Sampling Methods for the von Mises-Fisher Distributions and Smoothed Bootstrap

Yikun Zhang

May 25, 2022

In this note, we elucidate some feasible algorithms for sampling data points that are von Mises-Fisher (vMF) distributed. Though a rejection sampling is always applicable for an arbitrary number of dimensions, an analytic approach can be formulated for sampling vMF random data on $\Omega_2 \subset \mathbb{R}^3$, where $\Omega_q = \{\boldsymbol{x} \in \mathbb{R}^{q+1} : \|\boldsymbol{x}\|_2^2 = x_1^2 + \cdots + x_{q+1}^2 = 1\}$ is the q-dimensional sphere and $||\cdot||_2$ is the usual Euclidean norm. The analytic method is shown to be more efficient in practice. Later, we will leverage this sampling scheme to conduct the smoothed bootstrap with directional kernel density estimators (KDEs) given a von Mises kernel.

1 Reviews on the von Mises-Fisher Distribution

The von Mises-Fisher distribution (or von Mises distribution in Ω_1) is a probability distribution supported on $\Omega_q \subset \mathbb{R}^{q+1}$ with density

$$f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \cdot \exp(\kappa \boldsymbol{\mu}^T \boldsymbol{x}), \tag{1}$$

where $C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)}$, $\boldsymbol{\mu} \in \Omega_q$ is the directional mean, $\kappa \geq 0$ is the concentration parameter, and

$$\mathcal{I}_\alpha(\kappa) = \frac{\left(\frac{\kappa}{2}\right)^\alpha}{\pi^{\frac{1}{2}} \Gamma\left(\alpha + \frac{1}{2}\right)} \int_{-1}^{1} (1-t^2)^{\alpha - \frac{1}{2}} \cdot e^{\kappa t} dt$$

is the modified Bessel function of the first kind at order κ. We denote it by vMF$(\boldsymbol{\mu}, \kappa)$. The vMF density can be derived from the density of a normal/Gaussain distribution $\mathcal{N}(\boldsymbol{\mu}, \sqrt{\kappa} I_{q+1})$, where I_{q+1} is the identity matrix in $\mathbb{R}^{(q+1)\times(q+1)}$. Starting from the normal density

$$f_N(\boldsymbol{x}) = \left(\sqrt{\frac{\kappa}{2\pi}}\right)^{p+1} \exp\left(-\kappa \frac{\|\boldsymbol{x} - \boldsymbol{\mu}\|_2^2}{2}\right),$$

one can obtain the vMF density by using the fact that $\|\boldsymbol{x} - \boldsymbol{\mu}\|_2^2 = 2 - 2\boldsymbol{\mu}^T \boldsymbol{x}$ on Ω_q and rescaling the normalizing constant. For this reason, the vMF distribution is regarded as the normal distribution on the unit hypersphere.

2 Naive Rejection Sampling

The most naive rejection sampling for vMF-distributed data points is to set the proposal density p to be uniformly distributed on Ω_q, *i.e.*,

$$p(\boldsymbol{x}) = \frac{1}{\omega_q(\Omega_q)} \cdot \mathbb{1}_{\Omega_q}(\boldsymbol{x}) = \frac{\Gamma\left(\frac{q+1}{2}\right)}{2\pi^{\frac{q+1}{2}}} \cdot \mathbb{1}_{\Omega_q}(\boldsymbol{x}),$$

where ω_q is the Lebesgue measure on Ω_q. There are multiple ways to randomly sample data points from the uniform distribution on Ω_q. For instance, one can utilize the isotropic property of the multivariate standard normal distribution and generate data points from $\mathcal{N}(\boldsymbol{0}, I_{q+1})$ with an extra L_2-normalization step. A small advantage of using the uniform distribution as the proposal density in rejection sampling is that there is no need to calculate out the normalizing constant $C_q(\kappa)$ for the von Mises-Fisher density in order to choose the maximum value $M \geq \sup_{\boldsymbol{x}\in\Omega_q} \frac{f_{\text{vMF}}(\boldsymbol{x})}{p(\boldsymbol{x})}$. This is because we may take

$$M = \sup_{\boldsymbol{x}\in\Omega_q} \frac{f_{\text{vMF}}(\boldsymbol{x})}{p(\boldsymbol{x})} = \sup_{\boldsymbol{x}\in\Omega_q} \frac{C_q(\kappa)\cdot\exp(\kappa\boldsymbol{\mu}^T\boldsymbol{x})}{1/\omega_q(\Omega_q)} = \omega_q(\Omega_q)\cdot C_q(\kappa)\cdot e^{\kappa}$$

and thus $\frac{f(\boldsymbol{Y})}{M\cdot p(\boldsymbol{Y})} = \exp\left[\kappa(\boldsymbol{\mu}^T\boldsymbol{Y} - 1)\right]$. In a nutshell, the rejection sampling for a random vector $\boldsymbol{X} \sim \text{vMF}(\boldsymbol{\mu},\kappa)$ with the uniform distribution on Ω_q as the proposal density is given by

1. Generate a random vector $\boldsymbol{Y}$ from the uniform distribution p on Ω_q (*e.g.*, draw a random vector $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_{q+1})$ and let $\boldsymbol{Y} = \frac{\boldsymbol{Z}}{\|\boldsymbol{Z}\|_2}$) and another random number U from Uniform$[0,1]$.
2. If $U < \frac{f(\boldsymbol{Y})}{M\cdot p(\boldsymbol{Y})} = \exp\left[\kappa(\boldsymbol{\mu}^T\boldsymbol{Y} - 1)\right]$, we set $\boldsymbol{X} = \boldsymbol{Y}$. Otherwise go back to the previous step to draw another new pair of $\boldsymbol{Y}$ and U.

3 Analytic vMF Sampling Strategy on Ω_2

Without the loss of generality, consider generating $\boldsymbol{X}$ from vMF$(\boldsymbol{\mu}_0,\kappa)$ with $\boldsymbol{\mu}_0 = (0,...,0,1)^T \in \Omega_q \subset \mathbb{R}^{q+1}$. Based on the results in Ulrich (1984); Wood (1994); Kurz and Hanebeck (2015), we know that $\boldsymbol{X}$ follows vMF$(\boldsymbol{\mu}_0,\kappa)$ if and only if

$$\boldsymbol{X} = \left(\sqrt{1-W^2}\cdot\boldsymbol{V}, W\right)^T, \tag{2}$$

where $\boldsymbol{V}$ is uniformly distributed on Ω_{q-1} and $W \in [-1,1]$ has its probability density function as

$$f_W(w) = C_q \cdot \left(1-w^2\right)^{\frac{q}{2}-1}\exp(\kappa w) \tag{3}$$

with $C_q^{-1} = \int_{-1}^{1}\left(1-w^2\right)^{\frac{q}{2}-1}\exp(\kappa w)dw$. In general, the cumulative distribution function (CDF) of W has no closed forms, and Ulrich (1984) proposed another rejection sampling technique to randomly generate points from f_W with the proposal density Beta$\left(\frac{q}{2},\frac{q}{2}\right)$. (See also Wood (1994) and Section 5.1 in Dhillon and Sra (2003).)

However, when $q = 2$, the density of W becomes

$$f_W(w) = \frac{\kappa}{e^{\kappa} - e^{-\kappa}} \cdot \exp(\kappa w)$$

and the corresponding CDF is

$$F_W(t) = \frac{e^{\kappa t} - e^{-\kappa}}{e^{\kappa} - e^{-\kappa}}.$$

Some algebra will give rise to the inverse of the CDF as

$$F_W^{-1}(y) = \frac{1}{\kappa} \log\left[y(e^{\kappa} - e^{-\kappa}) + e^{-\kappa}\right]. \tag{4}$$

To avoid numerical overflow for large values of κ, one can replace (4) with the following equivalent expression

$$F_W^{-1}(y) = 1 + \frac{1}{\kappa} \log\left[y + (1-y)e^{-2\kappa}\right]. \tag{5}$$

The sampling of $\boldsymbol{V}$ that is uniformly distributed on Ω_1 is easy. For instance, one can sample $U \sim \text{Uniform}[0,1]$ and $\boldsymbol{V} = (\cos U, \sin U)$ will be a uniform data sample on Ω_1. Thus, the analytic strategy for sampling $\boldsymbol{X} \sim \text{vMF}(\boldsymbol{\mu}_0, \kappa)$ is clear.

In order to handle other values of $\boldsymbol{\mu} \in \Omega_2$, one can apply a rotation matrix $R \in \mathbb{R}^{3\times 3}$ to the resulting sample points. One remarkable fact for writing down the closed form of R is that, for the sake of bringing a normalized vector $\boldsymbol{\mu}_0$ into coincidence with another normalized vector $\boldsymbol{\mu}$, we simply need to rotate $\boldsymbol{\mu}_0$ about $\boldsymbol{k} = \frac{\boldsymbol{\mu}_0 + \boldsymbol{\mu}}{2}$ by the angle π. With Rodrigues's rotation formula[1], one gets the beautiful form

$$R = 2\frac{(\boldsymbol{\mu}_0 + \boldsymbol{\mu})(\boldsymbol{\mu}_0 + \boldsymbol{\mu})^T}{(\boldsymbol{\mu}_0 + \boldsymbol{\mu})^T(\boldsymbol{\mu}_0 + \boldsymbol{\mu})} - I_3. \tag{6}$$

This rotation formula works in any dimension. A small note to this rotation matrix is that it will indeed rotate the data point about $\boldsymbol{\mu}_0$ by the angle π when $\boldsymbol{\mu}_0 = \boldsymbol{\mu}$.

In short, the analytic algorithm for randomly sampling a data point $\boldsymbol{X} \sim \text{vMF}(\boldsymbol{\mu}, \kappa)$ with $\boldsymbol{\mu}$ is given by

1. Sample two independent data points Y and U from $\text{Uniform}[0,1]$.
2. Compute $W = 1 + \frac{1}{\kappa}\log\left[Y + (1-Y)e^{-2\kappa}\right]$ and $\boldsymbol{V} = (\cos U, \sin U)$.
3. Obtain $\boldsymbol{X} = \left(\sqrt{1-W^2}\boldsymbol{V}, W\right)^T$ and rotate it as $\boldsymbol{X}_{\text{vMF}} = R\boldsymbol{X}$.

Simulation studies show that this analytic sampling strategy is more than 100 times faster than the previous naive rejection sampling method.
When $q \neq 2$, one can sample $\boldsymbol{X} \sim \text{vMF}(\boldsymbol{\mu}, \kappa)$ using a faster rejection sampling approach [2]; see also Algorithm 1.

[1] http://en.wikipedia.org/wiki/Rodrigues'_rotation_formula and also, https://en.wikipedia.org/wiki/Rotation_matrix

[2] https://dlwhittenbury.github.io/ds-2-sampling-and-visualising-the-von-mises-fisher-distribution-in-p-dimensions.html

Algorithm 1 Fast Rejection Sampling of a vMF$(\boldsymbol{\mu}, \kappa)$ on Ω^q

Input: The mean vector $\boldsymbol{\mu} \in \Omega_q$ and concentration parameter $\kappa > 0$.
Step 1: Let $d \leftarrow \dim(\boldsymbol{\mu})$, i.e., $q+1$.
Step 2: Sample data points from the marginal distribution (3) as follows:

Step 2-1: $b \leftarrow \frac{\sqrt{4\kappa^2+(d-1)^2}-2\kappa}{d-1}$, $x_0 \leftarrow \frac{1-b}{1+b}$, $m \leftarrow \frac{d-1}{2}$, and $c \leftarrow \kappa x_0 + (d-1)\log(1-x_0^2)$.

Step 2-2: $t \leftarrow -1000$ and $U \leftarrow 1$.
while $t < \log(U) + c$:
(i) $z \leftarrow \text{Beta}(m, m)$, *i.e.*, z is a random variable with $\text{Beta}(m, m)$ distribution.
(ii) $U \leftarrow \text{Uniform}[0, 1]$.
(iii) $W \leftarrow \frac{1-(1+b)z}{1-(1-b)z}$ and $t \leftarrow \kappa W + (d-1)\log(1-x_0 W)$.
endwhile
Step 3: $\boldsymbol{V} \leftarrow \mathcal{N}(\boldsymbol{0}, I_{d-1})$ and $\boldsymbol{V} \leftarrow \frac{\boldsymbol{V}}{\|\boldsymbol{V}\|_2}$.
Step 4: Obtain $\boldsymbol{X} = \left(\sqrt{1-W^2}\boldsymbol{V}, W\right)^T$ and rotate it as $\boldsymbol{X}_{\text{vMF}} = R\boldsymbol{X}$ with (6).
Output: Point $\boldsymbol{X}_{\text{vMF}} \in \Omega_q$ that follows the vMF$(\boldsymbol{\mu}, \kappa)$ distribution.

4 Smoothed Bootstrap on Directional KDE with von Mises Kernel

The idea of (nonparametric) bootstrap stems from Efron's seminal work in 1979 Efron (1979). A variant of the original bootstrap is the so-called smoothed bootstrap, where the bootstrap sample is drawn from the kernel density estimate $\hat{f}_h$ instead of resampling from the original data set Silverman and Young (1987). At variance with the original bootstrap, the smoothed bootstrap takes into account both the variance and bias of the statistical quantity/functional estimated by KDE, though it will be less precise in variance estimation compared to the original bootstrap Chen et al. (2015).

Whereas the smoothed bootstrap has been widely used under the Euclidean KDE scenario, to the best of our knowledge, it has no previous applications with directional KDEs. Here we summarize the procedure of carrying out the smoothed bootstrap with directional KDEs, given the preceding discussion on sampling vMF-distributed data. Suppose that the original directional data sample is $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$ and the von Mises kernel $L(r) = e^{-r}$ is applied. Then the directional KDE becomes

$$\hat{f}_h(\boldsymbol{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{X}_i}{h^2}\right) = \frac{1}{n}\sum_{i=1}^{n} f_{\text{vMF}}\left(\boldsymbol{x}; \boldsymbol{X}_i, \frac{1}{h^2}\right), \tag{7}$$

where $c_{h,q}(L)$ is a normalizing constant. Thus, the smoothed bootstrap procedure for directional KDEs with the von Mises kernel is given by conducting the following two-step procedure repeatedly (which is identical to the one for regular KDEs)

- Sample a data point $\boldsymbol{X}^*$ uniformly from $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$.
- Sample $\boldsymbol{X}_i^{(B)}$ from vMF $\left(\boldsymbol{X}^*, \frac{1}{h^2}\right)$.

The smoothed bootstrap for directional KDE with the von Mises kernel is most efficient in Ω_2, since the second step is accomplished with the aforementioned analytic sampling scheme.

Remark 1. Note that both the original bootstrap and smoothed one can fitted into the online learning scenario, where we only observe the streaming data. For the original bootstrap, what we need is to maintain a (small) data set which is sampled uniformly from the whole streaming data. One notable approach to deal with this task is the well-known reservoir sampling Vitter (1985). For the smoothed bootstrap in online learning, it only requires one extra step to sample from the (directional) kernel function.

References

Y.-C. Chen, S. Ho, P. E. Freeman, C. R. Genovese, and L. Wasserman. Cosmic web reconstruction through density ridges: method and algorithm. *Monthly Notices of the Royal Astronomical Society*, 454 (1):1140–1156, 09 2015. URL https://doi.org/10.1093/mnras/stv1996.

I. Dhillon and S. Sra. Modeling data using directional distributions. *Technical Report*, 08 2003. URL https://www.cs.utexas.edu/users/inderjit/public_papers/tr03-06.pdf.

B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1–26, 01 1979. URL https://doi.org/10.1214/aos/1176344552.

G. Kurz and U. D. Hanebeck. Stochastic sampling of the hyperspherical von mises–fisher distribution without rejection methods. In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6, 2015.

B. W. Silverman and G. A. Young. The bootstrap: To smooth or not to smooth? *Biometrika*, 74(3): 469–479, 1987.

G. Ulrich. Computer generation of distributions on the m-sphere. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(2):158–163, 1984.

J. S. Vitter. Random sampling with a reservoir. *ACM Transactions on Mathematical Software*, 11(1): 37–57, 1985.

A. T. Wood. Simulation of the von mises fisher distribution. *Communications in Statistics - Simulation and Computation*, 23(1):157–164, 1994.