Simulation - Lectures 4 - Importance sampling

Lecture version: Sunday 9th February, 2020, 15:52

Robert Davies

Part A Simulation and Statistical Programming

Hilary Term 2020

Recap from previous lecture

- Monte Carlo can estimate integrals when we can simulate random variables
- We have seen how inversion, transformation, and rejection sampling can generate random variables for different distributions, conditional on being able to draw random uniforms
- Today, we have our first of two lectures on **Importance Sampling**

Importance Sampling

- We want to estimate
$$\theta = \mathbb{E}(\phi(X))$$
where X is a rv with pdf or pmf p and $\phi : \Omega \to \mathbb{R}$.
- The Monte Carlo estimator uses samples from p to estimate θ, but this choice is in general suboptimal
- Importance sampling uses samples from another distribution q, called importance or proposal distribution, and reweight them
- It is also useful when we need to make an accurate estimate of the probability that a random variable exceeds some very high threshold.
- In this context it is referred to as a *variance reduction* technique.

Importance Sampling Identity

Importance sampling identity

Let $Y \sim q$ and $X \sim p$ be continuous or discrete rv on Ω. Assume $p(x) > 0 \Rightarrow q(x) > 0$, then for any function $\phi : \Omega \to \mathbb{R}$ we have

$$\mathbb{E}_p(\phi(X)) = \mathbb{E}_q(\phi(Y)w(Y))$$

where $w : \Omega \to \mathbb{R}^+$ is the importance weight function

$$w(x) = \frac{p(x)}{q(x)}.$$

Importance Sampling Identity

- Proof: We have

$$\begin{aligned}\mathbb{E}_p(\phi(X)) &= \int_\Omega \phi(x)p(x)dx \\ &= \int_\Omega \phi(x)\frac{p(x)}{q(x)}q(x)dx \\ &= \int_\Omega \phi(x)w(x)q(x)dx \\ &= \mathbb{E}_q(\phi(Y)w(Y)).\end{aligned}$$

- Similar proof holds in the discrete case.

Importance Sampling Estimator

Definition

Let q and p be pdfs or pmfs on Ω. Assume $p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0$. Let $\phi : \Omega \to \mathbb{R}$ and $X \sim p$ such that $\theta = \mathbb{E}_p(\phi(X))$ exists.
Let $Y_1, ..., Y_n$ be a sample of independent random variables distributed according to q. The importance sampling estimator is defined as

$$\hat{\theta}_n^{\mathsf{IS}} = \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) w(Y_i).$$

Properties

The IS estimator is

- Unbiased: $\mathbb{E}[\hat{\theta}_n^{\mathsf{IS}}] = \theta$
- (Weakly and strongly) consistent: $\hat{\theta}_n^{\mathsf{IS}} \longrightarrow \theta$ a.s. as $n \to \infty$.

Importance Sampling Estimator

- Proof.

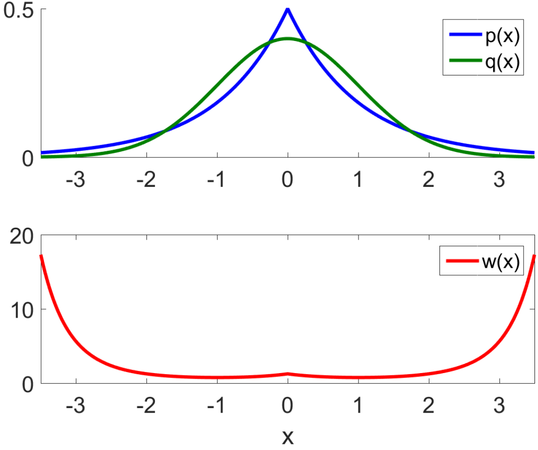
$$\mathbb{E}[\hat{\theta}_n^{\mathsf{IS}}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\phi(Y_i)w(Y_i))$$
$$= \mathbb{E}(\phi(Y_1)w(Y_1))$$
$$= \mathbb{E}(\phi(X)) = \theta$$

Let $Z_i = \phi(Y_i)w(Y_i)$. $Z_1, \ldots, Z_n$ are iid with mean $\mathbb{E}(Z_i) = \mathbb{E}(\phi(Y_i)w(Y_i)) = \theta$. From the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \to \theta \quad \text{a.s. as } n \to \infty$$

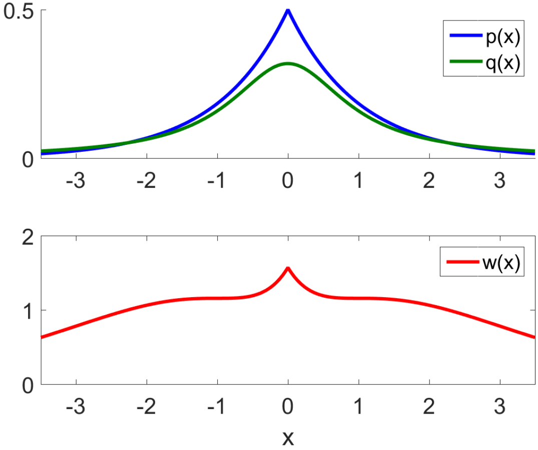
Target and Proposal Distributions

- Target: $p(x) = \frac{1}{2}e^{-|x|}$
- Proposal: $q(x) = 1/\sqrt{2\pi}\ e^{-\frac{x^2}{2}}$
- Weight function: $w(x) = \sqrt{\pi/2}\ e^{-|x|+\frac{x^2}{2}}$



Target and Proposal Distributions

- Target: $p(x) = \frac{1}{2}e^{-|x|}$
- Proposal: $q(x) = 1/(\pi(1 + x^2))$
- Weight function: $w(x) = \pi/2\ (1 + x^2)\ e^{-|x|}$



Example: Gamma Distribution

- Say we have simulated $Y_i \sim$Gamma(a, b) and we want to estimate $\mathbb{E}_p(\phi(X))$ where $X \sim$Gamma(α, β).
- Recall that the Gamma(α, β) density is

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

so

$$w(x) = \frac{p(x)}{q(x)} = \frac{\Gamma(a)\beta^\alpha}{\Gamma(\alpha)b^a} x^{\alpha-a} e^{-(\beta-b)x}$$

- Hence

$$\hat{\theta}_n^{\mathsf{IS}} = \frac{\Gamma(a)\beta^\alpha}{\Gamma(\alpha)b^a} \; \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) \; Y_i^{\alpha-a} e^{-(\beta-b)Y_i}$$

is an unbiased and consistent estimate of $\mathbb{E}_p(\phi(X))$.

Variance of the Importance Sampling Estimator

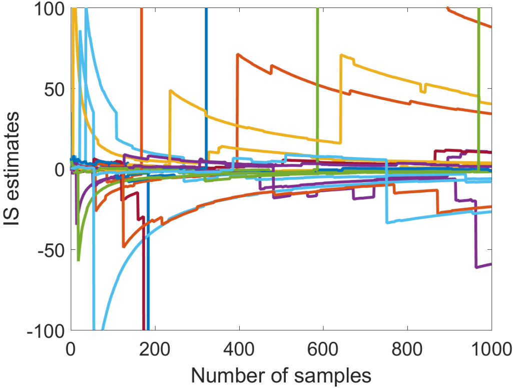
- **Proposition**. Assume $\theta = \mathbb{E}_p(\phi(X))$ and $\mathbb{E}_p(w(X)\phi^2(X))$ are finite. Then $\hat{\theta}_n^{\mathsf{IS}}$ satisfies

$$
\begin{aligned}
&\mathbb{E}\left(\left(\hat{\theta}_n^{\mathsf{IS}} - \theta\right)^2\right) = \mathbb{V}\left(\hat{\theta}_n^{\mathsf{IS}}\right) = \frac{1}{n}\mathbb{V}_q\left(w(Y_1)\phi(Y_1)\right) \\
&= \frac{1}{n}\left(\mathbb{E}_q\left(\frac{p^2(Y_1)}{q^2(Y_1)}\phi^2(Y_1)\right) - \mathbb{E}_q\left(\frac{p(Y_1)}{q(Y_1)}\phi(Y_1)\right)^2\right) \\
&= \frac{1}{n}\left(\mathbb{E}_p\left(w(X)\phi^2(X)\right) - \theta^2\right).
\end{aligned}
$$

- Each time we do IS we should check that this variance is finite, otherwise our estimates are somewhat untrustworthy! We check $\mathbb{E}_p(w(X)\phi^2(X))$ is finite.

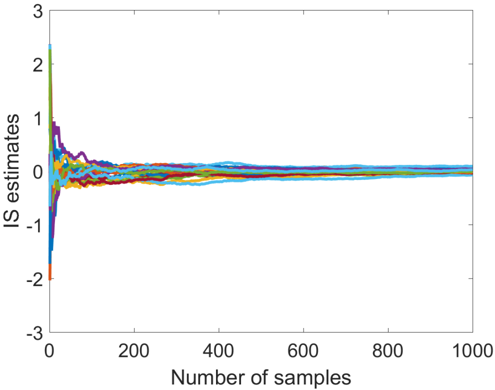
Variance of the Importance Sampling Estimator

- Target: $p(x) = \frac{1}{2}e^{-|x|}$
- Proposal: $q(x) = 1/\sqrt{2\pi}\ e^{-\frac{x^2}{2}}$
- $w(x) = \sqrt{\pi/2}\ e^{-|x|+\frac{x^2}{2}}$, $\phi(x) = x$
- $\mathbb{E}_p(w(X)\phi^2(X)) = \infty$



Variance of the Importance Sampling Estimator

- Target: $p(x) = \frac{1}{2}e^{-|x|}$
- Proposal: $q(x) = 1/(\pi(1+x^2))$
- $w(x) = \pi/2\ (1+x^2)\ e^{-|x|}$, $\phi(x) = x$
- $\mathbb{E}_p(w(X)\phi^2(X)) < \infty$



Variance of the Importance Sampling Estimator

- If $\mathbb{V}_p(\phi(X))$ is finite, a sufficient condition is that w is a bounded function: there is M such that $w(x) = \frac{p(x)}{q(x)} \leq M$ for all $x \in \Omega$
- Note that this is the same condition as for rejection sampling,
- For IS it is enough just for M to exist—we do not have to work out its value.
- Proof:

$$\begin{aligned}\mathbb{E}_p(w(X)\phi^2(X)) &\leq M\mathbb{E}_p(\phi^2(X)) \\ &< \infty\end{aligned}$$

as $\mathbb{V}_p(\phi(X)) < \infty$.

Example: Gamma Distribution

- Let us check that the variance of $\hat{\theta}_n^{\mathsf{IS}}$ in previous Example is finite if $\theta = \mathbb{E}_p(\phi(X))$ and $\mathbb{V}_p(\phi(X))$ are finite.
- It is enough to check that $\mathbb{E}_p\left(w(Y_1)\phi^2(Y_1)\right)$ is finite.
- The normalisation constants are finite so we can ignore those, and begin with

$$w(x)\phi^2(x) \propto x^{\alpha-a}e^{-(\beta-b)X}\phi^2(x).$$

- The expectation of interest is

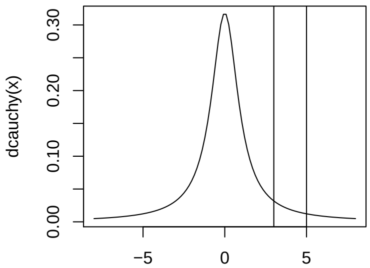
$$\begin{aligned}
\mathbb{E}_p\left(w(X)\phi^2(X)\right) &\propto \mathbb{E}_p\left(X^{\alpha-a}e^{-(\beta-b)X}\phi^2(X)\right) \\
&= \int_0^\infty p(x)\, x^{\alpha-a}\exp(-(\beta-b)x))\phi^2(x)\, dx \\
&\leq M\int_0^\infty p(x)\phi(x)^2\, dx = M\mathbb{E}_p(\phi^2(X)).
\end{aligned}$$

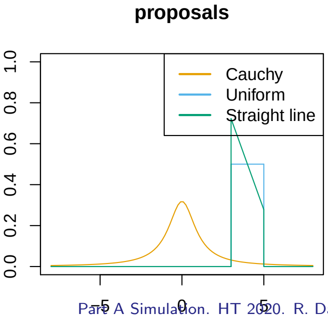
where $M = \max_{x>0} x^{\alpha-a}\exp(-(\beta-b)x)$ is finite if $a < \alpha$ and $b < \beta$ (see rejection sampling section).

- Since $\theta = \mathbb{E}_p(\phi(X))$ and $\mathbb{V}_p(\phi(X))$ are finite, we have $\mathbb{E}_p(\phi^2(X)) < \infty$ if these conditions on a, b are satisfied. If not, we cannot conclude as it depends on ϕ.
- These same (sufficient) conditions apply to our rejection sampler for Gamma(α, β).

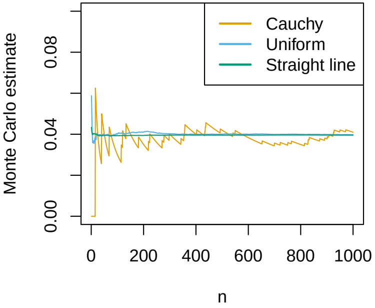
Estimate probability random variable is in a range

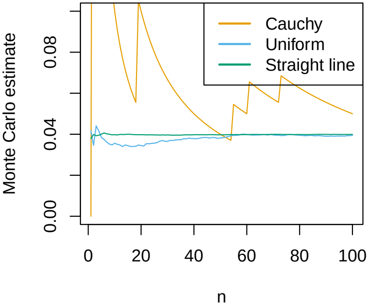
- Let X be a random variable from a Cauchy distribution with $f_X(x) = \frac{1}{\pi(1-x^2)}$
- Consider that we are interested in $P(X \in [3,5])$, *i.e.* that $\phi(X)$ is an indicator function
- We can calculate this analytically to yield $\frac{1}{\pi}(\tan^{-1}(5) - \tan^{-1}(3))$
- Let's consider the performance of three different estimators: one using normal Monte Carlo, and two using Importance sampling, with different simple proposals





Proposals that looks like the target times ϕ look promising





- Here we can see importance sampling is substantially more efficient than regular Monte Carlo for rare events
- What's more, we see as we try to make our proposal more similar to our target, we are becoming more efficient
- This is true more generally for trying to make our proposal look like the product of the target by ϕ, as we see in the next slides

Choice of the Importance Sampling Distribution

- While p is given, q needs to cover $p\phi$ (i.e. $p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0$) and be simple to sample.
- The requirement $\mathbb{V}\left(\hat{\theta}_n^{\mathsf{IS}}\right) < \infty$ further constrains our choice: we need $\mathbb{E}_p\left(w(X)\phi^2(X)\right) < \infty$.
- If $\mathbb{V}_p(\phi(X))$ is known finite then, it may be easy to get a sufficient condition for $\mathbb{E}_p\left(w(X)\phi^2(X)\right) < \infty$; e.g. $w(x) \leq M$. Further analysis will depend on ϕ.

Choice of the Importance Sampling Distribution

- What is the choice q_{opt} of q that actually minimizes the variance of the IS estimator? Consider for now $\phi : \Omega \to [0, \infty)$ then

$$q_{\text{opt}}(x) = \frac{p(x)\phi(x)}{\mathbb{E}_p(\phi(X))} \Rightarrow \mathbb{V}(\hat{\theta}_n^{\text{IS}}) = 0.$$

- This optimal zero-variance estimator cannot be implemented as

$$w(x) = p(x)/q_{\text{opt}}(x) = \mathbb{E}_p(\phi(X))/\phi(x)$$

where $\mathbb{E}_p(\phi(X))$ is the quantity we are trying to estimate! This can however be used as a guideline to select q.

Choice of the Importance Sampling Distribution

- For general function $\phi : \Omega \to \mathbb{R}$, the optimal importance distribution is

$$q_{\mathsf{opt}}(x) = \frac{p(x)|\phi(x)|}{\mathbb{E}_p\left(|\phi(X)|\right)}$$

with variance

$$\mathbb{V}(\hat{\theta}_n^{\mathsf{IS}}) = \frac{1}{n}\left(\mathbb{E}_p\left(|\phi(X)|\right)^2 - \theta^2\right).$$

Choice of the Importance Sampling Distribution

- Proof:

$$\begin{aligned}
\mathbb{E}_p\left(w(X)\phi^2(X)\right) &= \mathbb{E}_q\left(\frac{p^2(Y_1)}{q^2(Y_1)}\phi^2(Y_1)\right) \\
&= \mathbb{V}_q\left(\frac{p(Y_1)}{q(Y_1)}|\phi(Y_1)|\right) + \left(\mathbb{E}_q\left(\frac{p(Y_1)}{q(Y_1)}|\phi(Y_1)|\right)\right)^2 \\
&\geq \left(\mathbb{E}_q\left(\frac{p(Y_1)}{q(Y_1)}|\phi(Y_1)|\right)\right)^2 \\
&= \left(\mathbb{E}_p\left(|\phi(X)|\right)\right)^2
\end{aligned}$$

where the lower bound does not depend on q. This lower bound is achieved for $q = q_{\text{opt}}$

$$\mathbb{E}_p\left(\frac{p(X)}{q_{\text{opt}}(X)}\phi^2(X)\right) = \left(\mathbb{E}_p\left(|\phi(X)|\right)\right)^2$$

Recap

- Importance sampling uses draws from a proposal and re-weights them according to a weight that is the ratio of the target and proposal pdfs/pmfs
- It is unbiased and consistent
- Importance sampling can be used in place of procedures like rejection sampling when direct sampling is difficult
- Intelligent choice of the proposal can lead to Monte Carlo estimators with lower variance as a function of n, making them more efficient and generally more desirable