Part A Simulation and Statistical Programming HT14

Lecturer: Geoff Nicholls

University of Oxford

Lecture 6: Importance sampling

Notes and Problem sheets are available at

www.stats.ox.ac.uk\ ~nicholls\PartASSP

Importance Sampling Estimator

Slight revision on usual story: we can sample $Y \sim q, Y \in \Omega$. We want to estimate $\theta = E_p(\phi(X))$ where $X \sim p, X \in \Omega$ and ϕ is some given function $\phi : \Omega \longrightarrow \Re$.

Idea: simulate $Y1, Y_2, Y_3, ..., Y_n \sim q$ iid and form the weighted average

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

with $w(Y_i) = p(Y_i)/q(Y_i)$.

Proposition: If $p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0$ and $\mathbb{E}_p(\phi(X))$ exists then $\hat{\theta}_n^{\text{IS}}$ is unbiased and consistent.

Proposition: If $p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0$ and $\mathbb{E}_p(\phi(X))$ exists then $\hat{\theta}_n^{\text{IS}}$ is unbiased and consistent.

Proof. Unbiasedness:

$$\begin{aligned} E_q(\hat{\theta}_n^{\text{IS}}) &= \frac{1}{n}\sum_{i=1}^{n} E_q\left(\phi(Y_i)\frac{p(Y_i)}{q(Y_i)}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_\Omega \phi(y_i)\frac{p(y_i)}{q(y_i)}q(y_i)dy_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_\Omega \phi(x_i)p(x_i)dx_i \\ &= \frac{1}{n}\sum_{i=1}^{n} E_p(\phi(X)) \\ &= E_p(\phi(X)) \end{aligned}$$

so $\hat{\theta}_n^{\text{IS}}$ is unbiased.

Proof continued. Consistency: show that for each $\epsilon > 0$,

$$\Pr(|\hat{\theta}_n^{\mathsf{IS}} - \theta)| \geq \epsilon) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty.$$

By the Markov inequality for rv $Z \geq 0$, $\Pr(Z \geq a) \leq E(Z)/a$.

$$\begin{aligned}
\Pr(|\hat{\theta}_n^{\mathsf{IS}} - \theta| \geq \epsilon) &= \Pr(|\hat{\theta}_n^{\mathsf{IS}} - \theta|^2 \geq \epsilon^2) \\
&\leq \frac{E_q(|\hat{\theta}_n^{\mathsf{IS}} - \theta|^2)}{\epsilon^2} \\
&= \frac{\mathrm{var}_q(\hat{\theta}_n^{\mathsf{IS}})}{\epsilon^2} \\
&= \frac{\mathrm{var}_q\left(\frac{1}{n}\sum_{i=1}^{n} \phi(Y_i)\frac{p(Y_i)}{q(Y_i)}\right)}{\epsilon^2} \\
&= \frac{\mathrm{var}_q\left(\phi(Y)\frac{p(Y)}{q(Y)}\right)}{n\epsilon}
\end{aligned}$$

so the probability for a large error tends to zero as $n \longrightarrow \infty$.

Example: Gamma Distribution

Earlier on we used the transformation method to simulate

$$Y \sim \Gamma(a, b)$$

for $a = 1, 2, 3, ...$ and $b > 0$ by summing exponentials. Suppose we have simulated $Y_i, i = 1, 2, ..., n$ iid $\Gamma(a, b)$ rv, but want to estimate the expectation of $\phi(X)$ in some rv

$$X \sim \Gamma(\alpha, \beta)$$

for some $\alpha, \beta > 0$.

The Gamma(α, β) density is

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

so

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

Hence

$$\begin{aligned}\hat{\theta}_n^{\text{IS}} &= \frac{1}{n}\sum_{i=1}^{n} \phi(Y_i) w(Y_i) \\ &= \frac{\Gamma(a)\beta^\alpha}{\Gamma(\alpha)b^a} \frac{1}{n}\sum_{i=1}^{n} \phi(Y_i)\ Y_i^{\alpha-a} \exp(-(\beta-b)Y_i)\end{aligned}$$

is an unbiased and consistent estimate of $E_p(\phi(X))$. We can actually "recycle" the Y's and compute $E_{\alpha,\beta}(\phi(X))$ for lots of α's and β's.

So far so good.

Variance of the Importance Sampling Estimator

Proposition: If $\theta = E_p(\phi(X))$ and $E_p(w(X)\phi^2(X))$ are finite then

$$\mathrm{var}_q(\hat{\theta}_n^{\mathrm{IS}}) = \frac{1}{n}\left(\mathbb{E}_p\left(w(X)\phi^2(X)\right) - \theta^2\right).$$

Each time we do IS we should check that this variance is finite (and ideally small), otherwise our estimates have infinite variance and are somewhat untrustworthy! We check $E_p(w\phi^2)$ is finite.

How can we show $E_p(w\phi^2)$ is finite? We often know that $\phi(X)$ has finite mean and variance. That means $E_p(\phi^2)$ must be finite.

If $w(x)$ is bounded $w(x) \leq M$ for all $x \in \Omega$ then

$$E_p(w\phi^2) \leq M E_p(\phi^2) \leq \infty.$$

But that is just the same condition we needed for rejection,

$$p(x)/q(x) \leq M \qquad \text{for all } x \in \Omega$$

for some M (at least here we only have to show M exists).

However, it may be that $w(x)$ is not bounded, but $E_p(w\phi^2)$ is finite (if for example $\phi(x)$ gets small when $w(x)$ gets big). Importance sampling has a wider domain of application than rejection. It is also statistically more efficient (hardish proof - lecturer's prize if you can show this).

Proof:

$$\begin{aligned}
\mathrm{var}_q(\hat{\theta}_n^{\mathrm{IS}}) &= \mathrm{var}_q\left(\frac{1}{n}\sum_{i=1}^{n}\phi(Y_i)w(Y_i)\right) \\
&= \frac{1}{n}\mathrm{var}_q\left(\phi(Y_1)w(Y_1)\right) \\
&= \frac{1}{n}\left(E_q\left(w(Y_1)^2\phi(Y_1)^2\right) - E_q\left(w(Y_1)\phi(Y_1)\right)^2\right).
\end{aligned}$$

The second expectation is $E_q(\phi(Y_1)p(Y_1)/q(Y_1)) = \theta$ as we saw earlier. The first expectation can also be converted into an expectation in $X \sim p$.

$$\begin{aligned} E_q\left(w(Y_1)^2\phi(Y_1)^2\right) &= \int_\Omega \frac{p(y)^2}{q(y)^2}\phi(y)^2 q(y)dy \\ &= \int_\Omega \frac{p(y)}{q(y)}\phi(y)^2 p(y)dy \\ &= E_p\left(w(X)\phi(X)^2\right) \end{aligned}$$

and hence

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

Example: Gamma Distribution (continued)

Check that the variance of of our IS-estimator $\hat{\theta}_n^{\text{IS}}$ for the Gamma dbn is finite. I will assume $E_p(\phi)$ and $\text{var}_p(\phi)$ are finite.

We need sufficient conditions for $\mathbb{E}_p\left(w(Y)\phi(Y)^2\right)$ to be finite.

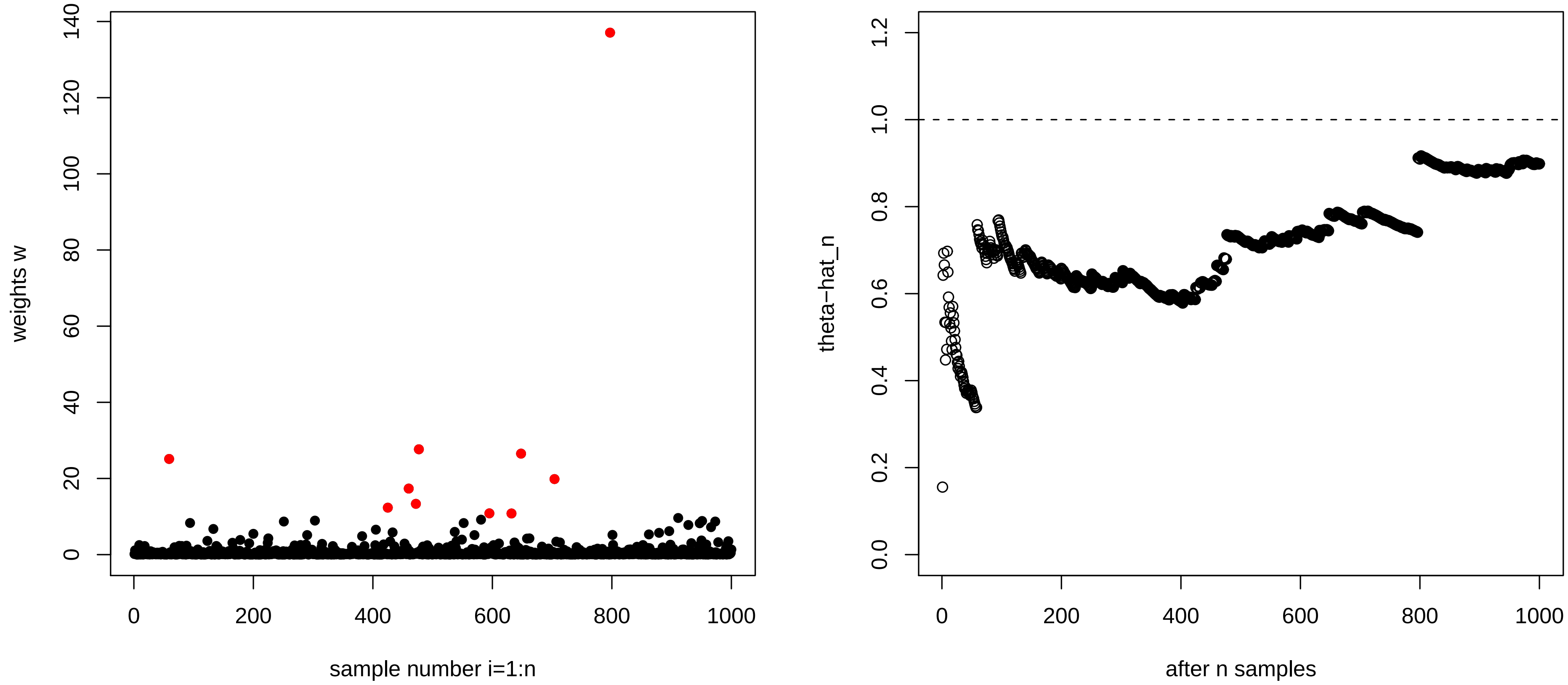
$$\begin{aligned} w(x)\phi(x)^2 &= \frac{\Gamma(x;\alpha,\beta)}{\Gamma(x;a,b)}\phi(x)^2 \\ &= \frac{\Gamma(a)\beta^\alpha}{\Gamma(\alpha)b^a}x^{\alpha-a}\exp(-(\beta-b)x)\phi(x)^2, \end{aligned}$$

so

$$\begin{aligned} E_p\left(w(X)\phi(X)^2\right) &\propto E_p\left(X^{\alpha-a}\exp(-(\beta-b)X)\phi(X)^2\right) \\ &= \int_0^\infty p(x)\, x^{\alpha-a}\exp(-(\beta-b)x)\phi(x)^2\, dx \end{aligned}$$

$x^{\alpha-a}\exp(-(\beta-b)x)$ bounded iff $\alpha > a$ and $\beta > b$. Unless $\phi(x)$ saves us, $\text{var}(\hat{\theta}_n^{\text{IS}}) = \infty$ when this condition is not satisfied.

Try $a = 2, b = 2$ and $\beta = 2.5$, $\alpha = 0.5$ (*ie* α less than a) and $\phi(x) = 1$. Monitor the weights $w(y_i)$ and the sequence of estimates $\hat{\theta}_m^{\mathrm{IS}}$, $m = 1, 2, ...n$.



The estimator is hit by occasional huge weights.

Exercise: What would happen if we used $\phi(x) = x$?

Rare Event Estimation and variance reduction

One important class of applications of IS is to problems in which we estimate the probability for a rare event. In such scenarios, we may be able to sample from p directly but this doesn't help.

For example, suppose $X \sim p$ and we want to estimate

$$P(X > x_0) = E_p\left(\mathbb{I}[X > x_0]\right)$$

with x_0 in the extreme upper tail of $p(x)$. We may not get any samples $X_i > x_0$ and the usual estimate

$$\hat{\theta}_n = \sum_i \mathbb{I}(X_i > x_0)/n$$

is simply zero. We can take a q-dbn that puts more probability at large Y, and then reweight to get expectations in X. By using IS, we can actually reduce the variance of our estimator.

Example

Say $p(x) = N(x; \mu, \sigma^2)$ and we want to estimate $\theta = \mathbb{P}(X > x_0)$ for some $x_0 \gg \mu + 3\sigma$.

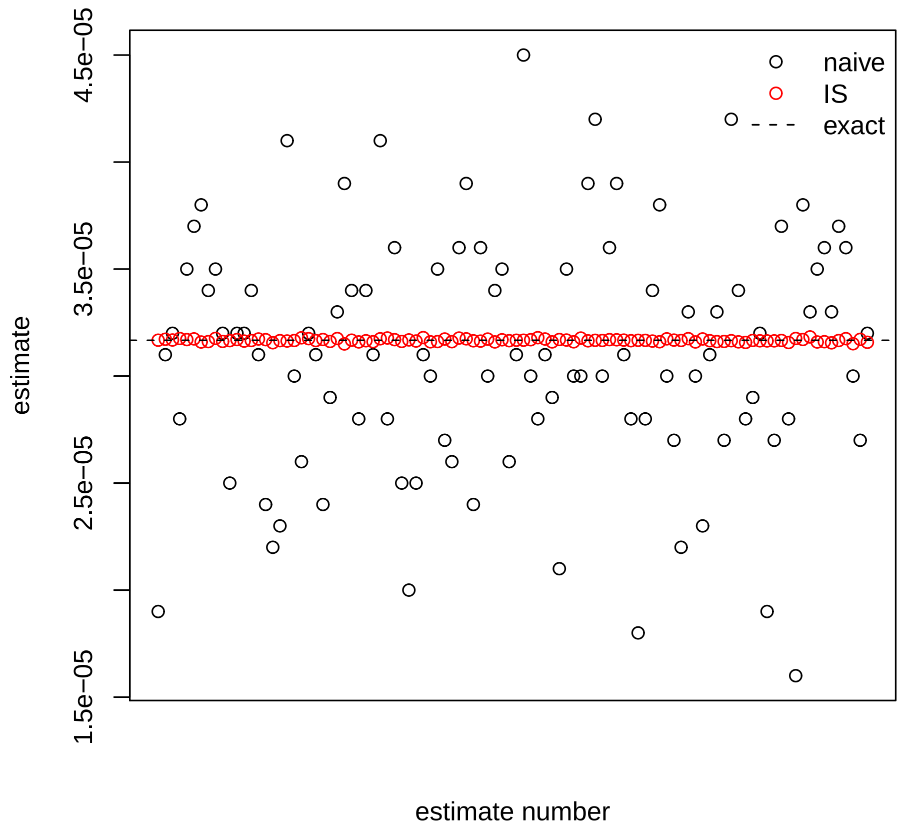
Take q to be some simple distribution that sits over x_0. A natural choice is $q(y) = N(y; x_0, \sigma^2)$.

The weights $w = p/q$ are

$$\begin{aligned} w(y) &= \frac{N(y; \mu, \sigma^2)}{N(y; x_0, \sigma^2)} \\ &= \exp(-(y-\mu)^2/2\sigma^2 + (y - x_0)^2/2\sigma^2) \end{aligned}$$

and the IS estimator is $\hat{\theta}_n^{\text{IS}} = \frac{1}{n}\sum_{i=1}^n w(Y_i)\mathbb{I}_{Y_i > x_0}$.

The variance reduction can be dramatic. Here are 100 estimates of $\Pr(X > 4)$ for $X \sim N(0, 1)$ using $q(y) = N(y; 4, 1)$.



Unnormalized Importance sampling

Recall $p(x) = \tilde{p}(x)/Z_p$, $q(x) = \tilde{q}(x)/Z_q$ with Z_p, Z_q commonly intractable.

Same issue as for rejection. The IS weights are $w = p/q$ so need q and p normalized.

Let $\tilde{w} = \tilde{p}/\tilde{p}$. If we use $\frac{1}{n}\sum_{i=1}^{n} \tilde{w}(Y_i)\phi(Y_i)$ then we find

$$
\begin{aligned}
E_q\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{p}(Y_i)}{\tilde{q}(Y_i)}\phi(Y_i)\right) &= E_q\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Z_p}{Z_q}\frac{p(Y_i)}{q(Y_i)}\phi(Y_i)\right) \\
&= \frac{Z_p}{Z_q}E_p(\phi(X)).
\end{aligned}
$$

We need to estimate Z_p/Z_q and divide. $\frac{1}{n}\sum_{i=1}^{n}\tilde{w}(Y_i)$ is the estimator we need.

$$
\begin{aligned}
E_q\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{p}(Y_i)}{\tilde{q}(Y_i)}\right) &= E_q\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Z_p}{Z_q}\frac{p(Y_i)}{q(Y_i)}\right) \\
&= \frac{Z_p}{Z_q}E_q\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p(Y_i)}{q(Y_i)}\right) \\
&= Z_p/Z_q
\end{aligned}
$$

since $\sum_{i=1}^{n} w(Y_i)/n$ is the IS estimator for $\phi = 1$. We will see next week that indeed

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

is consistent for $E_p(\phi(X))$.

Example: we saw that if $Y_i \sim \Gamma(a, b)$ and

$$w(y) = \frac{\Gamma(a)\beta^{\alpha}}{\Gamma(\alpha)b^{a}} y^{\alpha - a} \exp(-(\beta - b)y)$$

then

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

is unbiased and consistent for $E_p(\phi(X))$ with $X \sim \Gamma(\alpha, \beta)$. From above, if

$$\tilde{w}(y) = y^{\alpha - a} \exp(-(\beta - b)y)$$

then

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

is a consistent estimator for $E_p(\phi(X))$.