Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex

Nicolas Dobigeon and Jean-Yves Tourneret
E-mail : {Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr

TECHNICAL REPORT – 2007, March

IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

I. Problem statement

Let $\mathbb{S}$ denote the following simplex defined on $\mathbb{R}^{R-1}$:

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \middle| \alpha_r \geq 0,\ \forall r = 1, \ldots, R-1,\ \sum_{r=1}^{R-1} \alpha_r \leq 1 \right\}, \tag{1}$$

Denote $\mathcal{N}_{\mathbb{S}}(\boldsymbol{A}, \boldsymbol{B})$ as the truncated multivariate Gaussian distribution (MGD) defined on the simplex $\mathbb{S}$ with mean vector $\boldsymbol{A}$ and covariance matrix $\boldsymbol{B}$. The probability density function (pdf) of an MGD $\mathcal{N}_{\mathbb{S}}(\boldsymbol{A}, \boldsymbol{B})$ denoted as $\phi_{\mathbb{S}}(\cdot|\boldsymbol{A}, \boldsymbol{B})$ satisfies the following relation:

$$\phi_{\mathbb{S}}(\boldsymbol{\alpha}|\boldsymbol{A}, \boldsymbol{B}) \propto \phi(\boldsymbol{\alpha}|\boldsymbol{A}, \boldsymbol{B}) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \tag{2}$$

where

- $\phi(\cdot|\boldsymbol{A}, \boldsymbol{B})$ is the standard Gaussian pdf with mean vector $\boldsymbol{A}$ and covariance matrix $\boldsymbol{\Sigma}$,
- $\mathbf{1}_{\mathbb{S}}(\cdot)$ is the indicator function defined on $\mathbb{S}$,
- $\propto$ stands for "proportional to".

This report proposes an efficient strategy to generate samples $\widetilde{\boldsymbol{\alpha}}^{(t)}$ distributed according to $\mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is a $(R-1) \times 1$ mean vector and Σ is a $(R-1) \times (R-1)$ covariance matrix.

II. Gibbs sampler for truncated MGDs

A simple way to sample according to $\phi_{\mathbb{S}}(\boldsymbol{\alpha}|\boldsymbol{\mu}, \Sigma)$ is to use a standard rejection-sampling method [1, p. 49] that consists of simulating $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ until the sample $\boldsymbol{\alpha}$ belongs to $\mathbb{S}$. However, for high dimension spaces, this strategy can be inefficient as the acceptation ratio, related to the probability $\mathrm{P}\left[\boldsymbol{\alpha} \in \mathbb{S} \middle| \boldsymbol{\alpha} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)\right]$, can result in rejecting $\boldsymbol{\alpha}$ too often. To overcome this issue, we propose an alternative algorithm derived from the strategy presented in [2] and based on a Gibbs sampling approach. The generation of a sample $\widetilde{\boldsymbol{\alpha}}^{(t)}$ distributed according to the truncated MGD $\mathcal{N}_{\mathbb{S}}\left(\boldsymbol{\mu}, \Sigma\right)$ is achieved by drawing successively the components $\widetilde{\alpha}_r^{(t)}$ $(r = 1, \ldots, R-1)$ from their conditional distributions, i.e. at the step $t+1$:

1. $\widetilde{\alpha}_1^{(t+1)} \sim f\left(\widetilde{\alpha}_1^{(t+1)} \middle| \widetilde{\alpha}_2^{(t)}, \ldots, \widetilde{\alpha}_{R-1}^{(t)}\right)$,
2. $\widetilde{\alpha}_2^{(t+1)} \sim f\left(\widetilde{\alpha}_2^{(t+1)} \middle| \widetilde{\alpha}_1^{(t+1)}, \widetilde{\alpha}_3^{(t)}, \ldots, \widetilde{\alpha}_{R-1}^{(t)}\right)$,
3. $\widetilde{\alpha}_3^{(t+1)} \sim f\left(\widetilde{\alpha}_3^{(t+1)} \middle| \widetilde{\alpha}_1^{(t+1)}, \widetilde{\alpha}_2^{(t+1)}, \widetilde{\alpha}_4^{(t)}, \ldots, \widetilde{\alpha}_{R-1}^{(t)}\right)$,

⋮

$R-1$. $\widetilde{\alpha}_{R-1}^{(t+1)} \sim f\left(\widetilde{\alpha}_{R-1}^{(t+1)} \middle| \widetilde{\alpha}_1^{(t+1)}, \ldots, \widetilde{\alpha}_{R-2}^{(t+1)}\right)$,

By denoting $\boldsymbol{\alpha}_{-r}$ the vector $\boldsymbol{\alpha}$ in which the r^{th} element has been removed, the conditional distributions $f\left(\alpha_r \middle| \boldsymbol{\alpha}_{-r}\right)$ are:

$$\alpha_r \middle| \boldsymbol{\alpha}_{-r} \sim \mathcal{N}_-^+\left(\mu_r^\star, \sigma_r^{2\star}, \alpha_r^-, \alpha_r^+\right), \tag{3}$$

where $\mathcal{N}_-^+\left(\mu_r^\star, \sigma_r^{2\star}, \alpha_r^-, \alpha_r^+\right)$ denotes a two-sided truncated Gaussian distribution whose density is given by:

$$\phi_-^+\left(x|\mu_r^\star, \sigma_r^{2\star}, \alpha_r^-, \alpha_r^+\right) \propto \phi\left(x|\mu_r^\star, \sigma_r^{2\star}\right) \mathbf{1}_{\alpha_r^- \le x \le \alpha_r^+}(x) \tag{4}$$

The mean $\mu_r^\star$ and the variance $\sigma_r^{2\star}$ of the univariate double-truncated Gaussian distributions are given by introducing partitioned covariance matrix and mean vector [3, p. 324]:

$$\begin{cases} \mu_r^\star & = \mu_r + \mathbf{s}_r^{\mathsf{T}} \boldsymbol{\Sigma}_{-r}^{-1}\left(\boldsymbol{\alpha}_{-r} - \boldsymbol{\mu}_{-r}\right) \\ \sigma_r^{2\star} & = \sigma_{r,r}^2 - \mathbf{s}_r^{\mathsf{T}} \boldsymbol{\Sigma}_{-r}^{-1} \mathbf{s}_r, \end{cases} \tag{5}$$

where

- μ_r denotes the r^{th} element of $\boldsymbol{\mu}$,
- s_r is the $(R-2) \times 1$ vector derived from the r^{th} column of $\boldsymbol{\Sigma}$ by removing the r^{th} row term,
- $\boldsymbol{\Sigma}_{-r}$ is the $(R-2) \times (R-2)$ matrix derived from $\boldsymbol{\Sigma} = \left(\sigma_{i,j}^2\right)_{i,j}$ by removing the r^{th} row and the r^{th} column,
- $\boldsymbol{\mu}_{-r}$ is the $(R-2) \times 1$ vector derived from $\boldsymbol{\mu}$ by removing the r^{th} element.

For the MGD truncated on the simplex given in Eq. (1), the edges α_r^- and α_r^+ of the truncature are:

$$\begin{cases} \alpha_r^- &= 0, \\ \alpha_r^+ &= 1 - \sum_{i \neq r} \alpha_i. \end{cases} \tag{6}$$

Note that sampling according to the univariate two-sided Gaussian distribution can be easily achieved with the algoritm described in [2].

III. Simulations

As an example, consider the generation of variables $\boldsymbol{\alpha} = [\alpha_1, \alpha_2]^{\mathsf{T}}$ distributed according to the MGD $\mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ truncated on the simplex $\mathbb{S}$ with $\boldsymbol{\mu} = [0.2, 0.2]^{\mathsf{T}}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 0.13 & 0.08 \\ 0.08 & 0.13 \end{pmatrix}$. The histogram of the variables generated by the algorithm proposed above is depicted in Figure 1. Note that a standard accept-reject procedure leads to a acceptance ratio about $\rho = 0.59$.

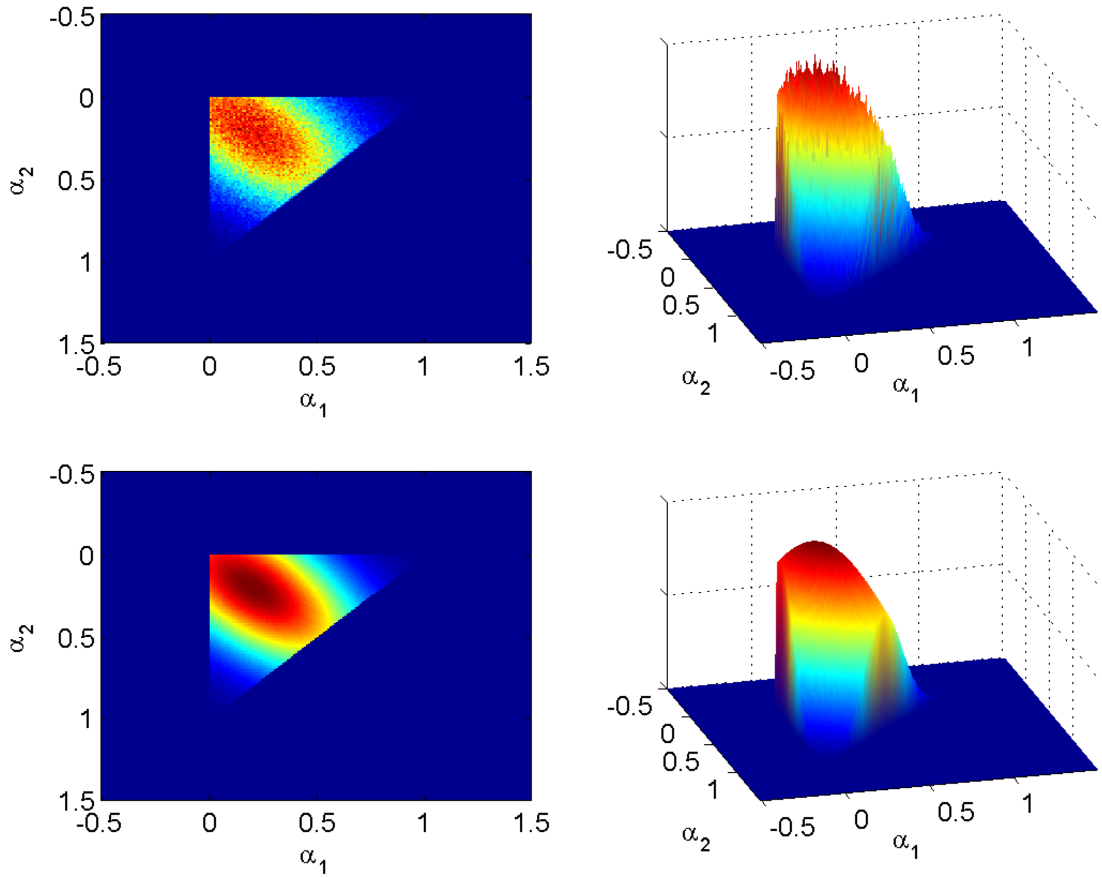

Fig. 1. Top: Histogram of $N = 500000$ variables simulated with the proposed Gibbs sampler for a MGD truncated on the simplex $\mathbb{S}$ ($R = 3$). Bottom: corresponding theoretical pdf.

References

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New-York: Springer-Verlag, 1999.

[2] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, pp. 121–125, 1995.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*. Englewood Cliffs NJ: Prentice Hall, 1993.