Probabilistic Programs with Stochastic Conditioning

arXiv:2010.00282

David Tolpin
Yuan Zhou
Hongseok Yang

Motivation

In a probabilistic program

$$p(x|y) \propto p(x)p(y|x)$$

'usual' conditioning is **deterministic:** $p(x|y = c)$.

Works when observations

- are samples from joint data distribution.

Won't work when observations

- are summarized or obfuscated
- are collected by multiple parties,
- are noisy and obtained online,
- reflect partial knowledge about future.

Definition

Probabilistic program computes

$$p(x, z) = p(x)p(z|x)$$

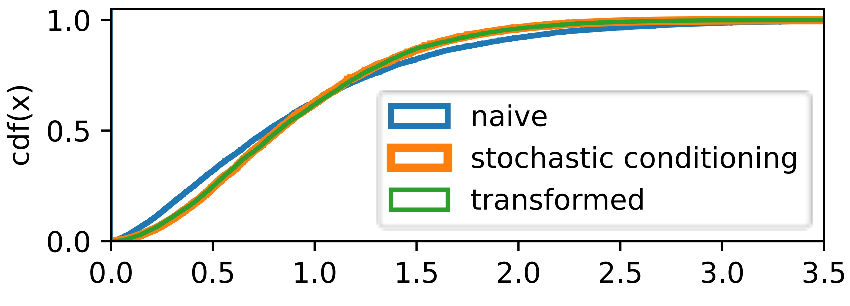
Our objective is to infer $p(x|D_z)$ when $z \sim D_z$.

Conditioning on D_z $\equiv$ conditioning on **all** values:

$$\begin{aligned} p(x|D_z) \propto p(x, D_z) &= p(x) \prod_{z \in \Omega_D} (p(z|x))^{p_D(z)dz} \\ &= \exp\left(\log p(x) + \int_{z \in \Omega_D} p_D(z) \log p(z|x) dz\right) \\ &\propto \exp\left(\log p(x) - \mathrm{KL}[p_D(z) \| p(z|x)]\right), \end{aligned}$$

Intuition

- We know that $z \sim \mathcal{N}(0, 1)$
- We want to infer x such that $y \approx x + z$



Naive	Stochastic	Transformed
$z \sim \mathcal{N}(0, 1)$ $x \sim \mathrm{Gamma}(2, 2)$ $y \mid x, z \sim \mathcal{N}(x + z, 1)$	$z \leftarrow \mathcal{N}(0, 1)$ ——— $x \sim \mathrm{Gamma}(2, 2)$ $y \mid x, z \sim \mathcal{N}(x + z, 1)$	$x \sim \mathrm{Gamma}(2, 2)$ $y \mid x \sim \mathcal{N}(x, 1)$

Population of New York

Data

	Population (N=804)	Sample1 (n=100)	Sample2 (n=100)
mean	17,135	19,667	38,505
sd	139,147	142,218	228,625
0%	19	164	162
5%	336	308	315
25%	800	891	863
50%	1,668	2,081	1,740
75%	5,050	6,049	5,239
95%	30,295	25,130	41,718
100%	2,627,319	1,424,815	1809578

Model

$$z_{1\ldots n} \leftarrow \text{Quantiles}$$

$$\begin{aligned} m &\sim \mathrm{Normal}(\text{mean}, \frac{\text{sd}}{\sqrt{\text{n}}}) \\ s^2 &\sim \mathrm{InvGamma}(\frac{\text{n}}{2}, \frac{\text{n}}{2}\text{sd}^2) \\ \sigma &= \sqrt{\log\left(s^2/m^2 + 1\right)} \\ \mu &= \log m - \frac{\sigma^2}{2} \\ z_{1\ldots\text{n}} | m, s^2 &\sim \mathrm{LogNormal}(\mu, \sigma) \end{aligned}$$