Divergence Frontiers for Generative Models: Sample Complexity, Quantization Effects, and Frontier Integrals

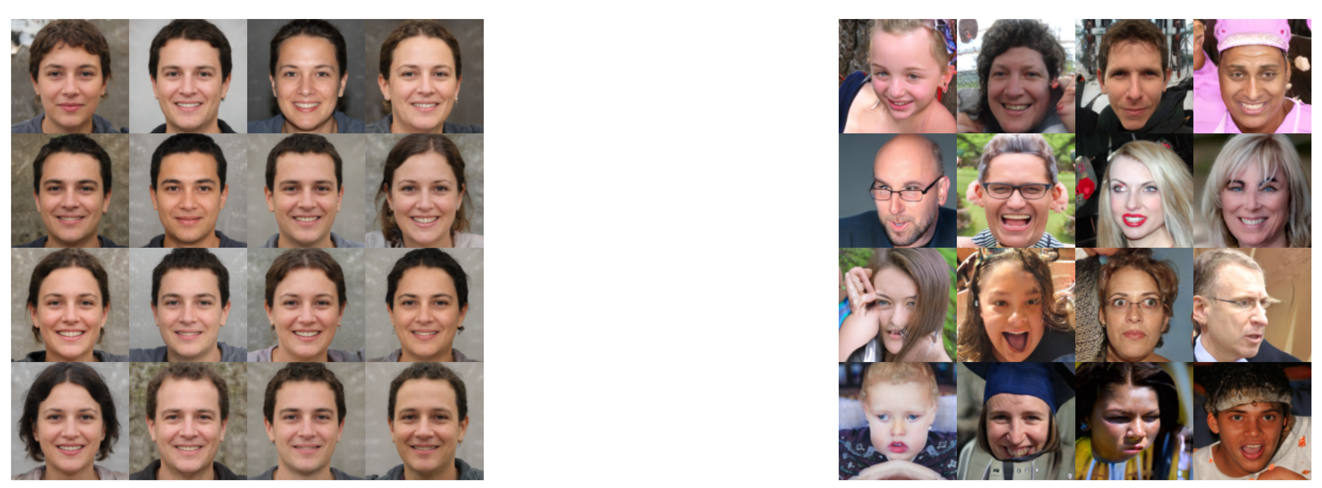
Lang Liu[1], Krishna Pillutla[2], Sean Welleck[2,3], Sewoong Oh[2], Yejin Choi[2,3], Zaid Harchaoui[1]

[1] Department of Statistics, University of Washington
[2] Paul G. Allen School of Computer Science & Engineering, University of Washington
[3] Allen Institute for Artificial Intelligence

Overview

- The spectacular success of deep generative models calls for **quantitative tools** to measure their performance.
- **Divergence frontiers** have recently been proposed as an evaluation framework for generative models. In practice, they are estimated from data via **quantization** and **empirical estimation**.
- We establish **non-asymptotic bounds** for the estimation procedure, characterizing the sample complexity of divergence frontiers.

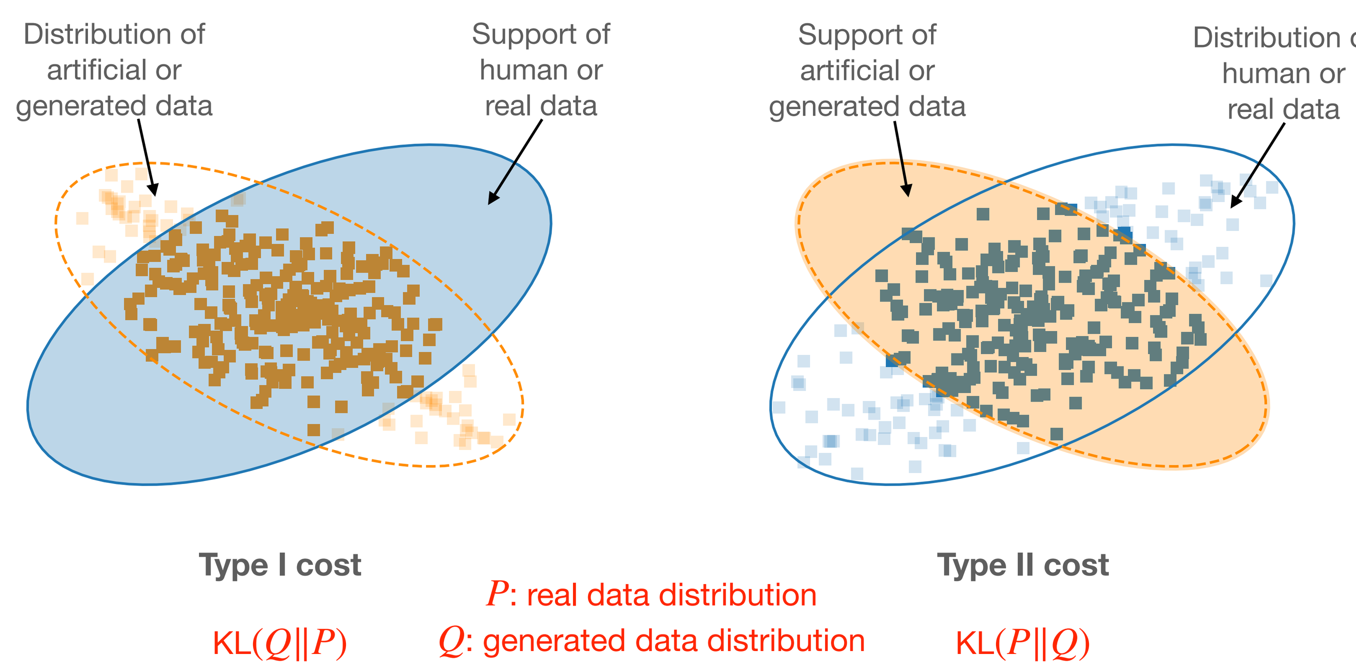
Image and Text Generation



...the techniques we used when cleaning out my mom's fabric stash last week... Next, you need to get **a small, sharp knife**. I like to use **a small, sharp knife**. I like to use **a small, sharp knife**.

...the techniques we used when cleaning out my mom's fabric stash last week... I had a great deal of **décor management** and was able to **stash the excess items away for safekeeping**.

Type I and Type II Costs



Divergence Frontiers

Divergence frontiers (Djolonga et al. '20). Define the mixture $R_\lambda = \lambda P + (1-\lambda)Q$. Let

$$\mathcal{F}(P,Q) := \{(\mathrm{KL}(Q\|R_\lambda), \mathrm{KL}(P\|R_\lambda)) : \lambda \in (0,1)\}.$$

Statistical summary.

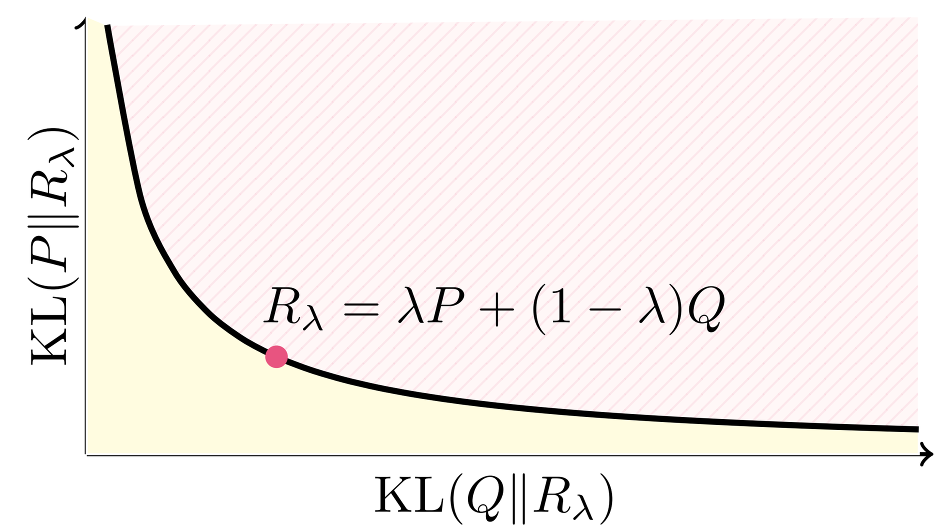
- The linearized cost (λ-skew Jensen-Shannon divergence)

$$\mathcal{L}_\lambda(P,Q) := \lambda \mathrm{KL}(P\|R_\lambda) + (1-\lambda)\mathrm{KL}(Q\|R_\lambda).$$

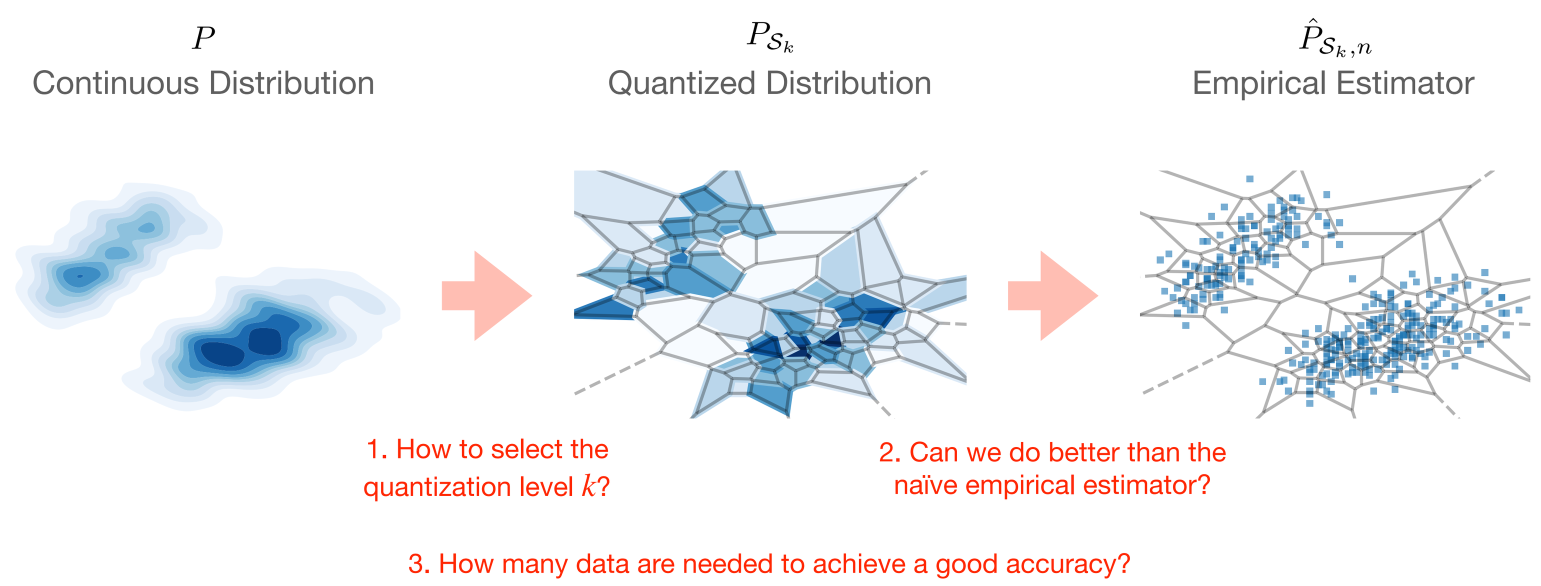
- **Frontier integral**—statistical summary

$$\mathrm{FI}(P,Q) := 2\int_0^1 \mathcal{L}_\lambda(P,Q)\,\mathrm{d}\lambda.$$

  - Symmetric divergence, i.e., $\mathrm{FI}(P,Q) = 0$ iff $P = Q$.
  - Taking values in $[0,1]$.



Esimation Procedure



Main Results

Statistical error. Assume P and Q are discrete with support size k. With probability at least $1-\delta$,

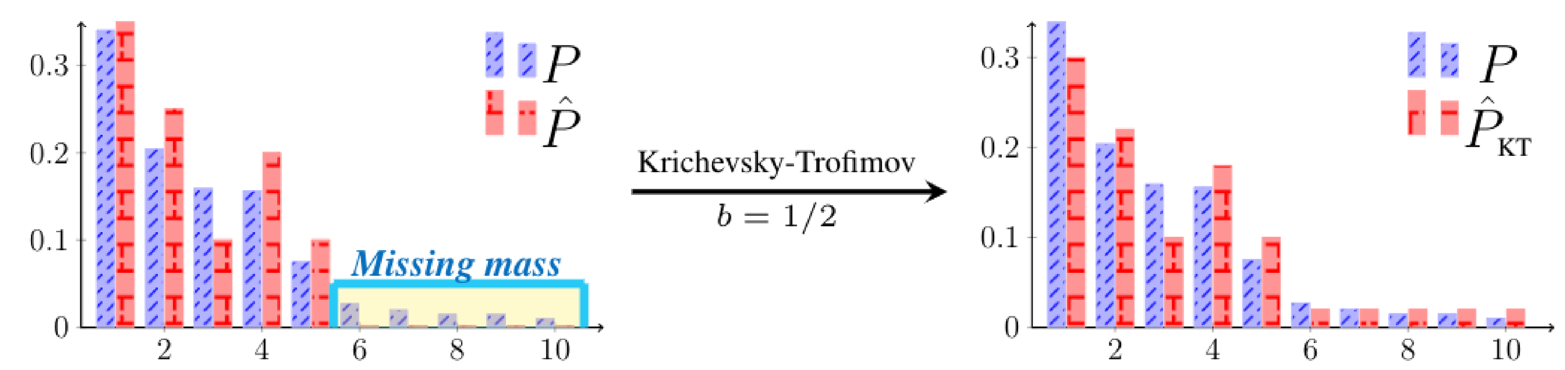
$$\left|\mathrm{FI}(\hat{P}_n, \hat{Q}_n) - \mathrm{FI}(P,Q)\right| \lesssim \sqrt{\frac{\log 1/\delta}{n}} + \sqrt{\frac{k}{n}} + \frac{k}{n}.$$

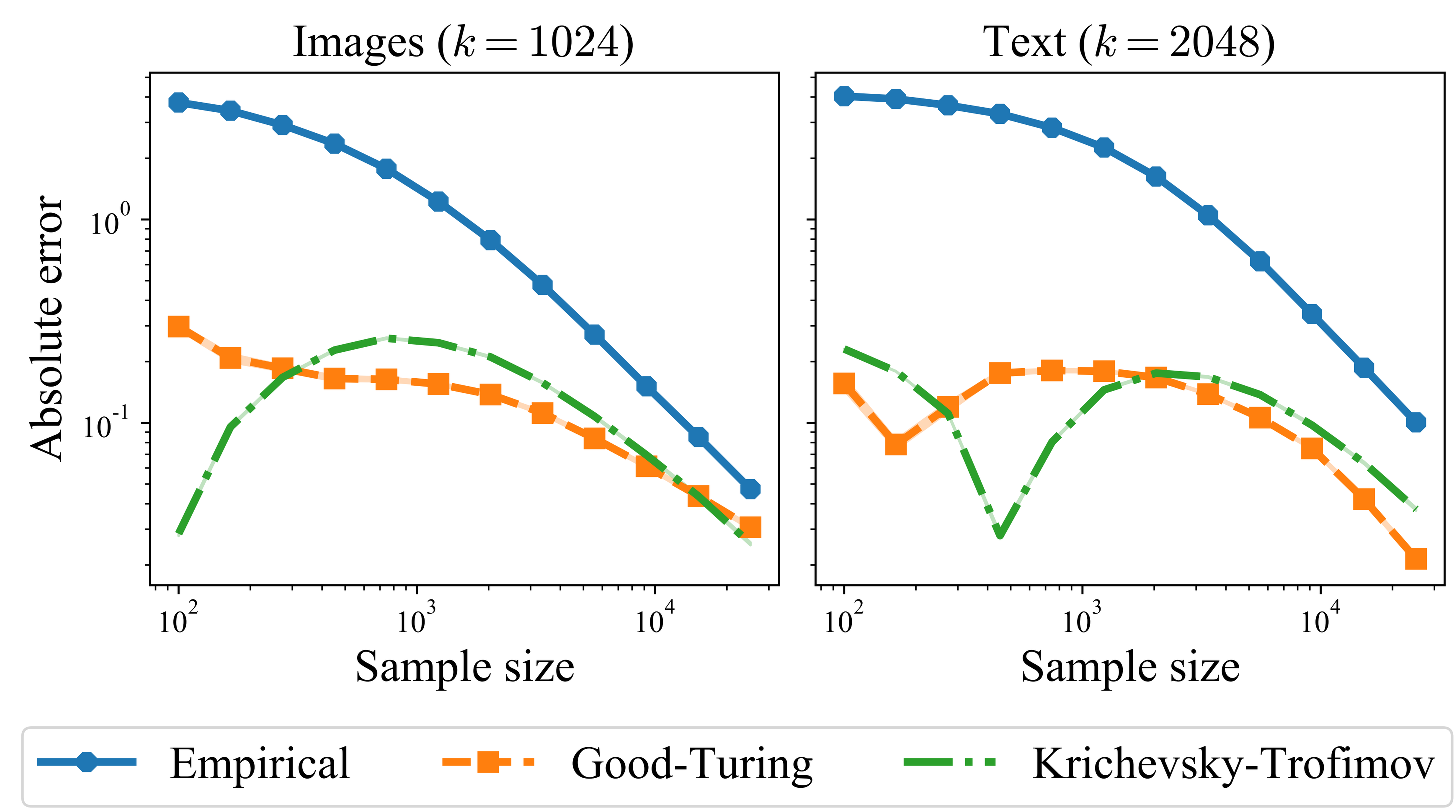
Total error. For **arbitrary** P and Q and any k, there exists a partition $\mathcal{S}_k$ of size k such that

$$\mathbb{E}\left|\mathrm{FI}(\hat{P}_{\mathcal{S}_k,n}, \hat{Q}_{\mathcal{S}_k,n}) - \mathrm{FI}(P,Q)\right| \lesssim \sqrt{\frac{k}{n}} + \frac{k}{n} + \frac{1}{k}.$$

Smoothed estimators. Let $\hat{P}_{\mathcal{S}_k,n,b}$ be the add-b estimator of $P_{\mathcal{S}_k}$.

$$\mathbb{E}\left|\mathrm{FI}(\hat{P}_{\mathcal{S}_k,n,b}, \hat{Q}_{\mathcal{S}_k,n,b}) - \mathrm{FI}(P,Q)\right| \lesssim \frac{\sqrt{nk} + bk}{n + bk} + \frac{1}{k}.$$





Code available at *https://github.com/langliu95/divergence-frontier-bounds*.