Efficient implementation of MVN and multivariate-t distributions

Chi Feng (chifeng@mit.edu) June 26, 2015

Multivariate normal distribution

Evaluating the (log) density

A d-dimensional multivariate normal distribution has density

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

and log density

$$\log f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \underbrace{-\tfrac{1}{2} d \log(2\pi)}_{\text{constant}} - \tfrac{1}{2}\log[\det(\boldsymbol{\Sigma})] - \tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \quad (1)$$

Using the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ of the positive definite (semi-definite) covariance matrix, we can rewrite the determinant in (1) as

$$\det(\boldsymbol{\Sigma}) = \det(\mathbf{L}) \cdot \det(\mathbf{L}^T) = \det(\mathbf{L})^2,$$

and since $\mathbf{L}$ is lower-triangular, its determinant is just the product of its diagonal terms, so the log of the determinant can be written as a sum (ℓ_1-norm) of logs

$$\log[\det(\mathbf{L})^2] = 2\|\log[\text{diag}(\mathbf{L})]\|_1 \quad \Longrightarrow \quad \log[\det(\boldsymbol{\Sigma})] = 2\|\log[\text{diag}(\mathbf{L})]\|_1 \quad (2)$$

The Cholesky factorization also leads to a Cholesky-like factorization of the inverse $\boldsymbol{\Sigma}^{-1} = \mathbf{S}^T\mathbf{S}$ with the lower triangular matrix $\mathbf{S} = \mathbf{L}^{-1}$. This lets us rewrite the $\boldsymbol{\Sigma}^{-1}$ term as

$$\begin{aligned}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= (\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{L}^{-T}\mathbf{L}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ &= \|\mathbf{L}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\|_2. \quad (3)\end{aligned}$$

By combining (2) and (3), the log density in a more computationally efficient form

$$\log f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{constant} - \|\log[\text{diag}(\mathbf{L})]\|_1 - \tfrac{1}{2}\|\mathbf{L}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\|_2 \quad (4)$$

where the triangular system $\mathbf{L}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ can be solved efficiently using forward substitution, e.g., by using the `dtrsv` subroutine from Level 2 BLAS.

Sampling

We can also draw samples of a random variable $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ according to

$$\boldsymbol{X} = \boldsymbol{\mu} + \mathbf{L}\boldsymbol{Z} \quad (5)$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ is a d-dimensional random vector with Z_i being independent standard normal random variables.

Multivariate-t distribution

A d-dimensional multivariate-t distribution with ν degrees of freedom has density

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma[(\nu+d)/2]}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}\det(\boldsymbol{\Sigma})^{1/2}[1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^{(\nu+d)/2}}$$

and log density

$$\begin{aligned}\log f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = &\overbrace{\text{gammaln}[(\nu+d/2)] - \text{gammaln}(\nu/2) - \tfrac{d}{2}\log\nu - \tfrac{d}{2}\log\pi}^{\text{constant}} \\ &- \tfrac{1}{2}\log[\det(\boldsymbol{\Sigma})] - \tfrac{\nu+d}{2}\log\big[1+\tfrac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big]\end{aligned}$$

Using the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$, we can rewrite the log density as

$$\log f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \text{constant} - \|\log[\text{diag}(\mathbf{L})]\|_1 - \tfrac{\nu+d}{2}\,\text{log1p}\big[\tfrac{1}{\nu}\|\mathbf{L}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\|_2\big] \tag{6}$$

where log1p(·) is a function found in most numerical libraries that uses the Taylor expansion

$$\log(1+p) = p - \frac{p^2}{2} + \mathcal{O}(p^3)$$

for very small p to avoid numerical underflow.

Sampling

We can draw samples $\boldsymbol{X} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ similarly to (5), but with an inverse χ^2 scaling factor

$$\boldsymbol{X} \sim \boldsymbol{\mu} + \sqrt{\nu/Y}\,\mathbf{L}\,\boldsymbol{Z} \tag{7}$$

where Y is a χ^2-distributed random variable with ν degrees of freedom and $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ is a d-dimensional random vector with Z_i being independent standard normal random variables.

Note: The matrix $\boldsymbol{\Sigma}$ is *not* the covariance matrix of the multivariate-t distribution. Instead, $\boldsymbol{\Sigma}$ is the "scale matrix." The actual covariance is $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$ if $\nu > 2$, otherwise it is undefined.