The Power Spherical distribution

Nicola De Cao and Wilker Aziz

A new reparameterizable and stable location-scale distribution on the n-sphere

https://nicola-decao.github.io

https://twitter.com/nicola_decao

Code at https://github.com/nicola-decao/power_spherical

Motivation and Contributions

The von Mises-Fisher is unstable and hard to reparameterize

- The **von Mises-Fisher** distribution (vMF; Mardia & Jupp,2009) is a popular choice as a location-scale distribution on the **n-sphere**
- Unfortunately it is i) **unstable** in high dimensions and concentration, and ii) difficult to reparameterize since it requires **rejection sampling**
- **We propose a simple alternative that is much more stable and easily reparameterizable** 😍

The Power Spherical distribution

Functional form

Let's define a distribution that follows a **power law** and is proportional to the **dot-product** with a direction:

$$p(x;\mu,\kappa) = \underbrace{\left\{ 2^{\alpha+\beta}\pi^{\beta}\frac{\Gamma(\alpha)}{\Gamma\left(\alpha+\beta\right)} \right\}^{-1}}_{=\text{normalizer } N_X(\kappa,d)} \left(1+\mu^\top x\right)^{\kappa} \quad \text{with} \quad \alpha = \frac{d-1}{2}+\kappa, \quad \beta = \frac{d-1}{2} \;,$$

with direction $\mu \in \mathbb{S}^{d-1}$, concentration $\kappa \in \mathbb{R}_{\geq 0}$, and argument $x \in \mathbb{S}^{d-1}$.

The Power Spherical distribution

Sampling

Sampling is obtained with an **invertible transformation (🎉):**

1. From **a Beta distribution**, and

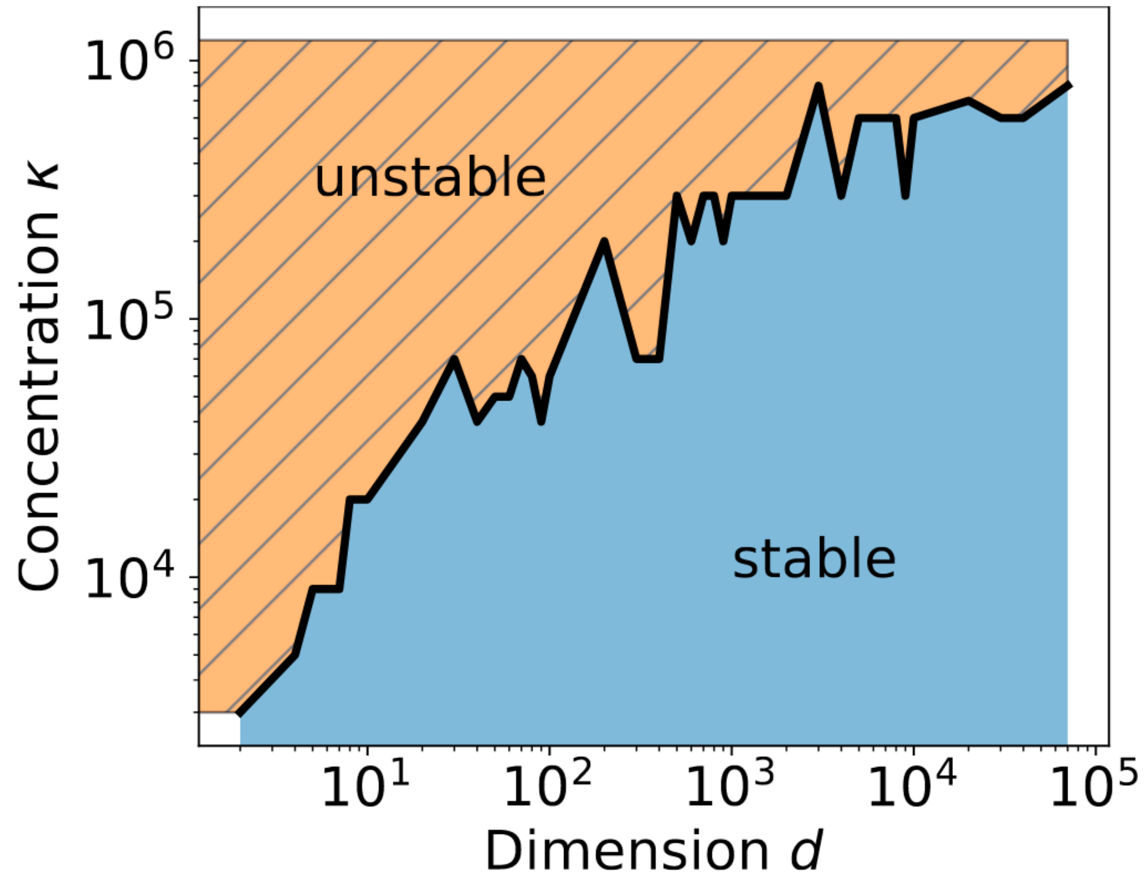
2. **an uniform distribution** on the (n-1)-sphere

Algorithm 1 Power Spherical sampling

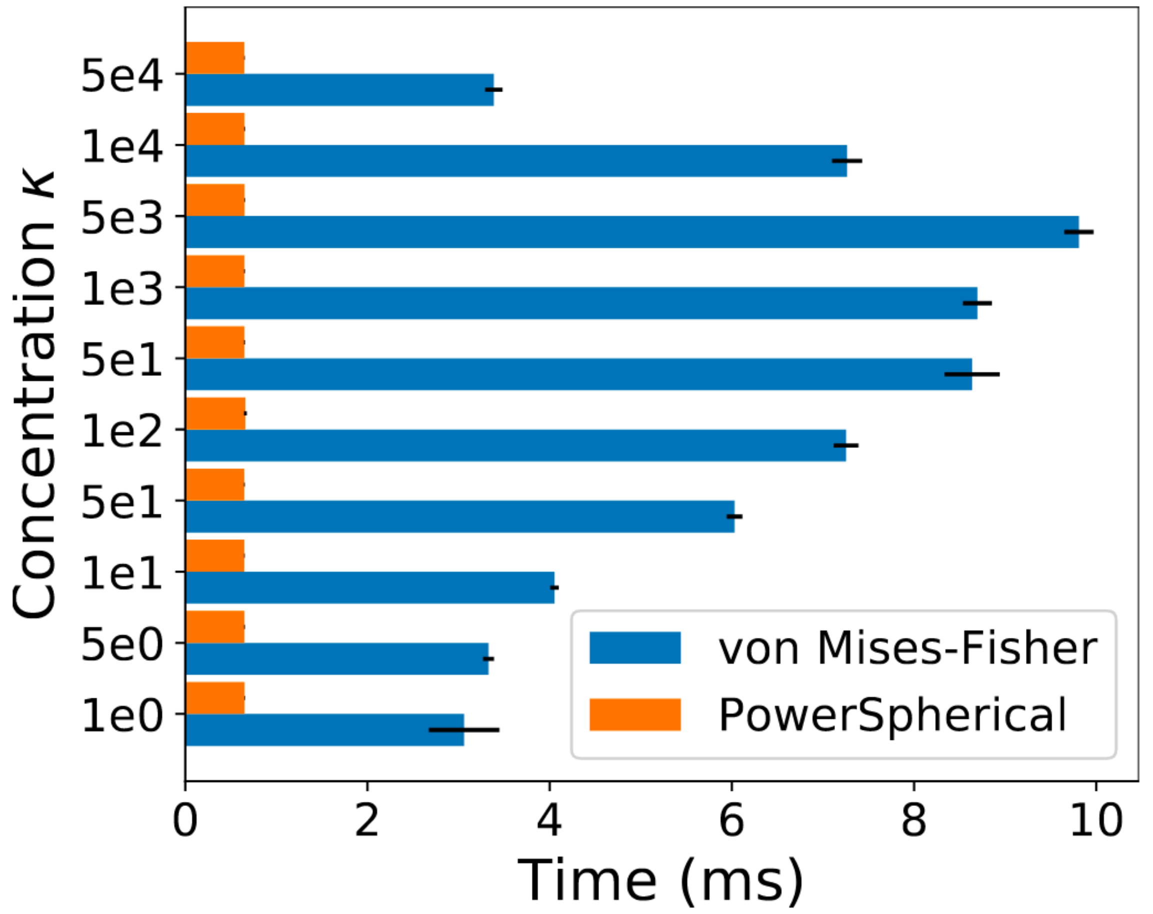
Input: dimension d, direction μ, concentration κ
sample $z \sim \text{Beta}\left(Z; (d-1)/2 + \kappa, (d-1)/2\right)$
sample $v \sim \mathcal{U}(\mathcal{S}^{d-2})$
$t \leftarrow 2z - 1$
$y \leftarrow [t; (\sqrt{1-t^2})v^\top]^\top$ {concatenation}
$\hat{u} \leftarrow e_1 - \mu$ {e_1 is the base vector $[1, 0, \cdots, 0]^\top$}
$u = \frac{\hat{u}}{\|\hat{u}\|_2}$
$x \leftarrow (I_d - 2uu^\top)y$ {I_d is the identity matrix $d \times d$}
Return: x

Experiments

Demonstrating high stability and equivalence to the vMF



(a) Stability of the vMF distribution. Ours does not have numerical issues in these intervals.



(b) Sampling time (on GPU) with $d = 64$ of a batch of 100 vectors of varius concentrations.

Figure 2. Comparing stability (a) and running time (b) of the von Mises-Fisher and the Power Spherical distribution.

Method	**vMF**		**Power Spherical**	
	LL	ELBO	LL	ELBO
$d = 5$	-114.51	-117.68	-114.49	-118.01
$d = 10$	-97.37	-101.78	-97.46	-101.86
$d = 20$	-93.80	-99.38	-93.70	-99.27
$d = 40$	-98.64	-108.44	-98.63	-108.32

Table 2. Comparison between the vMf and Power Spherical distributions in a VAE on MNIST with different dimensional latent spaces $\mathbb{S}^{d-1}$. We show estimated (with 5k Monte Carlo samples) log-likelihood (LL) and evidence lower bond (ELBO) on test set.

https://nicola-decao.github.io

Code at https://github.com/nicola-decao/power_spherical

The Power Spherical distribution

Useful proprieties

Property	Value
$\mathbb{E}[X]$	$\mu(\alpha-\beta)/(\alpha+\beta)$
$\mathrm{var}(X)$	$\frac{2\alpha}{(\alpha+\beta)^2(\alpha+\beta+1)}\big((\beta-\alpha)\mu\mu^\top+(\alpha+\beta)I_d\big)$
Mode	μ (for $\kappa > 0$)
$\mathrm{H}(X)$	$\log N_X(\kappa, d) - \kappa\big(\log 2 + \psi(\alpha) - \psi(\alpha+\beta)\big)$

Table 1. Properties of $X \sim \mathrm{PowerSpherical}(\mu, \kappa)$. Recall that $\alpha = (d-1)/2 + \kappa$ and $\beta = (d-1)/2$.

Theorem 16. *The Kullback–Leibler divergence* D_{KL} *(Definition 11) between a Power Spherical distribution P with parameters μ_p, κ_p and von Mises-Fisher and Q with parameters μ_q, κ_q is* $\mathrm{D}_{\mathrm{KL}}[P\|Q] =$

$$-\mathrm{H}(P) + \log C_X(\kappa_q, d) - \kappa_q \mu_q^\top \mu_p \left(\frac{\alpha-\beta}{\alpha+\beta}\right), \quad (67)$$

with $\alpha = \frac{d-1}{2} + \kappa$, $\beta = \frac{d-1}{2}$.

Theorem 17. *The Kullback–Leibler divergence* D_{KL} *(Definition 11) between a Power Spherical distribution P and a uniform distribution on the sphere $Q = \mathcal{U}(\mathbb{S}^{d-1})$ is*

$$\mathrm{D}_{\mathrm{KL}}[P\|Q] = -\mathrm{H}(P) + \mathrm{H}(Q) \quad (72)$$

https://nicola-decao.github.io

Code at https://github.com/nicola-decao/power_spherical