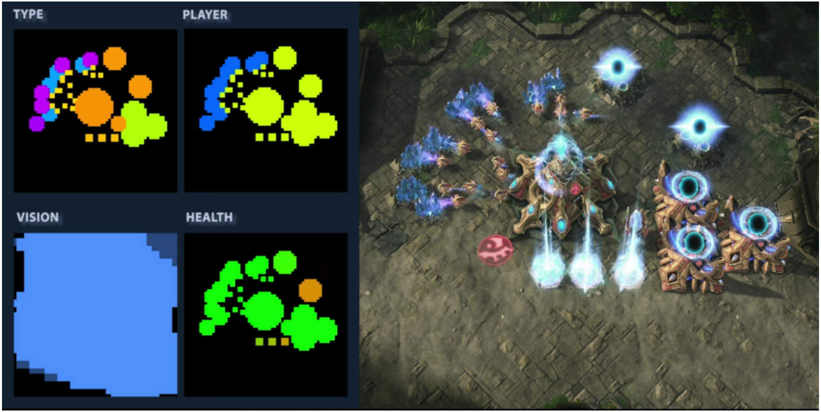
Reparametrisation Tricks
or
How to Differentiate Through Samples from Probability Distributions

Adam Kosiorek

June 8, 2017

Table of Contents

Playing Starcraft



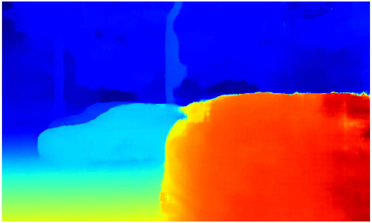
$$\theta_t = f_\phi\left(\mathbf{x}_{1:t}\right) \quad \text{compute parameters}$$

$$a_t \sim p(a \mid \theta_t) \quad \text{sample action from a distribution}$$

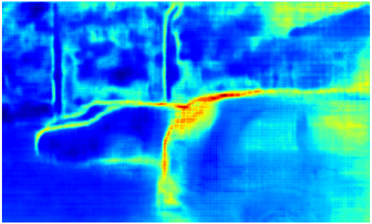
$$R = R\left(\mathbf{a}_{1:T}, \mathbf{x}_{1:T}\right) \quad \text{compute reward based on actions and states}$$

Alex Kendall's work on uncertainty

Input image $\mathbf{I}$



Depth estimate $\mathbf{d}$



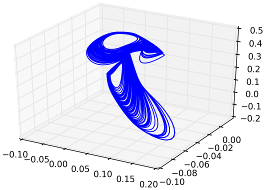
Uncertainty $\boldsymbol{\sigma}^2$

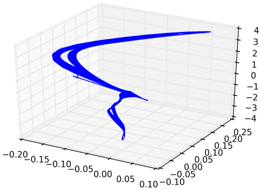
$$\boldsymbol{\mu}, \boldsymbol{\sigma}^2 = f_\phi\left(\mathbf{I}\right)$$

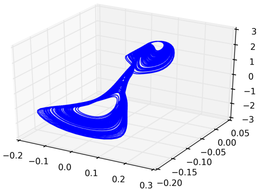
$$\mathbf{d} \sim \mathcal{N}\left(\mathbf{d} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$$

Stochastic Attractors of Chaotic Systems

Neil Dhir, Adam Kosiorek, Michael Osborne, Ingmar Posner







Dimensionality of the latent space?

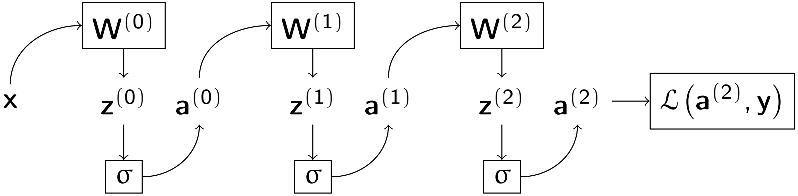
$$d_E \sim q_\phi(d_E \mid \mathbf{x}_{1:T})$$

$$f_{d_E} : \mathbb{R}^N \to \mathbb{R}^{d_E}$$

$$f_{d_E} : \mathbf{x}_{1:T} \mapsto \mathbf{z}_{1:T}$$

Table of Contents

General Computation Graph



Backprop Equations

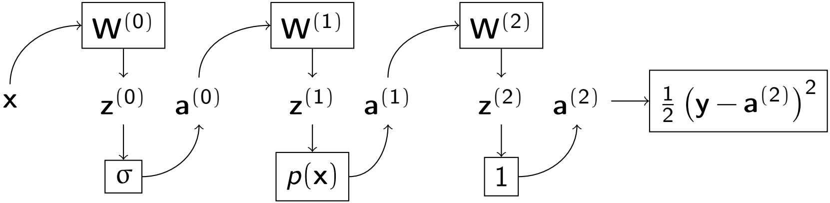
Let $\delta^{(l)} = \frac{\partial \mathcal{L}(\cdot)}{\partial \mathbf{z}^{(l)}}$, then backprop is given by:

$$\delta^{(L)} = \nabla_{\mathbf{a}^{(L)}} \mathcal{L}\left(\mathbf{a}^{(L)}, \mathbf{y}\right) \odot \sigma'\left(\mathbf{z}^{(L)}\right), \tag{1}$$

$$\delta^{(l)} = \left(\left(\mathbf{W}^{(l+1)}\right)^{T} \delta^{(l+1)}\right) \odot \sigma'\left(\mathbf{z}^{(l)}\right), \tag{2}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} \left(\mathbf{a}^{(l-1)}\right)^{T}. \tag{3}$$

Stochastic Computation Graph



Forward pass

$$\mathbf{z}^{(0)} = \mathbf{W}^{(0)}\mathbf{x}$$

Forward pass

$$\mathbf{z}^{(0)} = \mathbf{W}^{(0)}\mathbf{x}$$

$$\mathbf{a}^{(0)} = \sigma\left(\mathbf{z}^{(0)}\right)$$

Forward pass

$$\mathbf{z}^{(0)} = \mathbf{W}^{(0)}\mathbf{x}$$
$$\mathbf{a}^{(0)} = \sigma\left(\mathbf{z}^{(0)}\right)$$
$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)}\mathbf{a}^{(1)}$$

Forward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right)
\end{aligned}
$$

Forward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta)
\end{aligned}
$$

Forward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)}
\end{aligned}
$$

Forward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)} \\
\mathbf{a}^{(2)} &= \mathbf{z}^{(2)}
\end{aligned}
$$

Forward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)} \\
\mathbf{a}^{(2)} &= \mathbf{z}^{(2)} \\
\mathcal{L}(\cdot) &= \frac{1}{2}\left(\mathbf{y} - \mathbf{a}^{(2)}\right)^2
\end{aligned}
$$

Backward pass

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)} \\
\mathbf{a}^{(2)} &= \mathbf{z}^{(2)} \\
\mathcal{L}(\cdot) &= \frac{1}{2}\left(\mathbf{y} - \mathbf{a}^{(2)}\right)^2
\end{aligned}
$$

$$
\delta^{(2)} = (\mathbf{y} - \mathbf{a}^{(2)})
$$

Backward pass

$$\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)} \\
\mathbf{a}^{(2)} &= \mathbf{z}^{(2)} \\
\mathcal{L}(\cdot) &= \frac{1}{2}\left(\mathbf{y} - \mathbf{a}^{(2)}\right)^2
\end{aligned}$$

$$\delta^{(2)} = (\mathbf{y} - \mathbf{a}^{(2)})$$

Backward pass

$$\begin{aligned}
\mathbf{z}^{(0)} &= \mathbf{W}^{(0)}\mathbf{x} \\
\mathbf{a}^{(0)} &= \sigma\left(\mathbf{z}^{(0)}\right) \\
\mathbf{z}^{(1)} &= \mathbf{W}^{(1)}\mathbf{a}^{(1)} \\
\theta &= \sigma\left(\mathbf{z}^{(1)}\right) \\
\mathbf{a}^{(1)} &\sim p(\mathbf{a} \mid \theta) \\
\mathbf{z}^{(2)} &= \mathbf{W}^{(2)}\mathbf{a}^{(1)} \\
\mathbf{a}^{(2)} &= \mathbf{z}^{(2)} \\
\mathcal{L}(\cdot) &= \frac{1}{2}\left(\mathbf{y} - \mathbf{a}^{(2)}\right)^2
\end{aligned}$$

$$\vdots$$

$$\delta^{(1)} = \left(\left(\mathbf{W}^{(2)}\right)^T \delta^{(2)}\right) \odot \color{red}{\frac{\partial \mathbf{a}^{(1)}}{\partial \mathbf{z}^{(1)}}}$$

$$\delta^{(2)} = (\mathbf{y} - \mathbf{a}^{(2)})$$

Gradient of a sample?

$$
\begin{aligned}
\frac{\partial \mathbf{a}^{(1)}}{\partial \mathbf{z}^{(1)}} &= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \mathbf{z}^{(1)}} \\
&= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{z}^{(1)}}
\end{aligned} \tag{4}
$$

Gradient of a sample?

$$
\begin{aligned}
\frac{\partial \mathbf{a}^{(1)}}{\partial \mathbf{z}^{(1)}} &= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \mathbf{z}^{(1)}} \\
&= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{z}^{(1)}}
\end{aligned} \tag{4}
$$

but...

$$\mathbf{a}^{(1)} \sim p(\mathbf{a} \mid \theta) \tag{5}$$

Gradient of a sample?

$$\begin{aligned}\frac{\partial \mathbf{a}^{(1)}}{\partial \mathbf{z}^{(1)}} &= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \mathbf{z}^{(1)}} \\ &= \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} \frac{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{z}^{(1)}}\end{aligned} \tag{4}$$

but...

$$\begin{gathered}\mathbf{a}^{(1)} \sim p(\mathbf{a} \mid \theta) \\ \frac{\mathbf{a}^{(1)}}{\partial p\left(\mathbf{a}^{(1)} \mid \theta\right)} = \nexists\end{gathered} \tag{5}$$

Table of Contents

(Continuous) Reparametrisation Trick

Let $\mathbf{x} \sim p(\mathbf{x} \mid \theta)$ be a random variable. Perform change of variables such that

$$\begin{aligned} \epsilon &\sim p(\epsilon), \\ \mathbf{x} &= g(\epsilon, \theta) \end{aligned} \tag{6}$$

and treat ϵ as if it was a constant.

(Continuous) Reparametrisation Trick

Let $\mathbf{x} \sim p(\mathbf{x} \mid \theta)$ be a random variable. Perform change of variables such that

$$\begin{aligned} \epsilon &\sim p(\epsilon), \\ \mathbf{x} &= g(\epsilon, \theta) \end{aligned} \tag{6}$$

and treat ϵ as if it was a constant.

Example:

$$\begin{aligned} &\epsilon \sim \mathcal{N}(0, 1) \\ &x \sim \mathcal{N}\left(2, \frac{1}{4}\right) = \mathcal{N}(\mu, \sigma^2) \\ &x = g(\epsilon, \mu, \sigma^2) = \mu + \sigma\epsilon = 2 + \frac{1}{2}\epsilon \end{aligned}$$

(Continuous) Reparametrisation Trick

Let $\mathbf{x} \sim p(\mathbf{x} \mid \theta)$ be a random variable. Perform change of variables such that

$$\begin{aligned} \epsilon &\sim p(\epsilon), \\ \mathbf{x} &= g(\epsilon, \theta) \end{aligned} \tag{6}$$

and treat ϵ as if it was a constant.
Example:

$$\begin{aligned} \epsilon &\sim \mathcal{N}(0, 1) \\ x &\sim \mathcal{N}\left(2, \frac{1}{4}\right) = \mathcal{N}\left(\mu, \sigma^2\right) \\ x &= g(\epsilon, \mu, \sigma^2) = \mu + \sigma\epsilon = 2 + \frac{1}{2}\epsilon \end{aligned}$$

$$\frac{\mathrm{d}x}{\mathrm{d}\mu} = 1 \qquad \frac{\mathrm{d}x}{\mathrm{d}\sigma} = \frac{1}{2}\epsilon \qquad \frac{\mathrm{d}x}{\mathrm{d}\epsilon} \stackrel{!}{=} 0$$

Why can we do this?

We are NOT computing $\mathcal{L}(x)$.

Why can we do this?

We are NOT computing $\mathcal{L}(x)$.
We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$.

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\, \mathrm{d}x$.
We get the gradient by the change of variables:

$$p(x) = \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,}$$

$$\implies |p(x)\, \mathrm{d}x| = |p(\epsilon)\, \mathrm{d}\epsilon| \qquad \text{Mass conservation.}$$

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$.
We get the gradient by the change of variables:

$$p(x) = \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,}$$

$$\implies |p(x)\,\mathrm{d}x| = |p(\epsilon)\,\mathrm{d}\epsilon| \qquad \text{Mass conservation.}$$

$$\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$$

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\, \mathrm{d}x$.
We get the gradient by the change of variables:

$$p(x) = \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,}$$
$$\implies |p(x)\,\mathrm{d}x| = |p(\epsilon)\,\mathrm{d}\epsilon| \qquad \text{Mass conservation.}$$

$$\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$$
$$= \nabla_\theta \int p(\epsilon)\mathcal{L}(x)\,\mathrm{d}\epsilon$$

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$.
We get the gradient by the change of variables:

$$p(x) = \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,}$$

$$\implies |p(x)\,\mathrm{d}x| = |p(\epsilon)\,\mathrm{d}\epsilon| \qquad \text{Mass conservation.}$$

$$\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$$

$$= \nabla_\theta \int p(\epsilon)\mathcal{L}(x)\,\mathrm{d}\epsilon = \nabla_\theta \int p(\epsilon)\mathcal{L}(g(\epsilon, \theta))\,\mathrm{d}\epsilon$$

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$.
We get the gradient by the change of variables:

$$
\begin{aligned}
p(x) &= \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,} \\
\implies |p(x)\,\mathrm{d}x| &= |p(\epsilon)\,\mathrm{d}\epsilon| \qquad \text{Mass conservation.}
\end{aligned}
$$

$$
\begin{aligned}
&\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x \\
&= \nabla_\theta \int p(\epsilon)\mathcal{L}(x)\,\mathrm{d}\epsilon = \nabla_\theta \int p(\epsilon)\mathcal{L}(g(\epsilon,\theta))\,\mathrm{d}\epsilon \\
&= \int p(\epsilon)\nabla_\theta \mathcal{L}(g(\epsilon,\theta))\,\mathrm{d}\epsilon
\end{aligned}
$$

Why can we do this?

We are computing $\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$.
We get the gradient by the change of variables:

$$p(x) = \left|\frac{\mathrm{d}\epsilon}{\mathrm{d}x}\right| p(\epsilon) \quad \text{Change of variables,}$$
$$\implies |p(x)\,\mathrm{d}x| = |p(\epsilon)\,\mathrm{d}\epsilon| \qquad \text{Mass conservation.}$$

$$\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \nabla_\theta \int p_\theta(x \mid \mathcal{D})\mathcal{L}(x)\,\mathrm{d}x$$
$$= \nabla_\theta \int p(\epsilon)\mathcal{L}(x)\,\mathrm{d}\epsilon = \nabla_\theta \int p(\epsilon)\mathcal{L}(g(\epsilon, \theta))\,\mathrm{d}\epsilon$$
$$= \int p(\epsilon)\nabla_\theta \mathcal{L}(g(\epsilon, \theta))\,\mathrm{d}\epsilon = \mathbb{E}_{p(\epsilon)}[\nabla_\theta \mathcal{L}(g(\epsilon, \theta))]$$

MC approximation

How to compute $\mathbb{E}_{p(\epsilon)}[\nabla_\theta \mathcal{L}(g(\epsilon, \theta))]$?

1. Sample $\epsilon^{(s)} \sim p(\epsilon)$.
2. $\nabla_\theta \mathcal{L}(x) \simeq \frac{1}{S} \sum_{s=1}^{S} \nabla_\theta \mathcal{L}\left(g(\epsilon^{(s)}, \theta)\right)$

MC approximation

How to compute $\mathbb{E}_{p(\epsilon)}[\nabla_\theta \mathcal{L}(g(\epsilon, \theta))]$?

1. Sample $\epsilon^{(s)} \sim p(\epsilon)$.
2. $\nabla_\theta \mathcal{L}(x) \simeq \frac{1}{S} \sum_{s=1}^{S} \nabla_\theta \mathcal{L}\left(g(\epsilon^{(s)}, \theta)\right)$

What happens to the multiplication by $p(\cdot)$?

MC approximation

How to compute $\mathbb{E}_{p(\epsilon)}[\nabla_\theta \mathcal{L}(g(\epsilon, \theta))]$?

1. Sample $\epsilon^{(s)} \sim p(\epsilon)$.
2. $\nabla_\theta \mathcal{L}(x) \simeq \frac{1}{S} \sum_{s=1}^{S} \nabla_\theta \mathcal{L}\left(g(\epsilon^{(s)}, \theta)\right)$

What happens to the multiplication by $p(\cdot)$?

$$\mathbb{E}_{p(x)}[f(x)] = \int p(x) f(x)\, \mathrm{d}x \simeq \frac{1}{S} \sum_{s=1}^{S} f(x^{(s)}) \tag{7}$$

Table of Contents

REINFORCE or the log-derivative trick

We can use REINFORCE for continuous variables, too, but one-liners are typically easier and implemented by default.

Recall that

$$\nabla \log f(x) = \frac{\nabla f(x)}{f(x)}.$$

REINFORCE

Again, we're computing the expectation of the gradient:

$$\nabla_\theta \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\mathcal{L}(x)] = \int \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x$$

REINFORCE

Again, we're computing the expectation of the gradient:

$$\nabla_\theta \mathbb{E}_{p_\theta(x|\mathcal{D})}[\mathcal{L}(x)] = \int \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x$$

$$= \int \frac{p_\theta(x \mid \mathcal{D})}{p_\theta(x \mid \mathcal{D})} \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x$$

REINFORCE

Again, we're computing the expectation of the gradient:

$$
\begin{aligned}
\nabla_\theta \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\mathcal{L}(x)] &= \int \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int \frac{p_\theta(x \mid \mathcal{D})}{p_\theta(x \mid \mathcal{D})} \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int p_\theta(x \mid \mathcal{D}) \nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x
\end{aligned}
$$

REINFORCE

Again, we're computing the expectation of the gradient:

$$
\begin{aligned}
\nabla_\theta \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\mathcal{L}(x)] &= \int \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int \frac{p_\theta(x \mid \mathcal{D})}{p_\theta(x \mid \mathcal{D})} \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int p_\theta(x \mid \mathcal{D}) \nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x)]
\end{aligned}
$$

REINFORCE

Again, we're computing the expectation of the gradient:

$$
\begin{aligned}
\nabla_\theta \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\mathcal{L}(x)] &= \int \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int \frac{p_\theta(x \mid \mathcal{D})}{p_\theta(x \mid \mathcal{D})} \nabla_\theta p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \int p_\theta(x \mid \mathcal{D}) \nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x) \, \mathrm{d}x \\
&= \mathbb{E}_{p_\theta(x \mid \mathcal{D})}[\nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x)] \\
&\simeq \frac{1}{S} \sum_{s=1}^{S} \nabla_\theta \log p_\theta(x \mid \mathcal{D}) \mathcal{L}(x)
\end{aligned}
$$