Annealed Importance Sampling with q-Paths

Rob Brekelmans*[1], Vaden Masrani*[2], Thang Bui[3], Frank Wood[2], Aram Galstyan[1], Greg Ver Steeg[1], Frank Nielsen[4]
{brekelma,galstyan,gregv}@isi.edu, {vadmas,fwood}@cs.ubc.ca, thang.bui@uber.com, frank.nielsen@acm.org

1 USC Viterbi School of Engineering Information Sciences Institute 2 UBC PLAI 3 UBER AI 4 Sony CSL

TL;DR

- We introduce q-Paths, a generalization of the geometric mixture path commonly used in MCMC and VI methods
- q-Paths can be formed between arbitrary densities, and have a simple closed form expression for the potential function
- We observe that the choice of q-path can affect accuracy and qualitative mixing behavior in AIS

Summary

$$\tilde{\pi}^{(q)}_{\beta_t}(z) = \left[(1-\beta_t)\, \tilde{\pi}_0(z)^{1-q} + \beta_t\, \tilde{\pi}_1(z)^{1-q} \right]^{\frac{1}{1-q}}$$

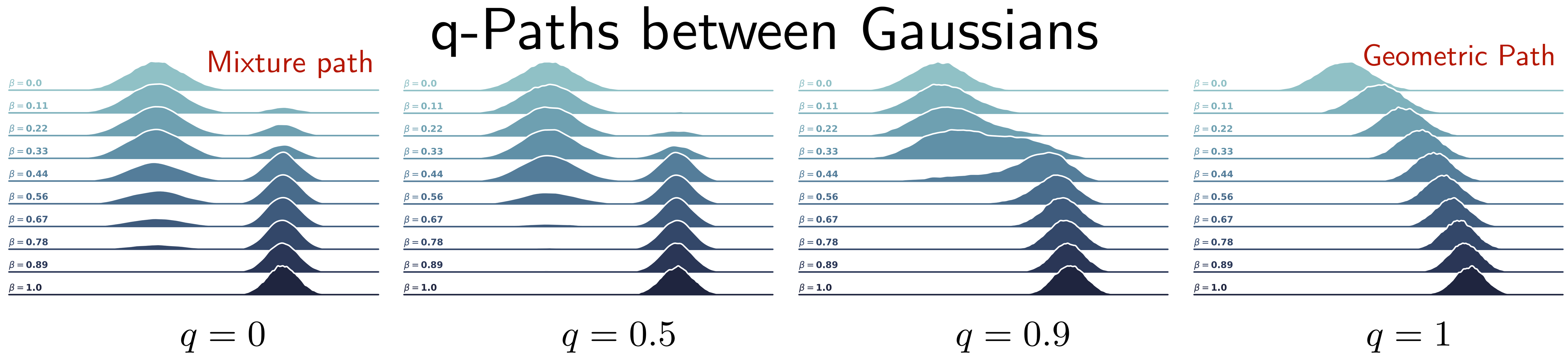
Geometric Path		q-Path
Log mixture	↔	q-Log mixture
Exponential Family	↔	q-Exponential Family
KL Divergence	↔	α-Divergence

q-Paths between Gaussians

$\pi_0 = \mathcal{N}(-4, 3)$

↓

$\pi_1 = \mathcal{N}(4, 1)$

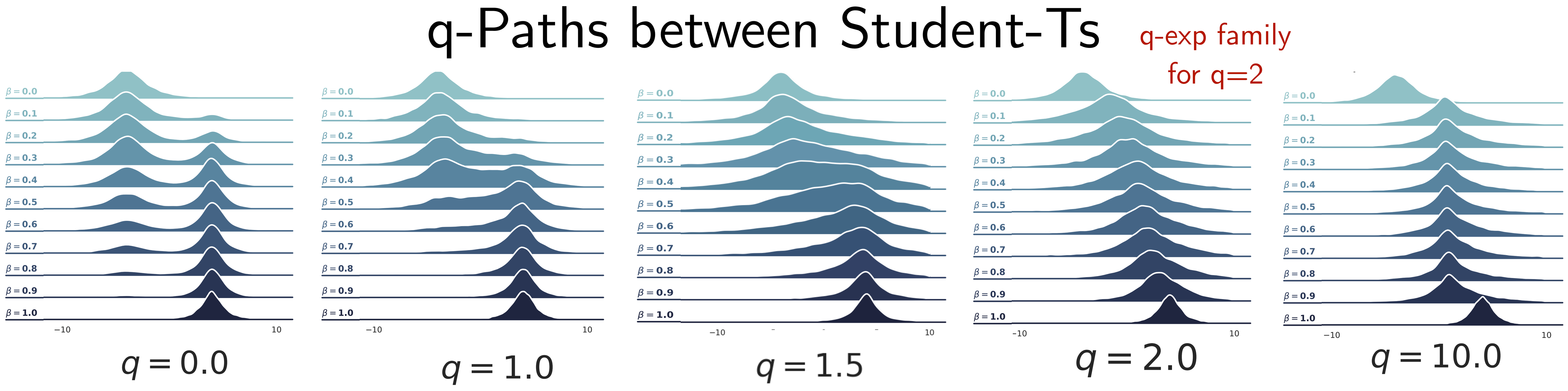


q-Paths between Student-Ts

$\mathrm{St}\,(\nu = 1, -4, 3)$

↓

$\mathrm{St}\,(\nu = 1, 4, 1)$



Derivation

- Interpret geometric mixture path as a log mixture

$$\tilde{\pi}_\beta(z) = \tilde{\pi}_0(z)^{1-\beta}\, \tilde{\pi}_1(z)^{\beta} = \exp\Big\{ (1-\beta)\cdot \log \tilde{\pi}_0(z) + \beta \log \tilde{\pi}_1(z) \Big\}$$

- We can recognize this as Kolmogorov's[1] generalization of the mean for monotonic function h

$$\mu_h = h^{-1}\Big(\sum_i w_i \cdot h(\tilde{\pi}_i) \Big)$$

inputs $= \tilde{\pi}_0, \tilde{\pi}_1$
weights $= \{1-\beta, \beta\}$
$h^{-1} = \exp \quad h = \log$

- Seek most general h that maintain *homogeneity property*[2]

$$\mu_h(\{w_i, c\cdot\tilde{\pi}_i\}) = c \cdot \mu_h(\{w_i, \tilde{\pi}_i\})$$

- This yields *q-deformed log*[3] with *q-exponential* as its inverse

$$h_q(\tilde{\pi}) = \ln_q(\tilde{\pi}) = \begin{cases} \frac{1}{1-q}\left(\tilde{\pi}(z)^{1-q} - 1\right) & q \neq 1 \\ \log \tilde{\pi}(z) & q = 1 \end{cases}$$

$$h_q^{-1}(\tilde{\pi}) = \exp_q(\tilde{\pi}) = \left[1 + (1-q)\,\tilde{\pi}(z)\right]_+^{\frac{1}{1-q}}$$

- Using q-log and q-exp in the generalized mean results in the q-path!

Variational Representations

- Distributions along the q-path minimize an expected Divergence[2] to each of the endpoints

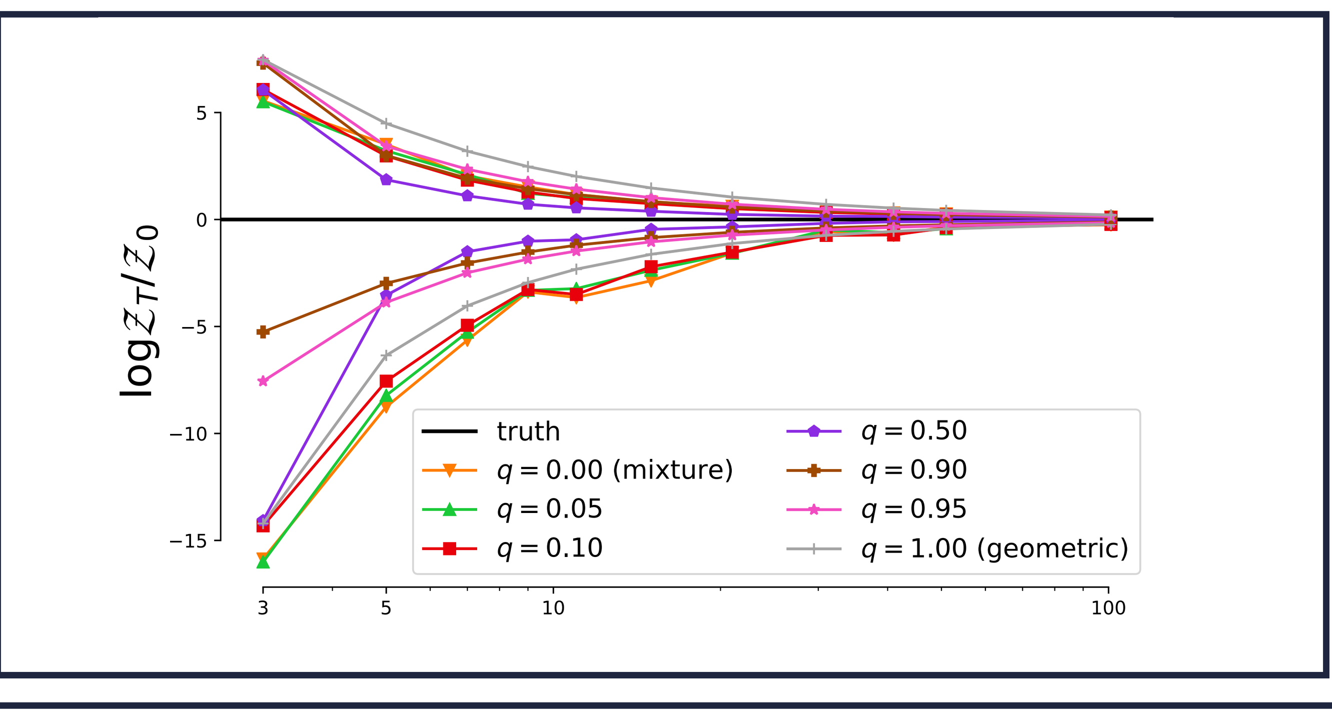
$$\pi^{(1)}_\beta = \arg\min_{r(z)} (1-\beta) D_{KL}[r(z)||\pi_0(z)] + \beta\, D_{KL}[r(z)||\pi_1(z)]$$

(with $1-q = (1-\alpha)/2$)

$$\tilde{\pi}^{(q)}_\beta = \arg\min_{\tilde{r}(z)} (1-\beta) D_\alpha[\tilde{\pi}_0(z) : \tilde{r}(z)] + \beta D_\alpha[\tilde{\pi}_1(z) : \tilde{r}(z)]$$

Results

- Bidirectional Monte Carlo[4] log partition bounds by # of intermediate distributions



1. Kolmogorov, *"On the notion of mean"*, 1930
2. Amari, "Integration of Stochastic Models by Minimizing α-Divergence", 2007.
3. Tsallis, "Introduction to nonextensive statistical mechanics", 2009.
4. Grosse et. al. "Sandwiching the marginal lkld using Bidirectional Monte Carlo", NeurIPS 2015.