

Gibbs Sampler for the Truncated Multivariate Normal Distribution

Stefan Wilhelm*

April 6, 2012

In this note we describe two ways of generating random variables with the Gibbs sampling approach for a truncated multivariate normal variable \mathbf{x} , whose density function can be expressed as:

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b}) = \frac{\exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}}{\int_{\mathbf{a}}^{\mathbf{b}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\} d\mathbf{x}}$$

for $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$ and 0 otherwise.

The first approach, as described by Kotecha and Djuric [1999], uses the covariance matrix $\boldsymbol{\Sigma}$ and has been implemented in the R package **tmvtnorm** since version 0.9 (Wilhelm and Manjunath [2010]). The second way is based on the works of Geweke [1991, 2005] and uses the precision matrix $\mathbf{H} = \boldsymbol{\Sigma}^{-1}$. As will be shown below, the usage of the precision matrix offers some computational advantages, since it does not involve matrix inversions and is therefore favorable in higher dimensions and settings where the precision matrix is readily available. Applications are for example the analysis of spatial data, such as from telecommunications or social networks.

Both versions of the Gibbs sampler can also be used for general linear constraints $\mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}$, what we will show in the last section. The function `rtmvtnorm()` in the package **tmvtnorm** contains the R implementation of the methods described in this note (Wilhelm and Manjunath [2011]).

1 Gibbs Sampler with covariance matrix $\boldsymbol{\Sigma}$

We describe here a Gibbs sampler for sampling from a truncated multinormal distribution as proposed by Kotecha and Djuric [1999]. It uses the fact that conditional distributions are truncated normal again. Kotecha use full conditionals $f(x_i | x_{-i}) = f(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$.

We use the fact that the conditional density of a multivariate normal distribution is multivariate normal again. We cite Geweke [2005], p.171 for the following theorem on the Conditional Multivariate Normal Distribution.

Let $\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}$

Denote the corresponding precision matrix

$$\mathbf{H} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{H}_{xx} & \mathbf{H}_{xy} \\ \mathbf{H}_{yx} & \mathbf{H}_{yy} \end{bmatrix} \quad (1)$$

*wilhelm@financial.com

Then the distribution of \mathbf{y} conditional on \mathbf{x} is normal with variance

$$\Sigma_{y.x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} = H_{yy}^{-1} \quad (2)$$

and mean

$$\mu_{y.x} = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) = \mu_y - H_{yy}^{-1}H_{yx}(\mathbf{x} - \mu_x) \quad (3)$$

In the case of the full conditionals $f(x_i|x_{-i})$, which we will denote as $i. - i$ this results in the following formulas: $\mathbf{z} = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_{-i} \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{pmatrix} \mu_i \\ \mu_{-i} \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{ii} & \Sigma_{i,-i} \\ \Sigma_{-i,i} & \Sigma_{-i,-i} \end{bmatrix}$. Then the distribution of i conditional on $-i$ is normal with variance

$$\Sigma_{i.-i} = \Sigma_{ii} - \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}\Sigma_{-i,i} = H_{ii}^{-1} \quad (4)$$

and mean

$$\mu_{i.-i} = \mu_i + \Sigma_{i,-i}\Sigma_{-i,-i}^{-1}(\mathbf{x}_{-i} - \mu_{-i}) = \mu_i - H_{ii}^{-1}H_{i,-i}(\mathbf{x}_{-i} - \mu_{-i}) \quad (5)$$

We can then construct a Markov chain which continuously draws from $f(x_i|x_{-i})$ subject to $a_i \leq x_i \leq b_i$. Let $\mathbf{x}^{(j)}$ denote the sample drawn at the j -th MCMC iteration. The steps of the Gibbs sampler for generating N samples $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$ are:

- Since the conditional variance $\Sigma_{i.-i}$ is independent from the actual realisation $\mathbf{x}_{-i}^{(j)}$, we can well precalculate it before running the Markov chain.
- Choose a start value $\mathbf{x}^{(0)}$ of the chain.
- In each round $j = 1, \dots, N$ we go from $i = 1, \dots, d$ and sample from the conditional density $x_i^{(j)} | x_1^{(j)}, \dots, x_{i-1}^{(j)}, x_{i+1}^{(j-1)}, \dots, x_d^{(j-1)}$.
- Draw a uniform random variate $U \sim \text{Uni}(0,1)$. This is where our approach slightly differs from Kotecha and Djuric [1999]. They draw a normal variate y and then apply $\Phi(y)$, which is basically uniform.
- We draw from univariate conditional normal distributions with mean μ and variance σ^2 . See for example Greene [2003] or Griffiths [2004] for a transformation between a univariate normal random $y \sim N(\mu, \sigma^2)$ and a univariate truncated normal variate $x \sim \text{TN}(\mu, \sigma^2, a, b)$. For each realisation y we can find a x such as $P(Y \leq y) = P(X \leq x)$:

$$\frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} = \Phi\left(\frac{y-\mu}{\sigma}\right) = U$$

- Draw $\mathbf{x}_{i.-i}$ from conditional univariate truncated normal distribution $\text{TN}(\mu_{i.-i}, \Sigma_{i.-i}, a_i, b_i)$ by

$$\mathbf{x}_{i.-i} = \mu_{i.-i} + \sigma_{i.-i}\Phi^{-1}\left[U\left(\Phi\left(\frac{b_i - \mu_{i.-i}}{\sigma_{i.-i}}\right) - \Phi\left(\frac{a_i - \mu_{i.-i}}{\sigma_{i.-i}}\right)\right) + \Phi\left(\frac{a_i - \mu_{i.-i}}{\sigma_{i.-i}}\right)\right] \quad (6)$$

2 Gibbs Sampler with precision matrix \mathbf{H}

The Gibbs Sampler stated in terms of the precision matrix $\mathbf{H} = \boldsymbol{\Sigma}^{-1}$ instead of the covariance matrix $\boldsymbol{\Sigma}$ is much easier to write and to implement: Then the distribution of i conditional on $-i$ is normal with variance

$$\Sigma_{i.-i} = H_{ii}^{-1} \quad (7)$$

and mean

$$\mu_{i,-i} = \mu_i - H_{ii}^{-1} H_{i,-i} (x_{-i} - \mu_{-i}) \quad (8)$$

Most importantly, if the precision matrix \mathbf{H} is known, the Gibbs sampler does only involve matrix inversions of \mathbf{H}_{ii} which in our case is a diagonal element/scalar. Hence, from the computational and performance perspective, especially in high dimensions, using \mathbf{H} rather than $\mathbf{\Sigma}$ is preferable. When using $\mathbf{\Sigma}$ in d dimensions, we have to solve for $d(d-1) \times (d-1)$ matrices $\mathbf{\Sigma}_{-i,-i}$, $i = 1, \dots, d$, which can be quite substantial computations.

3 Gibbs Sampler for linear constraints

In this section we present the Gibbs sampling for general linear constraints based on Geweke [1991]. We want to sample from $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$ subject to linear constraints $\mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}$ for a full-rank matrix \mathbf{D} .

Defining

$$\mathbf{z} = \mathbf{D}\mathbf{x} - \mathbf{D}\boldsymbol{\mu}, \quad (9)$$

we have $E[\mathbf{z}] = \mathbf{D}E[\mathbf{x}] - \mathbf{D}\boldsymbol{\mu} = \mathbf{0}$ and $Var[\mathbf{z}] = \mathbf{D}Var[\mathbf{x}]\mathbf{D}' = \mathbf{D}\mathbf{\Sigma}\mathbf{D}'$. Hence, this problem can be transformed to the rectangular case $\boldsymbol{\alpha} \leq \mathbf{z} \leq \boldsymbol{\beta}$ with $\boldsymbol{\alpha} = \mathbf{a} - \mathbf{D}\boldsymbol{\mu}$ and $\boldsymbol{\beta} = \mathbf{b} - \mathbf{D}\boldsymbol{\mu}$. It follows $\mathbf{z} \sim N(\mathbf{0}, \mathbf{T})$ with $\mathbf{T} = \mathbf{D}\mathbf{\Sigma}\mathbf{D}'$.

In the precision matrix case, the corresponding precision matrix of the transformed problem will be $\mathbf{T}^{-1} = (\mathbf{D}\mathbf{\Sigma}\mathbf{D}')^{-1} = \mathbf{D}'^{-1}\mathbf{H}\mathbf{D}^{-1}$. We can then sample from \mathbf{z} the way described in the previous sections (either with covariance or precision matrix approach) and then transform \mathbf{z} back to \mathbf{x} by

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{D}^{-1}\mathbf{z} \quad (10)$$

References

- John F. Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. <http://www.biz.uiowa.edu/faculty/jgeweke/papers/paper47/paper47.pdf>, 1991.
- John F. Geweke. *Contemporary Bayesian Econometrics and Statistics*. John Wiley and Sons, 2005.
- William H. Greene. *Econometric Analysis*. Prentice-Hall, 5 edition, 2003.
- William E. Griffiths. *Contemporary Issues In Economics And Econometrics: Theory and Application*, chapter A Gibbs' sampler for the parameters of a truncated multivariate normal distribution, pages 75 – 91. Edward Elgar Publishing, 2004.
- J. H. Kotecha and P. M. Djuric. Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In *ICASSP '99: Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*, pages 1757–1760, Washington, DC, USA, 1999. IEEE Computer Society. ISBN 0-7803-5041-3. doi: <http://dx.doi.org/10.1109/ICASSP.1999.756335>.
- Stefan Wilhelm and B G Manjunath. *tmvtnorm: Truncated Multivariate Normal Distribution*, 2010. URL <http://CRAN.R-project.org/package=tmvtnorm>. R package version 0.9-2.
- Stefan Wilhelm and B G Manjunath. *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*, 2011. URL <http://CRAN.R-project.org/package=tmvtnorm>. R package version 1.3-1.