Importance Sampling

Note Title 4/2/2006

Task: Want to estimate $\mu = \int_{\mathcal{D}} h(x)\pi(x)\,dx$.

Importance Sampling,

Draw i.i.d. samples $x^{(1)}, \dots, x^{(m)}$ from a trial distribution $g(x)$.

Calculate the importance weight

$$w^{(j)} = \frac{\pi(x^{(j)})}{g(x^{(j)})} \qquad j = 1 \text{ to } m.$$

Two possibilities:

(A) $$\bar{\mu}_m = \frac{1}{m} \sum_{i=1}^{m} w^{(i)} h(x^{(i)})$$

unknown.

$l(x) = c\pi(x)$

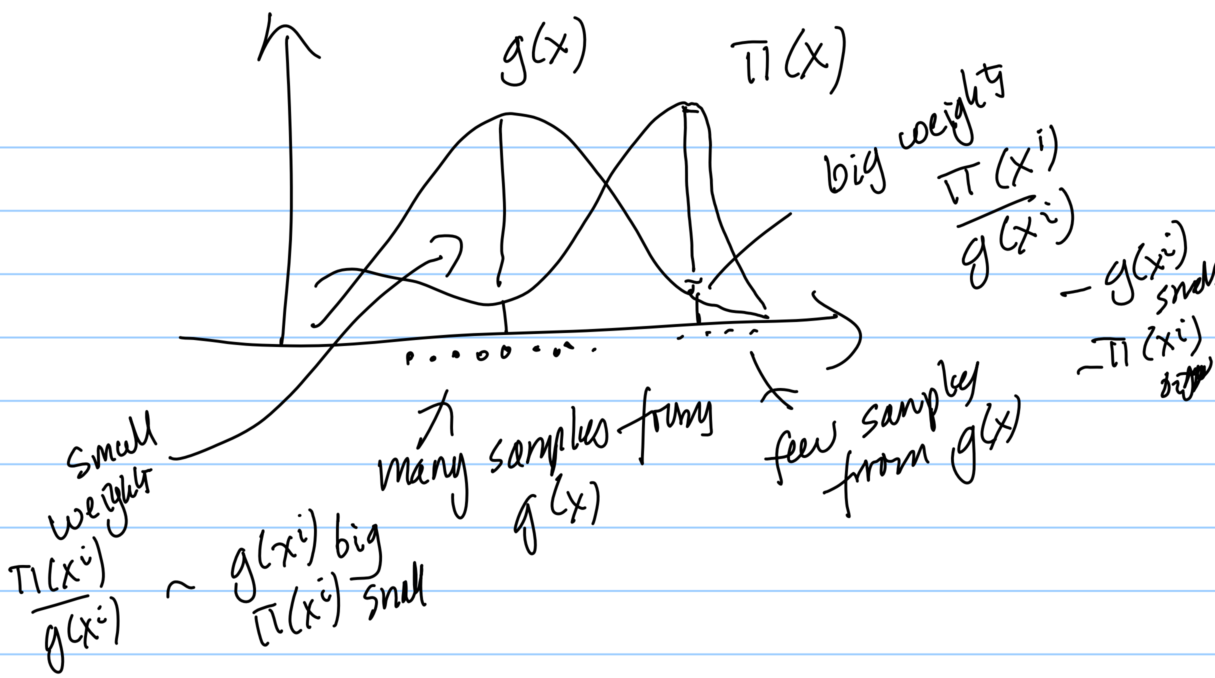
(B) $$\tilde{\mu}_m = \frac{\sum_{i=1}^{m} w^{(i)} h(x^{(i)})}{\sum_{i=1}^{m} w^{(i)}}$$

$w(x) = \frac{l(x)}{g(x)}$

Trade-offs:

$\tilde{\mu}_m$ only needs to know the ratios of the weights, so don't need to know the normalization of $\pi(x)$

$\bar{\mu}_m$ is unbiased, but $\tilde{\mu}_m$ may be biased.



- Why weight the samples like this?

Because it gives an unbiased estimator $\tilde{\mu}_m$
(see next page)

- What choice of $g(x)$ is best?

Best choice will make $g(x) \propto h(x)\pi(x)$
(∝: proportional to)

This makes the efficiency as good as possible. We will prove this later by computing the variance of the estimator.

(Page 3)

Check bias of $\tilde{\mu}_m$

$$E_\pi[\tilde{\mu}_m] = \sum_{x^{(1)},\dots,x^{(m)}} g(x^{(1)})\dots g(x^{(m)}) \frac{1}{m}\sum_{i=1}^{m} \frac{\pi(x^{(i)})\,h(x^{(i)})}{g(x^{(i)})}$$

$$E_\pi[\tilde{\mu}_m] = \sum_x \pi(x)\,h(x) = \mu$$ unbiased.

The bias of $\tilde{\tilde{\mu}}_m$ can be calculated by

$$E_\pi[\tilde{\tilde{\mu}}_m] = \sum_{x^{(1)},\dots,x^{(m)}} g(x^{(1)})\dots g(x^{(m)}) \frac{\sum_{i=1}^{m} \frac{\pi(x^{(i)})\,h(x^{(i)})}{g(x^{(i)})}}{\sum_{j=1}^{m} \frac{\pi(x^{(j)})}{g(x^{(j)})}}$$

Impossible to compute $E_\pi[\tilde{\tilde{\mu}}_m]$.
It will almost never equal μ exactly

But if m is large enough (law of large numbers)
then $\sum_{j=1}^{m} \frac{\pi(x^{(j)})}{g(x^{(j)})} \approx m \sum_x \frac{\pi(x)}{g(x)} \cdot g(x) = m$

So we expect $E_\pi[\tilde{\tilde{\mu}}_m] \approx \mu$.
(if m is large)

Variance (for no. of samples)
Empirical Claim: the variance of $\tilde{\mu}_m$ is often larger than $\tilde{\tilde{\mu}}_m$.

Page 4

To make estimation error small, we want to make $g(x)$ as close as possible in shape to $\pi(x)\, h(x)$. Intuitive.

<u>Why?</u>

$$\mathrm{Var}_g(\tilde{\mu}_m) = \sum_{x^1 \ldots x^m} \prod_{i=1}^{m} g(x^i) \frac{1}{m^2} \sum_{i,j=1}^{m} \omega^{(i)}\omega^{(j)} h(x^{(i)})\, h(x^{(j)}) - \mu^2$$

$$= \frac{1}{m^2}\, m \sum_x \frac{\{\pi(x)\}^2 \{h(x)\}^2 g(x)}{\{g(x)\}^2} + \frac{1}{m^2}(m^2 - m)\left\{\sum_x \frac{\pi(x) h(x) \cancel{g(x)}}{\cancel{g(x)}}\right\}^2 - \mu^2$$

$$\mathrm{Var}_g(\tilde{\mu}_m) = \frac{1}{m}\left\{\sum_x \frac{\{\pi(x)\}^2}{g(x)}\{h(x)\}^2\right\} - \frac{1}{m}\left\{\sum_x \pi(x) h(x)\right\}^2$$

By Cauchy-Schwarz inequality

$$\sum_x \frac{\{\pi(x)\}^2}{g(x)}\{h(x)\}^2 \geqslant \left\{\sum_x \pi(x) h(x)\right\}^2$$

$$|\underline{a}|^2 |\underline{b}|^2 \geqslant |\underline{a}\cdot\underline{b}|^2$$

$$\underline{a} = \frac{\pi(x)}{\{g(x)\}^{1/2}}\, h(x), \quad \underline{b} = \{g(x)\}^{1/2}$$

<u>Equality</u> only when $\underline{a} \propto \underline{b}$

if, and only, if. $g(x) \propto \pi(x) h(x)$

(5) Summary:

Importance Sampling:

Goal: estimate $\mu = \sum_x h(x)\pi(x)$

(o) Sample from $g(x)$ to get $x^1, x^2, \ldots x^m$

If $\pi(x)$ known (including normalization term)

$\hat{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} \omega^i h(x^i)$ $\quad \omega^i = \frac{\pi(x^i)}{g(x^i)}$

If $\ell(x) = \pi(x)/c$ known, but c unknown

$\tilde{\mu}_m = \frac{\sum_{i=1}^{m} \omega^i h(x^i)}{\sum_{i=1}^{m} \omega^i}$ → note: only need to know $\frac{\omega^i}{\sum_{i=1}^{m}\omega^i}$ ratio of weights

→ so, you don't need to know c.

Best Efficiency

for $\hat{\mu}_m$, best efficiency

if $g(x) \approx h(x)\pi(x)$

minimizes the variance of the estimator

Similarly for $\tilde{\mu}_m$, but can't obtain analytic results.