# Computational Statistics

*Hyperspherical VAE*

Victor Deng Inès Vati

École Normale Supérieure Paris-Saclay, Master MVA

$10^{\text{th}}$ JAN 2024

# Table of contents

# Table of contents

2018 paper from Tim R. Davidson *et al.* [DFDC$^+$]

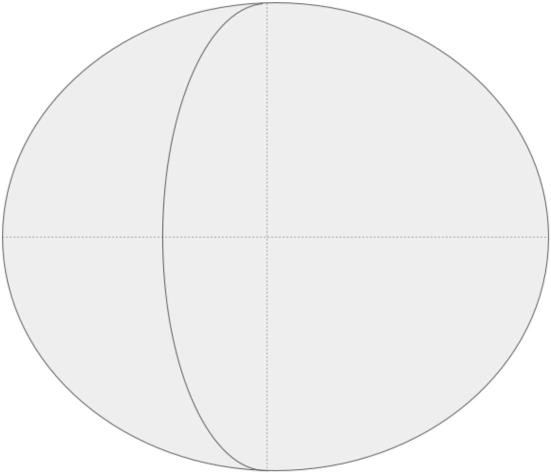
- Replacing the Gaussian prior and approximate posterior distributions with a von Mises-Fisher distribution
- Goal: better model data with a hyperspherical latent structure
- Various experiments, where the $\mathcal{S}$-VAE (von Mises-Fisher distributions) often outperforms the $\mathcal{N}$-VAE (Gaussian distributions) in low dimensions

# Table of contents

# Ines

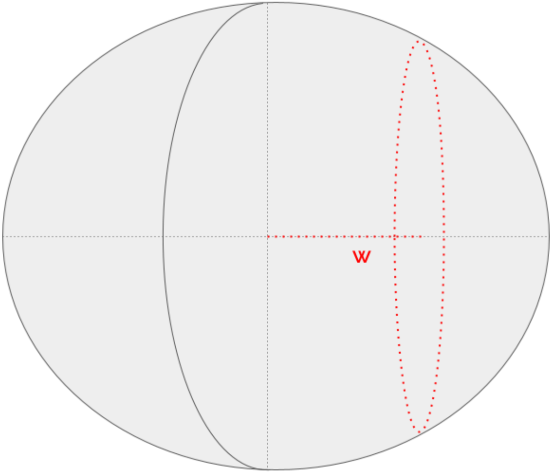
Démontrer que la méthode de sampling marche

# Sampling $w$ from $g(w|\kappa, \theta)$

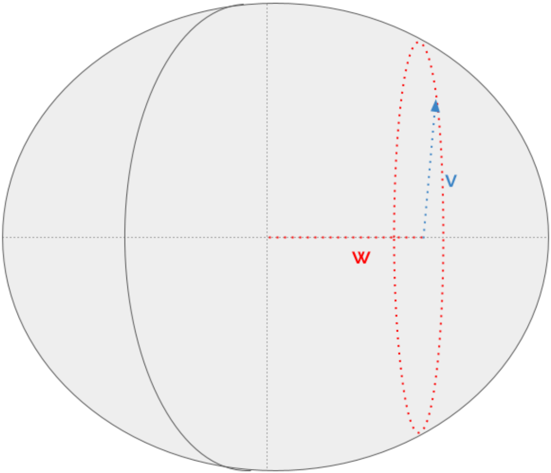


$S^2$ : unit sphere in $\mathbb{R}^3$

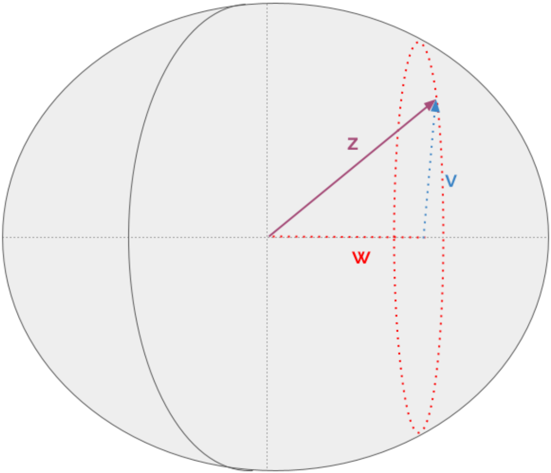
# Sampling $w$



# Sampling $w$



# Sampling $w$



# Table of contents

# Reparameterization Trick

The authors use a reparameterization trick that has been extended to distributions that can be sampled using rejection sampling [NRLB20].

**Algorithm 1** Reparameterized Rejection Sampling (from [NRLB20])

1: $i \leftarrow 0$
2: **repeat**
3: $\quad i \leftarrow i + 1$
4: $\quad$ Propose $\varepsilon_i \sim s(\varepsilon)$
5: $\quad$ Simulate $u_i \sim \mathcal{U}[0, 1]$
6: **until** $u_i < \frac{g(h(\varepsilon_i,\theta);\theta)}{r(h(\varepsilon_i,\theta);\theta)}$
7: **return** $\varepsilon_i$

# Monte Carlo estimation

By noting $\pi(\varepsilon|\theta)$ the distribution of the resulting $\varepsilon$, we have

$$\nabla_\theta \mathbb{E}_{g(\varepsilon|\theta)}[\ldots] = \mathbb{E}_{\pi(\varepsilon|\theta)}[\ldots] \text{ "} = \mathbb{E}_{(\varepsilon_i, U_i)_i}[\ldots]\text{"}$$

Problem: $(\varepsilon_i, U_i)_{i\in\mathbb{N}}$ is not a random variable (it is a stochastic process)
No reference to a convergence proof in [DFDC$^+$, NRLB20, PBJ12, MG14]

# Table of contents

# Experiments on link prediction

- Expliquer l'expérience (tâche à faire, données, negative sampling) (Victor)

reproduire l'experience

- data (Ines)
- implementer les modèles (Victor VGAE)
- gradients pour le hyperspherical VAE (Inès)
- courbes d'entraînement dans le cas normal (Victor)
- entrainement et evaluation

# Table of contents

# Conclusion and Discussion

- Quite meaningful contribution in low dimensions
- Algorithm not really useful in high dimensions, due to vanishing surface problem and soap bubble effect of the $\mathcal{N}$-VAE
- Much less variance parameters (1 vs. $d$ for $\mathcal{N}$-VAE), so possibly less expressivity
- vérifier différentes dimensions de l'espace latent
- et algo vraiment utile en petite ou moyenne dimension ?

# References


Tim R. Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M. Tomczak.
Hyperspherical variational auto-encoders.

Andriy Mnih and Karol Gregor.
Neural variational inference and learning in belief networks, 2014.

Christian A. Naesseth, Francisco J. R. Ruiz, Scott W. Linderman, and David M. Blei.
Reparameterization gradients through acceptance-rejection sampling algorithms, 2020.

John Paisley, David Blei, and Michael Jordan.
Variational bayesian inference with stochastic search, 2012.