Discrete Variables and Gradient Estimators

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

For all questions, the answers should be a few lines of derivations.

Problem 1 (Unbiasedness of the score-function estimator, 10 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) First, let's warm up with the score function. Prove that the score function has zero expectation, i.e. $\mathbb{E}_{p(x|\theta)}\left[\nabla_\theta \log p(x|\theta)\right] = 0$. Hint: Under some conditions, one can swap the derivative and integral operators.

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

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

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

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, 10 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 a sum of one-dimensional Gaussians with means given by a D-dimensional parameter vector θ and unit variance:

$$L(\theta) = \mathbb{E}_{\mathcal{N}(\theta,1)}\left[\sum_{d=1}^{D} X_d\right] \tag{2}$$

(a) As a warm-up, let's compute the variance of a Monte Carlo estimator of the objective $L(\theta)$:

$$\hat{L}_{MC} = \sum_{d=1}^{D} x_d, \qquad \text{where each } x_d \sim \mathcal{N}(\theta_d, 1) \tag{3}$$

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

(b) Next, we'll look at the variance of gradient estimators.

Recall the definition of the score-function, or REINFORCE estimator:

$$\hat{g}_{\text{SF}}[f] = f\,(b)\,\frac{\partial}{\partial\theta}\log p(b|\theta), \qquad b \sim p(b|\theta) \tag{4}$$

Because gradients are D-dimensional vectors, they actually have a $D \times D$ covariance matrix. For this question, we'll consider the variance of a gradient estimator to mean the sum of the diagonal of the covariance matrix.

Derive $\sum_{d=1}^{D} \text{Cov}\left[\hat{g}_{\text{SF}}\right]_{dd}$ as a function of D.

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

$$\hat{g}_{\text{REPARAM}}[f] = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta}, \qquad \epsilon \sim p(\epsilon) \tag{5}$$

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

Derive $\sum_{d=1}^{D} \text{Cov}\left[\hat{g}_{\text{REPARAM}}\right]_{dd}$ as a function of D.

(d) Finally, let's consider a more exotic gradient estimator, a variant of Evolution Strategies:

$$\hat{g}_{\text{ES}}[f] = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{f} - f(\theta_i)\right)\frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta}, \qquad x \sim p(x|\theta) \tag{6}$$

Compute $\mathbb{V}\left[\hat{g}_{\text{ES}}\right]$ as a function of D.

Problem 3 (The pointlessness of stochastic policies, 10 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, every stochastic policy is dominated by a non-stochastic policy.

Given the objective

$$L(\theta) = \mathbb{E}_{p}(b|\theta)\,[f(b)] \tag{7}$$

where b is a Bernoulli random variable,

(a) Show that $L(\theta)$ has local optima only for deterministic policies, i.e. values of $\theta = 0$ or $\theta = 1$.

(b) Show that if $L(\theta) = \mathbb{E}_{p}(b|\theta)\,[f(b, \theta)]$, then a non-deterministic policy can be 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 discrete variables, 10 points)

A Categorical distribution can be represented by a point in the simplex.

(a) If we parameterize a D-dimensional categorial distribution as

$$p(b|\theta) = \theta_b \tag{8}$$

what is the allowable range of θ?

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

$$\log p(b|\theta) = \theta_b \tag{9}$$

what is the allowable range of θ?

(c) 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

$$\log p(b|\theta) = \theta_b - \log \sum_{b'=1}^{D} \exp \theta_{b'} \tag{10}$$

what is the allowable range of θ?

(d) Now let's consider the analogous problem of parameterizing a Bernoulli random variable. Give a paramaterization for $logp(b|\theta)$ 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. 10 points)
Given the objective

$$L(\theta) = \mathbb{E}_p(b|\theta) \left[(b - t)^2 \right] \tag{11}$$

it would be informative to know the optimal surrogate.

(a) (Medium) Find the optimal temperature for REBAR as a function of t and θ. This requires only calculus.

(b) (Hard) Find the optimal control variate for LAX as a function of t and θ. This will require variational calculus, or solving a PDE.

(c) (Harder) Find the optimal control variate for RELAX as a function of t and θ.

Problem 6 (Bonus: Optimal Gaussian reparameterization, 10 points)
If

$$X \sim \mathcal{N}(\mu, \Sigma)$$

, we can draw samples in a reparameterizable way by

$$x = \mu + L\epsilon$$

where $\epsilon \sim \mathcal{N}(0, I)$, and $L^T L = \Sigma$.

(a) Given μ and Σ, what L minimizes the trace of the variance of $\nabla_L X$?