Importance sampling

(not a sampling metho
→ it used as an extensi
to MC for approximati
intractable integral)

Recall Monte-Carlo :

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)p(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) \quad , x_i \overset{i.i.d}{\sim} p(x)$$

- $f(x) = x$: expectation of x
- $f(x) = (x - \mathbb{E}(x))^2$: variance of x

Setup: Assume $p(x)$ is a density, i.e. $\int p(x)\,dx = 1$. Then:

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)\,p(x)\,dx = \int \left[f(x)\frac{p(x)}{q(x)}\right] q(x)\,dx$$

$$= \mathbb{E}_{x\sim q(x)}\left[f(x)\frac{p(x)}{q(x)}\right] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)} \quad , x_i \sim$$

$$\approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)\,w(x_i) \quad , \quad w(x) := \frac{p(x)}{q(x)} \quad \text{Importance weights.}$$

$q(x)$: proposal distribution
1. Easy to sample from
2. Easy to evaluate for a given x.

$p(x)$:
1. May not be easy to sample
2. Should be easy to evaluate

$\circledast$ $\forall$ density $q(x)$ s.t. $q(x)=0 \Rightarrow p(x)=0$

Remarks :

1.) $$\mathbb{E}_{x\sim q(x)}\left[\hat{\mu}_n^{IS}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{x\sim q(x)}\left[f(x_i)\frac{p(x_i)}{q(x_i)}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int f(x_i)\frac{p(x_i)}{\cancel{q(x_i)}}\cancel{q(x_i)}\,dx =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{x\sim p(x)}[f(x_i)] \xrightarrow{n\to\infty} \mathbb{E}_{x\sim p(x)}[f(x)]$$

$\Rightarrow$ Hence, the IS estimator is unbiased, consistent.

2.) $\mathbb{V}ar[\hat{\mu}_n^{IS}] = \underline{\frac{1}{n}} \underbrace{\mathbb{V}ar\left[f(x)\frac{p(x)}{q(x)}\right]}$ $\qquad \underset{\mathbb{V}ar[f(x)]}{\underline{MC}} >> \underset{\mathbb{V}ar\left[f(x)\frac{p(x)}{q(x)}\right]}{\underline{IS}}$

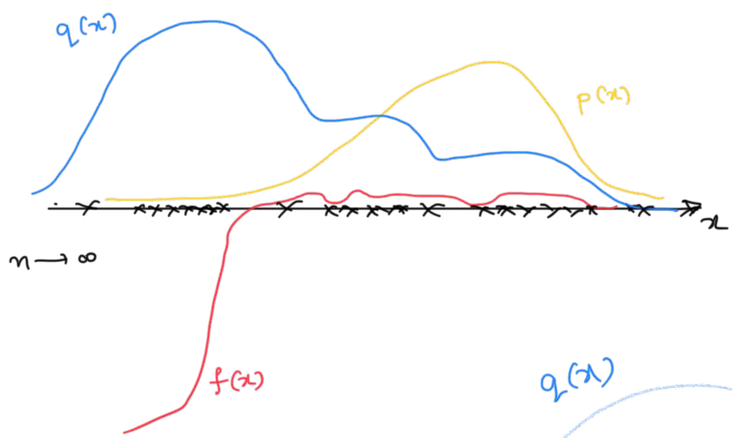
In fact, it is straightforward to solve for optimal $q^*(x)$, but in practice it may be difficult to sample from $q^*(x)$.

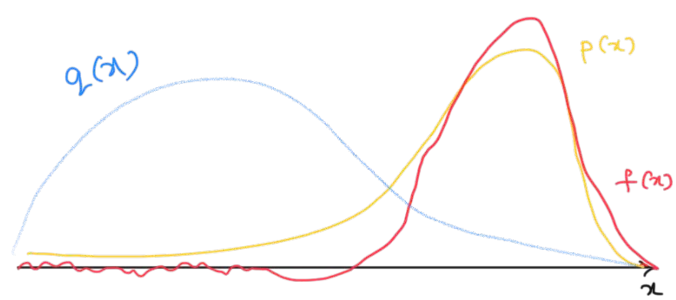
<u>Scenarios</u>:

1.) Can't sample from $p(x)$, use IS.

2.) Whether or not we sample from $p(x)$, use IS to improve upon the convergence of the vanilla MC estimator.

<u>Example</u>: Suppose $f(x)$ is the return of some investment x, we want to approximate the expected returns: $\mathbb{E}_{x\sim p(x)}[f(x)]$





Sampling from the wrong $q(x)$ can make things horribly wrong.

<u>How to choose $q(x)$?</u>

Choose $q(x)$ to be large where $|f(x)|p(x)$ is large

$\frac{1}{n}\sum^n f(x_i)\frac{p(x_i)}{\quad}, \; x_i \overset{i.i.d.}{\sim} q(x)$

$n \quad i=1 \qquad q(x_i)$

Limitations : 1. Our intuition on how to choose a good $q(x)$ breaks down in high dimensions.

2. It is difficult to assess the accuracy of our estimator.

Importance sampling for un-normalized distribution :

Often we may know $p(x)$ or $q(x)$ only up to a normalizing const

e.g. : Posterior inferece : $p(\theta|\mathcal{D}) = \dfrac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta} \propto p(\mathcal{D}|\theta)p(\theta)$

Setup: Assume : $p(x) = \dfrac{\tilde{p}(x)}{Z_p}$, $q(x) = \dfrac{\tilde{q}(x)}{Z_q}$, $\int \tilde{q}(x)dx = Z_q$, $\int \tilde{p}(x)dx = Z_p$.

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)p(x)dx \overset{I.S.}{=} \int \left[f(x)\frac{p(x)}{q(x)}\right]q(x)dx$$

$$= \int f(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\frac{Z_q}{Z_p}q(x)dx$$

Assume that we can efficiently sample from $q(x)$, and we can evaluate $\tilde{p}(x)$, $\tilde{q}(x)$.

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\frac{Z_q}{Z_p}q(x)dx \approx \frac{Z_q}{Z_p}\frac{1}{n}\sum_{i=1}^{n} f(x_i)\tilde{w}(x_i), \quad (1)$$

where $x_i \overset{i.i.d}{\sim} q(x)$, and $\tilde{w}(x_i) := \dfrac{\tilde{p}(x_i)}{\tilde{q}(x_i)}$

But we still don't know the normalizing constants ...

Then let's try to approximate them via Monte-Carlo :

$$\frac{Z_p}{Z_q} = \frac{1}{Z_q}\int \tilde{p}(x)dx = \int \frac{\tilde{p}(x)}{\tilde{q}(x)}q(x)dx = \mathbb{E}_{x\sim q(x)}\left[\frac{\tilde{p}(x)}{\tilde{q}(x)}\right]$$

$$\approx \frac{1}{n}\sum_{i=1}^{n}\tilde{w}(x_i) \qquad (2)$$

$$\Longrightarrow \mathbb{E}_{x \sim p(x)}\left[f(x)\right] \approx \frac{\frac{1}{n}\sum_{i=1}^{n} f(x_i)\,\tilde{w}(x_i)}{\frac{1}{n}\sum_{i=1}^{n} \tilde{w}(x_i)} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\,\hat{w}(\cdot$$

where $x_i \overset{i.i.d}{\sim} q(x)$, and $\hat{w}(x_i) := \dfrac{\tilde{w}(x_i)}{\frac{1}{n}\sum_{i=1}^{n}\tilde{w}(x_i)}$ "approx import weig

(*) Keep in mind that we made two MC approximations.

... hence, this is expected to perform worse than regular I.S.