VON MISES-FISHER SAMPLING OF GLOVE VECTORS

Walid Bendada[1,2], **Guillaume Salha-Galvan**[1], **Romain Hennequin**[1], **Théo Bontempelli**[1], **Thomas Bouabça**[1], **Tristan Cazenave**[2]
[1]Deezer Research, Paris, France
[2]LAMSADE, Université Paris Dauphine, PSL, Paris, France
research@deezer.com

ABSTRACT

A recent publication introduced von Mises-Fisher exploration (vMF-exp), a scalable sampling method for exploring large action sets in reinforcement learning problems where hyperspherical embedding vectors represent these actions. We present the first experimental validation of vMF-exp's key theoretical and scalability properties on a publicly available real-world dataset, confirming the potential of this method.

1 INTRODUCTION

In Bendada et al. (2024), we introduced von Mises-Fisher exploration (vMF-exp), a sampling method for exploring large action sets in reinforcement learning (RL) problems where hyperspherical embedding vectors represent these actions. We demonstrated that vMF-exp scales to millions of actions and exhibits several desirable properties. However, we did not test vMF-exp on publicly available real-world data, which is essential for reproducibility and deeper understanding of the method.

This paper addresses this limitation by experimentally validating the main properties of vMF-exp on a large-scale, publicly available real-world dataset of GloVe word embedding vectors (Pennington et al., 2014), reinforcing conclusions from the initial work. This paper is organized as follows: Section 2 covers preliminaries, Section 3 presents our experiments on GloVe vectors, and Section 4 concludes.

2 EXPLORING LARGE ACTION SETS USING VON MISES-FISHER SAMPLING

Context We study an RL agent sequentially selecting actions from a set $\mathcal{I}_n = \{1, 2, \ldots, n\}$ of $n \in \mathbb{N}^*$ actions. Each action $i \in \mathcal{I}_n$ is represented by an embedding vector $X_i \in \mathbb{R}^d$, where $2 \leq d \ll n$. All vectors have unit Euclidean norms, forming the set $\mathcal{X}_n = \{X_i\}_{1 \leq i \leq n} \in (\mathcal{S}^{d-1})^n$, where $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ is the d-dimensional unit hypersphere. At time t, the agent observes a current state vector $V_t \in \mathcal{S}^{d-1}$ and selects the next action for time $t+1$ in $\mathcal{I}_n$, with its relevance evaluated by a reward from the environment that the agent aims to maximize (Sutton and Barto, 2018). An instance of this setting is an agent sequentially recommending songs on a music streaming service, where each song is represented by some embedding vector. The agent updates its recommendations based on the previous song V_t and user feedback (likes, skips...) as the reward. In this example, the number of candidate actions can reach several millions (Bendada et al., 2023).

Objective In Bendada et al. (2024), we researched effective exploration strategies for such large-scale settings. Given V_t, an exploration strategy P assigns each action $i \in \mathcal{I}_n$ a sampling probability $P(i \mid V_t) \in [0, 1]$. We sought a strategy satisfying three crucial properties (**P1**, **P2**, **P3**). **P1** is scalability. It requires action sampling to have sublinear time complexity with respect to n, allowing the exploration of millions of actions. **P2** is unrestricted radius. Radius$(P \mid V_t)$ is the number of actions with a non-zero sampling probability given V_t. While exploring distant actions may be suboptimal (e.g., leading to poor song recommendations), restricting the radius by design would prevent the exploration of relevant actions beyond a certain range, limiting flexibility. **P3** is order preservation. The probability of sampling an action must increase with its embedding similarity to V_t, as measured, e.g., using inner products: $\forall (i, j) \in \mathcal{I}_n^2, \langle V_t, X_i \rangle > \langle V_t, X_j \rangle \implies P(i \mid V_t) > P(j \mid V_t)$. **P3** ensures the strategy leverages information from embedding vectors to assess action relevance.

vMF-exp Simultaneously satisfying **P1**, **P2**, and **P3** is challenging. As detailed in Appendix A, popular RL exploration methods fail to achieve this. The proposed solution, vMF-exp, consists in sampling a vector on the $\mathcal{S}^{d-1}$ hypersphere using a von Mises-Fisher distribution (Fisher, 1953) centered on V_t, then exploring its neighborhood. Due to space limitations, we present vMF-exp in Appendix A, explaining why we claim it scales to millions of actions (**P1**) with unrestricted radius (**P2**). Under some assumptions, vMF-exp also verifies **P3** and asymptotically matches the sampling probabilities of Boltzmann exploration (B-exp), a prevalent RL exploration method, hindered by scalability issues due to its reliance on softmax computations for all n actions (violating **P1**). This positions vMF-exp as a scalable alternative to B-exp. However, vMF-exp was only tested via simulations on synthetic data, and the brief report of its deployment on a private service.

3 von Mises-Fisher Sampling of GloVe Word Embedding Vectors

Experimental Setting We present the first vMF-exp experiments on real-world, publicly available data, comparing the behaviors of B-exp and vMF-exp on the GloVe-25 dataset of 1 million GloVe word embedding vectors with dimension $d = 25$ (Pennington et al., 2014). Each vector, learned using word2vec (Mikolov et al., 2013) from 2 billion tweets, represents a word token. We subtract the set's average from each vector and divide them by their norms. We obtain a vector set, denoted $\mathcal{G}$, with all vectors lying on the unit hypersphere, making GloVe-25 a relevant large-scale dataset for our study.

Our experiments follow the protocol outlined in Section 4.3 of Bendada et al. (2024), where we compared the empirical probabilities $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ of sampling an action a represented by a vector A given a state vector V_t, for varying action numbers n and inner products $\langle V_t, A\rangle$. In this paper, vectors are sampled from $\mathcal{G}$, with V_t and A also drawn from $\mathcal{G}$ such that $\langle V_t, A\rangle$ matches the pre-selected values. Our goal is to empirically compare $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$, while verifying the claims that **P1**, **P2**, and **P3** simultaneously hold for vMF-exp. Finally, Sections 4.2 and 4.3 of Bendada et al. (2024) provided analytical approximations of $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ in the presence of independent and identically distributed (i.i.d.) uniform embedding vectors. We will assess the usefulness of these approximations on GloVe vectors, which do not strictly satisfy these assumptions.

Results and Discussion We now discuss our results, with all figures, tables, and a GitHub link in Appendix B for brevity. We first focus on **P1**. While B-exp requires computing $\langle V_t, X_i\rangle$ and softmax values for all n vectors $X_i \in \mathcal{X}_n$, vMF-exp only involves sampling a d-dimensional vector (in constant time with respect to n) and finding its approximate nearest neighbor (ANN) in $\mathcal{X}_n$. Table 1 compares the performance of four popular ANN algorithms on GloVe-25. Following standard ANN literature (Simhadri et al., 2024), our performance metric is the maximum throughput, measured in queries per second (QPS), for which the average recall of the exact top-10 neighbors exceeds 90%. We also report the throughput of exhaustive search in $\mathcal{O}(n)$ time as an indicator of B-exp's inefficiency. Table 1 shows that exhaustive search yields throughput 2 to 3 orders of magnitude lower than ANN methods. This confirms the significantly better scalability (**P1**) of vMF-exp compared to B-exp.

Figure 1 compares $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ for increasing values of n and $\langle V_t, A\rangle$. Figure 1(a) highlights the two properties that make B-exp popular in RL: the ability to sample actions with unrestricted radius (**P2**) and the ordering of sampling probabilities based on action similarity to V_t (**P3**). Importantly, Figure 1(b) confirms that vMF-exp also satisfies both properties. In our tests, A always has a positive sampling probability, which strictly increases with $\langle V_t, A\rangle$. Thus, vMF-exp also satisfies **P2** and **P3** on GloVe-25. Finally, Figure 2 shows that, although GloVe vectors are not i.i.d. and uniform, the analytical approximations of Bendada et al. (2024) for $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ often remain accurate, particularly for B-exp. Also, vMF-exp closely matches B-exp's probabilities for low absolute values of $\langle V_t, A\rangle$. However, as $|\langle V, A\rangle|$ increases, the gap between $P_{\text{vMF-exp}}(a)$ and $P_{\text{B-exp}}(a)$ grows more rapidly than predicted by approximations, highlighting the limitations of the i.i.d. and uniform assumptions and opening the way for future research on more general expressions.

4 Conclusion

Our study confirmed the key theoretical and scalability properties of vMF-exp on a large-scale and publicly available real-world dataset. Our results highlight its potential as a practical solution for exploring large action sets when hyperspherical embedding vectors represent these actions.

REFERENCES

Susan Amin, Maziar Gomrokchi, Harsh Satija, Herke van Hoof, and Doina Precup. 2021. A Survey of Exploration Methods in Reinforcement Learning. *arXiv preprint arXiv:2109.00157* (2021).

Martin Aumüller, Erik Bernhardsson, and Alexander Faithfull. 2017. ANN-benchmarks: A Benchmarking Tool for Approximate Nearest Neighbor Algorithms. In *Proceedings of the 10th International Conference on Similarity Search and Applications*. Springer, 34–49.

Árpád Baricz. 2010. *Generalized Bessel Functions of the First Kind*. Springer.

Walid Bendada, Théo Bontempelli, Mathieu Morlon, Benjamin Chapus, Thibault Cador, Thomas Bouabça, and Guillaume Salha-Galvan. 2023. Track Mix Generation on Music Streaming Services using Transformers. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 112–115.

Walid Bendada, Guillaume Salha-Galvan, Romain Hennequin, Théo Bontempelli, Thomas Bouabça, and Tristan Cazenave. 2024. vMF-exp: von Mises-Fisher Exploration of Large Action Sets with Hyperspherical Embeddings. In *ICML 2024 Workshop: Aligning Reinforcement Learning Experimentalists and Theorists*.

Leonid Boytsov and Bilegsaikhan Naidan. 2013. Engineering Efficient and Effective Non-metric Space Library. In *Proceedings of the 6th International Conference on Similarity Search and Applications*. Springer, 280–293.

Nicolò Cesa-Bianchi, Claudio Gentile, Gábor Lugosi, and Gergely Neu. 2017. Boltzmann Exploration Done Right. *Advances in Neural Information Processing Systems* 30 (2017).

Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H Chi. 2019. Top-K Off-Policy Correction for a REINFORCE Recommender System. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. 456–464.

Minmin Chen, Can Xu, Vince Gatto, Devanshu Jain, Aviral Kumar, and Ed Chi. 2022. Off-Policy Actor-Critic for Recommender Systems. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 338–349.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. 2022. *Introduction to Algorithms*. MIT Press.

Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Mazaré, et al. 2024. The Faiss library. *arXiv preprint arXiv:2401.08281* (2024).

Ronald Aylmer Fisher. 1953. Dispersion on a Sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 217, 1130 (1953), 295–305.

Ruiqi Guo, Philip Sun, Erik Lindgren, Quan Geng, David Simcha, Felix Chern, and Sanjiv Kumar. 2020. Accelerating Large-Scale Inference with Anisotropic Vector Quantization. In *Proceedings of the 37th International Conference on Machine Learning*. 3887–3896.

Masajiro Iwasaki and Daisuke Miyazaki. 2018. Optimization of Indexing Based on k-Nearest Neighbor Graph for Proximity Search. *arXiv preprint arXiv:1810.07355* (2018).

Seungwoo Kang and Hee-Seok Oh. 2024. Novel Sampling Method for the von Mises–Fisher Distribution. *Statistics and Computing* 34, 3 (2024), 106.

Yu A Malkov and Dmitry A Yashunin. 2018. Efficient and Robust Approximate Nearest Neighbor Search using Hierarchical Navigable Small World Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 4 (2018), 824–836.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems* 26 (2013).

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. 1532–1543.

Carlos Pinzón and Kangsoo Jung. 2023. Fast Python Sampler for the von Mises Fisher Distribution. *HAL Id: hal-04004568* (2023).

Harsha Vardhan Simhadri, Martin Aumüller, Amir Ingber, Matthijs Douze, George Williams, Magdalen Dobson Manohar, Dmitry Baranchuk, Edo Liberty, Frank Liu, Ben Landrum, et al. 2024. Results of the Big ANN: NeurIPS'23 Competition. *arXiv preprint arXiv:2409.17424* (2024).

Richard S Sutton and Andrew G Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.

Yiwen Zhu, Jinyi Liu, Wenya Wei, Qianyi Fu, Yujing Hu, Zhou Fang, Bo An, Jianye Hao, Tangjie Lv, and Changjie Fan. 2024. vMFER: von Mises-Fisher Experience Resampling Based on Uncertainty of Gradient Directions for Policy Improvement of Actor-Critic Algorithms. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. 2621–2623.

APPENDIX

We begin this Appendix by providing technical details on action sampling methods in Appendix A, followed by a table, two figures, and additional insights from our experiments in Appendix B.

A TECHNICAL DETAILS ON ACTION SAMPLING METHODS

General Introduction Satisfying the three properties **P1**, **P2**, and **P3** from Section 2 is essential for effective exploration in RL problems with large action sets and embedding representations, but remains challenging. As an introductory example, the random policy $P_{\text{random}}(a \mid V_t) = \frac{1}{n}$, uniformly sampling actions, might be the simplest exploration method, with ε-greedy (Sutton and Barto, 2018) as a common variant. Both methods are scalable (**P1**), as actions can be uniformly sampled in $\mathcal{O}(1)$ time (Cormen et al., 2022). They satisfy **P2**, since all actions can be sampled. However, they overlook the relative positions of embedding vectors, failing to achieve order preservation (**P3**). This limitation is problematic. In the music recommendation example from Section 2, the agent would likely sample songs with minimal relevance to user preferences. The vMF-exp article provides a detailed discussion on how popular RL exploration methods fail to simultaneously meet **P1**, **P2**, and **P3**, highlighting the need for further research. Next, we focus on one of the most prevalent of these methods.

Boltzmann Exploration (B-exp) Boltzmann Exploration (B-exp) is a prevalent method in RL for sampling actions based on their embedding similarity to V_t (Cesa-Bianchi et al., 2017; Sutton and Barto, 2018; Chen et al., 2019; Amin et al., 2021). The embedding similarity is typically quantified using the inner product. B-exp leverages the Boltzmann distribution to guide the sampling process:

$$\forall i \in \mathcal{I}_n, P_{\text{B-exp}}(i \mid V_t) = \frac{\mathrm{e}^{\kappa \langle V_t, X_i \rangle}}{\sum_{j=1}^{n} \mathrm{e}^{\kappa \langle V_t, X_j \rangle}}, \tag{1}$$

for some hyperparameter $\kappa > 0$. B-exp samples actions based on a strictly increasing function of their inner product similarity with V_t, ensuring order preservation (**P3**). Adjusting κ excludes irrelevant actions effectively while assigning non-zero probabilities (**P2**). However, B-exp is not scalable to large action sets, as it requires computing inner products, softmax operations, and sampling probabilities for all n actions (violating **P1**), which is computationally prohibitive for large n. Similar scalability issues would arise with any other distribution requiring computing similarities for all n actions.

Prior work, aiming to address these scalability challenges, has sometimes employed a truncated version of B-exp (Chen et al., 2019; 2022), referred to as TB-exp in Bendada et al. (2024). In TB-exp, the sampling step from Equation (1) is restricted to a small candidate set of $m \ll n$ actions similar to V_t, selected via an ANN search engine. This approach, used in production systems like YouTube (Chen et al., 2019), significantly improves scalability, as its complexity depends on m rather than n, satisfying **P1** when m is small. However, TB-exp does not satisfy **P2** anymore, as it limits the radius to m actions based on technical constraints, risking the exclusion of relevant actions beyond this radius. This illustrates the challenge of designing a method that would satisfy **P1**, **P2**, and **P3** simultaneously – typically, a method with properties similar to B-exp but more scalable.

von Mises–Fisher Exploration (vMF-exp) B-exp's scalability is limited by the need to compute sampling probabilities for all n actions. The proposed solution, von Mises-Fisher Exploration (vMF-exp) (Bendada et al., 2024), addresses this limitation by employing a two-step process:

- Given a state vector $V_t \in \mathcal{S}^{d-1}$, the vMF-exp method first consists in sampling a hyperspherical vector $\tilde{V}_t \in \mathcal{S}^{d-1}$ from a von Mises-Fisher distribution (Fisher, 1953) centered on V_t.
- Next, vMF-exp selects the nearest neighbor of $\tilde{V}_t$ in $\mathcal{X}_n$ within the embedding space, according to the inner product similarity, as the action sampled for exploration.

The von-Mises Fisher distribution (Fisher, 1953) is a continuous vector distribution defined on the hypersphere $\mathcal{S}^{d-1}$. Recently, it has been applied in RL to evaluate the uncertainty of gradient directions (Zhu et al., 2024). The probability density function for any vector $\tilde{V}_t \in \mathcal{S}^{d-1}$ is as follows:

$$f_{\text{vMF}}(\tilde{V}_t \mid V_t) = C_d(\kappa) e^{\kappa \langle V_t, \tilde{V}_t \rangle}, \text{with } C_d(\kappa) = \frac{1}{\int_{\tilde{V}_t \in \mathcal{S}^{d-1}} e^{\kappa \langle V_t, \tilde{V}_t \rangle} \, \mathrm{d}\tilde{V}_t} = \frac{\kappa^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} I_{\frac{d}{2}-1}(\kappa)}, \quad (2)$$

for some hyperparameter $\kappa > 0$. $I_{d/2-1}$ is the modified Bessel function of the first kind at order $d/2-1$ (Baricz, 2010). $f_{\text{vMF}}(\tilde{V}_t \mid V_t)$ is proportional to $e^{\kappa \langle V_t, \tilde{V}_t \rangle}$, echoing B-exp. vMF-exp bypasses the need to manage a discrete distribution with n parameters by directly sampling a d-dimensional vector, which can be done in constant time relative to n. The second step, identifying the nearest neighbor of this vector, can be performed in sublinear time assuming the availability of an effective ANN search engine, making vMF-exp a scalable method (**P1**). Efficient algorithms for sampling from vMF distributions have been well-studied (Pinzón and Jung, 2023; Kang and Oh, 2024).

Moreover, Section 3.2 of Bendada et al. (2024) demonstrates that the probability of sampling any action is always strictly positive, thereby satisfying the unrestricted radius property (**P2**). Section 4 of Bendada et al. (2024) analyzes a theoretical setting where embedding vectors are i.i.d. and uniformly distributed on the unit hypersphere. Under these assumptions, we show that vMF-exp preserves order (**P3**) and asymptotically matches B-exp's sampling probabilities, positioning vMF-exp as a scalable alternative to B-exp.

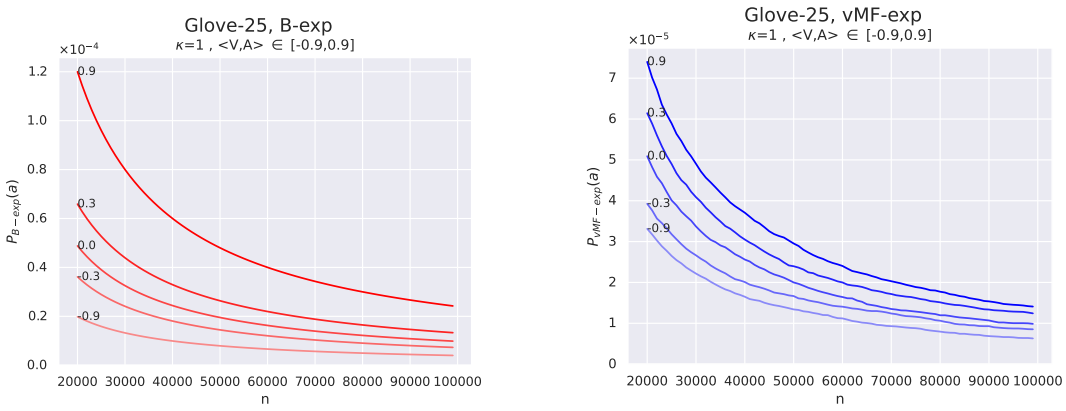
B Experimental Results

We present the table and two figures referenced in Section 3. Regarding our GloVe-25 experiments:

- The GloVe-25 word embedding dataset used in our experiments is publicly available for download at: `https://nlp.stanford.edu/projects/glove/`.
- In our experiments, we utilize the Python vMF sampler from Pinzón and Jung (2023) to efficiently explore large action sets.
- To ensure the reproducibility of our results, we have released our source code on GitHub. The repository has been anonymized for the review phase: `https://github.com/Anonymous-vmf/fpi-iclr_submission`.

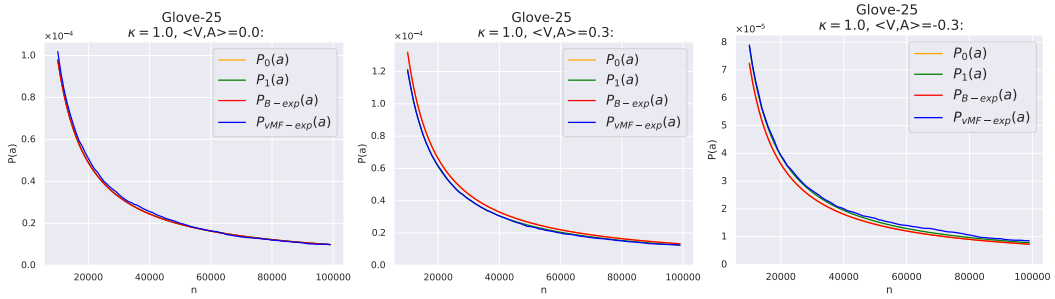
Table 1: Performance of popular ANN algorithms on GloVe-25, extracted from the benchmark of Aumüller et al. (2017). Following Simhadri et al. (2024), our performance metric is the maximum throughput, measured in Queries Per Second (QPS), for which the average recall of the exact top-10 neighbors exceeds 90%. The evaluated algorithms include two implementations of HNSW (Malkov and Yashunin, 2018), one from the Faiss library (Douze et al., 2024) and the other from NMSLIB (Boytsov and Naidan, 2013), as well as ScaNN (Guo et al., 2020) and NGT-QG (Iwasaki and Miyazaki, 2018). Exhaustive search is 2 to 3 orders of magnitude slower than ANN methods.

Algorithm	Exhaustive Search	HNSW (Faiss)	HNSW (NMSLIB)	ScaNN	NGT-QG
Maximum Throughput in Queries Per Second (QPS)	34	6197	14080	**23436**	**22733**

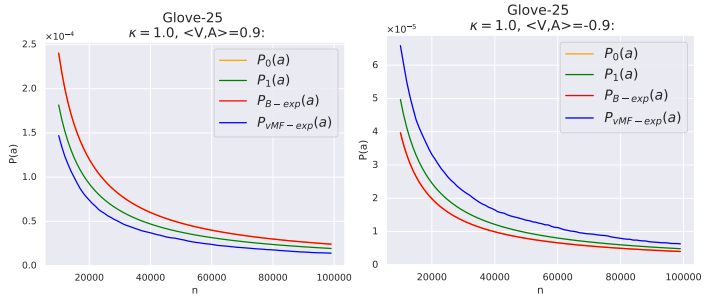


(a) (b)

Figure 1: We report the empirical probabilities $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ of sampling an action a represented by a GloVe embedding vector A, using B-exp and vMF-exp, respectively, given a state vector V, with $20000 \leq n \leq 100000$ and $\langle V, A \rangle \in \{0.9, 0.3, 0.0, -0.3, -0.9\}$, and with $d = 25$ and $\kappa = 1$. Sampling is repeated 30 million times and averaged to obtain precise estimates. For both methods, the probability of sampling a for exploration is strictly positive (**P2**) and is a strictly increasing function of the inner product similarity $\langle V, A \rangle$ (**P3**). Results remain consistent when κ is modified.



(a) (b) (c)



(d) (e)

Figure 2: We compare $P_{\text{B-exp}}(a)$ and $P_{\text{vMF-exp}}(a)$ on GloVe-25 to the analytical approximations $P_0(a)$ and $P_1(a)$ stated in Propositions 4.1 to 4.4 of Bendada et al. (2024) under the assumption of i.i.d. and uniformly distributed vectors. Our tests confirm the usefulness of these approximations on GloVe-25. The yellow curve ($P_0(a)$) is indistinguishable from the red curve ($P_{\text{B-exp}}(a)$), indicating that Proposition 4.2 holds across all configurations. Furthermore, $P_{\text{vMF-exp}}(a)$ (blue) remains close to $P_{\text{B-exp}}(a)$ (red) for low absolute values of $\langle V, A \rangle$ (Figures 2(a), 2(b), and 2(c)), as anticipated by Propositions 4.1 and 4.3. In Figures 2(b) and 2(c), the small difference between $P_{\text{vMF-exp}}(a)$ and $P_{\text{B-exp}}(a)$ aligns with the alternative expression $P_1(a)$ (green) derived in Proposition 4.4. However, as $|\langle V, A \rangle|$ increases (Figures 2(d) and 2(e)), the difference between vMF-exp and B-exp grows more rapidly than predicted by Proposition 4.4, highlighting the limitations of the i.i.d. and uniform assumptions.