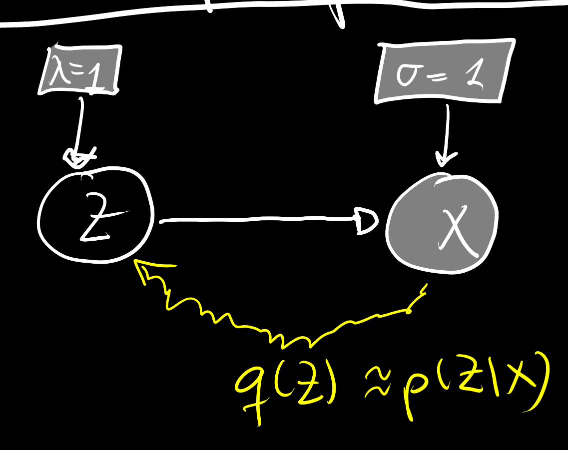
Simple Example for Variational Inference



$Z \sim Exp(Z; \lambda = 1)$

$X \sim \mathcal{N}(X; \mu = Z, \sigma = 1)$

↳ $p(Z|X)$ is intractable

Agenda

1. Trying to find a true posterior – and failing
2. Visualization: Joint, hypothetical posterior & surrogate posterior
3. Optimizing the ELBO (in closed-form) ← VI
4. Demo in Python with TensorFlow Probability

The joint distribution

$$p(Z) = Exp(Z; \lambda = 1) = \begin{cases} \lambda \exp(-\lambda Z) & , Z \geq 0 \\ 0 & , \text{else} \end{cases}$$

$$= \underbrace{\exp(-Z)\, \mathbb{I}(Z \geq 0)}_{}$$

Indicator Function

$$p(X|Z) = \mathcal{N}(X; \mu = Z, \sigma = 1)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X - \underset{\mu}{Z})^2\right)$$

$$p(X, Z) = p(Z)\, p(X|Z)$$

$$= \exp(-Z)\, \mathbb{I}(Z \geq 0) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X - Z)^2\right)$$

Bayes' Rule: $p(Z|X) = \dfrac{p(X, Z)}{p(X)}$ ← we don't have the marginal because it is intractable

$$p(X) = \int_0^\infty \underbrace{\exp(-Z) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X - Z)^2\right) dZ}_{\text{no closed-form solution, requires the erf}}$$

area under the curve

Let's instead go via the proportional

$$p(Z|X) \propto p(Z, X)$$

Posterior is proportional to the joint

↳ allows MAP

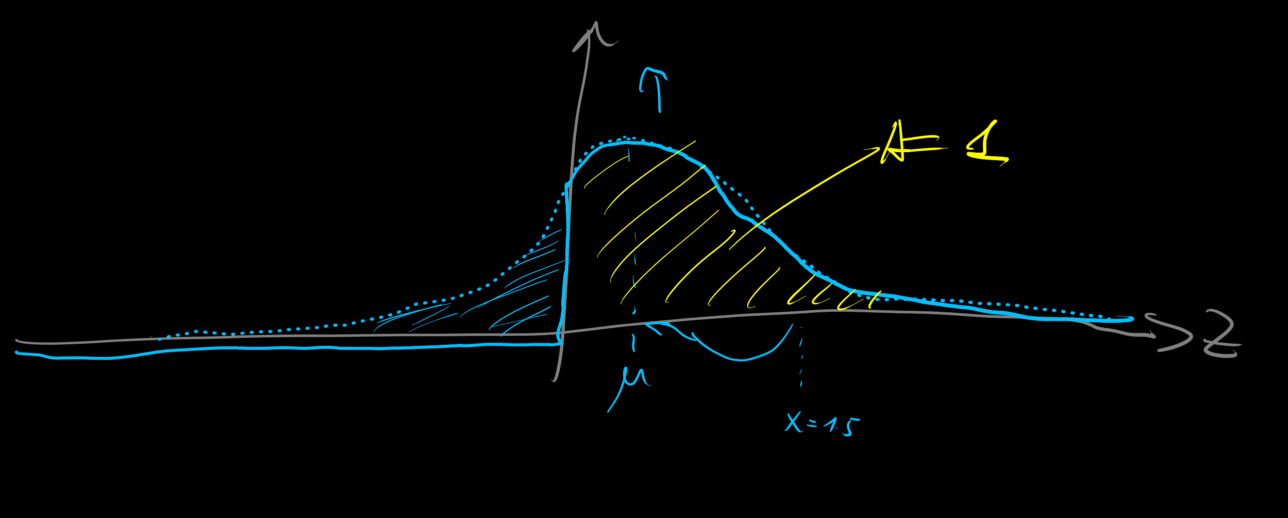
$$p(Z|X) \propto \exp(-Z)\, \mathbb{I}(Z \geq 0) \exp\left(-\frac{1}{2}(X - Z)^2\right)$$

$$= \exp\left(-Z - \cancel{\frac{1}{2}X^2} + XZ - \frac{1}{2}Z^2\right) \mathbb{I}(Z \geq 0)$$

$$\propto \exp\left(-\frac{1}{2}Z^2 + (X - 1)Z\right) \mathbb{I}(Z \geq 0)$$

$$= \exp\left(-\frac{1}{2}(Z - (X - 1))^2 + \cancel{\frac{1}{2}(X - 1)^2}\right) \mathbb{I}(Z \geq 0)$$

$$\propto \underbrace{\exp\left(-\frac{1}{2}(Z - \underbrace{(X - 1)}_{\mu})^2\right)}_{\text{a Normal Distribution}} \underbrace{\mathbb{I}(Z \geq 0)}_{\text{but only for } Z \geq 0}$$



Remedy: a Surrogate Posterior $q(Z)$

$$q^*(Z) = \arg\max\, (\mathcal{L}(q))$$

ELBO

$$\mathcal{L}(q) = \mathbb{E}_{Z \sim q(Z)}\left[\log\left(\frac{p(Z, X = D)}{q(Z)}\right)\right]$$

We need some data: e.g.: $D = \{ x^{[0]} = 1.3 \}$

A variational Optimization Problem: that's difficult

Let's introduce a parametric distribution

↳ Exponential Distribution

$$q_\theta(Z) = Exp(Z; \lambda = \theta)$$

$$= \begin{cases} \theta \exp(-\theta Z) & , Z \geq 0 \\ 0 & , \text{else} \end{cases}$$

$$= \theta \exp(-\theta Z)\, \mathbb{I}(Z \geq 0)$$

Surrogate is simpler than true posterior

$$\theta^* = \arg\max_{\theta \in \mathbb{R}_+} \underbrace{\mathbb{E}_{Z \sim q_\theta(Z)}\left[\log\left(\frac{p(Z, X = D)}{q_\theta(Z)}\right)\right]}_{\text{ELBO} : \mathcal{L}(\theta)}$$

[in real-world applications (e.g. VAE): you have approximate the ELBO by sampling]

$$\mathcal{L}(\theta) = \mathbb{E}_{Z \sim q_\theta(Z)}\left[\log(p(Z, X = D)) - \log(q_\theta(Z))\right]$$

$$= \mathbb{E}\Big[-Z + 0 - \frac{1}{2}\log(2\pi) - \frac{1}{2}(x^{[0]} - Z)^2$$
$$- (\log\theta - \theta Z + 0 \quad)\Big]$$

$$\log(\mathbb{I}(Z \geq 0)) = \begin{cases} 0 & , Z \geq 0 \\ -\infty & , \text{else} \end{cases}$$

$$= \mathbb{E}\left[-Z - \frac{1}{2}\log(2\pi) - \frac{1}{2}x^{[0]2} + x^{[0]}Z - \frac{1}{2}Z^2 - \log\theta + \theta Z\right]$$

$\mathbb{E}_{Z \sim Exp(\lambda)}[\ldots] = \ldots$

$\mathbb{E}_{Z \sim Exp(\lambda)}[\ldots Z] = \frac{1}{\lambda}\ldots$

$\mathbb{E}_{Z \sim Exp(\lambda)}[\ldots Z^2] = \frac{2}{\lambda^2}\ldots$

$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}x^{[0]2} - \log\theta - \frac{1}{\theta} + \frac{x^{[0]}}{\theta} + \cancel{\frac{\theta}{\theta}} - \frac{1}{\cancel{2}} \frac{\cancel{2}}{\theta^2}$$

$$= \cancel{-\frac{1}{2}\log(2\pi)} \cancel{- \frac{1}{2}x^{[0]2}} - \log\theta - \frac{1}{\theta} + \frac{x^{[0]}}{\theta} \cancel{+ 1} - \frac{1}{\theta^2}$$

$$= \mathcal{L}(\theta)$$

$$\boxed{\mathcal{L}(\theta) \stackrel{+}{=} -\log\theta + \frac{x^{[0]} - 1}{\theta} - \frac{1}{\theta^2}}$$

$$\theta^* = \arg\max_{\theta \in \mathbb{R}_+} \mathcal{L}(\theta)$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = -\frac{1}{\theta} - \frac{x^{[0]} - 1}{\theta^2} + \frac{2}{\theta^3} \stackrel{!}{=} 0 \qquad | \cdot \theta^3$$

$$-\theta^2 - (x^{[0]} - 1)\theta + 2 = 0 \qquad | \cdot (-1)$$

$$\Leftrightarrow \quad \theta^2 + (x^{[0]} - 1)\theta) - 2 = 0$$

$$\theta_{1,2} = -\frac{x^{[0]} - 1}{2} \pm \sqrt{\frac{(x^{[0]} - 1)^2}{4} + 2}$$

$$\boxed{\theta^* = -\frac{x^{[0]} - 1}{2} + \sqrt{\frac{(x^{[0]} - 1)^2}{4} + 2}}$$

The surrogate depends on the data