A Note on Angular Central Gaussian Distribution and its Matrix Variant

Kisung You

kisungyou@outlook.com

June 20, 2021

1 Introduction

Probability distributions with explicit forms of densities are core elements of statistical inference. In this note, we review Angular Central Gaussian (ACG) distribution on a hypersphere $\mathbb{S}^{p-1} \subset \mathbb{R}^p$ and its extension - Matrix Angular Central Gaussian (MACG) - defined on Stiefel $St(p,r)$ and Grassmann $Gr(p,r)$ manifolds.

2 Angular Central Gaussian Distribution

ACG distribution $ACG_p(A)$ has a density

$$f_{ACG}(x|A) = |A|^{-1/2}(x^\top A^{-1}x)^{-p/2}, \quad \text{for } x \in \mathbb{S}^{p-1}$$

where A is a symmetric positive-definite matrix - $A = A^\top, A \in \mathbb{R}^{p\times p}, \lambda_{\min}(A) > 0$ and $|\cdot|$ denotes a matrix determinant. Let's recap some properties of ACG distribution.

Property 1. $f_{ACG}(x|A) = f_{ACG}(-x|A)$.

This enables ACG as a distribution on the real projective space $\mathbb{R}P^{p-1} = \mathbb{S}^{p-1}/\{+1,-1\}$.

Property 2. $f_{ACG}(x|A) = f_{ACG}(x|cA),\ c > 0$.

Common convention is to *normalize* the matrix A by a constraint $\text{tr}(A) = p$, which is useful (or essential) in maximum likelihood estimation of the parameter to ensure algorithmic stability. If you want to show this property, simply use the fact that $|cA| = c^p|A|$.

Property 3. When $x \sim \mathcal{N}_p(0,A) \to x/\|x\| \sim ACG_p(A)$.

This property is indeed an intuition behind its origination from [4] and can be used for sampling.

Maximum Likelihood Estimation

Given random samples $x_1, x_2, \ldots, x_p \sim ACG_p(A)$, Tyler (1987) proposed an iterative updating scheme to estimate the parameter A by

$$\hat{A}_{k+1} = p\left\{\sum_{i=1}^{n} \frac{1}{x_i^\top \hat{A}_k^{-1} x_i}\right\}^{-1} \sum_{i=1}^{n} \frac{x_i x_i^\top}{x_i^\top \hat{A}_k^{-1} x_i} \tag{1}$$

where $\hat{A}_k$ is a k-th iterate of an estimator with an initial point $\hat{A}_0 = I_p$ is set as an identity matrix. While the formula (1) guarantees the convergence under mild conditions and abides by the constraint $\mathrm{tr}(\hat{A}_k) = p$, it is from the author's previous work on M-estimation of the scatter matrix. Here, we provide a naive derivation of 2-step fixed-point iteration algorithm for pedagogical purpose.

$$\hat{A}_{k'} = \frac{p}{n}\sum_{i=1}^{n} \frac{x_i x_i^\top}{x_i^\top \hat{A}_k^{-1} x_i} \quad \text{and} \quad \hat{A}_{k+1} = \frac{p}{\mathrm{tr}(\hat{A}_{k'})}\hat{A}_{k'} \tag{2}$$

First, let's write the log-likelihood

$$\log L = \log\left(\prod_{i=1}^{n} f(x_i|A)\right) = -\frac{n}{2}\log\det(A) - \frac{p}{2}\sum_{i=1}^{n}\log(x_i^\top \hat{A}_k^{-1} x_i)$$

and recall two facts from matrix calculus [3] that

$$\frac{\partial \log\det(A)}{\partial A} = A^{-1} \quad \text{and} \quad \frac{\partial x^\top A^{-1} x}{\partial A} = -A^{-1}xx^\top A^{-1}.$$

Then, the first-order condition for the log-likelihood can be written as

$$\begin{aligned}\frac{\partial \log L}{\partial A} &= -\frac{n}{2}A^{-1} + \frac{p}{2}\sum_{i=1}^{n}\frac{A^{-1}x_i x_i^\top A^{-1}}{x_i^\top A^{-1} x_i}\\ A^{-1} &= \frac{p}{n}\sum_{i=1}^{n}\frac{A^{-1}x_i x_i^\top A^{-1}}{x_i^\top A^{-1} x_i}\\ A &= \frac{p}{n}\sum_{i=1}^{n}\frac{x_i x_i^\top}{x_i^\top A^{-1} x_i}\end{aligned}$$

where the last equality comes from multiplying A from left and right. Therefore, $\hat{A}$ is a solution of the matrix equation in a form $X = f(X)$ where f is a contraction mapping under some conditions [4]. This leads to the formula (2) while projection step is added to keep $\mathrm{tr}(\hat{A}_k) = p$ for all $k = 1, 2, \cdots$.

3 Matrix Angular Central Gaussian Distribution

Chikuse (1990) [1] extended the distribution to the matrix case, namely Stiefel and Grassmann manifolds

$$\begin{aligned}St(p,r) &= \{X \in \mathbb{R}^{p\times r} \mid X^\top X = I_p\}\\ Gr(p,r) &= \{\mathrm{Span}(X) \mid X \in \mathbb{R}^{p\times r},\ \mathrm{rank}(X) = r\}\end{aligned}$$

which are sets of orthonormal k-frames and k-subspaces. The Matrix Angular Central Gaussian (MACG) distribution $MACG_{p,r}(\Sigma)$ has a density

$$f_{MACG}(X|\Sigma) = |\Sigma|^{-r/2}|X^\top\Sigma^{-1}X|^{-p/2}$$

where Σ is a symmetric positive-definite matrix. Note that the density is very similar to what we had before for vector-valued distribution. Likewise, it shares properties as before.

Property 1. $f_{MACG}(X|\Sigma) = f_{MACG}(-X|\Sigma)$.

Property 2. $f_{MACG}(X|\Sigma) = f_{MACG}(X|c\Sigma),\ c > 0$.

Property 3. $f_{MACG}(X|\Sigma) = f_{MACG}(XR|\Sigma)$ for $R \in O(r)$.

This property enables to consider MACG as a distribution on Grassmann manifold, which are quotient by modulo orthogonal transformation.

Sampling from MACG

In order to draw random samples from $MACG_{p,r}(\Sigma)$, we need the following steps, which are common in directional statistics with Stiefel/Grassmann manifolds [2]. First, draw r random vectors $x_1, \ldots, x_r \sim \mathcal{N}_p(0, \Sigma)$ and stack them as columns $X = [x_1|\cdots|x_r] \in \mathbb{R}^{p\times r}$. Then,

$$Y = X(X^\top X)^{-1/2} \sim MACG_{p,r}(\Sigma)$$

where the negative square root for a symmetric positive-definite matrix can be obtained from eigen-decomposition,

$$\Omega = UDU^\top \to \Omega^{-1/2} = UD^{-1/2}U^\top,\ \left[D^{-1/2}\right]_{ij} = \frac{1}{\sqrt{d_i j}} \text{ when } i = j \text{ and } 0 \text{ otherwise.}$$

Maximum Likelihood Estimation

Similarly, given random samples $X_1, X_2, \ldots, X_n \sim MACG_{p,r}(\Sigma)$, we can obtain a two-step iterative scheme to estimate the parameter Σ,

$$\boxed{\hat{\Sigma}_{k'} = \frac{p}{nr}\sum_{i=1}^{n} X_i(X_i^\top \Sigma^{-1} X_i)^{-1} X_i \quad \text{and} \quad \hat{\Sigma}_{k+1} = \frac{p}{\operatorname{tr}(\hat{\Sigma}_{k'})}\hat{\Sigma}_{k'}.} \tag{3}$$

Derivation of formula (3) follows the similar line as (2). We need another fact from matrix calculus that

$$\frac{\partial}{\partial \Sigma} \log\det(X^\top \Sigma^{-1} X) = -\Sigma^{-1}X(X^\top\Sigma^{-1}X)^{-1}X^\top\Sigma^{-1}.$$

First, log-likelihood is written as

$$\log L = \log\left(\prod_{i=1}^{n} f(X_i|\Sigma)\right) = -\frac{nr}{2}\log\det(\Sigma) - \frac{p}{2}\sum_{i=1}^{n}\log\det(X_i^\top\Sigma^{-1}X_i)$$

where the first-order condition gives

$$\begin{aligned}
\frac{\partial \log L}{\partial \Sigma} &= -\frac{nr}{2}\Sigma^{-1} + \frac{p}{2}\sum_{i=1}^{n}\left(\Sigma^{-1}X_i(X_i^\top\Sigma^{-1}X_i)^{-1}X_i^\top\Sigma^{-1}\right) \\
\frac{nr}{2}\Sigma^{-1} &= \frac{p}{2}\sum_{i=1}^{n}\left(\Sigma^{-1}X_i(X_i^\top\Sigma^{-1}X_i)^{-1}X_i^\top\Sigma^{-1}\right) \\
\Sigma &= \frac{p}{nr}\sum_{i=1}^{n} X_i(X_i^\top\Sigma^{-1}X_i)^{-1}X_i^\top
\end{aligned}$$

where the last equality comes from multiplying Σ from left and right. Therefore, $\hat{\Sigma}$ is a solution of the matrix equation, leading to the formula (3) with an additional projection step to keep $\mathrm{tr}(\hat{\Sigma}_k) = p$ for all $k = 1, 2, \cdots$. Note that this matrix equation, up to my knowledge, has not known whether the mapping is contraction or not.

4 Conclusion

ACG and MACG distributions are simple yet rather little used in directional statistics. We hope that this brief note boosts probabilistic inference on corresponding manifolds at ease. An R package Riemann, which is also available on CRAN, implements density evaluation, random sample generation, and maximum likelihood estimation of the scatter parameters A and Σ in the light of expecting handy utilization of the distributions we introduced.

References

[1] Yasuko Chikuse. The matrix angular central Gaussian distribution. *Journal of Multivariate Analysis*, 33(2):265–274, May 1990.

[2] Kanti V. Mardia and Peter E. Jupp, editors. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, January 1999.

[3] K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. Version 20121115.

[4] David E. Tyler. Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, 74(3):579–589, 1987.