Discrete Variables and Gradient Estimators

This assignment is designed to get you comfortable deriving gradient estimators, and optimizing distributions over discrete random variables.

For most questions, the answers should be a few lines of derivations. Throughout, you can assume that $p(b|\theta)$ is differentiable w.r.t. θ.

Problem 1 (Unbiasedness of the score-function estimator, 8 points)

Gradient estimators are simply functions that take a parameter vector θ and a (possibly stochastic) function $L(\theta)$, and return another vector $\hat{g}$ of the same size as θ. Gradient estimators are generally more useful the closer their estimate is to the true gradient, i.e. when $\hat{g}$ is close to $\frac{\partial}{\partial\theta}L(\theta)$.

All else being equal, it's useful for a gradient estimator to be unbiased. The unbiasedness of a gradient estimator guarantees that, if we decay the step size and run stochastic gradient descent for long enough (see Robbins & Monroe), it will converge to a local optimum.

The standard REINFORCE, or score-function estimator is defined as:

$$\hat{g}_{\text{SF}}[f] = f\,(b)\,\frac{\partial}{\partial\theta}\log p(b|\theta), \qquad b \sim p(b|\theta) \tag{1}$$

(a) **[2 points]** First, let's warm up with the score function. Prove that the score function has zero expectation, i.e. $\mathbb{E}_{p(x|\theta)}\left[\nabla_\theta \log p(x|\theta)\right] = 0$. Assume that you can swap the derivative and integral operators.

(b) **[2 points]** Show that $\mathbb{E}_{p(b|\theta)}\left[f(b)\frac{\partial}{\partial\theta}\log p(b|\theta)\right] = \frac{\partial}{\partial\theta}\mathbb{E}_{p(b|\theta)}\left[f(b)\right]$.

(c) **[2 points]** Show that $\mathbb{E}_{p(b|\theta)}\left[[f(b)-c]\frac{\partial}{\partial\theta}\log p(b|\theta)\right] = \frac{\partial}{\partial\theta}\mathbb{E}_{p(b|\theta)}\left[f(b)\right]$ for any fixed c.

(d) **[2 points]** If the baseline depends on b, then REINFORCE will in general give biased gradient estimates. Give an example where $\mathbb{E}_{p(b|\theta)}\left[[f(b)-c(b)]\frac{\partial}{\partial\theta}\log p(b|\theta)\right] \neq \frac{\partial}{\partial\theta}\mathbb{E}_{p(b|\theta)}\left[f(b)\right]$ for some function $c(b)$, and show that it is biased.

The takeaway is that you can use a baseline to reduce the variance of REINFORCE, but not one that depends on the current action.

Problem 2 (Comparing variances of gradient estimators, 16 points)

If we restrict ourselves to consider only unbiased gradient estimators, then the main (and perhaps only) property we need to worry about is the variance of our estimators. In general, optimizing with a lower-variance unbiased estimator will converge faster than a high-variance unbiased estimator. However, which estimator has the lowest variance can depend on the function being optimized. In this question, we'll look at which gradient estimators scale to large numbers of parameters, by computing their variance as a function of dimension.

For simplicity, we'll consider a toy problem. The goal will be to estimate gradients of the expectation of a sum of one-dimensional Gaussians. Each Gaussian has unit variance, and its mean is given by an element of the D-dimensional parameter vector θ:

$$f(\mathbf{x}) = \sum_{d=1}^{D} x_d \tag{2}$$

$$L(\theta) = \mathbb{E}_{\mathbf{x}\sim p(x|\theta)}\left[f(\mathbf{x})\right] = \mathbb{E}_{\mathbf{x}\sim\mathcal{N}(\theta,I)}\left[\sum_{d=1}^{D} x_d\right] \tag{3}$$

(a) **[2 points]** As a warm-up, compute the variance of a single-sample simple Monte Carlo estimator of the objective $L(\theta)$:

$$\hat{L}_{MC} = \sum_{d=1}^{D} x_d, \qquad \text{where each } x_d \sim_{\text{iid}} \mathcal{N}(\theta_d, 1) \tag{4}$$

That is, compute $\mathbb{V}\left[\hat{L}_{MC}\right]$ as a function of D.

(b) **[2 points]** Recall the definition of the score-function, or REINFORCE estimator:

$$\hat{g}^{\text{SF}}[f] = \left[f(x) - c(\theta)\right]\frac{\partial}{\partial\theta}\log p(x|\theta), \qquad x \sim p(x|\theta) \tag{5}$$

Derive a closed form for this gradient estimator as a deterministic function of ϵ, a D-dimensional vector of standard normals. Set the baseline to $c(\theta) = \sum_{d=1}^{D}\theta_d$.

(c) **[4 points]** Derive the variance of the above gradient estimator. Because gradients are D-dimensional vectors, their covariance is a $D \times D$ matrix. To make things easier, we'll consider only the variance of the gradient with respect to the first element of the parameter vector, θ_1. That is, derive the scalar value $\mathbb{V}\left[\hat{g}_1^{\text{SF}}\right]$ as a function of D. Hint: The third moment of a standard normal is 0, and the fourth moment is 3. As a sanity check, consider the case where $D = 1$.

(d) **[4 points]** Next, let's look at the gold standard of gradient estimators, the reparameterization gradient estimator, where we reparameterize $x = T(\theta, \epsilon)$:

$$\hat{g}^{\text{REPARAM}}[f] = \frac{\partial f}{\partial x}\frac{\partial x}{\partial\theta}, \qquad \epsilon \sim p(\epsilon) \tag{6}$$

In this case, we can use the reparameterization $x = \theta + \epsilon$, with $\epsilon \sim \mathcal{N}(0, 1)$.

Derive this gradient estimator, and give $\mathbb{V}\left[\hat{g}_1^{\text{REPARAM}}\right]$ as a function of D.

(e) **[2 points]** Finally, let's consider a more exotic gradient estimator, Evolution Strategies. Following [Salimans et. al., 2016], we'll choose a meta-policy $p(\theta|\psi) = \mathcal{N}(\theta|\psi, \sigma^2 I)$. In this case, $\hat{g}^{\text{ES}}$ estimates $\frac{\partial}{\partial\psi}\mathbb{E}_{p(\theta|\psi,\sigma)}\left[L(\theta)\right]$, where ψ specifies the mean of the distribution over θ.

$$\hat{g}^{\text{ES}}[f] = \left[f(x) - c(\psi)\right]\frac{\partial}{\partial\theta}\log p(\theta|\psi,\sigma), \qquad x \sim p(x|\theta), \quad \theta \sim p(\theta|\psi,\sigma) \tag{7}$$

$$= \left[f(x) - c(\psi)\right]\frac{\partial}{\partial\theta}\log\mathcal{N}(\theta|\psi,\sigma^2 I) \qquad x \sim \mathcal{N}(x|\theta, I), \quad \theta \sim \mathcal{N}(\theta|\psi,\sigma^2 I) \tag{8}$$

Set the baseline to $c(\psi) = \sum_{d=1}^{D}\psi_d$. Derive a closed form for this gradient estimator as a deterministic function of ψ and σ, and two D-dimensional vectors of standard normals, ϵ_x and ϵ_θ.

(f) **[2 points]** Compute $\mathbb{V}\left[\hat{g}_1^{\text{ES}}\right]$ as a function of D and σ.

Problem 3 (Sometimes stochastic policies are necessarily suboptimal, 8 points)

In reinforcement learning and hard attention models, we often parameterize a policy to stochastically choose actions according to the current state. However, in many situations, the best policy is necessarily deterministic.

Given the objective

$$L(\theta) = \mathbb{E}_{p(b|\theta)}\left[f(b)\right] \tag{9}$$

where b is a Bernoulli random variable,

(a) **[4 points]** Show that for any f, $L(\theta)$ has local optima only for deterministic policies, i.e. values of $\theta = 0$ or $\theta = 1$.

(b) **[4 points]** Show that if $L(\theta) = \mathbb{E}_{p(b|\theta)}\left[f(b, \theta)\right]$, give an example where a non-deterministic policy is optimal.

The takeaway is that, for a fully-observed state in a non-adversarial setting, the only reason to use a stochastic policy is to make optimization easier.

Problem 4 (Representing and computing Categorical distributions, 8 points)
There are several ways to parameterization Categorical distributions.

(a) **[1 point]** If we parameterize a D-dimensional categorial distribution using a D-dimensional vector θ as

$$p(c|\theta) = \theta_c \tag{10}$$

what is the allowable range of the vector θ?

(b) **[1 point]** One problem with this parameterization is that it is hard to represent very small probabilities. If we parameterize a D-dimensional categorial distribution as

$$\log p(c|\theta) = \theta_c \tag{11}$$

what is the allowable range of the vector θ? What is the computational cost of evaluating $\log p(c|\theta)$, for a particular value of c, as a function of D?

(c) **[2 points]** Another problem is that optimizing a point in the simplex requires using a constrained optimization routine.

If we parameterize a D-dimensional categorial distribution as

$$\log p(c|\theta) = \theta_c - \log \sum_{c'=1}^{D} \exp \theta_{c'} \tag{12}$$

what is the allowable range of the vector θ? What is the computational cost of evaluating $\log p(c|\theta)$, for a particular value of c, as a function of D?

(d) **[2 points]** Using the parameterization

$$\log p(c|\theta) = \theta_c - \log \sum_{c'=1}^{D} \exp \theta_{c'} \tag{13}$$

write the gradient of $\log p(c|\theta)$ w.r.t. the vector θ. What is the computational complexity of evaluating this gradient, for a particular value of c, as a function of D?

(e) **[2 points]** Now let's consider the analogous problem of parameterizing a Bernoulli random variable. Give a monotonic, numerically stable paramaterization for $\log p(b|\theta)$ such that all $\theta \in \mathbb{R}$ give valid probabilities, and there is a one-to-one correspondence between θ and $p(b)$.

Problem 5 (Bonus: Optimal surrogates. 15 points)
Consider the objective

$$L(\theta) = \mathbb{E}_{p(b|\theta)}\left[f(b)\right] = \mathbb{E}_{p(b|\theta)}\left[(b-t)^2\right] \tag{14}$$

where b is a single Bernoulli random variable.

(a) **[Hard 5 points]** The REBAR estimator is given by:

$$\hat{g}_{\text{REBAR}} = \left[f(b) - \eta f(\sigma_\lambda(\tilde{z}))\right]\frac{\partial}{\partial\theta}\log p(b|\theta) + \frac{\partial}{\partial\theta}\eta f(\sigma_\lambda(z)) - \frac{\partial}{\partial\theta}\eta f(\sigma_\lambda(\tilde{z})) \tag{15}$$
$$b = H(z), z \sim p(z|\theta), \tilde{z} \sim p(z|b,\theta)$$

For this objective, find the optimal temperature λ and scale η for REBAR as a function of t and θ.

(b) **[Harder, 5 points]** The discrete version of the LAX estimator is given by:

$$\hat{g}_{\text{DLAX}} = f(b)\frac{\partial}{\partial\theta}\log p(b|\theta) - c_\phi(z)\frac{\partial}{\partial\theta}\log p(z|\theta) + \frac{\partial}{\partial\theta}c_\phi(z), \qquad b = H(z), z \sim p(z|\theta). \tag{16}$$

Find the optimal surrogate $c_\phi(z)$ for DLAX as a function of t and θ. Compute the variance of this estimator as a function of t and θ.

(c) **[Hardest, 5 points]**

$$\hat{g}_{\text{RELAX}} = \left[f(b) - c_\phi(\tilde{z})\right]\frac{\partial}{\partial\theta}\log p(b|\theta) + \frac{\partial}{\partial\theta}c_\phi(z) - \frac{\partial}{\partial\theta}c_\phi(\tilde{z}) \tag{17}$$
$$b = H(z), z \sim p(z|\theta), \tilde{z} \sim p(z|b,\theta)$$

Find the optimal surrogate $c_\phi(z)$ for RELAX as a function of t and θ. Compute the variance of this estimator as a function of t and θ.