Basic indirect sampling methods

Outline

- Rejection Sampling
- Importance Sampling
  - estimating integrals and normalising constants
  - basic properties

Monte Carlo

- Consider an arbitrary distribution on $\mathcal{X}$ with a density π w.r.t to dx, such that
$$\pi(x) = \frac{\gamma(x)}{Z}$$
and is Z **unknown**
- Let $\varphi : \mathcal{X} \longrightarrow \mathbb{R}^{n_x}$, with $\overline{\varphi} = \sup \varphi < +\infty$
- We want to **compute**
$$\pi(\varphi) = \mathbb{E}_\pi[\varphi(X)] = \langle \pi, \varphi \rangle = \int_{\mathcal{X}} \varphi(x)\pi(dx) \tag{1}$$
and Z.

Indirect sampling

- There are indirect ways for sampling perfectly from π using
  - rejection sampling
  - importance sampling
  - Markov Chains
  - particle systems & methods

Rejection Sampling

- Basic idea:
  - sample from other proposal distribution q
  - accept or reject with a probability α (details below)
- Proposal distribution property 1
  - absolute continuity, $\frac{d\pi}{dq}(x)$ exists and is bounded
  - Let q also denote density, then $\pi \ll q$: $\gamma(x) > 0 \Rightarrow q(x) > 0$
  - i.e. q has heavier tails

Rejection Sampling

- Proposal distribution property 2
  - assume you know M such that for all x:

$$w(x) = \frac{\gamma(x)}{q(x)} < M$$

- Acceptance probability for a sample x' from q should be

$$\alpha = \frac{w(x')}{M}$$

Rejection Sampling

Accept Reject Procedure:

- Sample $X \sim q$
- Sample $U \sim U[0, 1)$
- **Accept** sample, $Y = X$ if $U < \frac{w(X)}{M}$

Rejection Sampling

- Procedure generates samples from π based on simple conditioning argument

$$
\begin{aligned}
\mathbb{P}[Y \in A] &= \mathbb{P}\left[X \in A \middle| U < \frac{w(X)}{M}\right] \\
&= \frac{\mathbb{P}\left[X \in A, U < \frac{w(X)}{M}\right]}{\mathbb{P}\left[U < \frac{w(X)}{M}\right]} \\
&= \frac{\int_A \int_0^1 q(x) 1_{u < w(x)/M} du dx}{\int_{\mathcal{X}} q(x) \left(\int_0^{w(x)/M} du\right) dx} \\
&= \frac{\int_A q(x) \frac{w(x)}{M} dx}{\int_{\mathcal{X}} q(x) \frac{w(x)}{M} dx} \\
&= \frac{\int_A \gamma(x) dx}{\int_{\mathcal{X}} \gamma(x) dx} = \pi(A)
\end{aligned}
$$

Rejection Sampling

- Issue is that

$$\mathbb{P}\left[U < \frac{w(X)}{M}\right] = \mathbb{E}_q\left[\mathbb{P}\left[U < \frac{w(X)}{M}\middle| X\right]\right] = \ldots = \frac{Z}{M}$$

  so method might not be very efficient if M high!
- So in practice need $M \approx Z$ i.e. $\pi \approx q$ which is not easy or realistic often
- There are also more advanced rejection methods
  - envelopes, adaptive accept-reject,...

Importance Sampling (IS)

- Will use a similar proposal q such that $\pi \ll q$. Recall $\pi(dx) = \frac{\gamma(x)dx}{Z}$ so

$$\begin{aligned}\pi(dx) &= \frac{d\pi}{dq}(x)q(dx) \\ &= \frac{1}{Z}w(x)q(dx)\end{aligned}$$

- q importance distribution

- When usual densities exist, the un-normalised importance weight is:

$$w(x) = \frac{\gamma(x)}{q(x)}$$

and

$$\pi(x) = \frac{w(x)q(x)}{\int w(x)q(x)dx}$$

Importance Sampling: known normalising constant

- When Z known, the following Monte Carlo approximation can be used

$$\widehat{\pi}(dx) = \frac{1}{N}\sum_{i=1}^{N} W^i \delta_{X^i}(dx)$$

where

$$W^i = \frac{w(X^i)}{Z}$$

- Approximate $\pi(\varphi)$ with

$$\widehat{\pi}(\varphi) = \frac{1}{N}\sum_{i=1}^{N} W^i \varphi(X^i)$$

Estimating normalising constant

Note

$$\mathbb{E}_q\left[\sum_{i=1}^N W^i\right] = \mathbb{E}_q\left[\sum_{i=1}^N \frac{\gamma(X^i)}{Zq(X^i)}\right] = \sum_{i=1}^N \frac{\int \gamma(x^1)dx^1}{Z} = N$$

In other words

$$\mathbb{E}_q\left[\frac{1}{N}\sum_{i=1}^N w(X^i)\right] = Z$$

So $\frac{1}{N}\sum_{i=1}^N w(X^i)$ is an unbiased estimator of Z.

QUIZ: Is $\hat{\pi}(\varphi)$ unbiased?

Variance of $\widehat{\pi}(\varphi)$

Importance Sampling: self normalising case

When Z unknown (most interesting cases), the following Monte Carlo approximation can be used

$$\widehat{\pi}(dx) = \sum_{i=1}^{N} W^i \delta_{X^i}(dx)$$

where

$$W^i = \frac{w(X^i)}{\sum_{i'=1}^{N} w(X^{i'})}$$

such that $\sum_{i=1}^{N} W^i = 1$, so for the integral:

$$\widehat{\pi}(\varphi) = \sum_{i=1}^{N} W^i \varphi(X^i)$$

Importance Sampling: self normalising case

- Z can be estimated **unbiasedly** as before using $\frac{1}{N}\sum_{i'=1}^{N} w(X^{i'})$
- Asymptotically the self normalising case behaves as the known Z case

QUIZ: Is $\hat{\pi}(\varphi)$ this time unbiased?

Self normalising case: some asymptotics

- Asymptotically consistent as $N \to \infty$. Asymptotic bias

$$(\widehat{\pi}(\varphi) - \pi(\varphi)) = -\frac{1}{N} \int_{\mathcal{X}} \frac{\pi^2(x)}{q(x)} (\varphi(x) - \pi(\varphi))\, dx$$

- Central Limit Theorem (CLT) holds:

$$\sqrt{N}(\widehat{\pi}(\varphi) - \pi(\varphi)) \Rightarrow \mathcal{N}\left(0, \sigma_{IS}^2\right)$$

where

$$\sigma_{IS}^2 = \left(\int_{\mathcal{X}} \frac{\pi^2(x)}{q(x)} (\varphi(x) - \pi(\varphi))^2\, dx \right)$$

Importance Sampling: choosing proposals

The asymptotic variance of the estimator $\hat{\pi}(\varphi)$ is minimised by

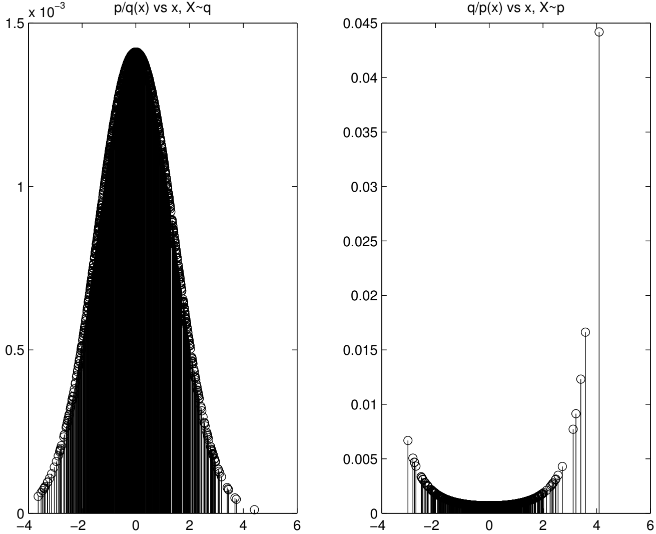
$$q(x) = \frac{|\varphi(x)| \, \pi(x)}{\int_{\mathcal{X}} |\varphi(x)| \, \pi(x) \, dx}.$$

but this is **not** very easy in practice!

QUIZ: Can you prove this? Hint: in variance calculation earlier consider only part that depends on q and apply Jensen's inequality.

A very simple example

How important is absolute continuity? Consider $p = \mathcal{N}(0,1)$, $q = \mathcal{N}(0,2)$



Tail estimation example: variance reduction

Problem: Consider the tail $A = \{x : x \leq c\}$ of a given density p. Estimate $p^* = \int_A p(x)dx$.

- Procedure 1:
  - Use perfect Monte Carlo (with $\pi = p$ and $\varphi = 1_A$)
  - Sample $X^i \sim p$ and $\widehat{p^*} = \frac{1}{N}\sum_{i=1}^{N} 1_A(X^i)$
  - Variance of estimator $\frac{p^* - {p^*}^2}{N}$

Tail estimation example: variance reduction

- Procedure 2:
  - change target distribution to

$$\pi = \frac{p(x)1_A(x)}{p^*} \quad p^* = \int_A p(x)dx$$

    and estimate normalising constant
  - Sampling directly from π unrealistic so aim is to find q with $w(x) < \infty$ and more "mass" in rare region A.

QUIZ: Can you write an expression for the Monte Carlo variance for Procedure 2? Can you use π above as a proposal for p?

Discussion

- We are typically interested in the expectations of several test functions
  - moments, simple functions for histograms or probabilities
- Results on with φ useful for understanding what types of functions will lead to good estimators
  - But not very useful in practice except when interested in specific test functions as in tail estimation above
- Can assess efficiency by looking at estimation of normalising constant

Importance Sampling: normalising constant

- Estimate normalising constant Z,

$$\widehat{Z} = \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma(X^i)}{q(X^i)}$$

- Variance:

$$\mathbb{V}ar\left[\widehat{Z}\right] = \frac{Z^2}{N}\left(\int \frac{\pi^2(x)}{q(x)}dx - 1\right)$$

Choosing importance proposals

- We can attempt to select q which minimises either
  - the variance of the importance weights.
  - the relative variance of $\hat{Z}$
  - in both cases q should be π.
- So one could **construct** q **similar or close** to π
- Can use other methods/approximations:
  - Laplace principle, Gaussian, Saddlepoint approximations etc.

The effective sample size (ESS)

- We can rescale the relative variance of the importance weights

$$ESS = \frac{N}{1 + \mathbb{V}ar_q\left[\frac{w(X)}{Z}\right]}$$

- The higher the ESS the better
  - ESS/N can be interpreted as an "efficiency" number compared to perfect Monte Carlo
- Can be monitored using Monte Carlo approximations:

$$ESS^N = \frac{1}{\sum_{i=1}^N (W^i)^2} = \frac{\left(\sum_{i=1}^N w(X^i)\right)^2}{\sum_{i=1}^N w(X^i)^2}$$

to get a number in $[1, N]$.

Discussion

- It is crucial to find a good q
  - cannot be easily automated and requires good understanding of the problem
  - key is to minimise dissimilarity between π and q
- Approach will degenerate for hard problems
  - high dimensional x or low p^* in tail estimation
  - this results to very low weights and high weight variance