Variance Reduction Part 1 notes

Our goal is to estimate expectations of functions

$$E_{p(x)}[f(x)] \approx \sum_i \frac{f(x_i)}{n} \quad \text{with } x_i \sim p(x) \text{ independent}$$

using monte Carlo estimators.

This is unbiased

$$E\left[\sum_i \frac{f(x_i)}{n}\right] = \sum_i \frac{1}{n} E[f(x)] = E[f(x)]$$

$$Var\left(\sum_i \frac{f(x_i)}{n}\right) = \frac{Var(f(x))}{n} = \frac{\sigma^2}{n}$$

We can reduce the variance by increasing n. Can we do better?

(Caveat: We have have to balance the cost of the extra work with the cost of drawing extra samples.)

Two motivating examples

1. Reinforcement learning - policy gradients
   Take an action according to $\pi(a)$ get stochastic reward $R(a)$, want to maximize

   $$E_\pi[R(a)] \quad \text{but} \quad \nabla E_\pi[R(a)] = E_\pi[R(a)\nabla \log \pi(a)]$$

   is high variance. We may be limited in the number of samples.
   We can use control variates to reduce the variance of the estimator.

2. Suppose you are responsible for Google Home metrics. You collect millions of interactions with the device and want to know what % are "satisfactory". You can evaluate the utterances by having someone listen to each interaction, but this is expensive. With a limited budget, what can you do?

George Tucker (gjt@google.com)

Antithetics

$$\mu = E[f(x)]$$

$$f(x_i) = \mu + (f(x_i) - \mu) = \mu + \varepsilon_i \Rightarrow \sum_i \frac{1}{n} f(x_i) = \mu + \sum_i \frac{\varepsilon_i}{n}$$

Each ε_i has variance σ^2 and the average has variance $\frac{\sigma^2}{n}$ due to independence. If we sample in a coupled fashion we might get even more cancellation.

$$X \xrightarrow{\text{an involution}} \tilde{X} \quad \text{s.t.} \quad f(x) \text{ and } f(\tilde{X})$$

somehow have opposite errors

For example

$$X \sim N(\mu, \sigma^2) \quad \text{then} \quad \tilde{X} = \mu - (x - \mu) = 2\mu - x$$

The effectiveness depends on f

$$Var(\hat{\mu}_{anti}) = Var\left(\frac{1}{N}\sum_i \frac{f(x_i) + f(\tilde{x}_i)}{2}\right) = \frac{1}{N^2}\, \frac{Var(f(x_i)) + Var(f(\tilde{x}_i)) + 2cov(f(x_i), f(\tilde{x}_i))}{2}$$

$$= \frac{1}{N^2}\left(N\sigma^2 + cov(f(x), f(\tilde{x}))\right) = \frac{1}{N}\sigma^2(1 + \rho)$$

So it is good if $f(x)$ & $f(\tilde{x})$ are anti correlated.

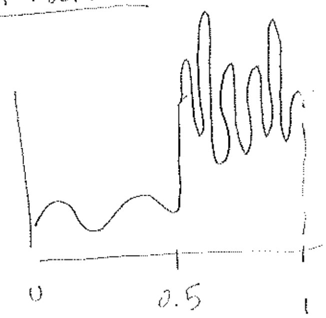
Q1
$$f(x) = \frac{f(x) + f(\tilde{x})}{2} + \frac{f(x) - f(\tilde{x})}{2} = f_E(x) + f_O(x)$$

Show that f_E and f_O are orthogonal and that

$$Var(\hat{\mu}_{anti}) = \frac{2Var(f_E(x))}{n} \quad \text{whereas} \quad Var(\hat{\mu}) = \frac{Var(f_E) + Var(f_O)}{n}$$

Sampling x & $\tilde{x}$ can sometimes be much cheaper than sampling x twice.

Stratification



might estimate the function in two pieces and possibly weight the right piece more

Partition space $D_1, \ldots, D_J$ w/ $p(x \in D_j) = w_j$ $\quad p_j(x) = p(x \mid x \in D_j)$

Then sample n_j samples $x_{ij} \sim p_j$ and form

$$\hat{\mu}_{strat} = \sum_{j=1}^{J} \frac{w_j}{n_j} \sum_{i=1}^{n_j} f(x_{ij})$$

Check: $\hat{\mu}_{strat}$ is unbiased

$$Var(\hat{\mu}_{strat}) = Var\left(\sum_{j=1}^{J} \frac{w_j}{n_j} \sum_{i=1}^{n_j} f(x_{ij}) \right) = \sum_{j=1}^{J} \frac{w_j^2}{n_j^2} n_j \sigma_j^2 = \sum_{j=1}^{J} \frac{w_j^2}{n_j} \sigma_j^2$$

Proportional allocation says $n_j = w_j n$, so

$$Var(\hat{\mu}_{prop}) = \sum_{j=1}^{J} \frac{w_j}{n} \sigma_j^2$$

Using the law of variance

$$Var(f(x)) = E[Var(f(x|y))] + Var(E[f(x|y)])$$

$$\sigma^2 = \underbrace{\sum_{j=1}^{J} w_j \sigma_j^2}_{\text{within strata}} + \sum_{j=1}^{J} w_j \underbrace{(\mu_j - \mu)^2}_{\text{between strata}}$$

So $Var(\hat{\mu}_{prop}) \leq Var(\hat{\mu})$ allowing us to drop the between strata variance term

Optimal allocation is possible, but tricky because σ_j^2 is unknown

Q1) For the metrics problem, how could you use stratified sampling? Make additional reasonable assumptions.

Q2) Can you construct $f = f_B + f_W$ similar to the antithetics where f_B contains the between strata variance and f_W contains the within strata variance

Common Random Numbers

Suppose we want

$$E[f(x) - g(x)] \quad \text{for } f \text{ and } g \text{ related}$$

e.g., $f(x) = h(x, \theta)$ and $g(x) = h(x, \tilde{\theta})$

Then $E[f(x)] - E[g(x)] = E[f(x) - g(x)]$

Which is better?

$$Var(\hat{\mu}_{common}) = \frac{Var(f(x)) + Var(g(x)) - 2cov(f(x), g(x))}{n}$$

$$Var(\hat{\mu}_{indep}) = \frac{Var(f(x)) + Var(g(x))}{n}$$

If f and g are closely related, this is great. Also useful for

$E[f(x) - f(\tilde{x})]$ if we can reparameterize from a common source of randomness.

Eg. $x \sim N(\mu, \sigma^2)$ then $z \sim N(0,1)$ $x = \mu + \sigma z$

$\tilde{x} \sim N(\tilde{\mu}, \tilde{\sigma}^2)$ $\tilde{x} = \tilde{\mu} + \tilde{\sigma} z$

For example,

$$\nabla_\theta E_{\pi(a;\theta)}[R(a)] \approx \frac{E_{\pi(a,\theta+\varepsilon)}[R(a)] - E_{\pi(a,\theta-\varepsilon)}[R(a)]}{2\varepsilon}$$

can get a much lower variance estimator w/ common random numbers.

Conditioning

If we want

$$E[f(x,y)] = E_x[\underbrace{E_y[f(x,y)|x]}_{h(x)}] = E_x[h(x)]$$

Now

$$\mathrm{Var}(h(x)) = \mathrm{Var}_x(f(x)) - E_x[\mathrm{Var}_y(f(x,y|x)] \geq \mathrm{var}_x(f(x))$$

Caveat: If $h(x)$ is expensive, then may not be useful.

(1) Suppose $\nabla E[f(h)]$ $\quad h = H(z) \quad z = \log\alpha + \log u - \log(1-u)$ (Logistic random variable), $u \sim U(0,1)$

Now

$$E_h[f(h)\nabla \log p(h)] = \nabla E_h[f(h)] = \nabla E_z[f(H(z))] = E_z[f(H(z))\nabla \log p(z)]$$

which is better?

(2) Suppose $\pi(a) = \pi(a_1)\pi(a_2|a_1)\ldots\pi(a_n|a_{1:n-1})$

Compute an estimator for

$$\nabla E_\pi[\log \pi]$$

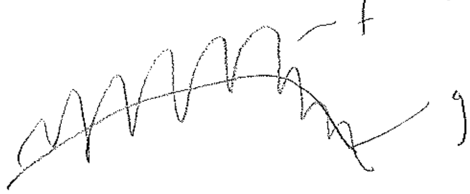
That does 1 step of conditioning - i.e. local conditional entropy.

George Tucker (gjt@google.com)

Control Variates - allowing us to use close formed solutions

We note

$$E[f(x)] = E[\underbrace{f(x) - g(x)}_{\text{complex residual}}] + \underbrace{E[g(x)]}_{\text{if this is tractable, this can be helpful}}$$



Check: unbiased

$$\hat{\mu}_{cv} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) - \eta g(x_i) + \eta \bar{g}$$

$$Var(\hat{\mu}_{cv}) = \frac{1}{n}\left(Var(f) + \eta^2 Var(g) - \eta 2 \, cov(f,g)\right)$$

So $$\eta_{opt} = \frac{cov(f,g)}{var(g)}$$

Thus $$var(\hat{\mu}_{cv}) = \frac{1}{n}\left(var(f) + \frac{cov(f,g)^2 var(g)}{var(g)^2} - 2\frac{cov(f,g)^2}{var(g)}\right)$$

$$= \frac{1}{n}\left(1 - \frac{cov(f,g)^2}{var(g)var(f)}\right) var(f)$$

So any CV is helpful if we can compute η optimally.

Note that

$$var(\hat{\mu}_{cv}) = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \eta(g(x_i) - \bar{g})\right)^2\right] - \underbrace{E\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right]^2}_{\text{constant in } \eta}$$

So the optimal η minimizes a least squares problem. Can estimate η_{opt} by least squares (or leave one out least squares for unbiasedness).

Ex 1 In the problem

$$E_\pi[R(a)\nabla \log \pi]$$

$\nabla \log \pi$ is a control variate because $E[\nabla \log \pi] = \nabla E_\pi[1] = 0$

So $E_\pi[(R(a) - b)\nabla \log \pi]$ is unbiased. b could be estimated from minibatch statistics.

George Tucker (gjt@google.com)

Q prop

Policy gradients in continuous control

$$E_s E_{a\sim\pi}\left[\hat{Q}\,\nabla\log\pi\right]$$ this is the key quantity to compute (on policy)

$\hat{Q}$ = sample of discounted returns.

1. $Q^{\pi}(a,s) = r(a,s) + \gamma E_{\pi}\left[Q^{\pi}(a',s')\right]$
   can be estimated off policy $\Rightarrow Q_w(a,s)$ parameterized critic

2. How can we use $Q_w(a,s)$ to reduce variance?

MuProp showed any linear function of a can be used

$$E_{a\sim\pi}\left[(ma+b)\nabla\log\pi(a)\right] = m\,E_{a\sim\pi}\left[a\,\nabla\log\pi(a)\right]$$
$$= m\,\underbrace{\nabla E_{a\sim\pi}[a]}_{\text{as long as this is tractable e.g. Gaussian policy}}$$

$$m = Q'_w(\bar{a},s)$$

So, we get

$$E_s\left[E_{a\sim\pi}\left[\left(\hat{Q} - \eta\left(Q_w(\bar{a},s) + (a-\bar{a})Q'_w(\bar{a},s)\right)\right)\nabla\log\pi\right] + \eta\,Q'_w(\bar{a},s)\,\nabla E_{\pi}[a]\right]$$

Alternatively

$$E_s\left[E_{a\sim\pi}\left[\left(\hat{Q}(a,s) - \eta Q_w(a,s)\right)\nabla\log\pi\right] + \eta\,\nabla \underbrace{E_{\pi}\left[Q_w(a,s)\right]}_{\text{tractable if } \pi \text{ is reparameterizable}}\right]$$

3. Can you use a linear approximation to reduce the variance further?

$$\nabla E_{\pi}\left[Q_w(a,s)\right] = \nabla E_{\varepsilon}\left[Q_w(a(\varepsilon,\theta),s)\right]$$
$$= E_{\varepsilon}\left[\nabla Q_w(a(\varepsilon,\theta),s)\right]$$
$$= E_{\varepsilon}\left[Q'_w(a,s)\big|_{a=a(\varepsilon,\theta)}\nabla a(\varepsilon,\theta)\right]$$

$- Q'_w(\bar{a},s)\,\nabla a(\varepsilon,\theta)$

George Tucker (gjt@google.com)

$\eta(s)$ can be learned online w/ least squares.

(E) How can you use a model-based system to reduce the variance of the model free policy gradient?

Rebar

h is discrete. Want to compute

$$E\left[f(h)\nabla\log p(h)\right]$$

$$= \nabla E_p\left[f(h)\right] = \nabla\left(E_h\left[f(h) - \eta E_{z|h}\left[f(\sigma_t(z))\right]\right] + \eta E_z\left[f(\sigma_t(z))\right]\right)$$

$$= E_h\left[\left(f(h) - \eta E_{z|h}\left[f(\sigma_t(z))\right]\right)\nabla\log p(h) \underbrace{- \eta\nabla E_{z|h}\left[f(\sigma_t(z))\right]\right] + \eta\nabla E_z\left[f(\sigma_t(z))\right]}_{\text{reparameterizable}}$$

$z = \log\alpha + \log u - \log(1-u) \qquad u \sim \text{Unif}(0,1)$

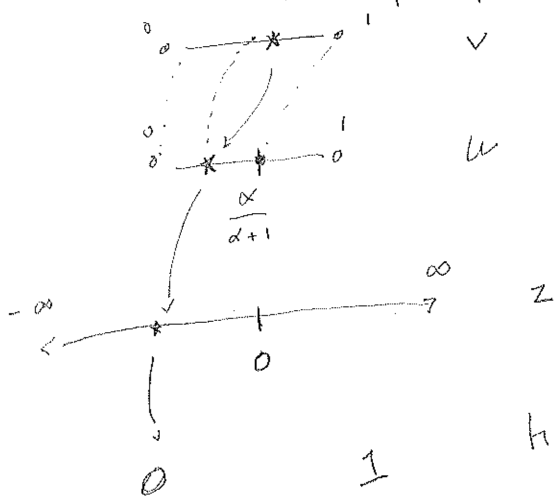
$h = H(z)$ so that $h \sim \text{Bern}\left(\frac{\alpha}{\alpha+1}\right)$

σ_t is a tempered sigmoid that approximates $H(z)$

$$h = H(z) \approx \sigma\left(\frac{z}{t}\right) = \sigma_t(z)$$

How can we use common random numbers?

Need h, z & $z|h$



(Q1) Why is this okay?

George Tucker (gjt@google.com)