Distributions for Positive Definite Matrices

Armin Schwartzman, Harvard Biostatistics

November 30, 2006

Outline

- Why Wishart is not enough
- Truncated normal
- Log-Euclidean log-normal
- Riemannian log-normal
- Geodesic normal

Wishart

Definition: A matrix $X_{p\times p} \in \mathrm{Sym}^+(p)$ has a Wishart distribution, $X \sim W_p(n,\Sigma)$, $n \ge p$, if it has density

$$f(X;n,\Sigma) = \frac{|X|^{(n-p-1)/2}}{2^{np/2}|\Sigma|^{n/2}\Gamma_p(n/2)} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}X)\right)$$

where

$$\Gamma_p(n/2) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma[(n-j+1)/2]$$

is the multivariate Gamma function.

Gamma

Let $X \sim W_p(n, \Sigma)$.

When $p = 1$ we have $x \sim \sigma^2 \chi^2(n)$ with density

$$f(x; n, \sigma) = \frac{|x|^{n/2-1}}{(2\sigma^2)^{n/2}\Gamma(n/2)} \exp\left(-\frac{x}{2\sigma^2}\right)$$

Replacing $n/2$ by $\nu \in \mathbb{R}$ gives the gamma density

$$f(x; \nu, \sigma) = \frac{|x|^{\nu-1}}{(2\sigma^2)^{\nu}\Gamma(\nu)} \exp\left(-\frac{x}{2\sigma^2}\right)$$

with shape parameter ν and scaling parameter $2\sigma^2$.

Multivariate Gamma

Similarly for the Wishart, replacing $n/2$ by $\nu \in \mathbb{R}$, $\nu > 2(p-1)$, gives the multivariate gamma density

$$f(X;\nu,\Sigma) = \frac{|X|^{\nu-(p+1)/2}}{2^{\nu p}|\Sigma|^{\nu}\Gamma_p(\nu)} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}X)\right)$$

where

$$\Gamma_p(\nu) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma[\nu-(j-1)/2]$$

A similar generalization can be done for the inverse Wishart.

Limitations

The multivariate Gamma has mean

$$E(X) = 2\nu\Sigma \in \mathrm{Sym}^+(p)$$

but the shape parameter ν is a scalar, so it creates a very specific covariance structure between the matrix elements.

Can we generalize?

Gaussian-based distributions

Strong connection between $\mathrm{Sym}^+(p)$ and $\mathrm{Sym}(p)$:

1. $\mathrm{Sym}(p)$ is a superset of $\mathrm{Sym}^+(p)$.
2. $\mathrm{Sym}(p)$ is the image under the matrix log of $\mathrm{Sym}^+(p)$.
3. $\mathrm{Sym}(p)$ is the tangent space to $\mathrm{Sym}^+(p)$.

Gaussian-based distributions

This gives many ways to construct probability distributions on $\mathrm{Sym}^+(p)$:

1. Define a multivariate normal distribution on $\mathrm{Sym}(p)$.
   (a) Vectorization
   (b) Matrix notation
2. Apply to $\mathrm{Sym}^+(p)$ by:
   (a) Truncation (Euclidean)
   (b) Straight exponential projection (Log-Euclidean)
   (c) Riemannian exponential projection (Riemannian)

Vectorized normal for symmetric matrices

Definition: $Y \in \mathrm{Sym}(p)$ has a *vectorized symmetric matrix variate normal distribution* with mean $M \in \mathrm{Sym}(p)$ and covariance $\Sigma_{q\times q} \in \mathrm{Sym}^+(q)$, with $q = p(p+1)/2$, if

$$\mathrm{vecd}(Y) \sim N_q\big(\mathrm{vecd}(M), \Sigma_{q\times q}\big).$$

The vectorization is such that the covariance decomposes nicely as

$$\Sigma = \begin{pmatrix} \Sigma_{\mathrm{diag}} & \Sigma_{\mathrm{diag,offdiag}} \\ \Sigma_{\mathrm{offdiag,diag}} & \Sigma_{\mathrm{offdiag}} \end{pmatrix}$$

Truncated vectorized normal for positive definite matrices

Definition: $X \in \mathrm{Sym}^+(p)$ has a *positive definite matrix variate truncated vectorized normal distribution* with mean $M \in \mathrm{Sym}^+(p)$ and $\sigma^2 > 0$, if it has density

$$f(X; M, \sigma^2) = C(M, \Sigma)\, \mathbf{1}(X \succ 0)\, N_q\big(\mathrm{vecd}(M), \Sigma_{q\times q}\big)$$

Truncated vectorized normal for positive definite matrices

Truncation:

- If most of the probability mass around M (as determined by Σ) is far away from the boundaries, then C may hopefully be close to 1 and the truncation may be ignored.
- C is difficult to compute because it is an integral over $\mathrm{Sym}^+(p)$, although Bayesians may not care.
- This setup imposes a Euclidean geometry on $\mathrm{Sym}^+(p)$.

Vectorization:

- The vectorization eliminates the symmetry and positive definiteness, cannot do eigenvalue decompositions, etc. Can we have a matrix notation?

Standard normal for symmetric matrices

Definition: We say that $Z \in \mathrm{Sym}(p)$ has the *standard symmetric matrix variate normal distribution* $N_{pp}(0, I_p)$ (also called Gaussian Orthogonal Ensemble) if it has density

$$\varphi_{pp}(Z) = \frac{1}{(2\pi)^{q/2}} \exp\left(-\frac{1}{2}\mathrm{tr}(Z^2)\right)$$

where $q = p(p+1)/2$.

This is the product of q univariate independent normals. For example, for $p = 3$, Z is constructed as

$$Z = \begin{pmatrix} N(0,1) & N(0,1/2) & N(0,1/2) \\ * & N(0,1) & N(0,1/2) \\ * & * & N(0,1) \end{pmatrix}$$

Non-standard normal for symmetric matrices

For $p = 1$, we create $y \sim N(\mu, \sigma^2)$ by defining

$$y = \sigma z + \mu, \qquad z \sim N(0,1)$$

For general p, start with $Z \sim N_{pp}(0, I_p)$ and define $Y \in \mathrm{Sym}(p)$ by the group action of $GL(p)$ plus an offset as

$$Y = GZG' + M$$

where $G \in GL(p)$ and $M \in \mathrm{Sym}(p)$ (Chikuse, 2000).

Non-standard normal for symmetric matrices

Replacing in the exponent of the GOE gives

$$\mathrm{tr}(Z^2) = \mathrm{tr}\big(G^{-1}(Y-M)(G')^{-1}\big)^2 = \mathrm{tr}\big((Y-M)(G'G)^{-1}\big)^2$$

with Jacobian

$$J(Z \to Y) = |G|^{-(p+1)} = |G'G|^{-(p+1)/2}$$

(Fang and Zhang, 1990).

Notice that both the exponent and the Jacobian depend on G through the single matrix $G'G = \Sigma \in \mathrm{Sym}^+(p)$.

Non-standard normal for symmetric matrices

Definition: We say that $Y \in \mathrm{Sym}(p)$ has the *symmetric matrix variate normal distribution* $N_{pp}(M, \Sigma)$, with $M \in \mathrm{Sym}(p)$ and $\Sigma \in \mathrm{Sym}^+(p)$, if it has density

$$f(Y; M, \Sigma) = \frac{1}{(2\pi)^{q/2} |\Sigma|^{(p+1)/2}} \exp\left(-\frac{1}{2} \mathrm{tr}\big((Y - M)\Sigma^{-1}\big)^2 \right)$$

Matrix exponential and logarithm

Definition: Let $Y \in GL(p)$. Then

$$X = \exp(Y) = \sum_{k=0}^{\infty} \frac{Y^k}{k!}$$

Consequences:

- If $Y \in \mathrm{Sym}(p)$ and $Y = VLV'$ is any eigen-decomposition, then

$$X = \exp(Y) = V \exp(L) V' \ \in \mathrm{Sym}^+(p)$$

- If $X \in \mathrm{Sym}^+(p)$ and $X = V\Lambda V'$ is any eigen-decomposition, then

$$Y = \log(X) = V \log(\Lambda) V' \ \in \mathrm{Sym}(p)$$

Both functions are inverse of each other and one-to-one.

Log normal for positive definite matrices

Definition: We say that $X \in \mathrm{Sym}(p)$ has the *positive definite matrix variate log normal distribution* $N_{pp}(M, \Sigma)$, with $M \in \mathrm{Sym}(p)$ and $\Sigma \in \mathrm{Sym}^+(p)$, if

$$Y = \log(X) \in \mathrm{Sym}(p) \sim N_{pp}(M, \Sigma)$$

with density

$$f(X; M, \Sigma) = \frac{J(X)}{(2\pi)^{q/2}|\Sigma|^{(p+1)/2}} \exp\left(-\frac{1}{2}\mathrm{tr}\big((\log X - M)\Sigma^{-1}\big)^2\right)$$

Log normal for positive definite matrices

The Jacobian of the transformation $Y = \log(X)$ is equal to

$$J(X) = \mathcal{J}(Y \to X) = \left| \frac{\partial Y}{\partial X} \right| = \frac{1}{\lambda_1 \ldots \lambda_p} \prod_{i<j} \frac{\log \lambda_j - \log \lambda_i}{\lambda_j - \lambda_i}$$

where $\lambda_1 > \ldots > \lambda_p$ are the eigenvalues of X.

Riemannian log normal for positive definite matrices

The standard matrix log transformation

$$Y = \log(X)$$

is a special case of the Riemannian log transformation

$$Y_M = \mathrm{Log}_M(X) = M^{1/2} \log(M^{-1/2} X M^{-1/2}) M^{1/2}$$

which is the projection of X onto the tangent space at M.

The standard log is the Riemannian log at $M = I$.

Riemannian log normal for positive definite matrices

Definition: We say that $X \in \mathrm{Sym}^+(p)$ has a *positive definite matrix variate Riemannian log normal distribution* with parameters $M \in \mathrm{Sym}^+(p)$ and covariance $\Sigma \in \mathrm{Sym}^+(q)$ if

$$Y_M = \mathrm{Log}_M(X) \sim N_{pp}(0, \Sigma)$$

Notes:

- The location offset M in the normal is no longer necessary, as it is replaced by the tangent point M.
- Interestingly, now there is redundancy in Σ.

Riemannian log normal for positive definite matrices

Consider the exponent of the density:

$$
\begin{aligned}
-\frac{1}{2}\mathrm{tr}\big(Y_M\,\Sigma^{-1}\big)^2 &= -\frac{1}{2}\mathrm{tr}\big(\mathrm{Log}_M(X)\,\Sigma^{-1}\big)^2 \\
&= -\frac{1}{2}\mathrm{tr}\big(\log(M^{-1/2}XM^{-1/2})(M^{-1/2}\Sigma M^{-1/2})^{-1}\big)^2
\end{aligned}
$$

So M seems to affect the covariance.

Consider the case $\Sigma = \sigma I$. The exponent becomes:

$$
\begin{aligned}
-\frac{1}{2}\mathrm{tr}\big(\mathrm{Log}_M(X)\,\Sigma^{-1}\big)^2 &= -\frac{1}{2\sigma^2}\mathrm{tr}\big(\mathrm{Log}_M(X)\big)^2 \\
&= -\frac{1}{2\sigma^2}d^2(M,X)
\end{aligned}
$$

This leads to the following general idea.

Geodesic normal for positive definite matrices

Definition: We say that $X \in \mathrm{Sym}^+(p)$ has a *positive definite matrix variate geodesic normal distribution* with parameter $M \in \mathrm{Sym}^+(p)$ if it has density

$$f(X; M) \propto \exp\left(-\frac{1}{2\sigma^2} d^2(M, X)\right)$$

The MLE of this distribution would be the intrinsic mean, the matrix $\hat{M}$ that minimizes the square geodesic distance.

Summary

Different distributions on the positive definite matrices may be obtained according to the geometry used.

- Vectorization is very general, but we loose track of the matrix.
- Euclidean geometry leads to the truncated normal.
- Log-Euclidean geometry leads to the log-normal using the standard log.
  - Well defined parameters M and Σ.
  - Need to better understand the covariance structure.
- Riemannian geometry leads to the Riemannian log-normal and the geodesic normal.
  - Not well defined parameters M and Σ.
  - Need to better understand the covariance structure.