Probabilistic Programs with Stochastic Conditioning

ICML 2021

David Tolpin
Yuan Zhou
Tom Rainforth
Hongseok Yang

PUB+ · Ben-Gurion University of the Negev · University of Oxford · KAIST

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

- independent samples from marginal distributions,
- summary statistics,
- distributions in closed form or as samplers,
- reflect partial knowledge about future.

Definition

Probabilistic program computes

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

Our objective is to infer

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

By definition, density of $y \sim D$ fiven x:

$$p(y \sim D|x) \propto \exp\left(\int_Y (\log p(y|x))\, q(y)dy\right)$$
$$= \prod_Y p(y|x)^{q(y)dy\,*}$$

*) Probability of observing *all* possible draws of y from D, each according to its probability $q(y)dy$.

Intuition

Coin flip

$$x \sim \mathrm{Beta}(\alpha, \beta)$$
$$y \sim \mathrm{Bernoulli}(x)$$

Observing a value

Posterior after n observations $y_n \circ \dots \circ y_2 \circ y_1$:

$$x|y_{1:n} \sim \mathrm{Beta}\left(\alpha + \sum_{i=1}^n y_i, \beta + n - \sum_{i=1}^n y_i\right)$$

Observing a distribution

Assume we just know that $\sum_{i=1}^n y_i = k$:

$$x|y_{1:n} \sim \mathrm{Beta}(\alpha + k, \beta + n - k)$$

Now y_i is an observation of Bernoulli $\left(\theta = \frac{k}{n}\right)$:

$$x|\left(y_{1:n} \sim \mathrm{Bernoulli}(\theta)\right) \sim \mathrm{Beta}(\alpha + n\theta, \beta + n(1 - \theta))$$

Conditional probability

Probabilistic programming requires computing $p(y|x)$:

$$p(y|x) = x^y(1-x)^{1-y}$$

For conditioning on distribution $y \sim \mathrm{Bernoulli}(\theta)$:

$$p(y \sim \mathrm{Bernoulli}(\theta)|x) = x^\theta (1-x)^{1-\theta}$$
$$= \exp\left(\sum_{y \in \{0,1\}} p_{\mathrm{Bern}(\theta)}(y) \cdot \log p_{\mathrm{Bern}(x)}(y)\right)$$

Population of New York

Data

	Population	Sample1	Sample2
	(N=804)	(n=100)	(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

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

$$m \sim \mathrm{Normal}\left(\text{mean}, \frac{\text{sd}}{\sqrt{\text{n}}}\right), s^2 \sim \mathrm{InvGamma}\left(\frac{\text{n}}{2}, \frac{\text{n}}{2}\text{sd}^2\right)$$

$$\sigma = \sqrt{\log\left(s^2/m^2 + 1\right)},\ mu = \log m - \frac{\sigma^2}{2}$$

$$z_{1\dots n}|m,\ s^2 \sim \mathrm{LogNormal}(\mu, \sigma)$$

Posterior