4]) for constructing confidence regions (intervals) for parameters of a specific distribution (assuming a *regular* case, meaning the distribution's support is *not* a function of any of the distribution's parameters) which rests on the fact that

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

has approximately the chi-square distribution with K degrees of freedom (K is

the number of parameters to be estimated), where

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

\mathbf{X} is the set of n observations, individually denoted \mathbf{x}_i (allowing for a possibility of a multivariate distribution), $f(\mathbf{x}; \boldsymbol{\theta})$ denotes the corresponding probability density function, $\hat{\boldsymbol{\theta}}$ is the vector of the resulting maximum-likelihood (ML) estimators of the parameters, and $\boldsymbol{\theta}_0$ represents their true (even though unknown) values.

The **proof** rests on expanding the LHS of following K -component equation

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

with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ to a linear (in $\boldsymbol{\theta} - \boldsymbol{\theta}_0$) accuracy, making the answer equal to $\mathbf{0}$ and solving for $\boldsymbol{\theta}$, thereby getting

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \simeq \mathbb{M}^{-1} \mathbf{Y} \tag{4}$$

where

$$\mathbf{Y} := \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \ln f(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{5}$$

and

$$\mathbb{M} := -\mathbb{E} \left(\left. \frac{\partial^2 \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \equiv \mathbb{E} \left(\left. \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \circ \frac{\partial \ln f(\mathbf{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \tag{6}$$

Note that \mathbf{Y} is a K -component vector, while \mathbb{M} represents a symmetric, positive-definite K by K matrix (the small circle stands for a direct product of two vectors).

Similarly expanding (1) and utilizing (4) we get

$$2n\mathbf{Y}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbb{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \dots = \sqrt{n}\mathbf{Y}^T \mathbb{M}^{-1} \mathbf{Y} \sqrt{n} \tag{7}$$

where $\mathbf{Y}\sqrt{n}$ has (by Central Limit Theorem), approximately, a K -variate Normal distribution with the mean of $\mathbf{0}$ and the variance-covariance matrix of \mathbb{M} ; this implies that (7) has, to the same level of approximation, the χ_K^2 distribution (since the components of $\mathbb{M}^{-1/2} \mathbf{Y}\sqrt{n}$ are then asymptotically independent, each having the mean of 0 and the variance of 1).

From all this, it then follows that an approximate confidence region is found by first finding the maximum likelihood estimators $\hat{\boldsymbol{\theta}}$, then making (1) equal to a critical χ_K^2 value and solving (usually, only graphically) for $\boldsymbol{\theta}_0$.

We demonstrate this by generating a random independent sample of 200 observations from a Negative Binomial distribution (with parameters α and p) and constructing the corresponding 90% confidence region for the two parameters; a simple, self-explanatory Mathematica code to do exactly that looks like this:

```

X = RandomVariate[NegativeBinomialDistribution[1.7, .4], 200];
lnf[x_] := x Log[1 - p] +  $\alpha$  Log[p] + LogGamma[x +  $\alpha$ ] - LogGamma[ $\alpha$ ]
- LogGamma[x + 1]
lnLF = Apply[Plus, Map[lnf, X]] // Together;
res = NMaximize[{lnLF,  $\alpha$  > 0, 0 < p < 1}, { $\alpha$ , p}]
{-417.369, { $\alpha$  -> 1.74844, p -> 0.40629}}

c = InverseCDF[ChiSquareDistribution[2], 0.90];
ContourPlot[2 (res[[1]] - lnLF) == c, { $\alpha$ , 1.1, 2.7}, {p, .3, .52},
ContourStyle -> Black]

```

The resulting confidence region is displaced in **Figure 1**.

Similarly, to test a null hypothesis which claims specific values for the K parameters, we evaluate (1) with θ_0 being the *hypothesized* (rather than the true) values, and check the result against the critical value of χ_k^2 ; something this article will not elaborate on any further.

2. Partial Confidence Regions

The aim of this article is to show how to construct a confidence region (called *partial*) for only *some* parameters of distribution, even though *all* of its parameters are unknown and need to be estimated by the maximum-likelihood technique.

We should mention that there is some existing literature on using partial likelihood functions (LF), for example [5], but its goals and results bear little resemblance to ours. Similarly, articles on marginal LF (for example [6]) and conditional LF deal with only rather specialized issues while our approach is fully general; the only shared feature is the occasional use of identical terminology.

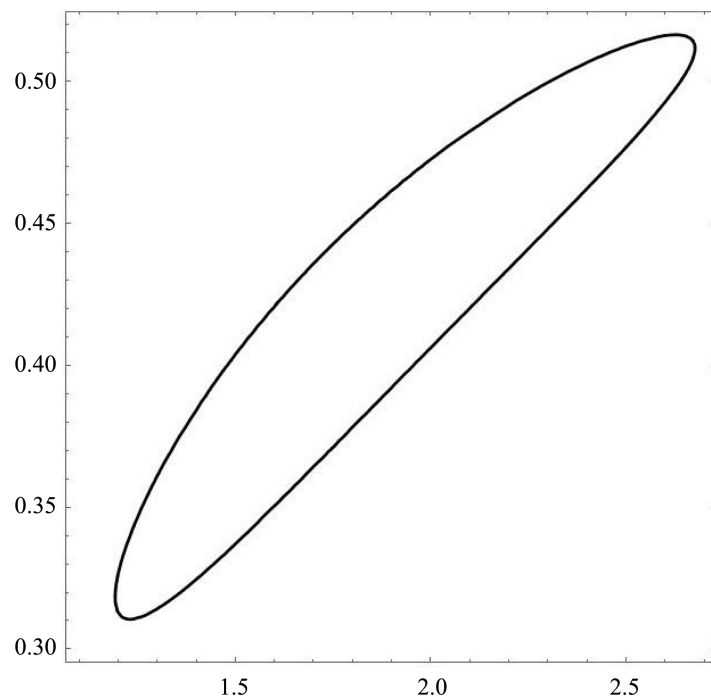


Figure 1. 90% confidence region for α and p .

Extending the technique delineated in our Introduction, we now need to find an approximate distribution of

$$2 \ln \mathcal{L}(\mathbf{X}; \hat{\boldsymbol{\theta}}) - 2 \ln \mathcal{L}(\mathbf{X}; \hat{\boldsymbol{\Theta}}_0) \tag{8}$$

where only some components of $\hat{\boldsymbol{\Theta}}_0$ are equal to the true values of the corresponding (we call them *pivotal*) parameters, while the rest (the *nuisance* parameters; a term introduced by [7] and further explored by [8]) are set to their $\hat{\boldsymbol{\theta}}$ values. Knowing this distribution will then enable us to construct confidence regions (or test hypotheses) for the pivotal parameters only while ignoring the estimates of the nuisance parameters.

Since we already have a good approximation to (1), namely (7), or equivalently

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbb{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sqrt{n} \tag{9}$$

we now need a similar approximation for

$$2 \ln L(\mathbf{X}; \hat{\boldsymbol{\Theta}}_0) - 2 \ln L(\mathbf{X}; \boldsymbol{\theta}_0) \tag{10}$$

and then, for the corresponding difference. To approximate (10), we go back to the LHS of (7) and replace $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\Theta}}_0 - \boldsymbol{\theta}_0$ (i.e. keeping the nuisance components of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ and setting the pivotal components to 0) while $\mathbf{Y} = \mathbb{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, computed from (4), remains unchanged. This results in

$$\begin{aligned} & 2n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbb{M}(\hat{\boldsymbol{\Theta}}_0 - \boldsymbol{\theta}_0) - n(\hat{\boldsymbol{\Theta}}_0 - \boldsymbol{\theta}_0)^\top \mathbb{M}(\hat{\boldsymbol{\Theta}}_0 - \boldsymbol{\theta}_0) \\ & := n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbb{M}_0(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \end{aligned} \tag{11}$$

where \mathbb{M}_0 is the original \mathbb{M} matrix with all pivotal-by-pivotal elements set to 0 (a notation to be used with other matrices as well).

This can be shown by rearranging the parameters to start with the pivotal and be followed by the nuisance ones, and visualizing the corresponding 2 by 2 block structure of the *symmetric* matrix \mathbb{M} . In such representation, the previous equation reads

$$\begin{aligned} & n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \begin{bmatrix} \textcircled{0} & \blacksquare \\ \textcircled{0} & \blacksquare \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \begin{bmatrix} \textcircled{0} & \textcircled{0} \\ \blacksquare & \blacksquare \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ & - n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \begin{bmatrix} \textcircled{0} & \textcircled{0} \\ \textcircled{0} & \blacksquare \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \begin{bmatrix} \textcircled{0} & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \end{aligned}$$

where a full square \blacksquare indicates keeping the original block of the \mathbb{M} matrix, while each $\textcircled{0}$ represents a zero sub-matrix of the corresponding dimensions.

Subtracting (11) from (9) then yields the desired approximation to (8), namely

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top (\mathbb{M} - \mathbb{M}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sqrt{n} \tag{12}$$

Introducing $\mathbf{U} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sqrt{n}$, which we know from our Introduction to be approximately K -variate Normal, having zero means and the variance-covariance matrix of \mathbb{M}^{-1} , we can now find the moment generating function (MGF) of (12)

by

$$\frac{\int \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{\mathbf{u}^T \mathbb{M} \mathbf{u}}{2} + t \mathbf{u}^T (\mathbb{M} - \mathbb{M}_0) \mathbf{u}\right) d\mathbf{u}}{(2\pi)^{K/2} \sqrt{\det(\mathbb{M}^{-1})}} \tag{13}$$

$$= \sqrt{\frac{1}{\det(\mathbb{I} - 2t\mathbb{I} + 2t\mathbb{M}_0\mathbb{M}^{-1})}} = \sqrt{\frac{1}{\det(\mathbb{I} - 2t\mathbb{I} + 2t\mathbb{M}(\mathbb{M}^{-1})_0)}}$$

where \mathbb{I} is the K by K identity matrix. The last equality follows from the fact that

$$\mathbb{M}_0\mathbb{M}^{-1} = \begin{bmatrix} \mathbb{M}_{\mathbf{p},\mathbf{i}}(\mathbb{M}^{-1})_{\mathbf{i},\mathbf{p}} & \cdots \\ \mathbb{O} & \mathbb{I} \end{bmatrix} = \begin{bmatrix} \mathbb{I} - \mathbb{M}_{\mathbf{p},\mathbf{p}}(\mathbb{M}^{-1})_{\mathbf{p},\mathbf{p}} & \cdots \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \tag{14}$$

and

$$\mathbb{M}(\mathbb{M}^{-1})_0 = \begin{bmatrix} \mathbb{M}_{\mathbf{p},\mathbf{i}}(\mathbb{M}^{-1})_{\mathbf{i},\mathbf{p}} & \mathbb{O} \\ \cdots & \mathbb{I} \end{bmatrix} \tag{15}$$

make the same contribution to the determinant in (13), where \mathbb{I} is now the L by L identity matrix (L being the number of pivotal parameters). In (14), we also use the fact that

$$\mathbb{M}_{\mathbf{p},\mathbf{i}}(\mathbb{M}^{-1})_{\mathbf{i},\mathbf{p}} + \mathbb{M}_{\mathbf{p},\mathbf{p}}(\mathbb{M}^{-1})_{\mathbf{p},\mathbf{p}} = \mathbb{I} \tag{16}$$

where the \mathbf{p} and \mathbf{i} subscripts refer to the corresponding pivotal and/or nuisance block of the matrix.

It is easy to see that the result of (13) does not change after replacing \mathbb{M} (the variance-covariance matrix of $\sqrt{n}\mathbf{Y}$) by the corresponding correlation matrix $\mathbb{C} = \mathbb{D}^{-1/2}\mathbb{M}\mathbb{D}^{-1/2}$, where \mathbb{D} is the main-diagonal matrix of the corresponding variances, and correspondingly replacing \mathbb{M}_0 by $\mathbb{C}_0 = (\mathbb{D}^{-1/2}\mathbb{M}_0\mathbb{D}^{-1/2})_0 = \mathbb{D}^{-1/2}\mathbb{M}_0\mathbb{D}^{-1/2}$ (recall that the 0 subscript indicates setting all pivotal-pivotal elements equal to 0), since clearly

$$\det(\mathbb{I} - 2t\mathbb{I} + 2t\mathbb{M}_0\mathbb{M}^{-1}) = \det(\mathbb{I} - 2t\mathbb{I} + 2t\mathbb{C}_0\mathbb{C}^{-1}) \tag{17}$$

Similarly, we can replace the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, namely \mathbb{M}^{-1} , by its correlation matrix $\tilde{\mathbb{C}} = \tilde{\mathbb{D}}^{-1/2}\mathbb{M}^{-1}\tilde{\mathbb{D}}^{-1/2}$ and \mathbb{M}_0 by $(\tilde{\mathbb{C}}^{-1})_0 = \tilde{\mathbb{D}}^{1/2}\mathbb{M}_0\tilde{\mathbb{D}}^{1/2}$ without affecting the value of the determinant.

Summary

Based on (14), the MGF of (12) is given by

$$\prod_{\ell=1}^L \frac{1}{\sqrt{1 - 2t \cdot e_{\ell}}} \tag{18}$$

where e_1, e_2, \dots, e_L are the eigenvalues of $\mathbb{M}_{\mathbf{p},\mathbf{p}}(\mathbb{M}^{-1})_{\mathbf{p},\mathbf{p}}$ (note that \mathbb{M} can be replaced by \mathbb{C} or $\tilde{\mathbb{C}}$, whichever is more convenient). The resulting PDF is then that of a convolution of the individual Gamma $(\frac{1}{2}, 2e_{\ell})$ distributions.

It is important to note that, when there are more pivotal than nuisance parameters (*i.e.* when $L > K - L$), the L by L matrix $\mathbb{M}_{p,i}(\mathbb{M}^{-1})_{i,p}$ in (16) is of rank $K - L$ only (this is now determined by the number of *columns* of $\mathbb{M}_{p,i}$), implying that $L - (K - L) = 2L - K$ of its eigenvalues are equal to 0, and correspondingly simplifying the eigenvalues of $\mathbb{M}_{p,p}(\mathbb{M}^{-1})_{p,p}$ ($2L - K$ of which will be equal to 1). Furthermore, the remaining eigenvalues of $\mathbb{M}_{p,i}(\mathbb{M}^{-1})_{i,p}$ are the same as those of $\mathbb{M}_{i,p}(\mathbb{M}^{-1})_{p,i}$; this is based on a general result stating that, when \mathbb{A} is an n by m matrix while \mathbb{B} is m by n , $\mathbb{A}\mathbb{B}$ and $\mathbb{B}\mathbb{A}$ will share all their non-zero eigenvalues, while the extra eigenvalues, if any, will be equal to 0.

This implies that, when $L > K - L$, we can replace the eigenvalues of $\mathbb{M}_{p,p}(\mathbb{M}^{-1})_{p,p}$ by the eigenvalues of $\mathbb{M}_{i,i}(\mathbb{M}^{-1})_{i,i}$ (a smaller matrix), knowing that each of the remaining $2L - K$ eigenvalues is equal to 1.

The final note: some steps of the resulting procedure for constructing partial confidence regions may be carried out analytically, while the rest require a numerical approach. This can be observed in our subsequent examples: some avoid explicit formulas entirely, performing each step numerically (always an available option), while our last example is almost completely analytical, to facilitate investigation of the technique’s accuracy.

3. Multivariate Normal Distribution

The most important multi-parameter distribution is the Normal distribution of several (say n) random variables X_1, X_2, \dots, X_n , collectively denoted \mathbf{X} . It is fully specified by the following parameters: n individual means (collectively denoted $\boldsymbol{\mu}$), n standard deviations $\boldsymbol{\sigma}$, and $\binom{n}{2}$ correlation coefficients ρ_{ij} ($1 \leq i < j \leq n$) usually collected in a symmetric matrix \mathbb{R} (each ρ_{ij} will appear twice, on both sides of the main diagonal, whose elements are all equal to 1). This section is a review of basic formulas relating to ML estimation of these parameters.

The natural logarithm of the corresponding PDF (aka likelihood function) is given by

$$-\frac{(\mathbf{x} - \boldsymbol{\mu})^T \mathbb{S}^{-1} \mathbb{R}^{-1} \mathbb{S}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2} - \frac{1}{2} \ln \det \mathbb{R} - \ln \det \mathbb{S} - \frac{n}{2} \ln(2\pi) \tag{19}$$

where \mathbb{S} is a main-diagonal matrix of the n values of σ . Note that $\mathbb{S}^{-1}(\mathbf{x} - \boldsymbol{\mu})$, explicitly expanded, yields the following vector $\left\langle \frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}, \dots, \frac{x_n - \mu_n}{\sigma_n} \right\rangle$.

To find the corresponding \mathbb{M} matrix, we use the following well-known formulas

$$\frac{d\mathbb{R}^{-1}}{d\rho_{kl}} = -\mathbb{R}^{-1} \frac{d\mathbb{R}}{d\rho_{kl}} \mathbb{R}^{-1} \tag{20}$$

$$\frac{d \ln \det \mathbb{R}}{d\rho_{kl}} = \text{Tr} \left(\frac{d\mathbb{R}}{d\rho_{kl}} \mathbb{R}^{-1} \right) \tag{21}$$

where Tr indicates taking the matrix' trace. Note that all but two elements of

$$\frac{d\mathbb{R}}{d\rho_{kl}} := \mathbb{H}^{[kl]} = \mathbf{v}^{[k]} \circ \mathbf{v}^{[l]} + \mathbf{v}^{[l]} \circ \mathbf{v}^{[k]} \tag{22}$$

are equal to 0 (the remaining two elements are equal to 1), where $\mathbf{v}^{[k]}$ stands for a vector of $n-1$ zeros with only its k^{th} component equal to 1.

Differentiating (19) twice with respect to $\boldsymbol{\mu}$ and changing the sign yields

$$\mathbb{S}^{-1}\mathbb{R}^{-1}\mathbb{S}^{-1} \tag{23}$$

which represents the $\boldsymbol{\mu}$ by $\boldsymbol{\mu}$ block of \mathbb{M} , while the $\boldsymbol{\mu}$ by $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ by ρ_{ij} blocks are both zero sub-matrices, since $\mathbb{E}(\mathbf{x}-\boldsymbol{\mu}) = \mathbf{0}$ —that goes for the $\boldsymbol{\sigma}$ by $\boldsymbol{\mu}$ and ρ_{ij} by $\boldsymbol{\mu}$ blocks as well.

To find the $\boldsymbol{\sigma}$ by $\boldsymbol{\sigma}$ block, we first differentiate (19) with respect to σ_i and then with respect to σ_j (assuming $i \neq j$), getting

$$-\frac{(x-\mu)_i(\mathbb{R}^{-1})_{ij}(x-\mu)_j + (x-\mu)_j(\mathbb{R}^{-1})_{ij}(x-\mu)_i}{2\sigma_i^2\sigma_j^2} \tag{24}$$

Reversing the sign and taking the expected value yields the following expression for the off-diagonal elements of the $\boldsymbol{\sigma}$ by $\boldsymbol{\sigma}$ block

$$\frac{\mathbb{R}_{ij}(\mathbb{R}^{-1})_{ij}}{\sigma_i\sigma_j} \tag{25}$$

since $\mathbb{E}((x_i - \mu_i)(x_j - \mu_j)) = \sigma_i\sigma_j\mathbb{R}_{ij}$.

When differentiating with respect of σ_i *twice*, the corresponding second derivative is

$$-\sum_{j=1}^n \frac{(x-\mu)_i(\mathbb{R}^{-1})_{ij}(x-\mu)_j + (x-\mu)_j(\mathbb{R}^{-1})_{ij}(x-\mu)_i}{\sigma_i^3\sigma_j} - \frac{(x-\mu)_i(\mathbb{R}^{-1})_{ii}(x-\mu)_i}{\sigma_i^4} + \frac{1}{\sigma_i^2} \tag{26}$$

Reversing the sign and taking the expected value then yields

$$2\sum_{j=1}^n \frac{(\mathbb{R}^{-1})_{ij}\mathbb{R}_{ji}}{\sigma_i^2} + \frac{\mathbb{R}_{ii}(\mathbb{R}^{-1})_{ii}}{\sigma_i^2} - \frac{1}{\sigma_i^2} = \frac{\mathbb{R}_{ii}(\mathbb{R}^{-1})_{ii}}{\sigma_i^2} + \frac{1}{\sigma_i^2} \tag{27}$$

for the main-diagonal elements of the $\boldsymbol{\sigma}$ by $\boldsymbol{\sigma}$ block (note that $\sum_{j=1}^n (\mathbb{R}^{-1})_{ij}\mathbb{R}_{ji} = \mathbb{I}_{ii} = 1$).

This means that

$$\frac{\mathbb{R}_{ij}(\mathbb{R}^{-1})_{ij} + \mathbb{I}_{ij}}{\sigma_i\sigma_j} \tag{28}$$

is the resulting expression for both the diagonal and off-diagonal elements of the $\boldsymbol{\sigma}$ by $\boldsymbol{\sigma}$ block.

To find the $\rho_{kl} - \rho_{mp}$ element of the \mathbb{M} matrix, we first need the corresponding second derivative of (19), namely

$$-\frac{2(\mathbf{x}-\boldsymbol{\mu})^T \mathbb{S}^{-1} \mathbb{R}^{-1} \mathbb{H}^{[kl]} \mathbb{R}^{-1} \mathbb{H}^{[mp]} \mathbb{R}^{-1} \mathbb{S}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2} + \frac{1}{2} \text{Tr} \left(\mathbb{H}^{[kl]} \mathbb{R}^{-1} \mathbb{H}^{[mp]} \mathbb{R}^{-1} \right) \quad (29)$$

Reversing the sign and taking the expected value results in

$$\frac{1}{2} \text{Tr} \left(\mathbb{H}^{[kl]} \mathbb{R}^{-1} \mathbb{H}^{[mp]} \mathbb{R}^{-1} \right) = \left(\mathbb{R}^{-1} \right)_{km} \left(\mathbb{R}^{-1} \right)_{lp} + \left(\mathbb{R}^{-1} \right)_{kp} \left(\mathbb{R}^{-1} \right)_{lm} \quad (30)$$

Finally, to get the $\sigma_i - \rho_{kl}$ element of \mathbb{M} , the corresponding differentiation of (19) yields

$$-\frac{(x-\mu)_i \mathbb{S}^{-1} \mathbb{R}^{-1} \mathbb{H}^{[kl]} \mathbb{R}^{-1} \mathbb{S}^{-1} (\mathbf{x}-\boldsymbol{\mu}) + (\mathbf{x}-\boldsymbol{\mu})^T \mathbb{S}^{-1} \mathbb{R}^{-1} \mathbb{H}^{[kl]} \mathbb{R}^{-1} \mathbb{S}^{-1} (x-\mu)_i}{2\sigma_i} \quad (31)$$

which, after changing the sign and taking the expected value, results in

$$\sum_{j=1}^n \frac{\mathbb{R}_{jj} \left(\mathbb{R}^{-1} \mathbb{H}^{[kl]} \mathbb{R}^{-1} \right)_{ji}}{\sigma_i} = \frac{\left(\mathbb{H}^{[kl]} \mathbb{R}^{-1} \right)_{ii}}{\sigma_i} = \frac{\left(\mathbb{R}^{-1} \right)_{ik} \mathbb{I}_{il} + \left(\mathbb{R}^{-1} \right)_{il} \mathbb{I}_{ik}}{\sigma_i} \quad (32)$$

having a non-zero value only when $i = k$ or $i = l$.

4. Case of Asymptotic Independence

This section discusses the situation in which all elements of the pivotal-nuisance and (consequently) nuisance-pivotal blocks of \mathbb{M} (and, correspondingly, of \mathbb{M}_0 and \mathbb{M}^{-1}) are equal to 0. In that case, $\mathbb{M}_0 \mathbb{M}^{-1}$ has the following simple form: its nuisance-nuisance block is the identity matrix, the remaining three blocks are all zero sub-matrices (implying that $\mathbb{I} - \mathbb{M}_0 \mathbb{M}^{-1}$ of (13) has an identity matrix in the pivotal-pivotal block, and zero sub-matrices elsewhere). The resulting MGF of (8) thus equals to $(1 - 2t)^{-L}$, where L is the number of pivotal parameters, which means that the asymptotic distribution of (8) is χ_L^2 , making a construction of the corresponding partial confidence region fairly routine.

Let us add that such asymptotic independence is not uncommon; for example, when a symmetric (with respect to its mean) distribution has a location and scale parameters only, the corresponding ML estimators are then always asymptotically independent.

To see why, note that, in this case, the PDF of the sampled distribution can be expressed as

$$f(x) = \frac{1}{\sigma} \cdot f_0 \left(\frac{x-\mu}{\sigma} \right) \quad (33)$$

where μ and σ are the location and scale parameters respectively, and $f_0(y)$ is a parameter-free PDF, symmetric with respect to 0. The off-diagonal element of \mathbb{M} is then proportional to

$$\int_{-\infty}^{\infty} f_0'(y) \cdot \frac{y \cdot f_0'(y) + f_0(y)}{f_0(y)} dy \quad (34)$$

due to (6); the integrand is clearly an anti-symmetric (odd) function of y , resulting in a zero integral. This enables us to find a partial confidence interval for either μ or for σ (while ignoring the other parameter) by using the χ_1^2 dis-

tribution for (8).

Similarly, in the case of a bivariate Normal distribution, the ML estimators of the two means on one hand, and of the two standard deviations and the correlation coefficient on the other, form two such mutually independent sets as well; this is clear from what we learned in the previous section but, due to the importance of this example, let us be more explicit and quote the well-known asymptotic correlation matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$, namely

$$\tilde{C} = \begin{bmatrix} 1 & \rho & 0 & 0 & 0 \\ \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho^2 & -\frac{\rho}{\sqrt{2}} \\ 0 & 0 & \rho^2 & 1 & -\frac{\rho}{\sqrt{2}} \\ 0 & 0 & -\frac{\rho}{\sqrt{2}} & -\frac{\rho}{\sqrt{2}} & 1 \end{bmatrix} \tag{35}$$

where the five parameters (collectively denoted θ) are the usual $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ (in that order); the respective asymptotic variances are $\sigma_1^2/n, \sigma_2^2/n, \sigma_1^2/(2n), \sigma_2^2/(2n)$ and $(1 - \rho^2)^2/n$.

Constructing a confidence region for either set of parameters is then quite simple (knowing that \tilde{C} has the above block-diagonal structure is all we need, the individual elements are irrelevant) as the following Mathematica program demonstrates.

```
X = RandomVariate[MultinormalDistribution[{2.3, -1.7},
  {{3.2, -2.9}, {-2.9, 4.9}}, 120]; (* random sample of 120 *)
lnf[x_] := -((x[[1]] - μ1)^2/σ1^2 + (x[[2]] - μ2)^2/σ2^2 - 2(x[[1]] - μ1)(x[[2]] - μ2)/(σ1σ2))/
  2/(1 - ρ^2) - Log[2π√(1 - ρ^2)σ1σ2] (* ln of Normal PDF *)
lnLF = Apply[Plus, Map[lnf, X]]//Together;
res = NMaximize[{lnLF, σ1 > 0, σ2 > 0, -1 < ρ < 1}, {μ1, μ2, σ1, σ2, ρ}]
{-463.11, {μ1 -> 2.374, μ2 -> -1.655, σ1 -> 1.768, σ2 -> 2.163, ρ -> -0.688}}
aux = lnLF /. res[[2, {3, 4, 5}]]; (* leave μ1 and μ2 unevaluated *)
ContourPlot[2 res[[1]] - 2 aux == InverseCDF[ChiSquareDistribution[2], 0.95],
  {μ1, 1.9, 2.9}, {μ2, -2.2, -1.1}, ContourStyle -> Black]
aux = lnLF /. res[[2, {1, 2}]]; (* leave σ1, σ2, ρ unevaluated *)
ContourPlot3D[2 res[[1]] - 2 aux == InverseCDF[ChiSquareDistribution[3], 0.9],
  {σ1, 1.4, 2.2}, {σ2, 1.8, 2.6}, {ρ, -0.8, -0.5}, ContourStyle -> Black]
```

The program produces the output presented in **Figure 2** and **Figure 3**; the first graph is the resulting partial 95% confidence region for μ_1 and μ_2 , the second one is the 90% confidence region for σ_1, σ_2 and ρ . This assumes that one is interested only in one or the other (*not* both)—if a confidence region for all five parameters is desired, one would use the basic procedure of our Introduction.

5. General Case

In general (and with no asymptotic independence to help) things get more complicated. We now go over all possible cases involving up to five parameters.

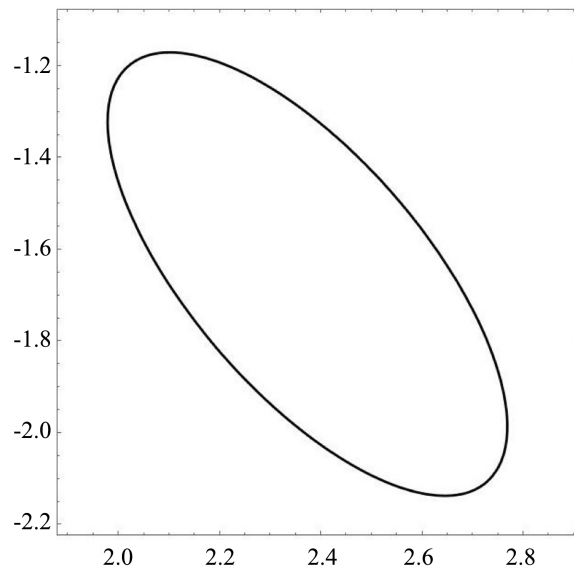


Figure 2. Confidence region for μ_1 and μ_2 .

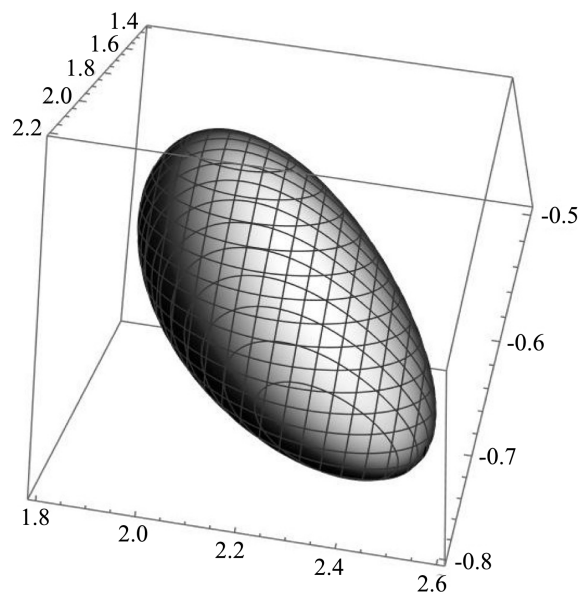


Figure 3. Confidence region for σ_1 , σ_2 and ρ .

5.1. Two Parameters

In this situation, the only possibility is constructing a confidence interval for one of the two parameters, while ignoring the other. Since the \mathbb{C} (or $\tilde{\mathbb{C}}$) matrix will always have the following form

$$\mathbb{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{36}$$

we get

$$\mathbb{C}_{1,1}(\mathbb{C}^{-1})_{1,1} = \frac{1}{1-\rho^2} \tag{37}$$

which implies that the distribution of (8) is that of a χ_1^2 random variable, further divided by $1 - \rho^2$. This means that, to construct a $1 - \alpha$ confidence interval (CI) for the pivotal parameter, we make (8) equal to the corresponding critical value of χ_1^2 , also divided by $1 - \rho^2$, substitute the ME estimate of the nuisance parameter, and solve for the pivotal one to get the two CI boundaries.

The following Mathematica program demonstrates the construction of a 90% confidence interval for the α parameter of the Gamma (α, β) distribution.

```
X = RandomVariate[GammaDistribution[2.3, 4.1], 50]; (* sample of 50 *)
lnf[x_] := (alpha - 1)Log[x] - x/beta - LogGamma[alpha] - alpha Log[beta];
lnLF = Apply[Plus, lnf[X]];
res = NMaximize[{lnLF, alpha > 0, beta > 0}, {alpha, beta}]
{-157.417, {alpha -> 1.98449, beta -> 4.83872}}
tau = 1 - (alpha PolyGamma[1, alpha])^-1 /. res[[2, 1]];
NSolve[2 res[[1]] - 2(lnLF /. res[[2, 2]]) ==
  InverseCDF[ChiSquareDistribution[1], 0.90]/tau, alpha, Reals]
{{alpha -> 1.41835}, {alpha -> 2.62821}} (* CI boundaries *)
```

Note that in this case $\rho = 1/\sqrt{\alpha\psi_1}$, based on

$$\mathbb{M} = \begin{bmatrix} \psi_1 & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix} \tag{38}$$

where ψ_1 is the second derivative of $\ln\Gamma(\alpha)$, called ‘‘PolyGamma[1, α]’’ by Mathematica; to evaluate it, we had to use the ML estimate of α , instead of its true value. That is how we deal with this problem in general; this does not change the asymptotic distribution of (8).

5.2. Three Parameters

A **three-parameter** situation requires discussing two possibilities:

When constructing a CI for a single parameter (say θ_1), (8) has again the $\chi_1^2 \cdot (\mathbb{C}^{-1})_{1,1} \equiv \chi_1^2 \cdot (\tilde{\mathbb{C}}^{-1})_{1,1}$ distribution; this follows from $\mathbb{C}_{1,1}(\mathbb{C}^{-1})_{1,1}$ having only one element (its only eigenvalue), and from $\mathbb{C}_{1,1} \equiv 1$. Note that this result (when interested in only *one* parameter) is true for any K .

When finding a confidence region for θ_2 and θ_3 , the distribution of (8) is a convolution (*i.e.* an independent sum) of χ_1^2 and another χ_1^2 , the latter multiplied $s := (\mathbb{C}^{-1})_{1,1}$; this is based on the arguments following (18). The corresponding PDF is

$$\frac{1}{2\sqrt{s}} \exp\left(-\frac{s+1}{4s} \cdot y\right) I_0\left(\frac{s-1}{4s} \cdot y\right) \tag{39}$$

where I_0 denotes the modified Bessel function of the first kind; note that using this PDF to find a critical value can be done only numerically (there is no analytic expression for the corresponding CDF, let alone for its inverse).

As an example, we assume sampling a distribution with the following PDF

$$f(x; \alpha, \beta, \gamma) = \frac{\gamma x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \quad \text{when } x > 0 \quad (40)$$

(each of the three parameters must be positive) and constructing a 95% confidence region for β and γ only. Leaving out routine details, the expression for s turns out to be

$$s = 1 + \frac{\alpha}{\gamma} + \frac{\alpha^2 \psi_1 - (\alpha + \gamma) \gamma}{\gamma^2 + \psi_1 (\gamma^2 - \alpha^2 \psi_1)} \quad (41)$$

where ψ_1 is the second derivative of $\ln \Gamma$, evaluated at $\frac{\alpha}{\gamma}$.

The following Mathematica code demonstrates the algorithm.

```
X = RandomVariate[GammaDistribution[0.32/2.6, 1.8, 2.6, 0], 1000];
Inf[x_] := -α Log[β] + Log[γ] - (x/β)^γ + (α-1) Log[x] - LogGamma[α/γ]
lnLF = Apply[Plus, Map[Inf, X]] // Together;
res = NMaximize[{lnLF, α>0, β>0, γ>0}, {α, β, γ}]
{554.79, {α → 0.3140, β → 1.6588, γ → 2.0671}}
aux = lnLF /. res[[2, 1]];
s = (α + γ) / γ - (γ(α + γ) - α^2 PolyGamma[1, α/γ]) / (γ^2 + PolyGamma[1, α/γ] (γ^2 - α^2 PolyGamma[1, α/γ])) /. res[[2]];
g = Exp[-(s+1)/(4s) y] BesselI[0, (s-1)/(4s) y];
c = c /. FindRoot[Integrate[g, {y, 0, c}] / (2√s) == 0.95, {c, 4}];
ContourPlot[2(res[[1]] - aux) == c, {β, 1.4, 1.92}, {γ, 1.4, 3}]
```

The confidence region for the β and γ thus produced is displayed in **Figure 4**.

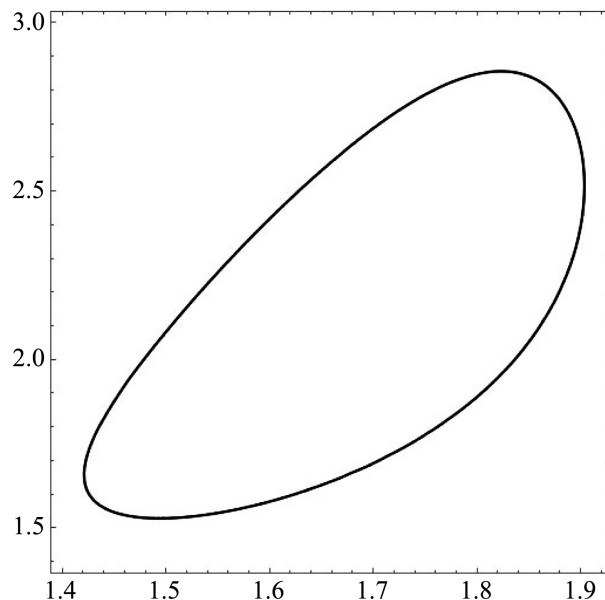


Figure 4. Confidence region for β and γ .

5.3. Four Parameters

A CI for θ_1 results in (8) having the $\chi_1^2 \cdot (\mathbb{C}^{-1})_{1,1}$ distribution, as discussed previously.

The complementary task of constructing a confidence region for θ_2, θ_3 and θ_4 then leads to a convolution of the previous distribution and that of χ_2^2 (to account for the two extra eigenvalues of $\mathbb{I} - \mathbb{C}_{p,1}(\mathbb{C}^{-1})_{1,p}$, both equal to 1); this convolution has a PDF given by

$$\frac{1}{2\sqrt{1-s}} \exp\left(-\frac{y}{2}\right) \operatorname{erf}\left(\sqrt{\frac{(1-s) \cdot y}{2s}}\right) \tag{42}$$

Finally, to build a confidence region for θ_1 and θ_2 requires using the following PDF for (8):

$$\frac{\sqrt{t_1 t_2}}{2} \exp\left(-\frac{t_1 + t_2}{4} \cdot y\right) I_0\left(\frac{t_1 - t_2}{4} \cdot y\right) \tag{43}$$

where t_1 and t_2 are the two eigenvalues of $\mathbb{M}_{p,p}(\mathbb{M}^{-1})_{p,p}$; using \mathbb{C} instead of \mathbb{M} is still possible, but has no longer any advantage, since critical values of (42) and (43) can again be found only numerically.

To show how to use the last formula, we assume sampling a *mixture* of Exponential and Normal distributions (the four parameters are: the mean of the Exponential distribution, denoted b , and its weight in the mixture a , followed by the mean c and standard deviation d of the Normal distribution). Note that in this example we also bypass an analytic solution for elements of the \mathbb{M} matrix—these can be easily computed by substituting values of the ML estimates for θ_0 in (6), and only then computing, by *numerical* integration, the corresponding expected value. The following Mathematica program demonstrates the complete algorithm.

```

aux=RandomVariate[BinomialDistribution[450,0.4],1][[1]];
X=Join[RandomVariate[ExponentialDistribution[1/3.1],aux],
RandomVariate[NormalDistribution[14.2,2.9],450-aux];
Inf:=Log[a Exp[-x/b]/b+(1-a)Exp[-(x-c)^2/2/d^2]/sqrt(2pi)/d]
lnLF=Apply[Plus,Map[Inf,X]];
res=NMaximize[{lnLF,0.1<a<0.9,b>0.9,d>0.1},{a,b,c,d}]
{-1332.77,{a->0.4544,b->3.85,c->14.5182,d->3.0303}
aux={D[lnf[x],a],D[lnf[x],b],D[lnf[x],c],D[lnf[x],d]};
M=NIntegrate[Outer[Times,aux,aux]Exp[lnf[x]]/.res[[2]],{x,0,inf}];
t=Eigenvalues[M[{{3,4},{3,4}}].Inverse[M][{{3,4},{3,4}}]];
z=z/.FindRoot[sqrt(t[[1]]t[[2]])/4NIntegrate[Exp[-(t[[1]]+t[[2]])/4 y]
BesselI[0,(t[[1]]-t[[2]])/4 y],{y,0,z}]==0.9,{z,4.};
ContourPlot[2(res[[1]]-lnLF/.res[[2,{1,2}]])=z,{c,14,15},{d,2.7,3.4}]

```

Producing the 90% confidence region displayed in **Figure 5** for the mean (horizontal scale) and standard deviation (vertical scale) of the Normal-distribution part of the mixture.

5.4. Five Parameters

As in all previous cases, a confidence interval for only one (say θ_1) of the parameters requires using the χ_1^2 distribution, further multiplied by $s := (\mathbb{C}^{-1})_{1,1}$.

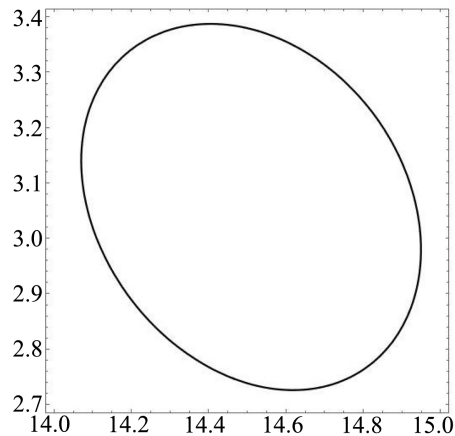


Figure 5. Confidence region for c and d .

For the complementary task of building a (four-dimensional) confidence region for $\theta_2, \theta_3, \theta_4$ and θ_5 , the distribution of (8) becomes a convolution of $\chi_1^2 \cdot s$ and χ_3^2 ; the corresponding PDF is

$$\frac{y}{4\sqrt{s}} \exp\left(-\frac{s+1}{4s} \cdot y\right) \left[I_0\left(\frac{s-1}{4s} \cdot y\right) + I_1\left(\frac{1-s}{4s} \cdot y\right) \right] \tag{44}$$

A confidence region for θ_1 and θ_2 is found using the PDF of (43), found in the case of four parameters.

The PDF used for the three-dimensional confidence region of the true values of θ_3, θ_4 and θ_5 is then a convolution of the last PDF and that of an independent χ_1^2 ; no explicit formula for the resulting PDF exists, but the following example demonstrates how to bypass this problem.

This time, we assume sampling a tri-variate Normal distribution having identical means (denoted μ) and identical standard deviations (denoted σ). We now construct a three-dimensional confidence region for μ, σ and ρ_{12} (while ignoring ρ_{13} and ρ_{23}), using the following Mathematica code.

```
X = RandomVariate[MultinormalDistribution[-2{1, 1, 1},
3^2{{1, 0.3, -0.5}, {0.3, 1, 0.4}, {-0.5, 0.4, 1}}, 100];
R = {{1, ρ1,2, ρ1,3}, {ρ1,2, 1, ρ2,3}, {ρ1,3, ρ2,3, 1}};
lnf[x_] := -(x - μ).Inverse[R].(x - μ)/2/σ^2 - 3/2 Log[2π] - 3Log[σ] - 1/2Log[Det[R]]
lnLF = Apply[Plus, Map[lnf, X]];
e = {Mean[X], StandardDeviation[X], Correlation[X]};
res = FindMaximum[{lnLF, ρ1,2^2 < 1, ρ1,2^2 + ρ1,3^2 + ρ2,3^2 - 2ρ1,2ρ1,3ρ2,3 < 1, σ > 0},
{{μ, e[[1,1]]}, {σ, e[[2,1]]}, {ρ1,2, e[[3,1,2]]}, {ρ1,3, e[[3,1,3]]}, {ρ2,3, e[[3,2,3]]}}];
h = {μ, σ, ρ1,2, ρ1,3, ρ2,3};
M = -(Table[D[lnf[{z1, z2, z3}], h[[i]], h[[j]]], {i, 5}, {j, 5}]) / (-μ + zk) -> Qk // Expand //
{Qk^2 -> σ^2, Qk -> Qm -> If[i < j, ρk,m, ρm,k]σ^2} / Qk -> 0 / . res[[2]];
t = Eigenvalues[M[[{4,5}], {4,5}]].Inverse[M[[{4,5}], {4,5}]];
g = Integrate[Sqrt[t[[1]]t[[2]]/8/π/q Exp[-y(t[[1]]+t[[2]])/4-q/2]
Bessel[0,y(t[[1]]-t[[2]])/4], {q, 0, Z-y} // PowerExpand;
z = W /. FindRoot[NIntegrate[g, Z -> W, {y, 0, W}] == 0.95, {W, 4}];
aux = 2(res[[1]] - lnLF /. res[[2, {4,5}]]) // Together;
ContourPlot3D[aux == z, {ρ1,2, .2, .4}, {μ, -2.5, -1.5}, {σ, 2.5, 3.5}]
```

The resulting confidence region is shown in **Figure 6**.

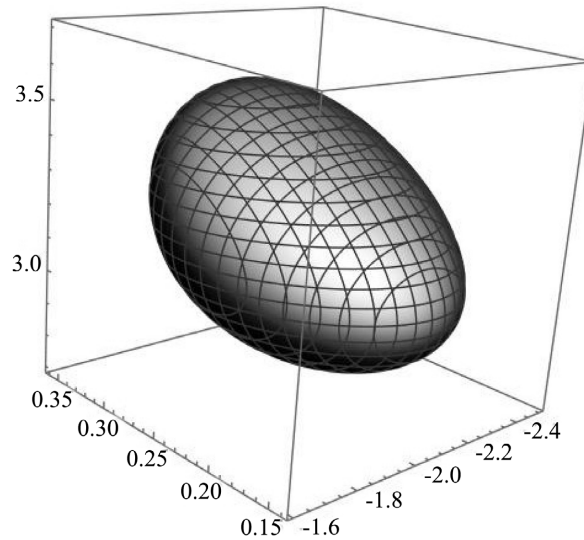


Figure 6. Confidence region for μ , σ and ρ_{12} .

6. Technique’s Accuracy

In this section, we go over a rather special example, allowing us to *analytically* find not only all ML estimators, but also an exact expression for (8). These are then used to build (still analytically) formulas for boundaries of a confidence interval for one of the parameters. Furthermore, the exact distribution of the estimator is also known, which gives us a unique opportunity to compute the error of our technique.

Let us now proceed with the actual example of constructing a confidence interval for the correlation coefficient of a bivariate Normal distribution, and computing its *exact* level of confidence (while ignoring the remaining four parameters).

Firstly, it is well known what the ML estimators of the two means (μ_x and μ_y), the two standard deviations (σ_x and σ_y), and the correlation coefficient ρ are; we will denote them $\bar{X}, \bar{Y}, S_x, S_y$ and r respectively. The likelihood function, once we replace all five parameters by their estimators, then reads

$$-\frac{1}{2(1-r^2)} \sum_{i=1}^n \left(\frac{(X_i - \bar{X})^2}{S_x^2} + \frac{(Y_i - \bar{Y})^2}{S_y^2} - 2r \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{S_x S_y} \right) \tag{45}$$

$$-\frac{n}{2} \ln(1-r^2) - n \ln S_x - n \ln S_y = -n - \frac{n}{2} \ln(1-r^2) - n \ln S_x - n \ln S_y$$

When similarly replacing the first four parameters only (keeping the exact value of ρ), the likelihood function becomes

$$-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{(X_i - \bar{X})^2}{S_x^2} + \frac{(Y_i - \bar{Y})^2}{S_y^2} - 2\rho \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{S_x S_y} \right) \tag{46}$$

$$-\frac{n}{2} \ln(1-\rho^2) - n \ln S_x - n \ln S_y = -n \frac{1-\rho r}{1-\rho^2} - \frac{n}{2} \ln(1-\rho^2) - n \ln S_x - n \ln S_y$$

making (8) equal to

$$-2n \frac{\rho r - \rho^2}{1 - \rho^2} - n \ln \frac{1 - r^2}{1 - \rho^2} \tag{47}$$

It is easy to find that $(\tilde{C}^{-1})_{5,5} = 1 + \rho^2$, where \tilde{C} is the matrix of (35); this means that the random variable (47) has, approximately, the χ^2_1 distribution further multiplied by $1 + \rho^2$. Making (47) less than the critical value of this distribution (say C_α) has, approximately, the probability of $1 - \alpha$ of being correct. Since the sampling distribution of r is known, we can also compute the exact probability of the same inequality, thus getting the error of the approximation. This is done by the following Mathematica program (for any chosen set of values for ρ, n and α)

```
f=(n-2)Gamma[n-1](1-ρ²)^(n-1)/2/√(2π)/Gamma[n-1/2](1-r²)^(n-2)/((1-ρ r)^(n-3/2)
Hypergeometric2F1[1/2,1/2,n-1/2,(1+ρ r)/2]
ρ=-0.9; n=30; α=0.05; aux=NSolve[2ρ(ρ-r)/(1-ρ²)-Log[1-r²]+Log[1-ρ²]==
InverseCDF[ChiSquareDistribution[1],1-α]/n(1+ρ²),r,Reals]
{{r->-0.9536},{r->-0.8159}}
NIntegrate[f,{r,r/.aux[[1]],(r/.aux[[2]])}]
0.9408
```

where the first two lines (a continuation of a single Mathematica statement) spell out the exact PDF of r . The program then specifies ρ, n and α , makes (47) equal to $1 - \alpha$, and solves for r (note that this is the reverse of what is done when finding boundaries of a confidence interval for ρ , given a value of r); the last line then evaluates the corresponding exact probability. Note that the PDF of r is notoriously slow in reaching its Normal limit (on which our approximation is based), making the errors of confidence intervals for ρ atypically large; this example is thus close to presenting the worst-case scenario.

By executing the program using various values of ρ, n and $1 - \alpha$, we get errors (in percent) presented in Table 1 (these are quoted for $n = 10, 30$ and 100 respectively).

Based on these results, we can make the following observations:

- The exact confidence level is always less than the claimed one.
- The error does not change much with the value of ρ (surprisingly, the error is usually the largest at $\rho = 0$ and decreases slightly towards both extremes).
- It decreases as the confidence level goes up, but increases relative to α .
- It decreases, to a good approximation, with $1/n$.

Table 1. List of the technique’s errors (in %).

$1 - \alpha =$	0.8	0.9	0.95	0.99
$\rho = 0 \downarrow$	6.6, 2.0, 0.57	5.4, 1.5, 0.43	3.9, 1.1, 0.29	1.5, 0.36, 0.10
± 0.5	6.2, 1.8, 0.53	4.8, 1.4, 0.39	3.4, 0.92, 0.26	1.2, 0.30, 0.08
± 0.9	5.3, 1.6, 0.46	4.2, 1.2, 0.36	3.2, 0.92, 0.27	1.7, 0.48, 0.14
± 0.99	5.0, 1.5, 0.44	4.0, 1.2, 0.35	3.1, 0.95, 0.28	1.9, 0.58, 0.17

7. Conclusion

We have shown how to construct confidence regions for parameters of interest while ignoring one or more additional (nuisance) parameters. This is done based on ML estimates of all parameters, utilizing the corresponding likelihood function and formulas of this article. The error of the resulting procedure behaves similarly (*i.e.* decreasing with the first power of the sample size) to the error of ordinary confidence regions based on the χ^2 distribution of (1). Our explicit examples have covered situations involving up to five parameters in total, but a general approach for dealing with *any* number of parameters has been clearly delineated as well. Future research will undoubtedly supply specific details of any such multi-parameter situation, and come up with a way of simplifying, at least numerically, the resulting distributions.

Conflicts of Interest

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

References

- [1] Bartlett, M.S. (1953) Approximate Confidence Intervals. *Biometrika*, **40**, 12-19. <https://doi.org/10.1093/biomet/40.1-2.12>
- [2] Bartlett, M.S. (1953) Approximate Confidence Intervals II. More Than One Unknown Parameter. *Biometrika*, **40**, 306-317. <https://doi.org/10.1093/biomet/40.3-4.306>
- [3] Bartlett, M.S. (1955) Approximate Confidence Intervals III. A Bias Correction. *Biometrika*, **42**, 201-204. <https://doi.org/10.1093/biomet/42.1-2.201>
- [4] Kendall, M.G. and Stuart, A. (1969) *The Advanced Theory of Statistics*, Vol. 1, Chapter 3, Hafner Publishing Company, New York.
- [5] Cox, D.R. (1975) Partial likelihood. *Biometrika*, **62**, 269-276. <https://doi.org/10.1093/biomet/62.2.269>
- [6] Kalbfleisch, J.D. and Sprott, D.A. (1973) Marginal and Conditional Likelihoods. *Sankhyā: The Indian Journal of Statistics, Series A*, **35**, 311-328.
- [7] Hotelling, H. (1940) The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters. *The Annals of Mathematical Statistics*, **11**, 271-283. <https://doi.org/10.1214/aoms/1177731867>
- [8] Basu, D. (1977) On the Elimination of Nuisance Parameters. *Journal of the American Statistical Association*, **72**, 355-366. <https://doi.org/10.1080/01621459.1977.10481002>

Appendix: Notation and Abbreviations

- $\mathcal{L}(\mathbf{X};\boldsymbol{\theta})$: the Likelihood Function; \mathbf{X} is the set of observations, $\boldsymbol{\theta}$ are the distribution parameters.
- LF: likelihood function.
- ML: maximum likelihood.
- LHS: left hand side.
- MGF: moment generatin function.
- PDF: probability density function.
- χ_K^2 : the chi-square distribution with K degrees of freedom.
- \mathbb{E} : a symbol for taking expected value.
- \mathbb{I} and \mathbb{O} : identity and zero matrix, respectively.
- \mathbb{M} : the variance-covariance matrix of multivariate Normal distribution.
- \mathbb{M}_0 : the previous matrix with all pivotal by pivotal elements set to 0.
- $\mathbb{M}_{p,i}$: the pivotal by nuisance (or incidental) block of \mathbb{M} .
- $\mathbf{v}^{[k]}$: a vector with k^{th} component equal to 1, the rest equal to 0.
- $\mathbb{H}^{[k]}$: a matrix with $(k, \ell)^{\text{th}}$ and $(\ell, k)^{\text{th}}$ components equal to 1, the rest equal to 0.