Diethe, T. (2015). A Note on the Kullback-Leibler Divergence for the von Mises-Fisher distribution. *arXiv*.

A Note on the Kullback-Leibler Divergence for the von Mises-Fisher distribution

T.R. Diethe
Department of Electrical and Electronic Engineering
University of Bristol

February 26, 2015

Abstract

We present a derivation of the Kullback Leibler (KL)-Divergence (also known as Relative Entropy) for the von Mises Fisher (VMF) Distribution in $d-$dimensions.

1 Introduction

The von Mises Fisher (VMF) Distribution (also known as the Langevin Distribution [8]) is a probability distribution on the $(d-1)$-dimensional hypersphere S^{d-1} in $\mathbb{R}^d$ [3]. If $d = 2$ the distribution reduces to the von Mises distribution on the circle, and if $d = 3$ it reduces to the Fisher distribution on a sphere. It was introduced by [3] and has been studied extensively by [6, 7]. The first Bayesian analysis was in [5] and recently it has been used for clustering on a hypersphere by [2].

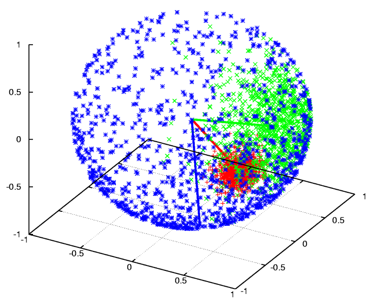

Figure 1: Three sets of 1000 points sampled from three VMF distributions on the 3D sphere with $\kappa = 1$ (blue), $\kappa = 10$ (green) and $\kappa = 100$ (red), respectively. The mean directions are indicated with arrows.

2 Preliminaries

2.1 Definitions

We will use $\log(z)$ to denote the natural logarithm of z throughout this article. Before continuing it will be useful to define the Gamma function $\Gamma(z)$,

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt, \qquad z \in \mathbb{C}, Re(z) > 0 \tag{1}$$

$$\Gamma(z) = (z-1)!, \qquad z \in \mathbb{Z}^+ \tag{2}$$

and its relation, the incomplete Gamma function $\Gamma(z,s)$,

$$\Gamma(z,s) = (s-1)!e^{-x}\sum_{m=0}^{s-1}\frac{z^m}{m!}, \quad z \in \mathbb{Z}^+ \tag{3}$$

and the Modified Bessel Function of the First Kind $I_\alpha(z)$,

$$I_\alpha(z) = \sum_{m=0}^{\infty}\frac{(z/2)^{2m+\alpha}}{m!\Gamma(m+\alpha+1)}, \tag{4}$$

which also has the following integral representations [1],

$$I_\alpha(z) = \frac{(z/2)^\alpha}{\sqrt{\pi}\Gamma(\alpha+1/2)}\int_0^\pi e^{\pm z\cos\theta}\sin^{2d}\theta \; d\theta, \qquad (\alpha \in \mathbb{R}) \tag{5}$$

$$= \frac{(z/2)^\alpha}{\sqrt{\pi}\Gamma(\alpha+1/2)}\int_{-1}^1 (1-t^2)^{(\alpha-1/2)}e^{\pm zt}\; dt. \qquad (\alpha \in \mathbb{R}, \alpha > -0.5) \tag{6}$$

Also of interest is the logarithm of this quantity (using the second integral definition (6)),

$$\begin{aligned}\log\left(I_\alpha(z)\right) &= \log\left[\frac{(\frac{z}{2})^\alpha}{\sqrt{\pi}\Gamma(\alpha+1/2)}\int_{-1}^1 (1-t^2)^{(\alpha-1/2)}e^{\pm zt}\; dt\right]\\ &= \log\frac{(\frac{z}{2})^\alpha}{\sqrt{\pi}\Gamma(\alpha+1/2)} + \log\left[\int_{-1}^1 (1-t^2)^{(\alpha-1/2)}e^{\pm zt}\; dt\right]\\ &= \log\left(\frac{z}{2}\right)^\alpha - \log\sqrt{\pi}\Gamma(\alpha+1/2) + \log\left[\int_{-1}^1 (1-t^2)^{(\alpha-1/2)}e^{\pm zt}\; dt\right].\end{aligned} \tag{7}$$

Note that the second term does not depend on z.

The Exponential Integral function $E_n(z)$ is given by,

$$\begin{aligned}E_\alpha(z) &= \int_1^\infty \frac{e^{-zt}}{t^\alpha}dt,\\ &= z^{\alpha-1}\Gamma(1-\alpha, z).\end{aligned} \tag{8}$$

An identity that will be useful is,

$$\int_{-1}^1 (1-t)^d e^{t\kappa} = -2^{d-1}E_{-d}(2\kappa)e^\kappa. \quad d > 0 \tag{9}$$

2.2 The von Mises Fisher (VMF) distribution

The probability density function (PDF) of the VMF distribution for a random d-dimensional unit vector $\mathbf{x}(\|\mathbf{x}\|_2 = 1)$ is given by:

$$M_d(\boldsymbol{\mu}, \kappa) = c_d(\kappa) e^{\kappa \boldsymbol{\mu}' \mathbf{x}}, \quad \mathbf{x} \in S^{d-1}, \tag{10}$$

where the normalisation constant $c_d(\kappa)$ is given by,

$$c_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}. \tag{11}$$

The (non-symmetric) Kullback Leibler (KL)-Divergence from one probability distributions $q(\mathbf{x})$ to another probability distribution $p(\mathbf{x})$ is defined as,

$$\mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) = \int_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} \, d\mathbf{x}, \tag{12}$$

$$= \mathbb{E}_x \left[\log \frac{q(\mathbf{x})}{p(\mathbf{x})} \right]. \tag{13}$$

Although this is general to any two distributions, we will assume that $p(\mathbf{x})$ is the "prior" distribution and $q(\mathbf{x})$ is the "posterior" distribution as commonly used in Bayesian analysis.

3 KL-Divergence for the VMF Distribution

3.1 General Case

We will assume that we have prior and posterior distributions defined over vectors $\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_2 = 1$ as follows,

$$\begin{aligned} p(\mathbf{x}) &\sim M_d(\boldsymbol{\mu}_p, \kappa_p), \\ q(\mathbf{x}) &\sim M_d(\boldsymbol{\mu}_q, \kappa_q). \end{aligned} \tag{14}$$

We will now derive the KL-Divergence for two VMF distributions. The main problem in doing so will be the the normalisation constants $c_d(\kappa_p)$ and $c_d(\kappa_q)$.

Theorem 3.1 *For prior and posterior distributions as defined above over vectors* $\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_2 = 1, d < \infty, d$ *odd*[1]*, we have*

$$\begin{aligned} \mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) \leq \kappa_q - \kappa_p \boldsymbol{\mu}_p' \boldsymbol{\mu}_q + d^{\bullet} \log(\kappa_q) + \sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \\ - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond} + 1) \log d^{\diamond} - d^{\diamond 2} + 1 \end{aligned} \tag{15}$$

Proof From (12), letting $d^{\star} = \frac{d}{2} - 1$, $d^{\diamond} = \frac{d-3}{2}$, and $d^{\bullet} = \frac{d-1}{2}$, we have,

$$\mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) = \int_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x},$$

[1] For even d we can simply add a "null" dimension

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\log c_d(\kappa_q) e^{\kappa_q \boldsymbol{\mu}_q' \mathbf{x}} - \log c_d(\kappa_p) e^{\kappa_p \boldsymbol{\mu}_p' \mathbf{x}}\right] d\mathbf{x}, \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\log c_d(\kappa_q) - \log c_d(\kappa_p) + \kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right] d\mathbf{x}, \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[d^\star \log(\kappa_q) - (d/2)\log(2\pi) - \log I_{d^\star}(\kappa_q)\right. \\
&\qquad \left. - d^\star \log(\kappa_p) + (d/2)\log(2\pi) + \log I_{d^\star}(\kappa_p) + \kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right] d\mathbf{x}, \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[d^\star \log\left(\frac{\kappa_q}{\kappa_p}\right) - \log I_{d^\star}(\kappa_q) + \log I_{d^\star}(\kappa_p) + \kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[d^\star \log\left(\frac{\kappa_q}{\kappa_p}\right) + \kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right. \\
&\qquad - \log\left(\frac{\kappa_q}{2}\right)^{d^\star} + \log\sqrt{\pi}\,\Gamma\left(d^\star + \frac{1}{2}\right) - \log\int_{-1}^{1} (1-t^2)^{(d^\star - 1/2)} e^{\pm\kappa_q t}\, dt \\
&\qquad \left. + \log\left(\frac{\kappa_p}{2}\right)^{d^\star} - \log\sqrt{\pi}\,\Gamma\left(d^\star + \frac{1}{2}\right) + \log\int_{-1}^{1} (1-t^2)^{(d^\star - 1/2)} e^{\pm\kappa_p t}\, dt\right] d\mathbf{x}
\end{aligned}
$$

(Using (7)) (16)

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[d^\star \log\left(\frac{\kappa_q}{\kappa_p}\right) + \kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} - d^\star \log\left(\frac{\kappa_q}{2}\right) + d^\star \log\left(\frac{\kappa_p}{2}\right)\right. \\
&\qquad \left. - \log\int_{-1}^{1} (1-t^2)^{d^\diamond} e^{\pm\kappa_q t}\, dt + \log\int_{-1}^{1} (1-t^2)^{d^\diamond} e^{\pm\kappa_p t}\, dt\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right. \\
&\qquad \left. - \log\int_{-1}^{1} (1-t^2)^{d^\diamond} e^{\pm\kappa_q t}\, dt + \log\int_{-1}^{1} (1-t^2)^{d^\diamond} e^{\pm\kappa_p t}\, dt\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x}\right. \\
&\qquad \left. - \log\left[-2^{\frac{d-3}{2}} E_{-d^\diamond}(2\kappa_q) e^{\kappa_q}\right] + \log\left[-2^{\frac{d-3}{2}} E_{-d^\diamond}(2\kappa_p) e^{\kappa_p}\right]\right] d\mathbf{x}
\end{aligned}
$$

(Using (9)) (17)

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} - \kappa_q + \kappa_p - \log\left[E_{-d^\diamond}(2\kappa_q)\right] + \log\left[E_{-d^\diamond}(2\kappa_p)\right]\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q(\boldsymbol{\mu}_q' \mathbf{x} - 1) - \kappa_p(\boldsymbol{\mu}_p' \mathbf{x} - 1)\right. \\
&\qquad \left. - \log\left(2\kappa_q^{-d^\bullet} \Gamma\left(d^\bullet, 2\kappa_q\right)\right) + \log\left(2\kappa_p^{-d^\bullet} \Gamma\left(d^\bullet, 2\kappa_p\right)\right)\right] d\mathbf{x}
\end{aligned}
$$

(Using the definition of the Exponential Integral function (8)) (18)

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q(\boldsymbol{\mu}_q' \mathbf{x} - 1) - \kappa_p(\boldsymbol{\mu}_p' \mathbf{x} - 1) + d^\bullet \log(2\kappa_q) - d^\bullet \log(2\kappa_q)\right. \\
&\qquad \left. - \log\left(\Gamma\left(d^\bullet, 2\kappa_q\right)\right) + \log\left(\Gamma\left(d^\bullet, 2\kappa_p\right)\right)\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q(\boldsymbol{\mu}_q' \mathbf{x} - 1) - \kappa_p(\boldsymbol{\mu}_p' \mathbf{x} - 1) + d^\bullet \log(2\kappa_q) - d^\bullet \log(2\kappa_q)\right. \\
&\qquad \left. - \log\left(d^\diamond! e^{-\kappa_q} \sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) + \log\left(d^\diamond! e^{-\kappa_p} \sum_{m=0}^{d^\diamond} \frac{\kappa_p^m}{m!}\right)\right] d\mathbf{x}
\end{aligned}
$$

(Using (3) and that $d^\bullet - 1 = d^\diamond$) (19)

$$= \int_{\mathbf{x}} q(\mathbf{x}) \big[\kappa_q(\boldsymbol{\mu}_q'\mathbf{x} - 1) - \kappa_p(\boldsymbol{\mu}_p'\mathbf{x} - 1) + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p) + \kappa_q - \kappa_p$$
$$- \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) + \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_p^m}{m!}\right)\Bigg] d\mathbf{x}$$
$$= \int_{\mathbf{x}} q(\mathbf{x}) \big[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p)$$
$$- \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) + \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_p^m}{m!}\right)\Bigg] d\mathbf{x}$$

Further simplifications: (20)

$$\leq \int_{\mathbf{x}} q(\mathbf{x}) \big[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p)$$
$$- \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) + \left(\sum_{m=0}^{d^\diamond} \log \frac{\kappa_p^m}{m!}\right)\Bigg] d\mathbf{x}$$

(by Jensen's inequality) (21)

$$= \int_{\mathbf{x}} q(\mathbf{x}) \big[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p)$$
$$+ \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) - \sum_{m=0}^{d^\diamond} (m \log(\kappa_p) - \log m!)\Bigg] d\mathbf{x}$$
$$\leq \int_{\mathbf{x}} q(\mathbf{x}) \big[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p)$$
$$+ \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right) - \sum_{m=1}^{d^\diamond} (m \log(\kappa_p) - m \log m + m - 1)\Bigg] d\mathbf{x}$$

(using $n \log \frac{n}{e} + 1 \leq \log n! \leq (n+1) \log \frac{n+1}{e} + 1$) (22)

$$= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p) + \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right)\right.$$
$$\left. - \sum_{m=1}^{d^\diamond} (m \log(\kappa_p) - m \log m) - d^\diamond(d^\diamond + 1) + (d^\diamond + 1)\right] d\mathbf{x}$$
$$= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p) + \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right)\right.$$
$$\left. - \sum_{m=1}^{d^\diamond} (m \log(\kappa_p) - m \log m) - d^{\diamond 2} + 1\right] d\mathbf{x}$$
$$= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q\boldsymbol{\mu}_q'\mathbf{x} - \kappa_p\boldsymbol{\mu}_p'\mathbf{x} + d^\bullet \log(\kappa_q) - d^\bullet \log(\kappa_p) + \log\left(\sum_{m=0}^{d^\diamond} \frac{\kappa_q^m}{m!}\right)\right.$$
$$\left. - \sum_{m=1}^{d^\diamond} (m \log(\kappa_p)) + d^\diamond(d^\diamond + 1) \log d^\diamond - d^{\diamond 2} + 1\right] d\mathbf{x}$$

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) - d^{\bullet} \log(\kappa_p) + \log \left(\sum_{m=0}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \right) \\
&\qquad -d^{\diamond}(d^{\diamond}+1)\log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) - d^{\bullet} \log(\kappa_p) + \log \left(\sum_{m=0}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \right) \\
&\qquad - \left(\frac{(d-3)^2}{4} + \frac{d-3}{2} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) + \log \left(\sum_{m=0}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \right) \\
&\qquad - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x} \\
&= \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) + \log \left(1 + \sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \right) \\
&\qquad - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x} \\
&\leq \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) + \sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \\
&\qquad - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x}
\end{aligned}
$$

$$
\text{(using } n \geq \log(1+n) \geq \tfrac{n}{1+n},\ (n > -1)\text{)} \tag{23}
$$

$$
\begin{aligned}
&= \int_{\mathbf{x}} q(\mathbf{x}) \Bigg[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - \kappa_p \boldsymbol{\mu}_p' \mathbf{x} + d^{\bullet} \log(\kappa_q) + \sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \\
&\qquad - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1 \Bigg] d\mathbf{x} \\
&= \kappa_q - \kappa_p \boldsymbol{\mu}_p' \boldsymbol{\mu}_q + d^{\bullet} \log(\kappa_q) + \sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \\
&\qquad - \left(\frac{d^2 - 2d + 1}{4} \right) \log(\kappa_p) + d^{\diamond}(d^{\diamond}+1)\log d^{\diamond} - d^{\diamond 2} + 1
\end{aligned}
$$

$$
\text{(as } \textstyle\int_{\mathbf{x}} q(\mathbf{x}) = 1, \text{ and } \mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}_q, \text{ and } \boldsymbol{\mu}_q' \boldsymbol{\mu}_q = 1\text{)} \tag{24}
$$

The term $\boldsymbol{\mu}_q' \boldsymbol{\mu}_p$ can be seen as the cosine distance between the prior and postieror mean vectors. For $0 < \kappa_q < 1$, the term $\sum_{m=1}^{d^{\diamond}} \frac{\kappa_q^m}{m!} \geq \kappa_q$. However for large κ_q and large d this term can grow very large.

Special case: uniform prior

Since the VMF distribution is defined on the S^{d-1}, hypersphere, which is actually a specific case of a Stiefel manifold where $r = 1$ is the radius. The Stiefel manifold has finite area,

$$\tau(d,r) = \frac{2^r \pi^{\frac{pr}{2}}}{\pi^{\frac{r(r-1)}{4}} \prod_{j=1}^{r} \Gamma\left(\frac{p-j+1}{2}\right)}, \tag{25}$$

and so,

$$\tau(d,1) = \frac{2\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)}, \tag{26}$$

For the special case of the uniform prior (more precisely $\lim_{\kappa_p \to 0}$), the prior PDF reduces to,

$$\begin{aligned} M_d(\boldsymbol{\mu}, \kappa) &= c_d(0) e^0 \\ &= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{\frac{d}{2}}}, \end{aligned} \tag{27}$$

which is simply one over the area on the manifold. This leads to a simpler form for the KL-divergence.

Corollary 3.2 *For prior and posterior distributions as defined above over vectors* $\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_2 = 1, d < \infty$, *we have*

$$\mathrm{KL}(q(\mathbf{x}) || p(\mathbf{x})) = \kappa_q - d^\star \log 2 \tag{28}$$

Proof

$$\begin{aligned} \mathrm{KL}(q(\mathbf{x}) || p(\mathbf{x})) &= \int_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\log c_d(\kappa_q) e^{\kappa_q \boldsymbol{\mu}_q' \mathbf{x}} - \log c_d(0) \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + \log c_d(\kappa_q) - \log \Gamma\left(\frac{d}{2}\right) + \log\left(2\pi^{\frac{d}{2}}\right) \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + \log c_d(\kappa_q) - \log(d^\star)! + (d/2) \log(2\pi) \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + d^\star \log(\kappa_q) - (d/2) \log(2\pi) \right. \\ &\qquad \left. - \log I_{d^\star}(\kappa_q) - \log(d^\star)! + (d/2) \log(2\pi) \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + d^\star \log(\kappa_q) - \log I_{d^\star}(\kappa_q) - \log(d^\star)! \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + d^\star \log(\kappa_q) - \log\left(\frac{\kappa_q}{2}\right)^{d^\star} + \log \Gamma\left(\frac{d}{2}\right) - \log(d^\star)! \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} + d^\star \log(\kappa_q) - d^\star \log\left(\frac{\kappa_q}{2}\right) \right] d\mathbf{x}, \\ &= \int_{\mathbf{x}} q(\mathbf{x}) \left[\kappa_q \boldsymbol{\mu}_q' \mathbf{x} - d^\star \log 2 \right] d\mathbf{x}, \\ &= \kappa_q - d^\star \log 2, \end{aligned} \tag{29}$$

For this special case, it can be seen that the dependence on the dimension is much more benign. This could prove useful for further computation (*e.g.* if the KL-divergence were to be used in a probably approximately correct (PAC)-Bayes bound [4]).

4 Conclusions

We have presented a derivation of the Kullback Leibler (KL)-divergence for the von Mises Fisher (VMF)-distribution, including the special case of a uniform prior over the hypersphere.

References

[1] Milton Abramowitz, Irene A Stegun, et al. *Handbook of mathematical functions*, volume 1. Dover New York, 1972.

[2] Arindam Banerjee, Inderjit S Dhillon, Joydeep Ghosh, and Suvrit Sra. Clustering on the unit hypersphere using von mises-fisher distributions. In *Journal of Machine Learning Research*, pages 1345–1382, 2005.

[3] Ronald Fisher. Dispersion on a sphere. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 217, pages 295–305. The Royal Society, 1953.

[4] John Langford. Tutorial on practical prediction theory for classification. In *Journal of machine learning research*, pages 273–306, 2005.

[5] Kanti V Mardia and SAM El-Atoum. Bayesian inference for the von mises-fisher distribution. *Biometrika*, 63(1):203–206, 1976.

[6] Kantilal Varichand Mardia. *Statistics of directional data*. Academic Press, 2014.

[7] KV Mardia and PJ Zemroch. Algorithm as 86: The von mises distribution function. *Applied Statistics*, pages 268–272, 1975.

[8] Yoko Watamori et al. Statistical inference of langevin distribution for directional data. *Hiroshima Mathematical Journal*, 26(1):25–74, 1996.