Motivation

1. Supposed x is a rv, $x \sim p(x;\theta)$, f is a function of x and we want to compute $\nabla_\theta E_x[f(x)]$.

2. If the analytical gradient is not available, then we can use the score function estimator:

$$\nabla_\theta \mathbb{E}_x[f(x)] = \mathbb{E}_x\left[f(x)\nabla_\theta \log(p(x;\theta))\right]$$

3. However, it is often more convenient to define an objective whose gradient is an estimate of the gradient of $\mathbb{E}_x[f(x)]$.

4. For example, we can define the objective to be $f(x)\log(p(x;\theta))$, and then take the gradient of this objective, instead of taking the gradient of $\log(p(x;\theta))$, and then multiple it by $f(x)$. Doing this works better with modern deep learning frameworks.

5. This approach can be generalized using stochastic computational graph (SCG).

6. A SCG is an acyclic directed graph with 4 types of nodes: input nodes, Θ; deterministic nodes, $\mathcal{D}$; cost nodes, $\mathcal{C}$; and stochastic nodes, $\mathcal{S}$. The set of cost nodes is associated with an objective function $\mathcal{L} = \mathbb{E}\left[\sum_{c\in\mathcal{C}} c\right]$.

7. Now, we would like to obtain $\nabla_\theta \mathcal{L} = \nabla_\theta \mathbb{E}\left[\sum_{c\in\mathcal{C}} c\right]$

8. For SGC, we can define a surrogate loss:

$$\mathrm{SL}(\Theta,\mathcal{S}) := \sum_{w\in\mathcal{S}} \log p(w|\,\mathrm{DEPS}_w)\hat{Q}_w + \sum_{c\in\mathcal{C}} c\,(\mathrm{DEPS}_c)$$

where DEPS_w denotes the set of nodes that influence w, either stochastically or deterministically.
and $\hat{Q}_w$ is the sum of sampled costs $\hat{c}$ corresponding to the cost nodes influenced by w.

9. The key point to note here is that $\hat{Q}_w$, with the hat notation, does not depend on θ. Thus, the gradient of the SL is an estimate of the gradient of $\mathcal{L}$. $\nabla_\theta \mathcal{L} = \mathbb{E}\left[\nabla_\theta \,\mathrm{SL}(\Theta,\mathcal{S})\right]$.

10. Notice that without removing the dependency of Q_w, $\nabla_\theta\,\mathrm{SL}(\Theta,\mathcal{S})$ would contain a term of the form $\log(p)\nabla_\theta Q$, and thus, would not be an unbiased estimate of the gradient of $\mathcal{L}$.

Higher Order derivatives

1. Supposed x is a rv, $x \sim p(x;\theta)$, f is a function of x and we want to compute $\nabla_\theta \mathcal{L} = \nabla_\theta E_x[f(x;\theta)]$. Notice that f depends on θ now.

2. It can be shown through the score function estimator approach that:

$$\nabla_\theta E_x[f(x;\theta)] = \mathbb{E}_x\left[f(x;\theta)\nabla_\theta \log(p(x;\theta)) + \nabla_\theta f(x;\theta)\right] = \mathbb{E}_x[g(x;\theta)]$$

3. Now in the SL approach, if we set:

$$\begin{aligned}
\mathrm{SL}(f(x;\theta)) &= \log p(x;\theta)\hat{f}(x) + f(x;\theta) \\
(\nabla_\theta \mathcal{L})_{\mathrm{SL}} &= \mathbb{E}_x\left[\nabla_\theta \mathrm{SL}(f(x;\theta))\right] \\
&= \mathbb{E}_x\left[\hat{f}(x)\nabla_\theta \log p(x;\theta) + \nabla_\theta f(x;\theta)\right] \\
&= \mathbb{E}_x\left[g_{\mathrm{SL}}(x;\theta)\right]
\end{aligned}$$

Thus, $\nabla_\theta \mathrm{SL}(f(x;\theta))$ is an unbiased estimate of $\nabla_\theta \mathcal{L}$.

4. The key point to note is that in the SL approach, the dependency of f on θ has to be removed, for $\nabla_\theta \mathrm{SL}(f(x;\theta))$ to be an unbiased estimate of $\nabla_\theta \mathcal{L}$.

5. Notice that $g_{\mathrm{SL}}(x;\theta)$ and $g(x;\theta)$ evaluate to the same unbiased estimate of $\nabla_\theta \mathcal{L}$, which is the first order derivative.

6. Trouble appears when we try to compute the second order derivative $\nabla^2_\theta \mathcal{L}$ with the SL approach.

7. By applying the score function trick to $\nabla_\theta \mathcal{L}$, it can be shown that:

$$\begin{aligned}\nabla^2_\theta \mathcal{L} &= \nabla_\theta \mathbb{E}_x[g(x;\theta)] \\ &= \mathbb{E}_x\left[g(x;\theta)\nabla_\theta \log p(x;\theta) + \nabla_\theta g(x;\theta)\right]\end{aligned}$$

8. Applying the SF approach:

$$\begin{aligned}\mathrm{SL}\left(g_{\mathrm{SL}}(x;\theta)\right) &= \log p(x;\theta)\hat{g}_{\mathrm{SL}}(x) + g_{\mathrm{SL}}(x;\theta) \\ \left(\nabla^2_\theta \mathcal{L}\right)_{\mathrm{SL}} &= \mathbb{E}_x\left[\nabla_\theta \,\mathrm{SL}\left(g_{\mathrm{SL}}\right)\right] \\ &= \mathbb{E}_x\left[\hat{g}_{\mathrm{SL}}(x)\nabla_\theta \log p(x;\theta) + \nabla_\theta g_{\mathrm{SL}}(x;\theta)\right]\end{aligned}$$

9. Now, notice that:

$$\begin{aligned}\nabla_\theta g(x;\theta) =& \nabla_\theta f(x;\theta)\nabla_\theta \log(p(x;\theta)) \\ &+ f(x;\theta)\nabla^2_\theta \log(p(x;\theta)) \\ &+ \nabla^2_\theta f(x;\theta)\end{aligned}$$

$$\begin{aligned}\nabla_\theta g_{\mathrm{SL}}(x;\theta) = \hat{f}(x)\nabla^2_\theta \log(p(x;\theta)) \\ + \nabla^2_\theta f(x;\theta)\end{aligned}$$

10. Thus, by using the SL approach, we have lost the term $\nabla_\theta f(x;\theta)\nabla_\theta \log(p(x;\theta))$ in the second order derivative, leading to biased gradient estimate. This issue occurs because $\nabla_\theta \hat{f}(x) = 0$.

Correct Higher Order Gradient estimate with DICE

1. What we would like to have is an objective that can be differentiated n times to obtain an estimate of the n^{th}-order derivative of $\mathcal{L}$.

2. To accomplish this, the paper defines a new operator called MAGIC BOX M (not typesetted the same ways as the papers here).M takes a set $\mathcal{W}$ of stochastic nodes w as input and has the following 2 properties:

$$\begin{aligned}&1.M(\mathcal{W}) \mapsto 1 \\ &2.\nabla_\theta M(\mathcal{W}) = M(\mathcal{W})\sum_{w\in\mathcal{W}} \nabla_\theta \log(p(w,\theta))\end{aligned}$$

where $\mapsto$ indicates "evaluate to", and not full equality, which would have indicated equality if gradient was taken on both side.

3. The paper then defines the DICE objective:

$$\mathcal{L}_M = \sum_{c\in C} M(\mathcal{W}_c)c$$

where:

$$\mathcal{W}_c = \{w | w \in S, w \prec c, \theta \prec w\}$$

, i.e. the set of stochastic nodes that influences c and is influenced by the input θ.

4. They then prove that:

$$\mathbb{E}\left[\nabla^n_\theta \mathcal{L}\right] \mapsto \nabla^n_\theta \mathcal{L}, \forall n \in \{0,1,2,\ldots\}$$

that is, the n^{th}-order derivative of the DICE objective is an unbiased estimate of the the n^{th}-order derivative of the original objective $\mathcal{L}$.

5. The DICE operator M can be implemented as:

$$\begin{aligned} M(\mathcal{W}) &= \exp(\tau - \perp(\tau)) \\ \tau &= \sum_{w \in \mathcal{W}} \log(p(w;\theta)) \end{aligned}$$

where $\perp$ is an operator that sets the gradient of the operand to 0.

$$\nabla_x \perp(x) = 0$$