Supplementary provements to "vMF-Contact: Uncertainty-aware Evidential Learning for Probabilistic Contact-grasp in Noisy Clutter"

Yitian Shi, Rania Rayyes

March 10, 2025

1 Introduction to 3D von-Mises-Fisher Distribution

The von Mises-Fisher (vMF) distribution is one of the exponential family distributions on the $(p-1)$-dimensional unit sphere in $\mathbb{R}^p$. In the area of robotic manipulation, Liu et al. [LQY+24] utilized the Power Spherical (PS) distribution [DCA20] to model the baseline vectors of contact grasp representation [SMTF21] as a substitute for the vMF distribution, owing to its stability and faster sampling speed compared to Gibbs sampling. Here we consider modeling the 3-D ($p = 3$) vMF distribution, which is systematically analyzed in [Str17]:

$$\mathrm{vMF}(x; \mu, \kappa) = Z(\kappa) \exp(\kappa \mu^\top x), Z(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)}.$$

Given random mean parameter μ and constant κ, the conjugate prior of vMF distribution is another vMF distribution as $\mathrm{vMF}(\mu|\mu_0, \kappa_0)$ [NAGP05]. When directional data $\mathbf{x} = \{x_i\}_{i=1}^N$ is collected, the posterior distribution can be derived by:

$$p(\mu \mid \mathbf{x}; \kappa, \mu_0, \kappa_0) \propto \mathrm{vMF}(\mu|\mu_0, \kappa_0) \prod_{i=1}^{N} \mathrm{vMF}(x_i|\mu, \kappa).$$

In addition, in cases where numerical Monte Carlo (MC) sampling: $\mathbb{E}_{\theta \sim \mathbb{Q}_{\text{post.}}}[f(\theta^s)] \approx \frac{1}{S} \sum_{s=1}^{S} f(\theta^s)$ is unavoidable due to analytical intractability, PS distributions can be ideal surrogates with the same parameterizations as the vMF posterior $\mathbb{Q}_{\text{post.}} = \mathrm{vMF}(\mu \mid \mathbf{x}; \kappa, \mu_0, \kappa_0)$. However, it is important to note that the conjugate prior of the PS distribution can only be derived in close form [DCA20], which may impose limitations on analytical stability.

For joint conjugate prior between μ and κ one can only derive $p(\mu, \kappa \mid \mathbf{x}; \mu_0, a, b)$ with $0 < a < b$ up to proportionality. While there's no analytical solution for normalization of the posterior density as well as statistics in terms of e.g. marginalization, entropy, maximum likelihood, etc [NAGP05]. For this reason, we only consider vMF with a fixed concentration parameter κ in the following derivations.

2 Evidential learning with natural posterior networks (NatPN)

We follow the same regime to formulate the evidential learning with natural posterior update as Charpentier et al. [CBZ+22]. Since vMF belongs to exponential family distribution the corresponding natural parameters can be interesting. Consider the following formulation as the general expression of the exponential family [BN06]:

$$f(x; \theta) = h(x) \exp\left(\eta(\theta)^\top T(x) - A(\theta)\right).$$

We can rewrite 3D vMF density as:

$$f(x; \kappa\mu) = \exp(\kappa \mu^\top x + \log Z(\kappa)),$$

with natural parameter: $\eta(\kappa\mu) = \kappa\mu$, sufficient statistic: $T(x) = x$, log-partition function: $A(\kappa\mu) = -\log Z(\kappa)$. The posterior distribution:

$$Q(\theta|\chi^{(\cdot)}, m^{(\cdot)}) = \eta(\chi^{(\cdot)}, n) \exp\left(m^{(\cdot)}\theta^\top \chi^{(\cdot)} - m^{(\cdot)} A(\theta)\right),$$

with the corresponding update:

$$\chi^{\text{post},(i)} = \frac{m^{\text{prior}}\chi^{\text{prior}} + m^{(i)}\chi^{(i)}}{m^{\text{prior}} + m^{(i)}},\ m^{\text{post},(i)} = m^{\text{prior}} + m^{(i)}$$

exits for vMF as a member of the exponential family. With $\chi^{\text{prior}} = \mu_0$, $m^{\text{prior}} = \kappa_0$ and $\eta(\chi^{(\cdot)}, m)$ as normalization factor, we may perform the following posterior update for vMF as well:

$$\mu_i^{\text{post}} = \frac{\kappa_0\mu_0 + m_i x_i}{\kappa_0 + m_i},\ \kappa_i^{\text{post}} = \kappa_0 + m_i,\ m_i \equiv N_H p(x_i).$$

Here m_i represents the evidence (or pseudo-count) given N_H as the scaling factor (or certainty budget). The evidence inherently contains epistemic uncertainty. Intuitively, the posterior update can be considered as the linear interpolation between the prior and observed data.

2.1 Maximum a-posterior

One important statistic is the maximum a-posterior (MAP). Here the posterior distribution for μ and κ given the priors μ_0 and κ_0 can be written as:

$$p(\mu, \kappa \mid \mathbf{x}; \mu_0, \kappa_0) \propto p(\mathbf{x} \mid \mu, \kappa) p(\mu \mid \mu_0, \kappa_0).$$

Given the vMF likelihood $p(\mathbf{x}; \mu, \kappa) = Z(\kappa)\exp(\kappa\mathbf{x}^\top\mu)$ and prior $p(\mu; \mu_0, \kappa_0) = Z(\kappa_0)\exp(\kappa_0\mu_0^\top\mu)$

$$\log p(\mu \mid \mathbf{x}; \kappa, \mu_0, \kappa_0) \propto \kappa\mathbf{x}^\top\mu + \kappa_0\mu_0^\top\mu,$$

with $\log Z(\kappa_0), \log Z(\kappa)$ as constants. Since μ is a unit vector, the maximum occurs when:

$$\vartheta_N = \kappa_0\mu_0 + \kappa\mathbf{x},\ \mu_{\text{MAP}} = \frac{\vartheta_N}{\|\vartheta_N\|_2}.$$

This aligns with the posterior mean from [Str17] , where $p(\mu \mid \mathbf{x}; \kappa, \mu_0, \kappa_0) = \text{vMF}\left(\mu; \frac{\vartheta_N}{\|\vartheta_N\|_2}, \|\vartheta_N\|_2\right)$.

2.2 Bayesian loss

In evidential deep learning, one needs to optimize the following objective as "Bayesian loss":

$$\mathcal{L}_i^{post} = -\underbrace{\mathbb{E}_{\boldsymbol{\theta}\sim\mathbb{Q}_{\text{post.}}}[\log \mathbb{P}(x_i \mid \boldsymbol{\theta})]}_{\text{(i)}} - \underbrace{\mathbb{H}[\mathbb{Q}_{\text{post.}}]}_{\text{(ii)}}, \tag{1}$$

where (i) is the expected log-likelihood and (ii) denotes the entropy of the predicted posterior distribution $\mathbb{Q}_{\text{post.}}$. In terms of vMF posterior, the entropy can be derived by [CBZ+22]:

$$\mathbb{H}(\text{vMF}(\cdot; \mu_0, \kappa_0)) = -\log Z(\kappa_0) - \frac{\kappa_0}{\tanh(\kappa_0)} + 1. \tag{2}$$

This represents the epistemic uncertainty, which is independent of the mean direction μ.

As for the expected log-likelihood of vMF distribution, we would like to derive the analytical formulation of expected log-likelihood κ:

$$\begin{aligned}
&\mathbb{E}_{\mu\sim\mathbb{Q}(\mu_0,\kappa_0)}[\log \mathbb{P}(x_i \mid \mu)] \\
=&\int_{\mu\in S^2} \text{vMF}(\mu; \mu_0, \kappa_0) \log \text{vMF}(x_i; \mu, \kappa)\, d\mu \\
=&\int_{\mu\in S^2} Z(\kappa_0)\exp(\kappa_0\mu_0^\top\mu)\left(\log Z(\kappa) + \kappa x_i^\top\mu\right) d\mu \\
=&Z(\kappa_0)[\log Z(\kappa) \underbrace{\int_{\mu\in S^2} \exp(\kappa_0\mu_0^\top\mu)\, d\mu}_{①} + \kappa \underbrace{\int_{\mu\in S^2} \exp(\kappa_0\mu_0^\top\mu) x_i^\top\mu\, d\mu}_{②}].
\end{aligned}$$

Due to rotational symmetry, we can assume $\mu_0 = (0,0,1)^\top$. Transform vector μ to cartesian coordinates using $\mu = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^\top$, the term ① gives:

$$\begin{aligned}
&\int_0^{2\pi}\int_0^{\pi} \exp(\kappa_0\cos\theta)\sin\theta\, d\theta\, d\phi \\
=&2\pi\int_0^{\pi} \exp(\kappa_0\cos\theta)\sin\theta\, d\theta = 4\pi\frac{\sinh(\kappa_0)}{\kappa_0}.
\end{aligned}$$

For generality, we need to apply the same transformation to the data x_i as we would transform any $\mu_0 \in S^2$ to align with the z-axis. This is not necessary for μ due to the integral over the whole sphere. Suppose $x_i^{\text{proj}} = (x_{ia}, x_{ib}, x_{ic})^\top$, the term ② with x_i^{proj} as the data after transformation:

$$\begin{aligned}
&\int_0^{2\pi}\int_0^{\pi} \exp(\kappa_0\cos\theta)x^\top \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix} \sin\theta\, d\theta\, d\phi \\
=&\int_0^{2\pi}\int_0^{\pi} \exp(\kappa_0\cos\theta)\Big(\underbrace{\sin\theta\cos\phi x_{ia} + \sin\theta\sin\phi x_{ib}}_{\text{integral as 0 over }\phi} + \cos\theta x_{ic}\Big)\sin\theta\, d\theta\, d\phi \\
=&2\pi x_{ic}\int_0^{\pi} \exp(\kappa_0\cos\theta)\cos\theta\sin\theta\, d\theta \\
=&4\pi x_{ic}\left(\frac{\cosh(\kappa_0)}{\kappa_0} - \frac{\sinh(\kappa_0)}{\kappa_0^2}\right)
\end{aligned}$$

We can see that terms x_{ia} and x_{ib} are removed since the integral over $\sin(\phi)$ and $\cos(\phi)$ cancel out the positive and negative halves in range $0-2\pi$. Finally, combine ① and ② together:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}\sim\mathbb{Q}_{\text{post.}}}\left[\log \mathbb{P}(x_i \mid \boldsymbol{\theta})\right] \\
=&Z(\kappa_0)\left[4\pi\log Z(\kappa)\frac{\sinh(\kappa_0)}{\kappa_0} + 4\pi\kappa x_{ic}\left(\frac{\cosh(\kappa_0)}{\kappa_0} - \frac{\sinh(\kappa_0)}{\kappa_0^2}\right)\right] \\
=&\frac{4\pi Z(\kappa_0)}{\kappa_0}\left[\log Z(\kappa)\sinh(\kappa_0) + \kappa x_{ic}\cosh(\kappa_0) - \kappa x_{ic}\frac{\sinh(\kappa_0)}{\kappa_0}\right] \\
=&\log Z(\kappa) - \frac{\kappa x_{ic}}{\kappa_0} + \frac{\kappa x_{ic}}{\tanh(\kappa_0)}
\end{aligned}$$

and x_{ic} is the projection of x_i on any $\mu_0 \in S^2$ as $x_{ic} = \frac{x_i^\top\mu_0}{\|\mu_0\|}$ with $\|\mu_0\| = 1$. This finally gives:

$$\begin{aligned}
&\mathbb{E}_{\mu\sim\mathbb{Q}(\mu_0,\kappa_0)}\left[\log \mathbb{P}(x_i \mid \mu)\right] \\
=&\log Z(\kappa) + \left(\frac{1}{\tanh(\kappa_0)} - \frac{1}{\kappa_0}\right)\kappa x_i^\top\mu_0
\end{aligned}$$

3 Short analysis of Bayesian loss

As in the aforementioned Bayesian loss (Eq.1), if we first put our attention on the expected log-likelihood, where, intuitively, the normalizer $\log Z(\kappa)$ achieves maximum when $\kappa = 0$, imposing high aleatoric uncertainty for small κ_0. But when $\kappa_0 >> 0$, $\frac{1}{\tanh(\kappa_0)} - \frac{1}{\kappa_0} \approx 1$ (which also means low epistemic uncertainty due to the posterior update $\kappa_0 \leftarrow \kappa_0 + m^i$), to maximize the expected log-likelihood, there exists a balance between this normalizer $\log Z(\kappa)$ and $\kappa x_i^\top\mu_0$, where κ is weighted by the alignment between data x_i with the prior/posterior mean direction μ_0. While in case the posterior has a small concentration κ_0 (or high epistemic uncertainty), the second term will also be down-weighted.

In terms of learning κ_0 (or n_i in specific), the expected log-likelihood intends to reduce the epistemic uncertainty by maximizing $\frac{1}{\tanh(\kappa_0)} - \frac{1}{\kappa_0}$, while in Bayesian loss, this is balanced through the maximization of the entropy $\mathbb{H}(\text{vMF}(\cdot;\mu_0,\kappa_0))$ in Eq.2.

References

[BN06] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[CBZ+22] Bertrand Charpentier, Oliver Borchert, Daniel Zügner, Simon Geisler, and Stephan Günnemann. Natural Posterior Network: Deep Bayesian Predictive Uncertainty for Exponential Family Distributions. In *International Conference on Learning Representations*, 2022.

[DCA20] Nicola De Cao and Wilker Aziz. The power spherical distribution. *arXiv preprint arXiv:2006.04437*, 2020.

[LQY+24] Yushi Liu, Alexander Qualmann, Zehao Yu, Miroslav Gabriel, Philipp Schillinger, Markus Spies, Ngo Anh Vien, and Andreas Geiger. Efficient end-to-end detection of 6-dof grasps for robotic bin picking. *arXiv preprint arXiv:2405.06336*, 2024.

[NAGP05] Gabriel Nunez-Antonio and Eduardo Gutiérrez-Pena. A bayesian analysis of directional data using the von mises–fisher distribution. *Communications in Statistics—Simulation and Computation®*, 34(4):989–999, 2005.

[SMTF21] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13438–13444. IEEE, 2021.

[Str17] Julian Straub. Bayesian inference with the von-mises-fisher distribution in 3d, 2017.