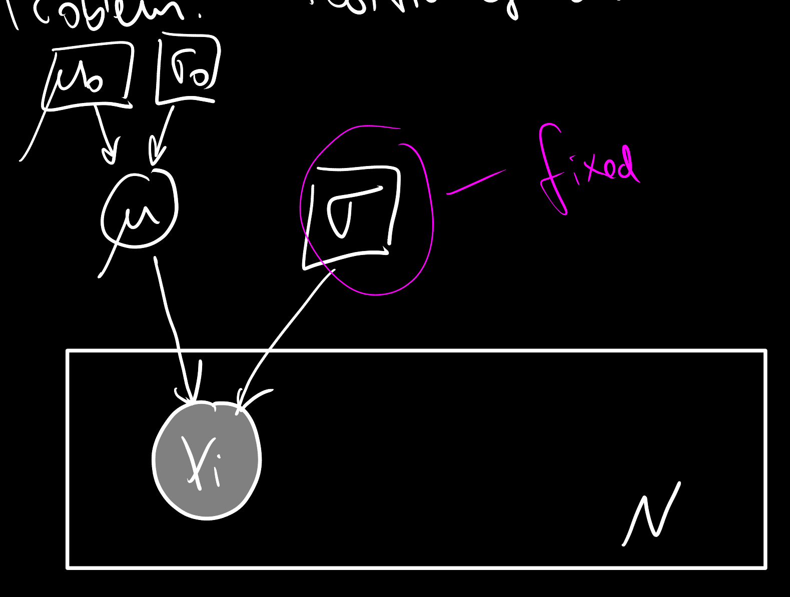
Variational Inference by Automatic Differentiation

$$q(z)^* = \underset{q(z)}{\operatorname{argmax}} \left(\underbrace{\mathbb{E}_{z\sim q(z)}\left[\log \frac{p(z, x=D)}{q(z)}\right]}_{\text{Evidence Lower Bound (ELBO)}} \right)$$

now propose a parametric distribution $q(z)$ and optimize its parameters

$\Rightarrow$ by AutoDiff in TensorFlow Probability

Problem: Posterior of a Normal with unknown mean



$$\mu \sim \mathcal{N}(\mu; \mu_0, \sigma_0)$$

$$X_i \sim \mathcal{N}(X_i \mid \mu, \sigma)$$

we have some data $D = \underbrace{\{1.3, 1.5, 1.1, -0.1, \dots\}}_{N \text{ data points}}$

$z = \mu$

seek: posterior $p(\mu \mid D)$ $\quad p(\mu \mid \underline{X} = D)$

we know: $p(\mu \mid D) = \mathcal{N}(\mu; \mu_N, \sigma_N)$

$$\boxed{\mu_N = \frac{\sigma^2 \mu_0 + \sigma_0^2 \sum_{i=0}^{N-1} x^{[i]}}{\sigma^2 + N\sigma_0^2}}$$

$$\boxed{\sigma_N = \frac{\sigma_0 \sigma}{\sqrt{\sigma^2 + N\sigma_0^2}}}$$

Assume we don't know the true posterior

propose surrogate $q(\mu) = \mathcal{N}(\mu; \mu_S, \sigma_S)$

learnable parameters

Maximize ELBO $\quad q(\mu)^* = \underset{q(\mu)\in Q}{\operatorname{argmax}} \ (\text{ELBO})$

$$\mu_S^*, \sigma_S^* = \underset{\substack{\mu_S \\ \sigma_S > 0}}{\operatorname{argmax}} \ (\text{ELBO})$$

log joint prob : can be evaluated by the DGM

$p(\mu, \underline{X}=D) = p(\mu)\, p(\underline{X}=D \mid \mu)$

$$\mu_S^*, \sigma_S^* = \underset{\substack{\mu_S \\ \sigma_S > 0}}{\cancel{\operatorname{argmax}}\ \operatorname{argmin}} - \underbrace{\left(\mathbb{E}_{\mu\sim q(\mu)}\left[\log \frac{p(\mu, \underline{X}=D)}{q(\mu)}\right]\right)}_{\text{negative ELBO} \ \hat{=}\ \text{loss}}$$

$$-\mathcal{L}(\mu_S, \sigma_S)$$

Build a gradient tape / computational while evaluating ELBO

$\rightarrow$ then gradients

$$\left.\begin{aligned} \frac{\partial -\mathcal{L}(\mu_S, \sigma_S)}{\partial \mu_S} &= \dots \\ \frac{\partial -\mathcal{L}(\mu_S, \sigma_S)}{\partial \sigma_S} &= \dots \end{aligned}\right\} \text{by reverse-mode auto-diff ("back-propagation")}$$

$\hookrightarrow$ use gradient-based optimizer, like ADAM

iterative solution $\quad \mu_S^{[j]}, \sigma_S^{[j]}$

$$\mu_S^{[j]} \xrightarrow{j\to\infty} \mu_S^* \qquad \sigma_S^{[j]} \xrightarrow{j\to\infty} \sigma_S^*$$

(we expect $= \mu_N$, $= \sigma_N$)

Approximate Expectation (here: ELBO) by sampling

leave-out

$$-\mathcal{L}(\mu_S, \sigma_S) \approx -\frac{1}{L} \sum_{l=0}^{L-1} \left(q(\mu = \mu^{[l]})\right) \cdot \log \frac{p(\mu = \mu^{[l]}, \underline{X} = D)}{q(\mu = \mu^{[l]})}$$

$$\mu^{[l]} \sim q(\mu)$$

① Sample L samples from $q(\mu)$

② Evaluate approximate ELBO

μ_S & σ_S appear in:

$\hookrightarrow$ creation of the samples

$\hookrightarrow$ evaluation of the log-prob