Robustness of Lognormal Confidence Regions for Means of Symmetric Positive Matrices

Benoit Ahanda[1], Daniel E. Osborne[2], Leif Ellingson[1]
[1] Texas Tech University, Department of Mathematics and Statistics
[2] Florida Agricultural and Mechanical University, Department of Mathematics

Abstract
Symmetric positive definite (SPD) matrices arise in a wide range of applications including diffusion tensor imaging (DTI), cosmic background radiation, and as covariance matrices. A complication when working with such data is that the space of SPD matrices is a manifold, so traditional statistical methods may not be directly applied. However, there are nonparametric procedures based on resampling for statistical inference for such data, but these can be slow and computationally tedious. Schwartzman (2015) introduced a lognormal distribution on the space of SPD matrices, providing a convenient framework for parametric inference on this space. Our goal is to check how robust confidence regions based on this distributional assumption are to a lack of lognormality. The methods are illustrated in a simulation study by examining the coverage probability of various mixtures of distributions.

1. Introduction

For statisticians, symmetric positive definite (SPD) matrices appear most commonly as covariance matrices for multivariate data. In such cases, researchers typically study these matrices via properties of covariance and underlying properties of the underlying data. However, for other areas in which they arise, researchers must take a different approach. Some of these other applications include diffusion tensor imaging (DTI) and astronomy in the form of cosmic background radiation. In these cases, researchers must instead consider the SPD matrices as the data objects themselves, which is complicated by the fact that the space of SPD matrices is a manifold.

The field of statistics on manifolds initially arose largely to study directional data but has since been further developed for analyses on general manifolds. While many parametric methods were developed for directional data analysis, the trend since the turn of the century has been towards developing nonparametric procedures, starting with Hendriks and Landsman (1999) and Bhattacharya and Patrangenaru (2003), due to the fact that there exist no general goodness of fit tests for data on manifolds. These nonparametric methods are typically based either on asymptotic results or resampling techniques such as the nonparametric bootstrap. Since sample sizes are often relatively small compared to the dimensionality of the data, bootstrapping is often necessary for inference. While this certainly works well in many situations, bootstrapping on manifolds can often be quite computationally intensive, as discussed in Bhattacharya et al (2012).

For the specific case of SPD matrices, Osborne et al (2013) and Ellingson et al (2016) have developed nonparametric methods utilizing the bootstrap with illustrations of these methods towards DTI data analysis. However, those papers consider only single diffusion tensors while a typical diffusion tensor image will contain many thousands of tensors. As such, Schwartzman (2015) introduced lognormal distributions for the space of SPD matrices that allowed for the development of parametric inference procedures based on this

distributional assumption. While these methods are computationally efficient compared to the bootstrap methods, and thus more suitable for large-scale analyses, it remains to be seen how well these procedures hold up to violations in the distributional assumptions.

As such, it is the goal of this paper to examine the robustness of these procedures via simulation studies by examining effective coverage probabilities. More specifically, we will begin exploring how coverage probabilities of lognormal confidence regions are affected when the actual underlying probability distribution is a mixture of lognormal random matrices with different means.

The remainder of this paper will be organized as follows. In Section 2, we will present the relevant background on lognormal distributions and their associated confidence regions for the population mean. Then, in Section 3, we will present the methodology we used to perform simulations with the results of this preliminary study presented in Section 4. We will then discuss these results in Section 5.

2. Lognormal Distributions and Inference for Means

Schwartzman (2015) introduced parametric methods for performing inference for means of SPD matrices based on two types of lognormal distributions on this space, which consists of all matrices having positive eigenvalues. Schwartzman arrived at these distributions using different distances that can be utilized for analyzing SPD matrices. The Type I lognormal distribution is defined with respect to what is known as the log-Euclidean distance. If A and B are two $p \times p$ SPD matrices, then the log-Euclidean distance between them is

$$d_{LE}(A, B) = \| \log(A) - \log(B) \|,$$

where $\log(A) = \exp^{-1}(A)$ and $\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$.

Based on this, an SPD matrix X is said to follow a **Type I lognormal distribution** with mean μ and covariance Σ if $\mathrm{vecd}(\log X) \sim N_q(\log(\mu), \Sigma)$, where $q = p(p+1)/2$ and vecd is a vectorizing operator such that the l_2 norm of $\mathrm{vecd}(A)$ equals the Frobenius norm of A.

The Type II lognormal distribution is defined similarly using what is known as the canonical Riemannian distance on the space of $p \times p$ SPD matrices. For thew purposes of this initial study, we will only consider the Type I lognormal distribution, though, for computational and theoretical simplicity.

For the Type I lognormal distribution, the MLE for μ is

$$\hat{\mu} = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \log(X_i) \right).$$

The MLE of Σ is

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{vecd}(\log(X_i) - \log(\hat{\mu})) \mathrm{vecd}(\log(X_i) - \log(\hat{\mu}))'.$$

Using Slutsky's Theorem, Schwartzman (2015) defined an asymptotic $100(1-\alpha)\%$ confidence region for the mean SPD matrix μ to be

$$\mathcal{C}_1 = \{\mu : n \mathrm{vecd}(\log(\hat{\mu}) - \log(\mu))' \hat{\Sigma}^{-1} \mathrm{vecd}(\log(\hat{\mu}) - \log(\mu)) \leq \chi^2_q(\alpha)\}. \qquad (1)$$

Utilizing the Central Limit Theorem, this result will also apply for sufficiently large n even when X does not follow a Type I lognormal distribution. However, motivated by the high computational cost of using the nonparametric bootstrap for inference with small sample sizes, we choose to instead focus on two exact confidence regions.

If Σ is known, then

$$\mathcal{C}_2 = \{\mu : n\text{vecd}(\log(\hat{\mu}) - \log(\mu))'\Sigma^{-1}\text{vecd}(\log(\hat{\mu}) - \log(\mu)) \leq \chi^2_q(\alpha)\}. \quad (2)$$

is an exact $100(1-\alpha)\%$ confidence region for μ.

In the more realistic scenario, Σ is unknown, as in $\mathcal{C}_1$. However, multivariate normal distribution theory tells us that we can define another exact $100(1-\alpha)\%$ confidence region for μ to be

$$\mathcal{C}_3 = \{\mu : n\text{vecd}(\log(\hat{\mu}) - \log(\mu))'(\hat{\Sigma})^{-1}\text{vecd}(\log(\hat{\mu}) - \log(\mu)) \leq \frac{p}{n-p}F_{1-\alpha,p,n-p}\}. \quad (3)$$

3. Methodology

To examine the robustness of parametric confidence regions for data arising from this distribution, we performed a simulation study to compare the effective confidence level to the nominal confidence level under violations of the distributional assumption.

3.1 Simulation

More specifically, we decided to consider violations arising from the data coming from a mixture of Type I lognormal distributions. That is, for a fixed sample size n and means μ_1 and μ_2, we considered mixtures of the form $\beta LN_I(\mu_1, \Sigma) + (1-\beta)LN_I(\mu_2, \Sigma)$, where $0 \leq \beta \leq 1$ to analyze how the effective confidence level is altered as the level of mixing is changed. We then repeated this for increasingly large deviations between μ_1 and μ_2 to gain a better understanding of the impact of bimodality on the coverage probabilities. Finally, we repeated all of these calculations for increasingly large values of n to explore how many observations are needed for coverage probabilities to converge to the nominal levels even under severe model violations.

To simplify the simulations and interpretations of the results, we worked with 2×2 symmetric matrices, resulting in the vectorized form being 3-dimensional. If we define the the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix},$$

then the vectorized form will be defined as

$$\text{vecd}(A) = \begin{bmatrix} a_{11} \\ a_{22} \\ \sqrt{2}a_{12} \end{bmatrix}$$

so that, as described in the previous section, the Frobenius norm of A is equal to the l_2 norm of $\text{vecd}(A)$.

For the purposes of this initial study, we simulated data using vectorized means of the form $\mu_1 = (\delta, 0, b)$ and $\mu_2 = (-\delta, 0, b)$ for the following values of δ: 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 3.0, 4.0, 6.0, and 10 and a fixed value of b. This yielded the global mean

$\mu_0 = \beta * \mu_1 + (1 - \beta) * \mu_2$. We used 0.5, 0.6, 0.7, 0.8, 0.9, and 1 as the values for β with the values less than 0.5 accounted for by the symmetry in the form of the mixture. Finally, we considered the following sample sizes: 6, 20, 40, 80, and 500.

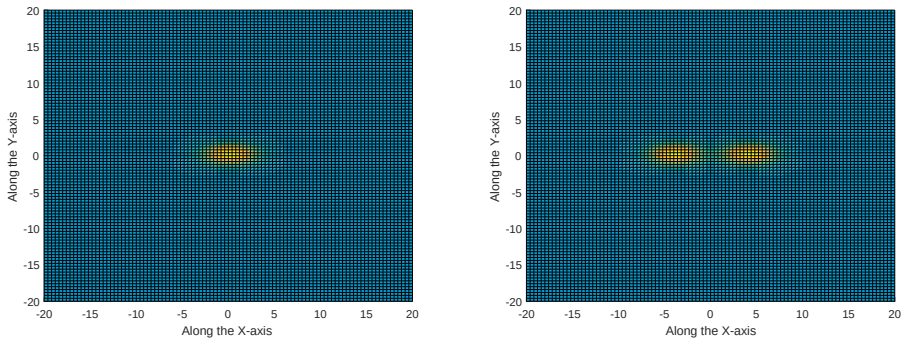
We simulated the data under two different scenarios involving the covariance matrix. For Case 1, the covariance matrix for both distributions is:

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & .01 \end{bmatrix}.$$

Then, for Case 2, the covariance matrix for both distributions is:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & .01 \end{bmatrix}$$

In Figure 1 are examples of how the densities are impacted (on the log-scale) when the data are simulated as described in Case 1. As we can see from this, when $\delta = 0.04$, the deviation from lonormality is small, while it is quite large for $\delta = 4$.



(a) $\delta = 0.4$ (b) $\delta = 4$

Figure 1: Densities of data (on the log-scale) simulated from Case 1

3.2 Evaluation

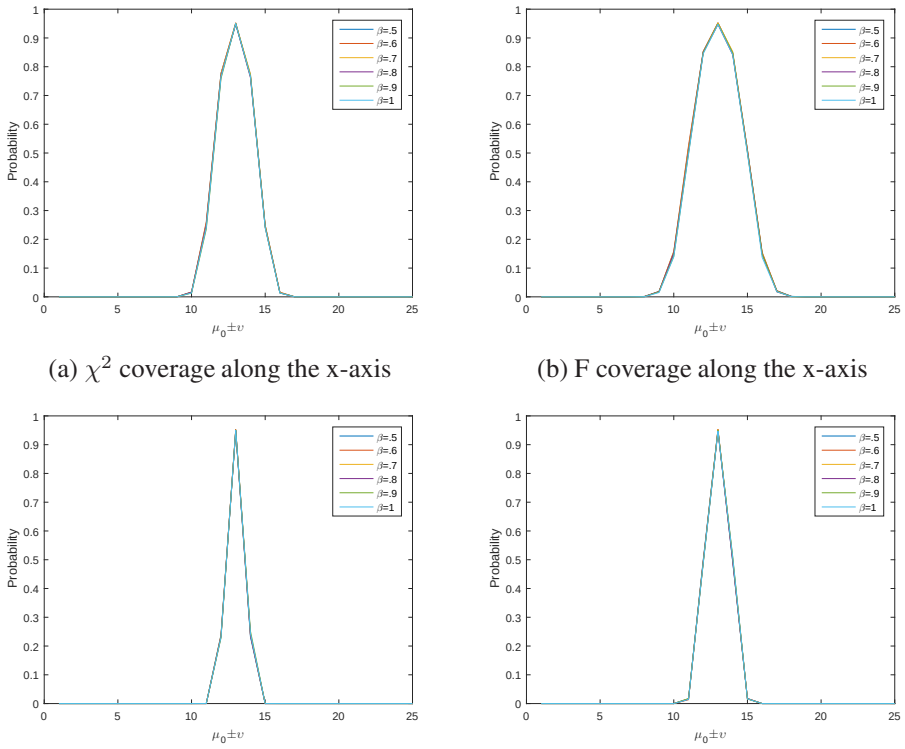
After simulating the data, our primary goal was to evaluate the robustness of the confidence regions $\mathcal{C}_2$ and $\mathcal{C}_3$ at a fixed nominal confidence level, which we chose to be 95%. Rather than just focusing on the true coverage probability, we also wanted to explore the impact of the model violation on false coverage probabilities.

To do this, we repeatedly simulated data sets for each set of values of β, δ, and n for both cases and recorded the proportion of times that a point μ was included in each confidence region. We used μ of the form $\mu_0 + (\upsilon, 0, 0)$ and $\mu_0 + (0, \upsilon, 0)$ for the following values of υ: -12,...,-4,-3,-2,-1,0,1,2,3,4,...,12. No changes were made in the third entry because so little variability was present in that direction. For the purposes of evaluating the performance of $\mathcal{C}_3$, we approximated the true value of Σ by calculating the sample covariance matrix of a sample of 30,000 observations for the given mixture of distributions.

4. Results

4.1 Case 1

Results of the simulations for Case 1 are shown in Figures 2, 3, 4, and 5. In all four scenarios, the true coverage probabilities are close to the nominal level of 95%. The effects of the mixtures on the coverage probabilities are not seen until false coverage probabilities are considered. In all cases, the coverage probabilities converge quickly to 0 while traveling along the y-axis because the lack of lognormality is not felt in that direction, even for the larger values of δ.



(a) χ^2 coverage along the x-axis

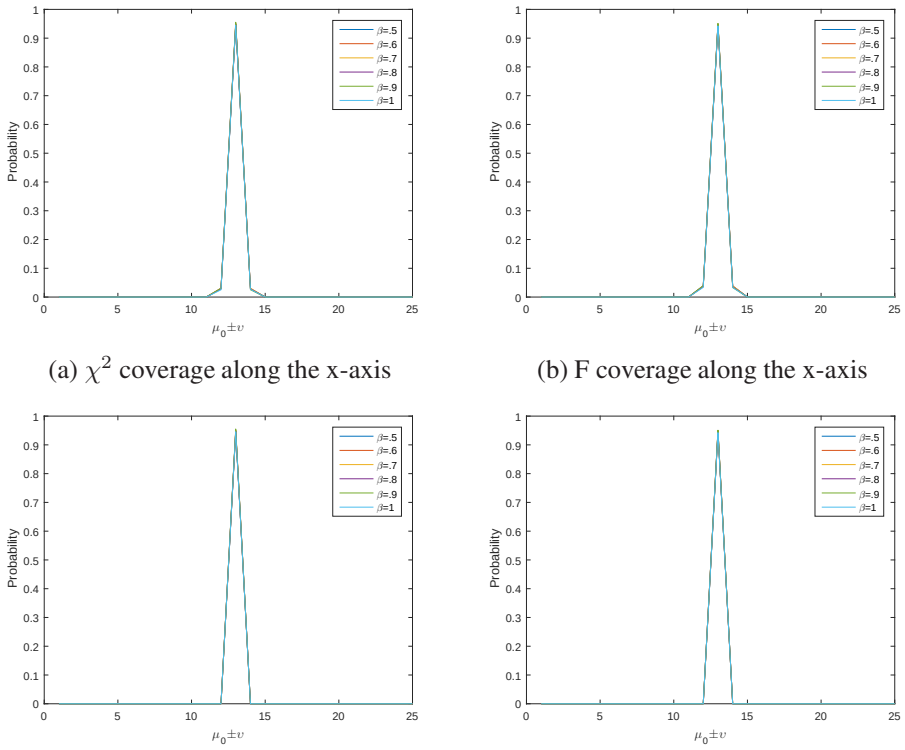
(b) F coverage along the x-axis

(c) χ^2 coverage along the y-axis

(d) F coverage along the y-axis

Figure 2: Case 1: $n = 6, \delta = 0.4$

More interestingly, as shown in Figures 2 and 3, we see that the coverage probabilities for the various mixtures are nearly identical to those of the lognormal data. This suggests that the procedure is robust to minor violations of normality, even for small sample sizes. However, as illustrated in Figures 4 and 5, the coverage probabilities along the x-axis for the mixtures are impacted considerably more by larg deviations from lognormality when the sample size is small, whereas they are not as highly impacted for larger sample sizes. This, then, suggests that the performance of these confidence regions are moderately robust for large deviations from lognormality along the primary direction of variability when the sample size is small, but is considerably more robust for larger sample sizes.



(a) χ^2 coverage along the x-axis

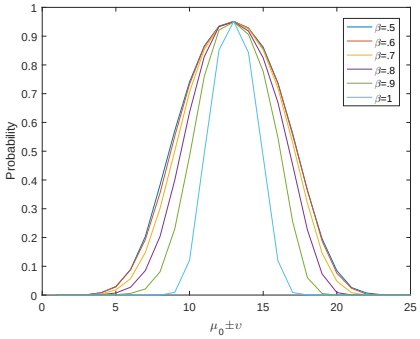
(b) F coverage along the x-axis

(c) χ^2 coverage along the y-axis

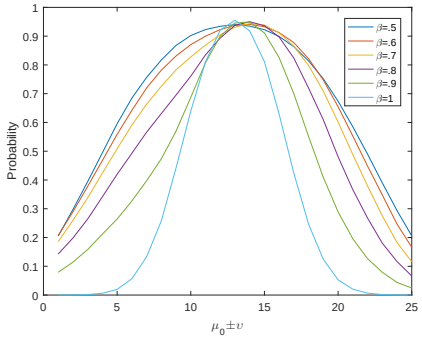
(d) F coverage along the y-axis

Figure 3: Case 1: $n = 80, \delta = 0.4$

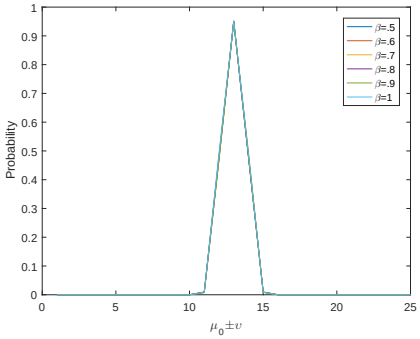
These patterns are heightened for the F confidence regions compared to the Chi-square confidence regions. This is due to the fact the latter confidence regions are calculated using the true covariance matrix for the data, whereas the former intervals are heavily impacted by the estimation of the covariance matrix in addition to the violations of lognormality. This result is intuitive based on the fact that the Chi-square regions are not plausible to use in practice since the covariance matrix is so rarely known.



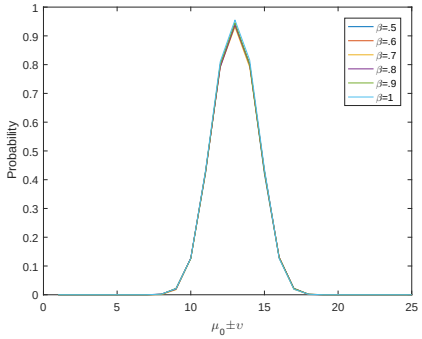
(a) χ^2 coverage along the x-axis



(b) F coverage along the x-axis

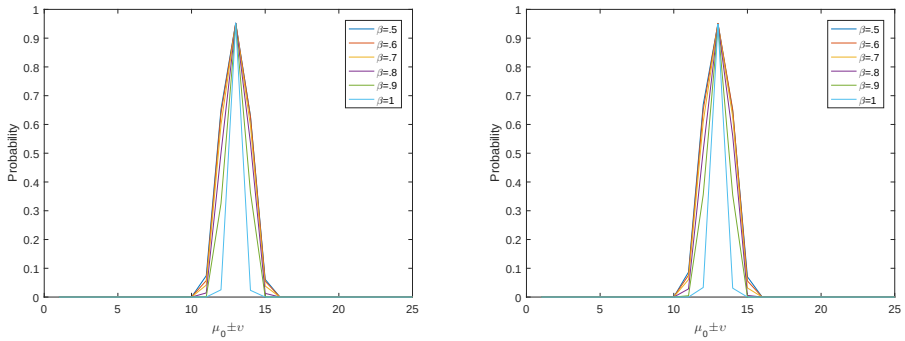


(c) χ^2 coverage along the y-axis



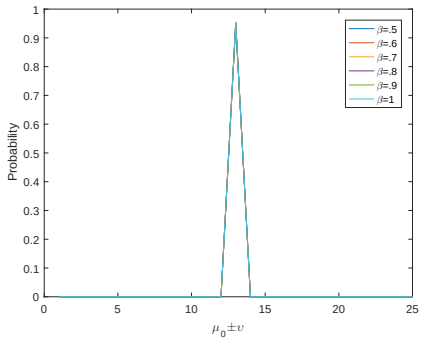
(d) F coverage along the y-axis

Figure 4: Case 1: $n = 6, \delta = 4$

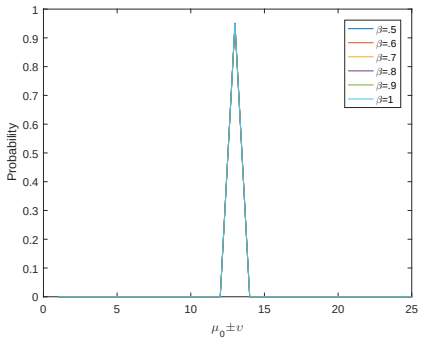


(a) χ^2 coverage along the x-axis

(b) F coverage along the x-axis



(c) χ^2 coverage along the y-axis

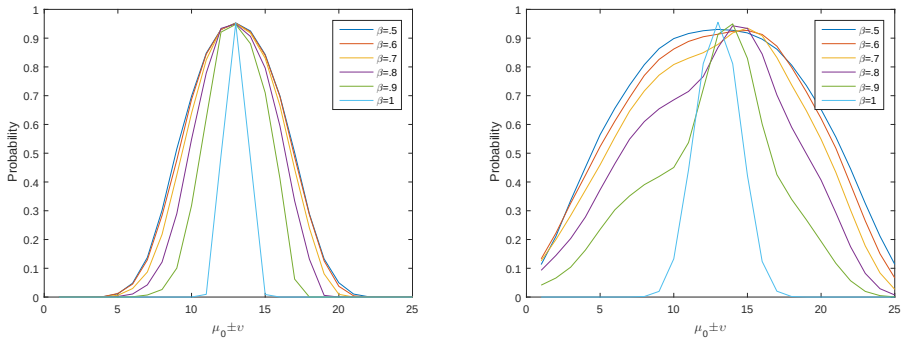


(d) F coverage along the y-axis

Figure 5: Case 1: $n = 80, \delta = 4$

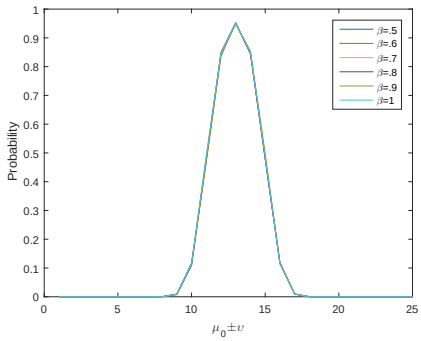
4.2 Case 2

For Case 2, we largely observed very similar behavior within each scenario as what we saw with Case 1. As a result, we present only an illustration of how these results differ from the former case. Notably, as shown in Figure 6, even the Chi-square coverage probabilities are considerably more impacted by large deviations from lognormality occurring in a direction orthogonal to the primary direction of variability when the sample size is small. The F coverage probabilities for the mixtures are hugely impacted here, as well. Also, the coverage probabilities along the y-axis for the F confidence regions are also impacted more noticeably in this case. All of this suggests that these confidence regions are not robust to moderate deviations in lognormality when the deviations occur in this manner.

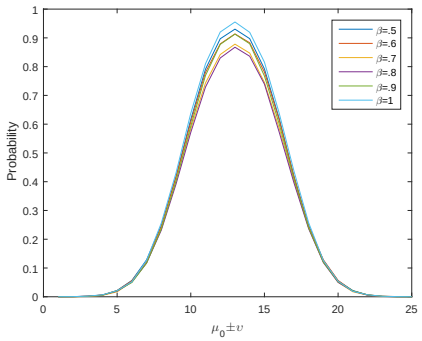


(a) χ^2 coverage along the x-axis

(b) F coverage along the x-axis



(c) χ^2 coverage along the y-axis



(d) F coverage along the y-axis

Figure 6: Case 2: $n = 6, \delta = 4$

4.3 Synthesis of Results

While the above results help us understand how the performances of the confidence regions are impacted by violations from the model assumptions, it is difficult to use those plots to concretely specify the patterns in the relationships between deviations from lognormality and coverage performance across different scenarios. To help us explore these relationships, we decided to quantify the differences between both the densities and the coverage probabilities.

More specifically, the L_2 distance between the distributions is

$$D_f = \left(\iint_L (f^* - f_\beta)^2 \right)^{\frac{1}{2}}$$

where f^* is the pdf of the nearest lognormal distribution to the mixture and f_β is the pdf of the mixture when $\beta = 0.5, 0.6, 0.7, 0.8, 0.9, 1$. The L_2 distance between the coverage probabilities is

$$D_c = \left(\iint_L (C_1 - C_\beta)^2 \right)^{\frac{1}{2}}$$

where C_1 is the F coverage probability at $\beta = 1$ of the mixture and C_β is the F coverage probability of the mixture when $\beta = 0.5, 0.6, 0.7, 0.8, 0.9, 1$

We plotted both distances for each case and we obtained the following figures. An important aspect of these plots is that every point represents a specific curve from our previous

results. In Figure 7, for example, the point (J) represents the F coverage curve along the x-axis when $\delta = 4$ and n=6 at $\beta = 0.5$.

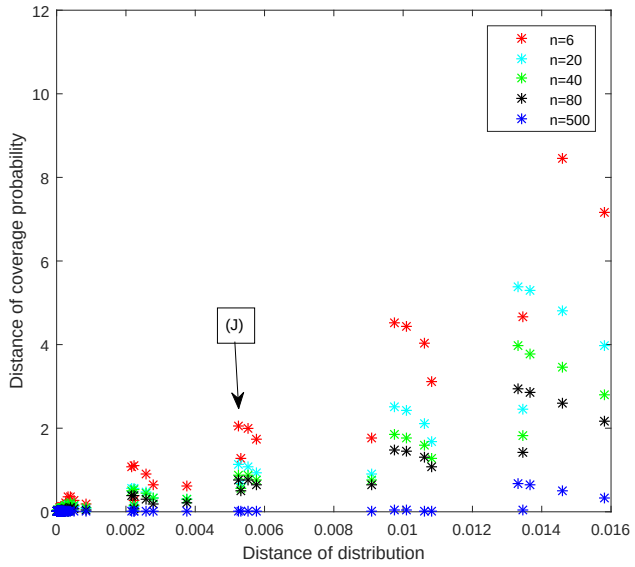


Figure 7: Distance between pdf vs the coverage probabilities: Case 1

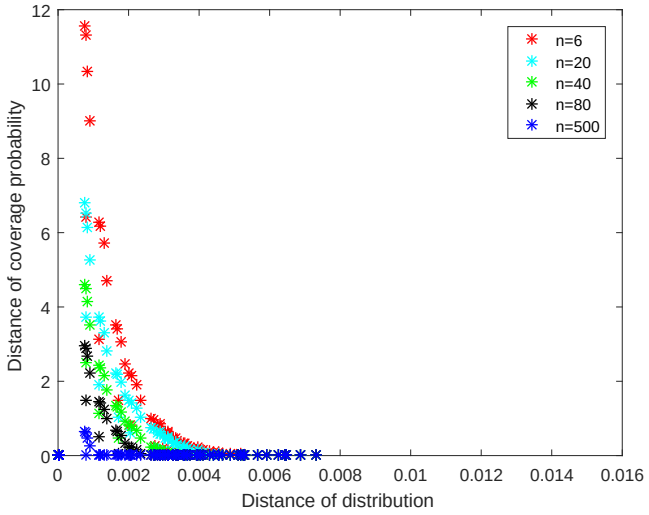


Figure 8: Distance between the pdf vs the coverage probabilities: Case 2

In Case 1, the distance for the coverage probabilities are near zero for large sample sizes and for small distances between the distributions. For Case 2, we observe similar behavior for large sample sizes. Note that for Case 2, though, that we have the surprising result that the greater distances between distributions are associated with lower distances in coverage probabilities, seemingly indicating that the procedure is more robust to larger violations of this type than small ones.

5. Discussion

In this paper, we have investigated the robustness of parametric confidence regions based on the Type I lognormal distribution for SPD matrices to mixtures of lognormal distributions with different means. From our simulation study, we can arrive at the following conclusions for the robustness of the Type I lognormal confidence regions to mixtures of the type described earlier.

First, unsurprisingly, due to the central limit theorem applied to the log-transformed data, true coverage probabilities are all near the nominal level for large sample sizes. Secondly, for similar reasons, for fixed δ and fixed β, false coverage probabilties decrease as n increases. Also, for fixed n and fixed β, false coverage probabilities increase as δ increases. Additionally, for fixed δ and fixed n, false coverage probabilities decrease as β increases. Finally, for fixed small n and large enough δ, it appears that the procedure in Case 1 of the first simulation is more robust than the one in Case 2, which appears, interestingly, to be more robust to larger deviations from lognormality than to smaller deviations.

In general, it appears that the procedure is somewhat robust to moderate violations of log-normality and for large sample size. However, more work remains to be done. For instance, we need to further investigate the surprising behavior shown in Figure 8. Additionally, we will examine robustness to mixtures of lognormal distributions with identical means but different covariance matrices because such mixtures violate the distributional assumptions in a fundamentally different manner than those investigated here.

References

[1] R. N. Bhattacharya and V. Patrangenaru, (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-Part I, *Ann. Statist.* **31**, no. 1, 1-29.

[2] R. N. Bhattacharya, L. Ellingson, X. Liu, V. Patrangenaru, and M. Crane (2012). Extrinsic Analysis on Manifolds is Computationally Faster than Intrinsic Analysis with Application to Quality Control by Machine Vision. Appl. Stochastic Models Bus. Ind., 28. 222-235.

[3] L. Ellingson, D. Groisser, D. E. Osborne, V. Patrangenaru, and A. Schwartzman (2016). Nonparametric Bootstrap of Sample Means of Positive Definite Matrices with an Application to Diffusion-Tensor-Imaging Data. *To appear in Communications in Statistics C Simulation and Computation*.

[4] H. Hendriks and Z. Landsman (1998). Mean Location and Sample Mean Location on Manifolds: Asymptotics, Tests, Confidence Regions. *Journal of Multivariate Analysis*, **67**, No. 2, p. 227-243.

[5] D. E. Osborne, V. Patrangenaru, L. Ellingson, D. Groisser, and A. Schwartzman (2013). Nonparametric Two-Sample Tests on Homogeneous Riemannian Manifolds, Cholesky Decompositions and Diffusion Tensor Image Analysis. *Journal of Multivariate Analysis*, 119. p. 163-175.

[6] A. Schwartzman (2015) Log-Normal Distributions and Geometric Averages of Positive Definite Matrices, *International Statistical Review*. **84**(3), p. 456-486.