A fast Metropolis-Hastings method for generating random correlation matrices

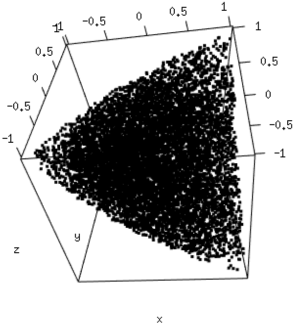
Irene Córdoba[1], Gherardo Varando[2], Concha Bielza[1], Pedro Larrañaga[1]

[1]Universidad Politécnica de Madrid
[2]University of Copenhagen

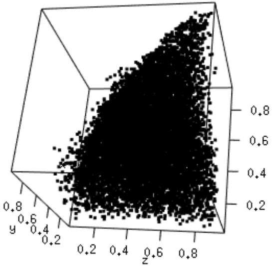
IDEAL 2018

Motivation

Uniform sampling



Naive sampling



Correlation matrices

- $\mathbb{S}^{>0}$: set of symmetric positive definite matrices

$$\mathcal{R} = \{\mathbf{R} \in \mathbb{S}^{>0} : \text{diag}(\mathbf{R}) = \mathbf{1}\}$$

A matrix $\mathbf{R}$ satisfying

- Symmetry: $r_{ij} = r_{ji}$
- Positive definiteness
- Unit diagonal: $r_{ii} = 1$
- $-1 \leq r_{ij} \leq 1$

$$\mathbf{R} = \begin{pmatrix} 1 & 0.45 & 0.25 & 0.12 \\ 0.45 & 1 & 0.65 & -0.27 \\ 0.25 & 0.65 & 1 & -0.034 \\ 0.12 & -0.27 & -0.034 & 1 \end{pmatrix}$$

Upper Cholesky decomposition

- $\mathcal{U}_1$: set of upper triangular matrices with positive diagonal entries and normalized rows
- $\mathbf{R} \in \mathcal{R}$ if and only if there exists a unique $\mathbf{U} \in \mathcal{U}_1$ such that $\mathbf{R} = \mathbf{U}\mathbf{U}^t$

Upper Cholesky decomposition

- $\mathcal{U}_1$: set of upper triangular matrices with positive diagonal entries and normalized rows
- $\mathbf{R} \in \mathcal{R}$ if and only if there exists a unique $\mathbf{U} \in \mathcal{U}_1$ such that $\mathbf{R} = \mathbf{U}\mathbf{U}^t$
- Therefore $\Phi : \mathcal{U}_1 \mapsto \mathbb{S}^{>0}$ such that $\Phi(\mathbf{U}) = \mathbf{U}\mathbf{U}^t$ is a **parametrization** of $\mathcal{R}$

$$\begin{aligned}\mathcal{R} &= \{\mathbf{R} \in \mathbb{S}^{>0} : \mathrm{diag}(\mathbf{R}) = \mathbf{1}\} \\ &= \Phi(\mathcal{U}_1)\end{aligned}$$

Upper Cholesky decomposition

- $\mathcal{U}_1$: set of upper triangular matrices with positive diagonal entries and normalized rows
- $\mathbf{R} \in \mathcal{R}$ if and only if there exists a unique $\mathbf{U} \in \mathcal{U}_1$ such that $\mathbf{R} = \mathbf{U}\mathbf{U}^t$
- Therefore $\Phi : \mathcal{U}_1 \mapsto \mathbb{S}^{>0}$ such that $\Phi(\mathbf{U}) = \mathbf{U}\mathbf{U}^t$ is a **parametrization** of $\mathcal{R}$

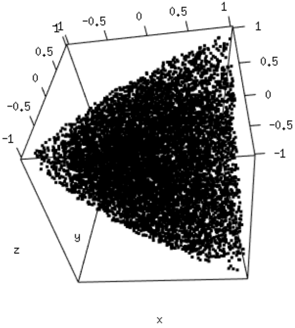
$$\begin{aligned} \mathcal{R} &= \{\mathbf{R} \in \mathbb{S}^{>0} : \text{diag}(\mathbf{R}) = \mathbf{1}\} \\ &= \Phi(\mathcal{U}_1) \end{aligned}$$

- The Jacobian determinant of Φ is [Eaton, 1983]

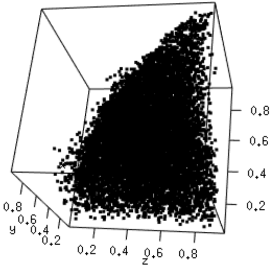
$$\det(J\Phi(\mathbf{U})) = 2^p \prod_{i=1}^{p-1} u_{ii}^i$$

Sampling correlation matrices

Uniform sampling



Sampling i.i.d. entries in $\mathbf{U}$



Our proposal

- When sampling in $\mathcal{U}_1$ from $f(\mathbf{U}) \propto \det(J\Phi(\mathbf{U}))$ we get the uniform density on $\mathcal{R}$ [Diaconis et al., 2013]

Our proposal

- When sampling in $\mathcal{U}_1$ from $f(\mathbf{U}) \propto \det(J\Phi(\mathbf{U}))$ we get the uniform density on $\mathcal{R}$ [Diaconis et al., 2013]
- Each row $\boldsymbol{u}_i$ can be sampled independently because of the form of

$$\det(J\Phi(\mathbf{U})) = 2^p \prod_{i=1}^{p-1} u_{ii}^i$$

Our proposal

- When sampling in $\mathcal{U}_1$ from $f(\mathbf{U}) \propto \det(J\Phi(\mathbf{U}))$ we get the uniform density on $\mathcal{R}$ [Diaconis et al., 2013]
- Each row $\boldsymbol{u}_i$ can be sampled independently because of the form of

$$\det(J\Phi(\mathbf{U})) = 2^p \prod_{i=1}^{p-1} u_{ii}^i$$

- $\boldsymbol{u}_i$ is unitary, has the first $i-1$ entries equal to zero, and the i-th entry positive $\rightarrow$ sampling the i-th row $\boldsymbol{u}_i$ is equivalent to sampling $\boldsymbol{v} \in \mathcal{S}_+^{p-i}$, where

$$\mathcal{S}_+^{p-i} = \{\boldsymbol{v} \in \mathbb{R}^{p-i+1} : \boldsymbol{v}\boldsymbol{v}^t = 1 \text{ and } v_1 > 0\}$$

is the $(p-i)$-**dimensional positive hemisphere**.

Algorithm for our proposal

Uniform sampling in $\mathcal{R}$

Input: Sample size N

Output: Uniform sample from $\mathcal{R}$ of size N

for $n = 1, \ldots, N$ **do**

  $\mathrm{U}^n \leftarrow \mathbf{0}_p$

  for $i = 1, \ldots, p$ **do**

    [Metropolis-Hastings]

    $\boldsymbol{u}_i^n \leftarrow$ sample from $f(\boldsymbol{u}_i) \propto u_{ii}^i$ on $\mathcal{S}_+^{p-i}$

  end for

end for

return $\{\Phi(\mathrm{U}^1), \ldots, \Phi(\mathrm{U}^N)\}$

Metropolis-Hastings on $\mathcal{S}_+^{p-i}$

General Metropolis-Hastings

Input: Sample size N

Output: Sample of size N from target distribution f

$\boldsymbol{v}_0 \leftarrow$ initial sample
for $t = 1, \ldots, N$ **do**
  $\tilde{\boldsymbol{v}} \leftarrow$ sample from proposal distribution $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$
  $r \leftarrow \min\left(1, (f(\tilde{\boldsymbol{v}})q(\boldsymbol{v}_{t-1} \mid \tilde{\boldsymbol{v}}))/(f(\boldsymbol{v}_{t-1})q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1}))\right)$
  $\delta \leftarrow$ random uniform observation on $[0, 1]$
  if $\delta \leq r$ **then**
    $\boldsymbol{v}_t \leftarrow \tilde{\boldsymbol{v}}$
  else
    $\boldsymbol{v}_t \leftarrow \boldsymbol{v}_{t-1}$
  end if
end for
return $\{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_N\}$

Our proposal distribution and theoretical properties

- Our target is $f(\boldsymbol{v}) \propto v_1^i$. At step t we add Gaussian noise with **variance** σ_ϵ^2 to $\boldsymbol{v}_{t-1}$ and normalize the result $\rightarrow$ proposal distribution $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is **projected Gaussian** [Mardia and Jupp, 1999]

Our proposal distribution and theoretical properties

- Our target is $f(\boldsymbol{v}) \propto v_1^i$. At step t we add Gaussian noise with **variance** σ_ϵ^2 to $\boldsymbol{v}_{t-1}$ and normalize the result $\rightarrow$ proposal distribution $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is **projected Gaussian** [Mardia and Jupp, 1999]
- We obtain its explicit density from Pukkila and Rao [1988]

Our proposal distribution and theoretical properties

- Our target is $f(\boldsymbol{v}) \propto v_1^i$. At step t we add Gaussian noise with **variance** σ_ϵ^2 to $\boldsymbol{v}_{t-1}$ and normalize the result $\rightarrow$ proposal distribution $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is **projected Gaussian** [Mardia and Jupp, 1999]
- We obtain its explicit density from Pukkila and Rao [1988]
- It can be seen that $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is symmetric $\rightarrow$ we may omit Hastings correction, and get the **acceptance probability**

$$r = \min\left(1, \frac{f(\tilde{\boldsymbol{v}})q(\boldsymbol{v}_{t-1} \mid \tilde{\boldsymbol{v}})}{f(\boldsymbol{v}_{t-1})q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})}\right) = \min\left(1, \frac{f(\tilde{\boldsymbol{v}})}{f(\boldsymbol{v}_{t-1})}\right)$$

Our proposal distribution and theoretical properties

- Our target is $f(\boldsymbol{v}) \propto v_1^i$. At step t we add Gaussian noise with **variance** σ_ϵ^2 to $\boldsymbol{v}_{t-1}$ and normalize the result $\rightarrow$ proposal distribution $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is **projected Gaussian** [Mardia and Jupp, 1999]
- We obtain its explicit density from Pukkila and Rao [1988]
- It can be seen that $q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})$ is symmetric $\rightarrow$ we may omit Hastings correction, and get the **acceptance probability**

$$r = \min\left(1, \frac{f(\tilde{\boldsymbol{v}})q(\boldsymbol{v}_{t-1} \mid \tilde{\boldsymbol{v}})}{f(\boldsymbol{v}_{t-1})q(\tilde{\boldsymbol{v}} \mid \boldsymbol{v}_{t-1})}\right) = \min\left(1, \frac{f(\tilde{\boldsymbol{v}})}{f(\boldsymbol{v}_{t-1})}\right)$$

- Good basic convergence properties for Markov chains

Empirical monitoring of row sampling

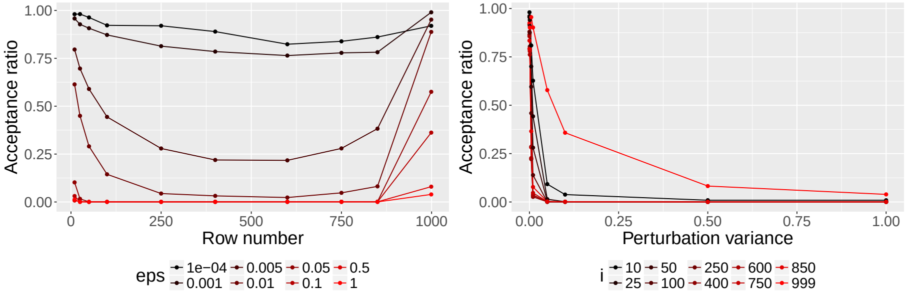


Figure: Acceptance ratio as a function of the row number i (left) and the perturbation variance σ_ϵ^2 (right). eps: σ_ϵ^2.

- Small values for $\sigma_\epsilon^2 \rightarrow$ high acceptance ratios
- As row number increases, $f(\boldsymbol{v}) \propto v_1^i$ approaches to a delta function $\rightarrow$ smaller σ_ϵ^2 for higher acceptance ratio

Empirical monitoring of row sampling

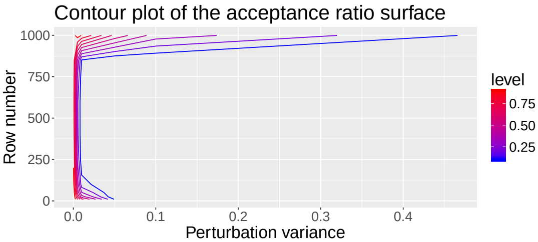


Figure: Contour lines of the acceptance ratio surface. `level`: magnitude of the acceptance ratio.

- Convergence is **guaranteed theoretically**, but this can help to choose the hyper parameters of the algorithm
- **Caution**: high acceptance ratio might also indicate slow convergence

Comparison to other algorithms for uniform sampling

- Pourahmadi and Wang [2015] parametrize again $\mathbf{U}$ using spherical coordinates, and sample using the Jacobian of the new parametrization
- Lewandowski et al. [2009] use two alternative representations of the correlation matrix: *vines* and *elliptical distributions*.

Comparison to other algorithms for uniform sampling

- Pourahmadi and Wang [2015] parametrize again $\mathbf{U}$ using spherical coordinates, and sample using the Jacobian of the new parametrization
- Lewandowski et al. [2009] use two alternative representations of the correlation matrix: *vines* and *elliptical distributions*.

Implementation

- All the experiments have been implemented and executed using R [R Core Team, 2018].
- We have implemented the method of Pourahmadi and Wang [2015]
- The methods in Lewandowski et al. [2009] are available in the CRAN package *clusterGeneration*
- Our method is available in the CRAN package *gmat*

Eigenvalue density of the samples

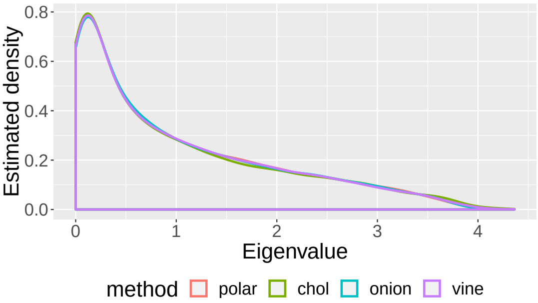


Figure: Eigenvalue densities for the different sampling methods. `chol`: our proposal; `c-vine`, `onion`: methods by Lewandowski et al. [2009]; `polar`: method by Pourahmadi and Wang [2015]. For our method, we have set $t_b = 1000$ and $\sigma_\epsilon^2 = 0.01$.

Computational performance analysis

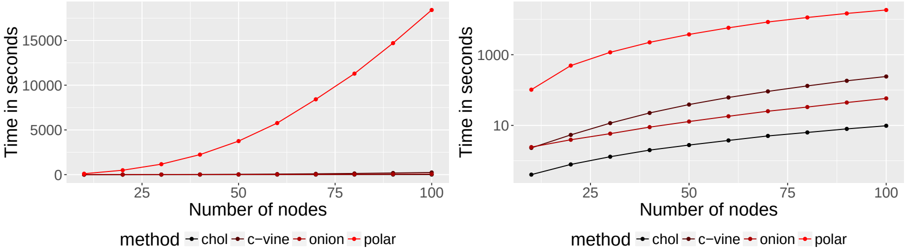


Figure: Execution time of available methods for uniform sampling of correlation matrices, both in linear (left) and logarithmic (right) scale. `chol`: our proposal; `c-vine`, `onion`: methods by Lewandowski et al. [2009]; `polar`: method by Pourahmadi and Wang [2015]. For our method, we have set $t_b = 1000$ and $\sigma_\epsilon^2 = 0.01$.

Conclusions and future work

Conclusions

- The repository `https://github.com/irenecrsn/rcor` contains the files and instructions for **replicating the experiments**
- Our method is faster than related state of the art algorithms, yet being conceptually simple

Conclusions and future work

Conclusions

- The repository `https://github.com/irenecrsn/rcor` contains the files and instructions for **replicating the experiments**
- Our method is faster than related state of the art algorithms, yet being conceptually simple

In the future, we would like to

- Explore other variants for our Markov chain: independent Metropolis, adaptive sampling, other proposal distributions, etc.
- Monitor further quantities besides the acceptance ratio.
- Explore other parametrizations of $\mathcal{R}$.

References

P. Diaconis, S. Holmes, and M. Shahshahani. *Sampling from a Manifold*, volume 10 of *Collections*, pages 102–125. Institute of Mathematical Statistics, 2013.

M. L. Eaton. *Multivariate Statistics: A Vector Space Approach*. Wiley, 1983.

D. Lewandowski, D. Kurowicka, and H. Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, 2009.

K. Mardia and P. Jupp. *Directional Statistics*. Wiley, 1999.

M. Pourahmadi and X. Wang. Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor. *Statistics & Probability Letters*, 106:5–12, 2015.

T. M. Pukkila and C. R. Rao. Pattern recognition based on scale invariant discriminant functions. *Information Sciences*, 45(3):379–389, 1988.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

A fast Metropolis-Hastings method for generating random correlation matrices

Irene Córdoba[1], Gherardo Varando[2], Concha Bielza[1], Pedro Larrañaga[1]

[1]Universidad Politécnica de Madrid
[2]University of Copenhagen

IDEAL 2018