Cholesky Normal Distribution in the Space of Symmetric Positive- Definite Matrices.

Benoit Ahanda[1], Leif Ellingson[2], Daniel E. Osborne[3]
[1] Bradley University, Department of Mathematics
[2] Texas Tech University, Department of Mathematics and Statistics
[3] Florida Agricultural and Mechanical University, Department of Mathematics

Abstract

The aim of this paper is to introduce a probability distribution in the space of symmetric positive definite (SPD) matrices called the Cholesky normal distribution. Because the space of SPD matrices is a non-Euclidean manifold, standard arithmetic and thusly standard statistical methods do not directly apply for data on this space. Instead, researchers typically either perform an intrinsic analysis by defining a Riemannian metric and then projecting the data onto a tangent space or an intrinsic analysis by embedding the space into the space of symmetric matrices. For both approaches, since there are not many probability distributions defined on the space of SPD matrices, researchers typically use nonparametric inference procedures, which may be too computationally expensive for practical use on large-scale data analyses. Following from Schwartzman (2015), we utilize the Cholesky metric on the space of SPD matrices to define the distribution, investigate some of its properties, and develop a parametric inference procedure for the mean of SPD matrices.

1. Introduction

Statisticians often encounter symmetric positive definite (SPD) matrices as covariance matrices of multivariate data, and rely rely on properties of covariance and the original data to study these matrices. However, there are many fields where SPD matrices have become the primary data objects, themselves, with each SPD matrix being an observation, such as medical imaging, physics, and astronomy in the form of cosmic background variation. In such cases, researchers must take a different approach when developing statistical inference procedures for these objects because the space of SPD matrices is a non-Euclidean manifold. As such,standard arithmetic does not directly apply because the space is locally but not globally homeomorphic to a Euclidean space.

Many nonparametric procedures were developed for statistical analysis on general manifolds, starting with Hendriks and Landsman (1999), Bhattacharya and Patrangenaru (2003) and Bhattacharya and Patrangenaru (2005), due to the fact that there exist no general goodness of fit tests for data on manifolds despite the need for inference procedures to be readily available. 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 the more tractable approach. 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) 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. Schwartzman (2015) introduced log-

normal distributions for the space of SPD matrices that allowed for the development of parametric inference procedures based on his distributional assumption. These methods have proven to be far more computationally efficient than the bootstrap methods, and thus more suitable for large-scale analyses. Inspired by Schwarztman's approach of defining probability distributions via two distances on the space of SPD matrices, we sought to develop additional probability distributions on the space. As such, we have considered a distribution that some researchers may have already utilized for modeling but, to our knowledge, has not been studied in-depth based on the Cholesky distance between SPD matrices. Using the naming convention of the lognormal distribution, we call this distance the Cholesky normal distribution.

The layout of this paper will be as follows. Section 2 presents the Cholesky distance and sample means for SPD matrices. Section 3 introduces the Cholesky normal distribution and some of its properties are investigated via simulation study. Section 4 introduces parametric confidence regions for the Cholesky mean of SPD matrices and the performance of these confidence regions is also evaluated through a simulation study. In Section 5, we discuss the results and some of the limitations of this study.

2. The Cholesky Distance and Associated Means for SPD Matrices

2.1 Cholesky Distance Between SPD Matrices

A $p \times p$ SPD matrix A may often be associated with and visualized by an ellipsoid corresponding to all points $y \in \mathbb{R}^p$ that satisfy the quadratic form $y'Ay = c$ for some constant c. As discussed in Dryden et al. (2009), due to the geometry of the space of $p \times p$ SPD matrices within the space of symmetric $p \times p$ matrices, when the Euclidean distance is used to travel between SPD matrices, the resulting paths are prone to swell the volume of the ellipsoids associated with the matrices along these paths. This property is inherited by sample means when calculated using the usual Euclidean arithmetic and therefore also affects the calculation of covariance operators for these data objects, as well. Among the many distance metrics described in Dryden et al. (2009) that do not have this volume-swelling effect is the Cholesky distance, which is defined as follows.

Definition 2.1. If A and B are two $p \times p$ SPD matrices, then the *Cholesky distance* between A and B is

$$d_C(A, B) = \|chol(A) - chol(B)\|,$$

where $chol(A)$ and $chol(B)$ are the lower Cholesky decomposition matrices of A and B and $\|\cdot\|$ denotes the Frobenius norm.

For the purposes of characterizing variability and performing inference, we will find it convenient to vectorize the matrices of Cholesky decompositions via an invertible mapping that preserves the distance between observations. Let Vec denote a mapping from $\mathbb{L}(p)$, the space of lower-triangular $p \times p$ matrices, to $\mathbb{R}^q$, where $q = \frac{p(p+1)}{2}$. There are multiple equivalent ways to define this function, but we will use the following form:

$$Vec(L) = \left[\mathrm{diag}(L)', \mathrm{offdiag}(L)'\right]',$$

where $\mathrm{diag}(L)$ is a $p \times 1$ column vector containing the diagonal entries of L and $offdiag(L)$ is a $(q - p) \times 1$ column vector containing the off-diagonal lower-triangular entries of L.

For example, if $p = 2$, and we define

$$L = \begin{bmatrix} l_{11} & 0 \\ l_{12} & l_{22} \end{bmatrix},$$

then the $vec(L)$ is as follows:

$$Vec(L) = V = \begin{bmatrix} l_{11} \\ l_{22} \\ l_{12} \end{bmatrix} \Rightarrow Vec^{-1}(V) = L$$

2.2 Cholesky Means and Covariances of SPD Matrices

To define a mean with respect to the Cholesky metric, we utilize the definition of a Fréchet mean, which is defined as

$$\mu_F = \arg\min_P [E(d^2(X, P))],$$

for some distance d. Similarly, if we have a sample $X_1, X_2, \ldots, X_n$, then the sample Frechet mean is

$$\hat{\mu}_F = \arg\min_P \sum_{i=1}^{n} d^2(X_i, P)$$

Following from Dryden et al. (2009), if $d = d_C$, it can be shown that the minimizer of the expected squared distance, which we will call the Cholesky mean μ_C, is

$$\mu_C = (E[chol(X)]) (E[chol(X)])'.$$

Likewise, our sample Fréchet mean, which we will refer to as the sample Cholesky mean $\hat{\mu}_C$ is

$$\hat{\mu}_C = \left[\frac{1}{n}\sum_{i=1}^{n} chol(X_i)\right] \left[\frac{1}{n}\sum_{i=1}^{n} chol(X_i)\right]'.$$

To understand the variability in our data, we also need to define covariance matrices for our SPD matrices. To do so, we will need to use the vectorized form of our data. Let $X_1, \ldots, X_n$ be iid $p \times p$ SPD matrices and $L_i = chol(X_i)$. We can define our population covariance matrix as

$$\Sigma_C = E\left[(Vec(L_1) - Vec(\mu_C)) (Vec(L_1) - Vec(\mu_C))'\right].$$

The corresponding sample covariance matrix is

$$\hat{\Sigma}_C = \frac{1}{n-1}\sum_{i=1}^{n} \left[(Vec(L_i) - Vec(\hat{\mu}_C)) (Vec(L_i) - Vec(\hat{\mu}_C))'\right].$$

3. Cholesky Normal Random Matrices

With the above concepts, we can now define the Cholesky normal distribution on the space of $p \times p$ SPD matrices.

Definition 3.1. Let $X \sim N_q(\theta, \Sigma)$. Then the random $p \times p$ SPD matrix Y follows the Cholesky normal $cholN(\theta, \Sigma)$ distribution if $Y = (Vec^{-1}(X)) \cdot (Vec^{-1}(X))'$, where θ is a location parameter vector and Σ is a dispersion matrix.

Please note that μ_C and Σ_C are defined for the random matrix Y on the space of $p \times p$ SPD matrices while θ and Σ correspond to a random vector in $\mathbb{R}^p$. Furthermore, since $Vec^{-1}(X) \in \mathbb{L}(p)$ and $chol(Y) \in \mathbb{L}^+(p)$, the space of lower triangular $p \times p$ matrices with positive diagonal entries, we cannot conclude that $Vec(chol(\mu_C)) = \theta$ or $\Sigma_C = \Sigma$ due to the diagonal entries of X having positive density for negative values.

More generally, this means that $Vec(chol(Y))$ is non-normal. However, if $\theta >> 0$, then $Vec(chol(Y)) \approx X$ because the diagonal entries of X will have trivially positive density for negative values. In this case, $Vec(chol(\mu_c)) = E[Vec(chol(Y)) \approx \theta]$ and any inferences on $chol(\mu_c)$ will be approximately the same as inferences on θ. Likewise, inferences about Σ_C will be able to approximate those for Σ, though these fall outside the scope of this study. For the next portion of this subsection, we will study the impact of θ on the distribution of $chol(Y)$.

We first examine this by simulating $n = 1000$ observations Y_i from $cholN(\theta, \mathbf{I}_6)$, where, for simplicity we chose $\theta = (\alpha, \alpha, \alpha, \alpha, \alpha, \alpha)$ for $\alpha = 0$ and $\alpha = 3.2$. Histograms of the marginal entries of $chol(Y)$ for these choices of θ are shown in Figures 1 and 2 and Figures 3 and 4, respectively.

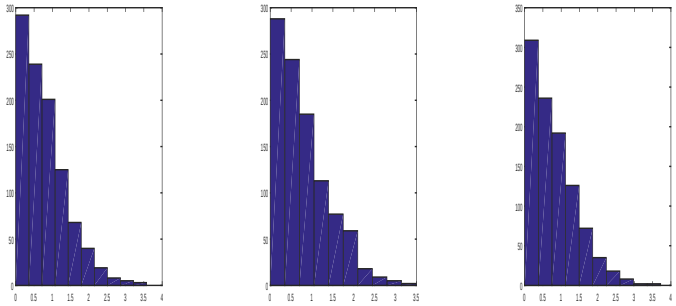

Figure 1: The diagonal marginal entries of $chol(Y)$ when $\theta = 0$

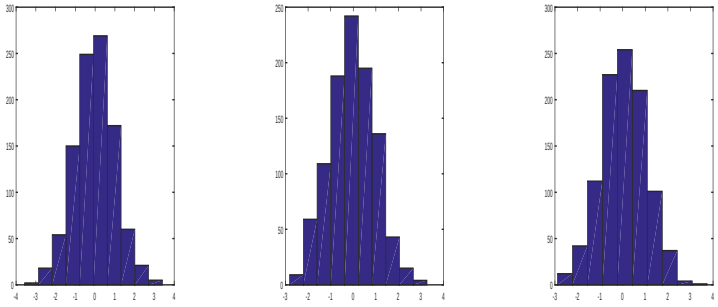

Figure 2: The off-diagonal marginal entries of $chol(Y)$ when $\theta = 0$

When we take the most extreme case where $\alpha = 0$, as in Figures 1 and 2, the distributions of the diagonal marginals are positively skewed and the distributions of the off-diagonal marginals appear to be normally distributed. However, when $\alpha = 3.2$, as in

Figures 3 and 4, while the distributions of the off-diagonal entries still appear normal, the same is also now true for the diagonal marginals. From this, then, it is clear that we should focus on the marginal distributions of the diagonal entries of $chol(Y)$ to study how large θ must be for $Vec(chol(Y))$ to be approximately multivariate normally distributed.

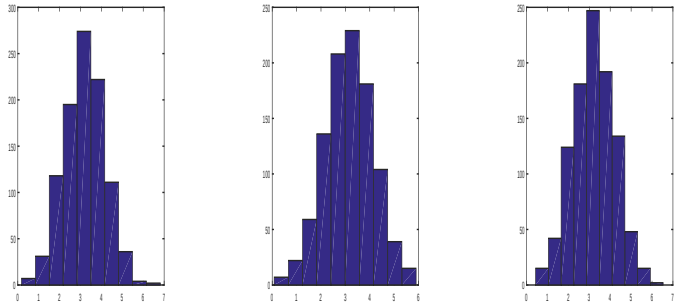

Figure 3: Marginals of the diagonal when $\theta = 3.2$

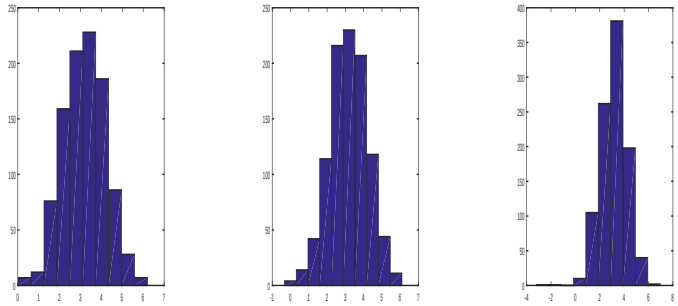

Figure 4: Marginals of the diagonal the off-diagonal when $\theta = 3.2$

To investigate further, we focus on the marginal distribution of the first diagonal entry and now consider $\alpha = 0, 1.5, 2.5, 3.2, 4.1, 5.5$. Kernel density estimates of the marginal distributions generated with these values of α are shown in Figure 5. We can see that the densities appear to become approximately normal for values of $\alpha \geq 2.5$. Since the marginal variance is 1, this makes sense because $P(Z < -2.5) = .0062$ for $Z \sim N(0, 1)$. To try to further establish a threshold for α to result in $Vec(chol(Y))$ being approximately multivariate normal, we compare the same marginal density alongside normal densities for $\alpha = 0, 1, 2.5$ in Figure 6. It is clear that when α is near 0, the density reflects to the right of zero meaning that the area gets added to the right of 0. Through 6, it is also obvious that as θ increases as the difference between the two areas to the left of zero of the two curves decreases.

So in general, if $Y \sim cholN(\theta, \Sigma)$, then $Vec(chol(Y)) \sim N(\theta, \Sigma)$ approximately if $P(\vec{X} > 0) = (1 - \beta)$ is large, where β is the the difference between the area to the left of 0 between the normal curve and the marginal density of $chol(Y)$. We can formalize this under specific assumptions in the following theorem.

Theorem 3.1. Let $\vec{X}$ be $1 \times p$ vector of the diagonal terms of the random vector used to generate Y such that its marginal entries are iid and β denote the difference between the area to the left of 0 between the normal curves and the diagonal marginal curves of $chol(Y)$.

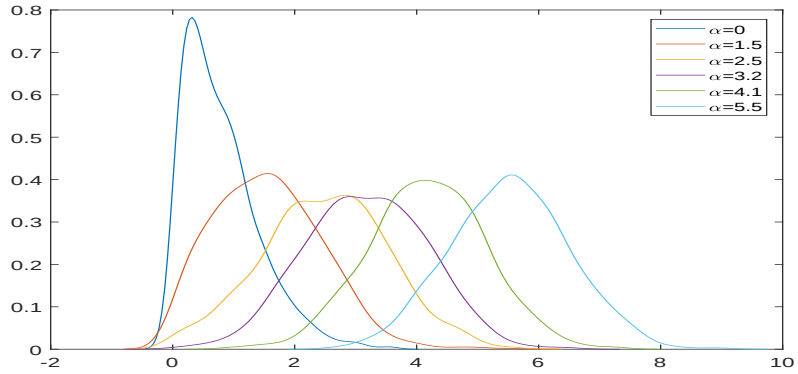

Figure 5: Density from different means

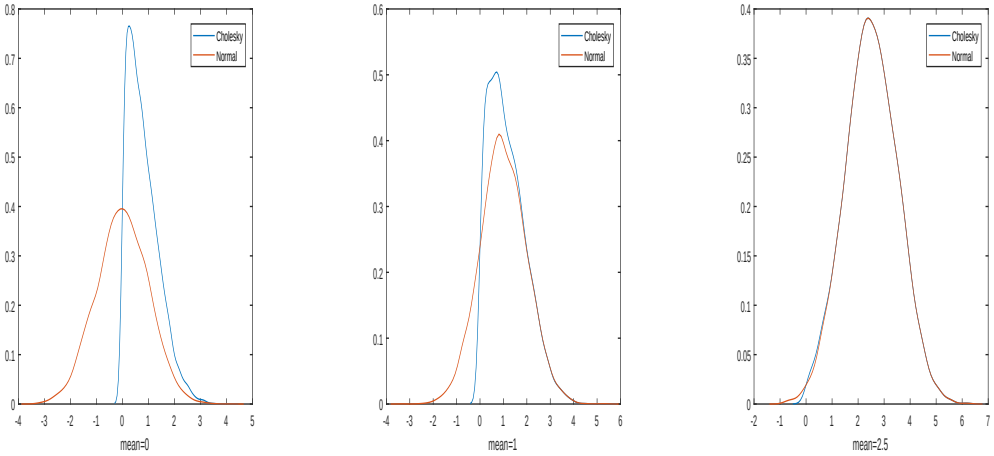

Figure 6: Cholesky vs Normal

The cdf of the diagonal entries of $chol(Y)$ will differ from that of $\vec{X}$ by β if and only if θ_j and Σ_{jj} are chosen to satisfy

$$\phi^{-1}[1-(1-\beta)^{\frac{1}{p}}] = \frac{-\theta_{jj}}{\sqrt{\Sigma_{jj}}},$$

where ϕ denotes the cdf of the standard normal distribution.

Proof. Since the diagonal entries of $chol(Y)$ must be positive, the difference between the cdfs of ($\vec{X}$ and these entries is $P(\vec{X} \leq 0) = \beta$. This yields the following result.

$$P(\vec{X} > 0) = (1-\beta) \tag{1}$$

$$\Leftrightarrow P(X_{11} > 0, ..., X_{pp} > 0) = (1-\beta) \tag{2}$$

$$\Leftrightarrow [P(X_{jj} > 0)]^p = (1-\beta) \tag{3}$$

$$\Leftrightarrow [P(X_{jj} > 0)] = (1-\beta)^{\frac{1}{p}} \tag{4}$$

$$\Leftrightarrow [P(X_{jj} < 0)] = 1-(1-\beta)^{\frac{1}{p}} \tag{5}$$

$$\Leftrightarrow P\left(Z < \frac{0-\theta_{jj}}{\sqrt{\Sigma_{jj}}}\right) = 1-(1-\beta)^{\frac{1}{p}} \tag{6}$$

$$\Leftrightarrow \phi^{-1}(1-(1-\beta)^{\frac{1}{p}}) = \left(\frac{-\theta_{jj}}{\sqrt{\Sigma_{jj}}}\right) \tag{7}$$

□

Corollary 3.1. In the univariate case, θ must be chosen so that

$$\phi^{-1}(\beta) = \frac{-\theta}{\sqrt{\Sigma}}$$

Proof. : The result follows by letting $p = 1$ in (7). □

3.1 Cholesky Normal and the Wishart Distribution

It is important to contrast the non central Wishart distribution to the Cholesky normal distribution. To generate a non-central Wishart, we must generate $X \sim N_{p,n}(\theta, \Sigma)$, $n \geq p$, then $S = XX^T \sim W_p(n, \Sigma, \Theta)$ is a non central Wishart with $\Theta = \Sigma^{-1} * \theta * \theta^T$ is called the noncentrality parameter.
Applying the Cholesky matrix on S doesn't neccessary map back to X and even when X is lower triangular.

The following results is very important in multivariate analysis and is known as the Bartlett's decomposition, Bartlett(1933) and see Gupta and Nagar(1999),Theorem 3.3.4, Matrix Variate Distributions, for a proof of this classical result.

Theorem 3.2. Let $S = XX^T \sim W_p(n, I_p)$ and X is lower triangular, where $X = (x_{ij})$ with $x_{ii} > 0$, $\Theta = 0$ and I_p is the $p \times p$ identity matrix then the x_{ij}, $i \geq j$, are independent random variables distributed. Then, $x_{ij} \sim N(0,1)$, $1 \leq j < i \leq p$ and $x_{ii}^2 \sim \chi_{n-i+1}$, $1 \leq i \leq p$.

Using this result clearly shows the difference beteween the Cholesky normal distribution in figures 7 and 8. The plot to the left represents the Cholesky normal distribution on the Cholesky scale and the plot to the right represents the Wishart distribution on the Cholesky scale. The multivarite normal for both was generated with the same parameters $\theta = [0, 0, 0, 0, 0, 0]$, Σ is 6×6 identity matrix, the sample size was of $n = 10000$, and we used 6 degree of freedom for the Wishart distribution.

In 7, the off-diagonal terms are normally distributed with mean 0 and standard deviation of 1 in both cases, and in 8 it is obvious that both distributions are different.The Wishart distribution in this case is centered at the degree of freedom and the Cholesky normal is centered at zero. Note that to generate the Wishart distribution, the degree of freedom has be 6 or higher. A more compelling argument of why both distributions are different is that for the Wishart distribution, the square of the diagonal entries in Theorem 4.1.4 , $x_{ii}^2 \sim \chi_{n-i+1}$ and for the Cholesky distribution the square of the diagonal entries follow a Chi-square distribution with the same degree of freedom.

The density of the Cholesky normal doesn't exist in the full Space, since in the Cholesky scale the distribution is singular normal.

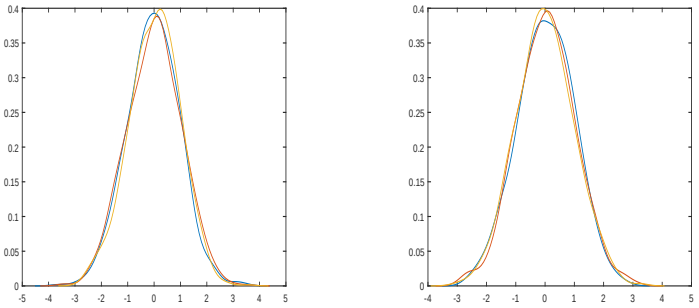

Figure 7: marginals of the off-diagonal entries

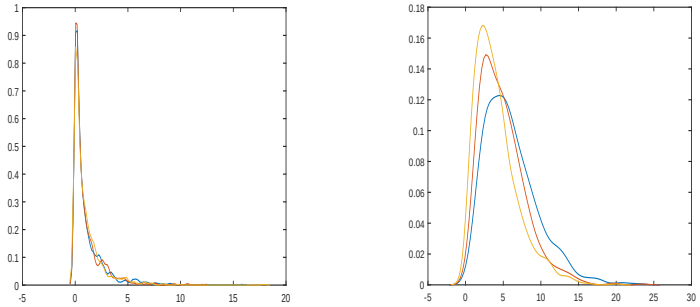

Figure 8: marginals of the square of the diagonal entries

4. Inferences for Means

4.1 Cholesky Confidence Regions

Let $Y_1, ..Y_n$ be $p \times p$ random matrices from a Cholesky normal distribution with mean μ_C and covariance Σ_C where the location parameter θ is large enough for the Cholesky decompositions to be approximately normal. Then $chol(\hat{\mu_C})$ is the MLE of $chol(\mu_C)$ and is also approximately the MLE of θ and $\hat{\Sigma}_C$ is the MLE of $Cov(Vec(chol(Y)) = \Sigma_C$ and is also the approximately the MLE of Σ.

Theorem 4.1. If Σ_C is known, then an $100(1-\alpha)\%$ confidence region for the mean SPD matrix μ_C is:

$$\mathcal{C}_5 = \{\mu_C : n\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C))'{\Sigma_C}^{-1}\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C)) \leq \chi^2_q(\alpha)\}. \tag{8}$$

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

Proof. Let $Y_1, ..Y_n$ be $p \times p$ matrix from Cholesky normal sample distribution with mean μ_C and covariance Σ_C, then if $\text{Vec}(chol(Y) \sim N_q(chol(\mu_c), \Sigma_C)$, where $q = p(p+1)/2$ and if $\hat{\mu_C}$ is the MLE of μ_C, then the MLE of $chol(\mu_c)$ is $chol(\hat{\mu_C})$ and,

$$n\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C))'{\Sigma_C}^{-1}\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C)) \sim \chi^2_q$$

Thus a $100(1-\alpha)\%$ confidence region for θ is :

$$n\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c))'{\Sigma_w}^{-1}\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c)) \leq \chi^2_q(\alpha)\}$$

□

Theorem 4.2. When Σ_C is unknown, however, multivariate normal distribution theory tells us that we can define another exact $100(1-\alpha)\%$ confidence region for μ_C to be

$$\mathcal{C}_6 = \{\mu_C : n\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C))'(\hat{\Sigma}_C)^{-1}\text{Vec}(chol(\hat{\mu_C}) - chol(\mu_C)) \leq \frac{q}{n-q}F_{1-\alpha,q,n-q}\}. \tag{9}$$

Proof. In the univariate case, if we let $X_1, ..X_n \sim N(\theta_1, \sigma)$ be from a sample of normal variables with the mean θ_1 and where $\hat{\theta_1}$ is the maximum likelihood estimator of θ_1 and s^2 is the sample variance. Then $n\frac{(\hat{\theta_1} - \theta_1)^2}{s^2} \sim F_{1,n-1}$ and is the same as :

$$n(\hat{\theta_1} - \theta_1)s^2(\hat{\theta_1} - \theta_1)' = T^2$$

Now, if we let $Y_1, ..Y_n$ be from a multivariate Cholesky normal sample distribution with mean μ_c and covariance Σ_c where $\hat{\mu_c}$ is the mle of μ_c and $\hat{\Sigma_w}$ is the mle of Σ_c .
Then $\text{vec}(cholY) \sim N_q(chol(\mu_c), \Sigma_w)$, where $q = p(p+1)/2$ and
replacing

$$n\frac{(\hat{\theta_1} - \theta_1)^2}{s^2}$$

by

$$n\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c))'(\hat{\Sigma_w})^{-1}\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c)) = T^2$$

gives

$$\frac{n-q}{(n-1)q}T^2 \sim F_{q,n-q}$$

and hence, it follows that

$$n\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c))'(\hat{\Sigma_w})^{-1}\text{vec}(chol(\hat{\mu_c}) - chol(\mu_c)) \leq \frac{q(n-1)}{n-q}F_{1-\alpha,q,n-q}$$

□

4.2 Performance of Cholesky Confidence Regions

We checked the performance of the Cholesky confidence regions in more realistic scenario with respect to (9). A simulation study was performed to compare the effective confidence level to the nominal confidence level when the data are from a Cholesky normal distribution and when the data are generated from a normal distribution with the same mean and covariance matrix as in the Cholesky scale.

4.2.1 Simulation

More specifically, for a fixed sample size n we decided to consider confidence regions in the Cholesky scale when the data are generated from a normal distribution with different means of the form $\theta = [\alpha_1, \alpha_1, \alpha_1, ...]$. We repeated this for increasingly large α to gain a better understanding the impact that the mean has when generating a Cholesky normal distribution and how it affects the coverage probabilities. Finally for fixed α_1, we repeated

all of these calculations for increasingly large values of n to explore how many observations are needed for true coverage probabilities to converge to the nominal levels and false coverage probabilities to converge to zero.

To simplify the simulations and interpretations of the results, we worked with 2×2 symmetric matrices, resulting in the vectorized form being 3-dimensional, then to further reduce the computational cost, we projected the vector onto the first two principal components. This is because our covariance matrix is chosen such as most the variation are explained by the first two principal components. Also, this allows us to look at our confidence regions along two principal directions.

For the purposes of this simulation, we simulated data using vectorized means of the form $\theta = [\alpha_1, \alpha_1, \alpha_1]$ for the following values of $\alpha_1 : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, and the covariance matrix of

$$\Sigma = \begin{bmatrix} 5 & .8\sqrt{(}15) & .7\sqrt{(}5) \\ .8\sqrt{(}15) & 3 & .5\sqrt{(}3) \\ .7\sqrt{(}5) & .5\sqrt{(}3) & 1 \end{bmatrix}.$$

Also, we considered the following samples sizes: $n = 6, 10, 20, 40, 80, 160$.

4.2.2 Evaluation

In the Cholesky scale, we evaluated the performance of confidence regions in 9 at a fixed nominal confidence level of 95% and we explored the impact that the mean of the data has on false coverage probabilities similar to the simulations in Chapter 2. We repeatedly simulated the data sets of values of α_1 and n and recorded the proportion of times that the $\mu_T = chol(\mu_c)$ falls in the confidence region. Along the first principal direction, we used μ_T of the form $\mu_1 + (\nu, 0)$ and along the second principal direction, we used μ_T of the form $\mu_1 + (0, \nu)$ where μ_1 is the projected mean onto the first two principal components and ν : $12, -9, -6, -3, -2.5, -2.25, -2, -1.75, -1.5, -1.25, -1, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 3, 6, 9, 12$. Finally, for fixed θ we quantified the distance between coverage probabilities in the Cholesky scale when data are from Cholesky normal and when the data are generated from a normal distribution with the same mean and same covariance matrix as the one that we used to generate the data before projecting it on the two first components. The following samples size were used n=6,10,20,40,80 and 160 and the L^2 distance between the coverage probabilities is defined as:

$$D = \left(\iint_L (C_{CholN} - C_N) \right)^{\frac{1}{2}}$$

C_{CholN} and C_N are respectively coverage probabilities for Cholesky normal in the Cholesky scale and normal distribution generated with the same parameters at any given α_1 value of μ.

4.2.3 Results

Results along the first two principal components for various sample sizes are shown below. The plot to the left of each figure represents the coverage probabilities along the first principal component direction and the plot to the right represents coverage probabilities along the second principal component directions:

In figure 9, when $n = 6$ along the first principal component direction, we clearly observed true coverage probabilities being around the nominal level when the entries of the

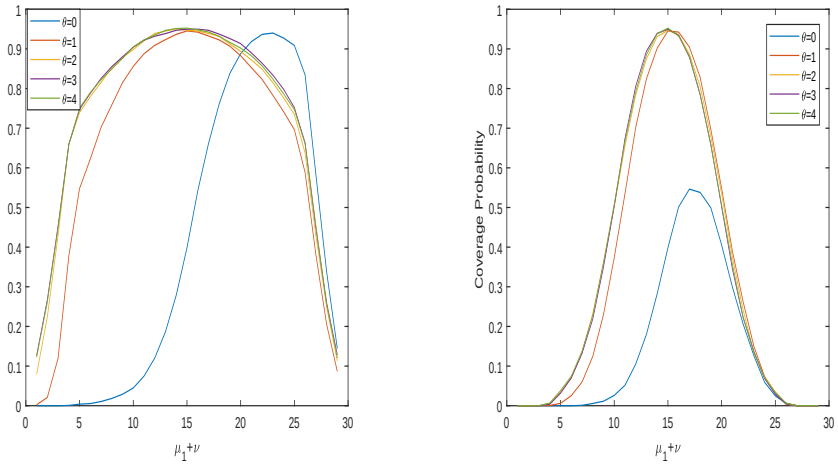

Figure 9: Coverage probabilities curves at different values of θ and when $n = 6$

mean are all one's or greater and false coverage probabilities are also high. The coverage probability curve when the entries of the mean are zeros has a true nominal probability less than 95%. Along the second principal component direction, false coverage probabilities are converging faster to zero because there is lesser variability, but likewise in the first principal component direction, true coverage are all near the nominal level when entries of the mean are all one's or greater and is less than zero when entries of the mean are all zeros.

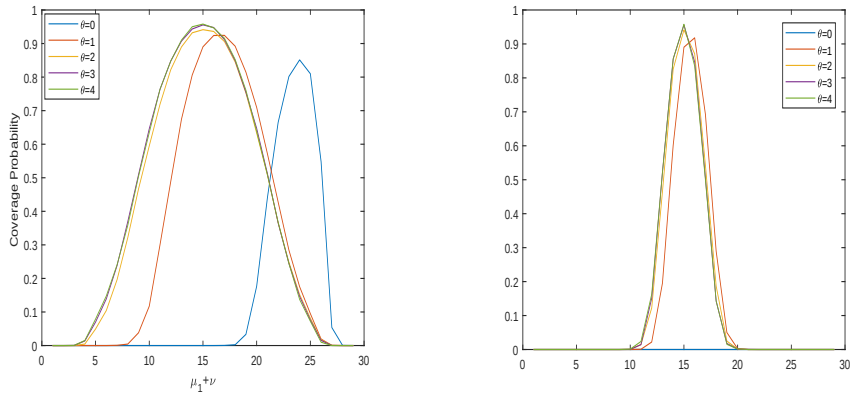

Figure 10: Coverage probabilities curves at different values of θ and when $n = 20$

In figure 10 and 11 when $n = 20$ and $n = 40$ along the first principal components direction, when entries of the mean are all 2 or greater true coverage probabilities are around the nominal level and when the entries of the mean are all less than one true coverage probabilities are less than 95%. On the other hand, false coverage probabilities are gradually converging to zero as the the sample size increases. We observed similar behavior of the coverage probability curve when the entries of the mean are zeros as when $n = 6$.
Along the second principal component direction, we also observed similar behavior as when $n = 6$ except that all coverage probabilities are all zeros even the nominal level when entries of the mean are all zeros.

In figure 12 and 13 when $n = 80$ and $n = 160$ along the two principal component

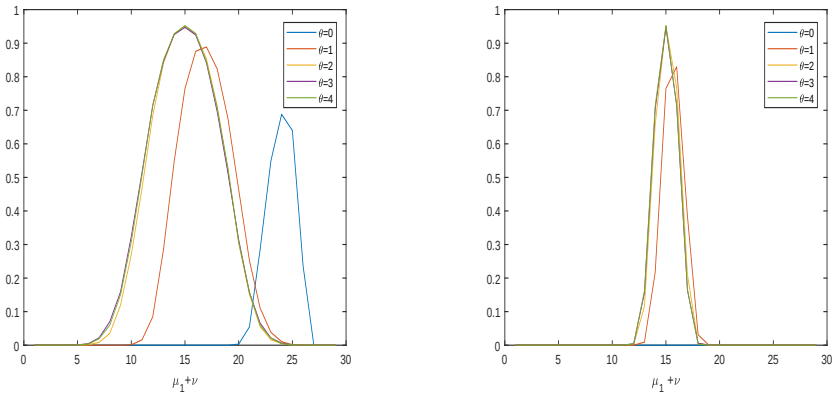

Figure 11: Coverage probabilities curves at different values of θ and when $n = 40$

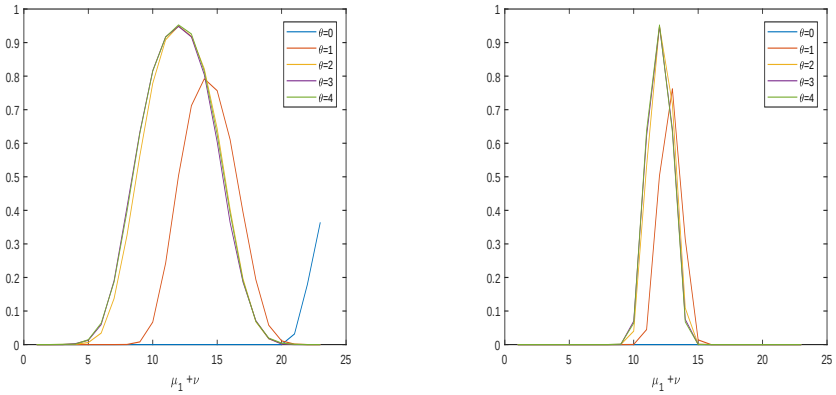

Figure 12: Coverage probability curves at different values of θ and when $n = 80$

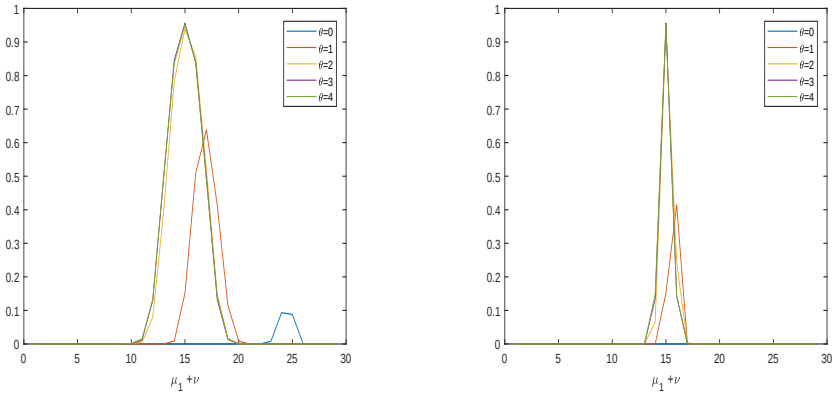

Figure 13: Coverage probabilities curves at different values of θ and when $n = 160$

direction, we observed similar behavior of coverage probabilities curves as when $n = 40$. Coverage probabilities along the second principal component direction are all converging to zero when the entries of the mean are all zeros and one.

To understand and explain the behavior of these coverage probability curves in the Cholesky scale, we decided to first plot the the coverage probability surfaces for a small sample says, $n = 6$ at $\theta = 10$ and at $\theta = 0$. As we see in figure 14 and in 15, when $\theta = 10$ the surface area appears to be elliptical and symmetrical along both directions and when $\theta = 0$ the surface area appears to be shrinking in one direction and it is not symmetrical. This is what motivated us to compare the distance between coverage probabilities in the Cholesky scale when data are from Cholesky normal and when the data are generated from a normal distribution with the same mean and covariance matrix as the one that was used to generate the data. The distance were computed for $\theta = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ and for $X = Y = -6. - 5 - 4. - 3. - 2.5 - 2.25 - 2 - 1.75 - 1.5 - 1.25 - 1. - 0.75 - 0.5 - 0.2500.250.50.7511.251.51.7522.252.53456$.

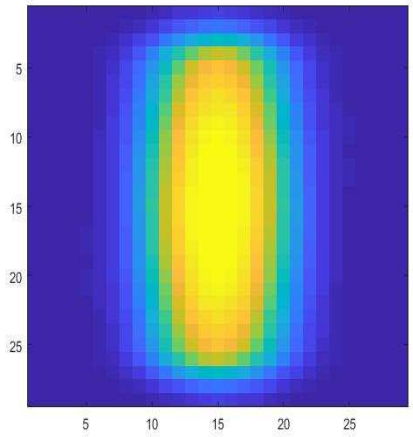

Figure 14: Coverage probability surface $n = 6, \theta = 10$

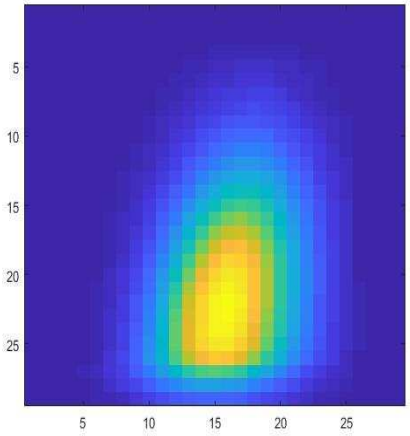

Figure 15: Coverage probability surface $n = 6, \theta = 0$

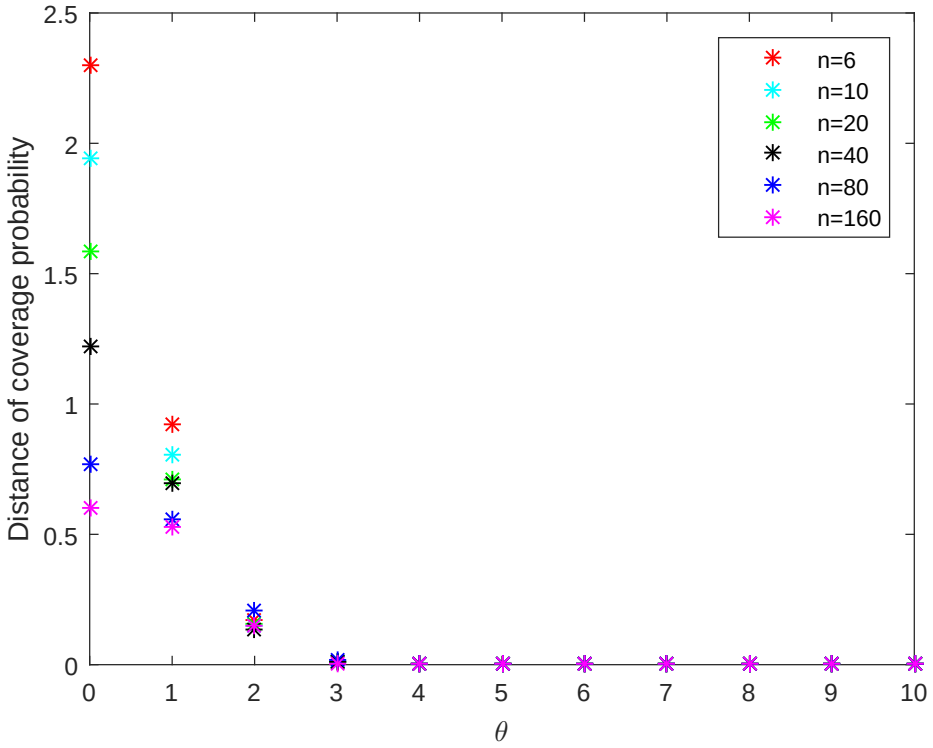

Figure 16: Coverage probability distance for different θ

As we see in figure 16, for fixed sample size and for values of θ near zero the distance between coverage probabilities are high and for large value of θ the distance between the coverage probabilities is zero.

More specifically, for θ around 3 or greater the distance between the coverage probabilities converges to zero, for fixed $\theta \leq 1$ the distance between coverage probabilities are high for small samples size and for fixed $\theta \geq 2$, sample size seems to not have an effect on the distance between coverage probabilities.

To visualize the area between te two curves, for fixed sample size of 6, we plotted the coverage probabilities curve along the x-axis at a given distance of 0.1 0.41 and 0.9 between the normal curve and the Cholesky curve, it obvious that two plots look very similar at a distance of 0.1 as we see in 17, 18 and 19

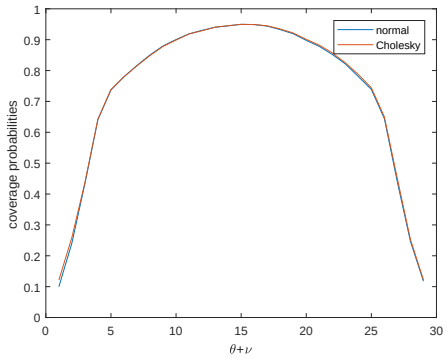

Figure 17: Normal and Cholesky coverage probability plots, distance=0.1

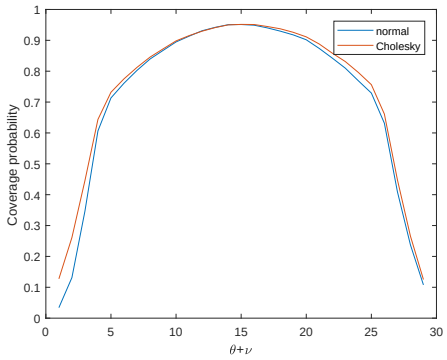

Figure 18: Normal and Cholesky coverage probability plots, distance=0.41

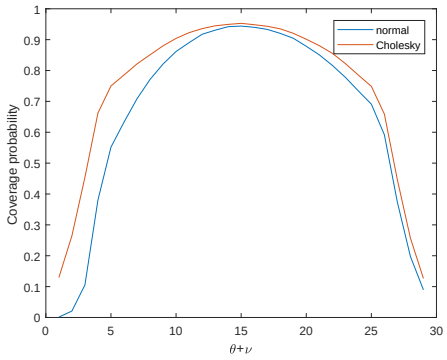

Figure 19: Normal and Cholesky coverage probability plots, distance=0.9

5. Conclusion

5.1 Discussion

From our simulation study we can arrive at the following conclusions about the Cholesky normal distribution assumptions. To generate the Cholesky normal distribution, we must generate a multivariate normal vector so that the location parameter of the Cholesky normal distribution on the Cholesky scale coincides with the mean of the multivariate normal vector. More importantly, the entries of this location parameter must be far from zero. Also, doing inferences on the mean of SPD matrices using Cholesky normal distribution on the Cholesky scale is the same as doing inferences on the mean of the multivariate normal from which the data are generated. So, on the the Cholesky scale, parametric confidence regions for the mean of SPD matrices can be defined as the one for multivariate normal.

Also, for the performance of the Cholesky normal confidence regions we arrived at the following conclusions. For fixed sample size, true coverage probabilities are all near the nominal level for large values of θ and moreover, the distance between coverage probabilities as defined earlier decrease as θ increases. For fixed small $\theta < 3$, distance between coverage probabilities decrease as n increases, and for fixed large $\theta \geq 3$, the distance between coverage probabilities are all zeros. This seemingly to indicate that in general, confidence regions defined via Cholesky normal performs well for large θ regardless of the sample size. Also, for any value of θ above 1.5, Cholesky confidence regions can still perform well. We may need the distance between coverage probability to be roughly around 0.41 or less for these confidence regions to perform well and ideally around 0.1. Finally, as expected for any value of θ between 0 and 1.5, the Cholesky confidence regions perform better for large sample size than small sample size.

5.2 Limitations and Future Work

For Cholesky normal distribution, one of the limitations is that we may not need to put a restriction on each entry of θ For Cholesky normal distribution, in the multivariate case, it will be very difficult the find a threshold for θ_{jj} the mean of the j^{th} diagonal entry of the normal distribution from which the data are generated if the diagonal entries are not independent as in Theorem 4.1.1. Also, unlike the lognormal distribution, for Cholesky normal we don't have enough flexibility when choosing θ which can be problematic for modeling when using real application data.

For the Cholesky normal distribution, more extensive computation may be needed to check if we can have less restriction on θ. We may only need to restrict the entries of θ that corresponds to the diagonal of $Chol(Y)$. We may also consider to check robustness of confidence regions when the distributions in the mixture are both Cholesky normal.

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, *To appear in International Statistical Review*.

[7] B. Ahanda , D. Osborne and L. Ellingson. Robustness of Lognormal Confidence Regions for Means of Symmetric Positive Matrices. *JSM Proceedings*, Institute of Mathematical Statistics Section. Alexandria, VA: American Statistical Association:2511-2521, 2017.

[8] A. Schwartzman (2008) Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices *Ann. Statist*.

[9] I. Dryden, A. Koloydenko, D. Zhou (2009). Non-Euclidean Statistics for Covariance Matrices, with Applications to Diffusion Tensor Imaging. The Annals of Applied Statistics, 3(3), 1102-1123

[10] Harrie Hendriks, Zinoviy Landsman, Mean Location and Sample Mean Location on Manifolds: Asymptotics, Tests, Confidence Regions, Journal of Multivariate Analysis, Volume 67, Issue 2, 1998,