We thank the reviewers for their comments. We will fix all minor issues and do not discuss them individually here.

Common Questions:

Q1: Use datasets with natural probability distributions?

A1: We appreciate your concern and will consider other datasets as well. However, we would like to remark that we used the same distributions used in Mao et al., 2017, which is the most relevant work. Mao et al. studies LSH of Jensen-Shannon distance (an information-theoretic distance) and normalizes the images in MNIST and CIFAR to probability distributions.

Q2: Krein-LSH (Sec. 4) vs. the Hellinger-approximation-based LSH (Sec. 3).

A2: Our Hellinger-approximation-based LSH is a more general method and applies to the family of f-divergences (including MIL). However, particularly focusing on MIL, Krein-LSH has a better approximation factor. In fact, our Krein-LSH is *lossless*, i.e., we do not lose anything on the approximation factor except for computing the integral numerically. We have a more detailed discussion in lines 68–75.

Response to Reviewer #1:

Q3: Data sets seem somewhat artificial in the context of probability distributions. **A3:** Please see **A1**.

Q4:- There are multiple papers on this topic not listed in the paper, see below.

A4: Thank you for providing the references. We will discuss and compare them in the related work section.

Q5: It would be helpful if the author included more "combinatorial" measures such as the number of examined points.

A5: Thank you for this great suggestion. We will definitely report the number of examined points in our final version.

Response to Reviewer #2:

Q6: Theorem 2: c should be fixed, possibly by replacing c with an upper bound. Is it a reasonable constant?

A6: Thanks for noting this. We agree with you. It is possible to drive an upper bound of $c \leq \sqrt{2}$. This is serves its purpose since one can always increase the code length and the number of hash functions (buckets) to separate close and faraway points apart. We will address this in the paper.

Q7: Sec. 4, Krein-LSH: it is not clear what the motivation is for this section. Are the guarantees for MIL-LSH based on the reduction from squared Hellinger distance not good enough? are the two methods comparable?

A7: Please see **A2** for comparison between MIL-LSH and Krein-LSH. We will make the motivation clearer.

Q8: Sec. 5: which MIL LSH is used? What were the Obtained guarantees? How do these affect the correctness-efficiency tradeoff? Do they seem to be tight?

A8: We used the MIL-LSH via Hellinger approximation here. To show the guarantees, we evaluate the performance by solving an end-to-end nearest neighbor search problem (this is a standard setup to evaluate LSH methods). All six subfigures in Fig. 2 illustrate the correctness-efficiency tradeoff. All figures seem as expected.

Q9: Sec. 5: the chosen datasets of distributions are artificially generated from image datasets. **A9:** Please see **A1**.

Response to Reviewer #3:

Q10: L^2 LSH would be a reasonable baseline.

A10: Thank you for the constructive comment. We have results of L^2-LSH. On MNIST, when its speedup factor is 2, the precision is only 0.7 (ours is 0.85). When its speedup factor is 3, the precision drops below 0.63 (ours is 0.73). Our method clearly outperforms it. Due to the space limit here, we will present them in the final version.

Q11: Section 4 is independent of the rest. The described algorithm is not readily connected with that task. Experiments are not presented either. What is the strength of this hashing scheme?

A11: Sec. 4 is indeed an improvement for the method described in Sec. 3. For the strength of Krein-LSH, please refer to **A2**. The described algorithm reduces the problem of finding small MIL to the problem of maximum inner product search and uses, say, SIMPLE-LSH (Neyshabur and Srebro, 2014) at the final stage (Line 5 of Algorithm 1). We will clarify these points in the final version. We have experiment results regarding Krein-LSH and will include them in the final version by moving some other content to the appendix.

Q12: Proposition 1 is tough to follow at first glance. If possible, some visualization helps the understanding.

A12: Thank you for this valuable comment. We will provide a proof sketch for this key proposition.