# An R Package for Fast Sampling from von Mises Fisher Distribution

Elysée Aristide Houndetoungan

2020-09-10

## Context

I build an R package named **vMF** which samples from the von Mises-Fisher distribution ($\mathcal{M}$) as **movMF**. However, unlike the **movMF** package (Hornik and Grün, 2018) which also simulates and estimates mixtures of $\mathcal{M}$, **vFM** instead focuses on fast sampling from $\mathcal{M}$. The package also computes the density and the normalization constant of the von Mises-Fisher distribution.

Note that **movMF** is more general and can be used for many other purposes. It cannot be replaced by **vMF** which is more specific.

The von Mises Fisher distribution is used to model coordinates on a hypersphere of dimension $p \geq 2$. It can be considered as the equivalent of the normal distribution on a hypersphere. The von Mises Fisher distribution is characterized by two parameters. The location (or mean directional) parameter $\boldsymbol{\mu}$ around which simulations from the distribution will be concentrated and the intensity parameter $\eta$ which measures the intensity of concentration of the simulations around $\boldsymbol{\mu}$. The higher $\eta$, the more the simulations are concentrated around $\boldsymbol{\mu}$. Comparing to the normal distribution, $\boldsymbol{\mu}$ is similar to the mean parameter of the normal distribution and $\dfrac{1}{\eta}$ is similar to the standard deviation.

There are several definitions of the density function of $\mathcal{M}$. In this package, the density is normalized by the uniform distribution without loss of generality. This is also the case in Mardia and Jupp (2009) and Hornik and Grün (2013).

Let $\mathbf{z} \sim \mathcal{M}\left(\eta, \boldsymbol{\mu}\right)$. Then,

$$f_p(\mathbf{z}|\eta, \boldsymbol{\mu}) = C_p(\eta)e^{\eta\mathbf{z}'\boldsymbol{\mu}},$$

where $C_p(x) = \left(\dfrac{x}{2}\right)^{\frac{p}{2}-1}\dfrac{1}{\Gamma\left(\frac{p}{2}\right)I_{\frac{p}{2}-1}(x)}$ is the normalization constant and $I_.(.)$ the Bessel function of the first kind defined by:

$$I_\alpha(x) = \sum_{m=0}^{\infty}\frac{\left(\frac{x}{2}\right)^{2m+\alpha}}{m!\Gamma(m+\alpha+1)}.$$

The normalization with respect to the uniform distribution simplifies some results. For example, $C_p(0) = 1$.

## Simulation from von Mises Fisher distribution

The following algorithm provides a rejection sampling scheme for drawing a sample from the $\mathcal{M}$ with mean directional parameter $\boldsymbol{\mu} = (0, ..., 0, 1)$ and concentration (intensity) parameter $\eta \geq 0$ (see Section 2.1 in Hornik and Grün, 2014).

- Step 1. Calculate $b$ using * Step 1. Calculate $b$ using

$$b = \frac{p-1}{2\eta + \sqrt{2\eta^2 + (p-1)^2}}.$$

Let $x_0 = (1-b)/(1+b)$ and $c = \eta x_0 + (p-1)\log\left(1-x_0^2\right)$.

- Step 2. Generate $Z \sim Beta((p-1)/2, (p-1)/2)$ and $U \sim Unif([0,1])$ and calculate

$$W = \frac{1-(1+b)Z}{1-(1-b)Z}.$$

- Step 3. If

$$\eta W + (p-1)\log(1-x_0W) - c < \log(U),$$

go to step 2.

- Step 4. Generate a uniform $(d-1)$-dimensional unit vector $V$ and return

$$X = \left(\sqrt{1-W^2}V', W\right)'$$

The uniform $(d-1)$-dimensional unit vector $V$ can be generated by simulating $d-1$ independent standard normal random variables and normalizing them. To get samples from $\mathcal{M}$ with arbitrary mean direction parameter $\boldsymbol{\mu}$, $X$ is multiplied from the left with a matrix where the first $d-1$ columns consist of unitary basis vectors of the subspace orthogonal to $\boldsymbol{\mu}$ and the last column is equal to $\boldsymbol{\mu}$.

### Package Installation

The code source of the package is written in C++ with the package **Rcpp** (Eddelbuettel et al., 2020). **vMF** is currently available on GitHub. Its installation requires to prior install **devtools** package (Wickham et al., 2020). Moreover, Windows users should install Rtools compatible with their R version.

**vMF** package can be installed from this GitHub repos using the `install_github` function of **devtools**. All the dependencies will also be installed automatically.

```r
library(devtools)
install_github("ahoundetoungan/vMF")
```

The option `build_vignettes = TRUE` can be added if one desires to install the vignettes.

### References

## References

Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Russell, N., Bates, D., and Chambers, J. (2020). *Rcpp: Seamless R and C++ Integration*. R package version 1.0.5.

Hornik, K. and Grün, B. (2013). On conjugate families and jeffreys priors for von mises–fisher distributions. *Journal of statistical planning and inference*, 143(5):992–999.

Hornik, K. and Grün, B. (2014). movmf: an r package for fitting mixtures of von mises-fisher distributions. *Journal of Statistical Software*, 58(10):1–31.

Hornik, K. and Grün, B. (2018). *movMF: Mixtures of von Mises-Fisher Distributions*. R package version 0.2-4.

Mardia, K. V. and Jupp, P. E. (2009). *Directional statistics*, volume 494. John Wiley & Sons.

Wickham, H., Hester, J., and Chang, W. (2020). *devtools: Tools to Make Developing R Packages Easier*. R package version 2.2.2.