Variational Inference

Ideas and Advancements

Ananyapam De

Indian Institute of Science Education and Research, Kolkata

A simple example

Suppose we generate our data in the following manner-

$$x \sim \mathcal{N}(\mu_0, {\sigma_0}^2)$$

How to infer μ_0 and σ_0 given $X = (x_1, x_2, ..., x_n)$?

A rather practical example

Now, suppose we generate our data as follows-

$$z \sim \mathcal{N}(\mu_0, {\sigma_0}^2)$$
$$x \sim Pois(e^z)$$

How to infer μ_0 and σ_0 given $X = (x_1, x_2, ..., x_n)$?

Setup

Let **x** be the set of observed variables and **z** be the set of latent variables with joint density **p(z,x)**. The inference problem is to compute the conditional density of the **p(z|x)**.

We can write the conditional density as

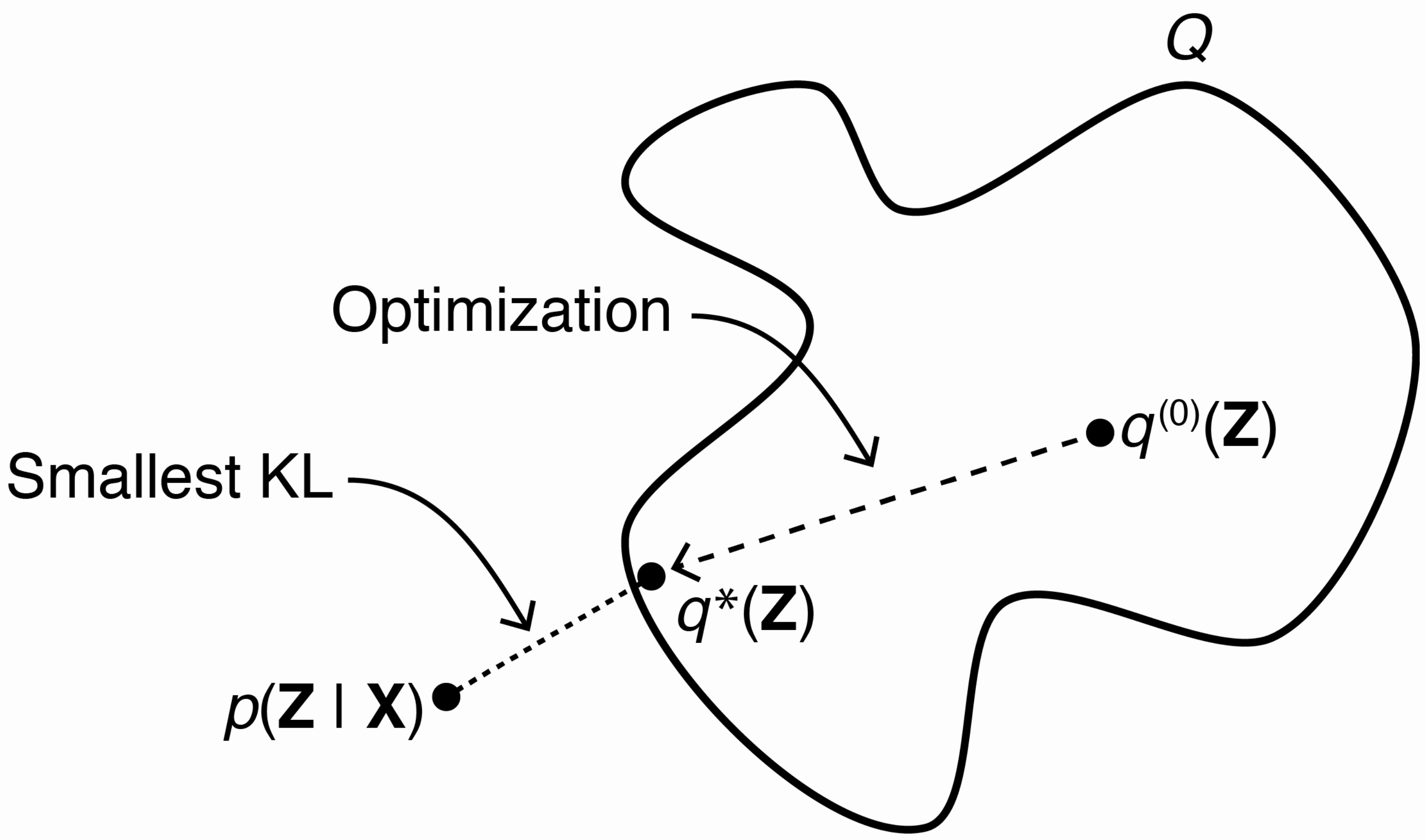
$$p(z|x) = \frac{p(z,x)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} \propto p(x|z)p(z)$$

Approximate Inference

1. The **Slow** Markov Chain Monte Carlo
2. The **OverPowered** Variational Inference!

Variational Inference

Instead of sampling from the exact conditional density $p(z|x)$, we approximate our conditional density using a "close-enough" density which is simple enough to sample from. Our problem now becomes an **optimization** problem!



But how does this help?

$$\mathrm{KL}(q(z|\phi)||p(z|x)) = \int q(z|\phi) \ln \left(\frac{q(z|\phi)}{p(z|x)} \right) dz$$

$$\underbrace{E_{q(z|\phi)} \left[\ln \left(\frac{p(z|x)}{q(z|\phi)} \right) \right]}_{\mathrm{ELBO}(\phi)} = \ln(p(x)) - \mathrm{KL}(q(z|\phi)||p(z|x))$$

Notice that minimizing the KL is same as maximizing the ELBO. Hence our goal now is to maximize this ELBO!

Maximizing the ELBO

$$\text{ELBO}(\phi) = E_{q(z|\phi)}\left[\ln\left(\frac{p(z|x)}{q(z|\phi)}\right)\right]$$

Impressive! But ...

- Not always available in closed form
- Not scalable
- Not fast **enough**

Advances leading to Stochastic Variational Inference

- Gradient estimators of the ELBO
- Robbin's Monro Algorithm
- Natural Gradient

Gradient Estimators of the ELBO

$$\begin{aligned}\nabla_\phi \text{ELBO}(\phi) &= E_{q(z|\phi)}\left[\nabla_\phi \ln q(z|\phi) \ln \frac{p(x,z|\theta)}{q(z|\phi)}\right] \\ &\approx \frac{1}{S}\sum_{s=1}^{S} \nabla_\phi \ln q(z^s|\phi) \ln \frac{p(x,z^s|\theta)}{q(z^s|\phi)} \\ &\text{where } z^s \sim q(z|\phi)\end{aligned}$$

Robbin's Monro Algorithm

Usage

Given a function $g(\phi)$ such that

$$E(g(\phi)) = f(\phi)$$

It helps us find ϕ^* such that $f(\phi) = 0$

For our case, $E_q(g(\phi)) = \nabla_\phi \text{ELBO}(\phi)$ where

$$g(\phi) = \nabla_\phi \ln q(z|\phi) \ln \frac{p(x, z|\theta)}{q(z|\phi)}$$

Algorithm

At every iteration t, we update ϕ as follows-

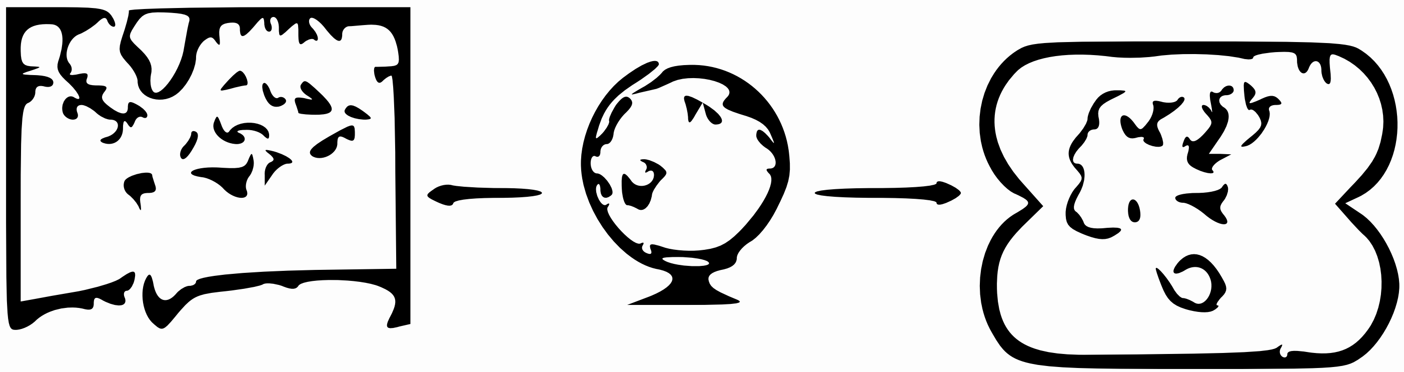
$$\phi^{(t)} = \phi^{(t-1)} + \rho_t g_t \left(\phi^{(t-1)}\right),$$

where g_t is an independent draw from $g(\phi)$ and the step sizes ρ_t follow-

$$\sum \rho_t = \infty \text{ and } \sum \rho_t^2 < \infty$$

Natural Gradient

- Euclidean gradient overshoots in regions of high curvature and becomes very slow in regions of low curvature
- Riemannian metric structure instead of the Euclidean metric structure
- Direction of steepest descent for Riemannian spaces is not the same



Mapping a manifold to a flat coordinate system distorts distances.

Natural Gradient

Solution: We compute the gradient directly on the globe!

Suppose we have a parameter space-

$\Theta = \{\boldsymbol{w} \in \boldsymbol{R}^n\}$

$$\text{Euclidean Distance: } |d\boldsymbol{w}|^2 = \sum_{i=1}^{n} (dw_i)^2$$

$$\text{Riemannian Distance: } |d\boldsymbol{w}|^2 = \sum_{i=1}^{n} g_{i,j}(w) dw_i dw_j \,,$$

So what do you think should be this distance metric?

Natural Gradient

The Fisher Information metric-

$$[\mathcal{I}(\phi)]_{i,j} = \mathrm{E}\left[\left(\frac{\partial}{\partial\phi_i}\log f(z;\phi)\right)\left(\frac{\partial}{\partial\phi_j}\log f(z;\phi)\right) \mid \phi\right]$$

These three ideas together give us **Stochastic Variational Inference** which is **fast** and **scalable**!

References

1. Stochastic Variational Inference
2. Robbin's Monro Algorithm
3. Natural Gradient