Numerical Methods II

Prof. Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Variance Reduction

Monte Carlo starts as a very simple method; much of the complexity in practice comes from trying to reduce the variance, to reduce the number of samples that have to be simulated to achieve a given accuracy.

- antithetic variables
- control variates
- importance sampling
- stratified sampling (lecture 4)
- Latin hypercube (lecture 4)
- quasi-Monte Carlo (lecture 5)

Review of elementary results

If a, b are random variables, and λ, μ are constants, then

$$\begin{aligned}
\mathbb{E}[a+\mu] &= \mathbb{E}[a]+\mu \\
\mathbb{V}[a+\mu] &= \mathbb{V}[a] \\
\mathbb{E}[\lambda\, a] &= \lambda\, \mathbb{E}[a] \\
\mathbb{V}[\lambda\, a] &= \lambda^2\, \mathbb{V}[a] \\
\mathbb{E}[a+b] &= \mathbb{E}[a]+\mathbb{E}[b] \\
\mathbb{V}[a+b] &= \mathbb{V}[a]+2\,\text{Cov}[a,b]+\mathbb{V}[b]
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{V}[a] &\equiv \mathbb{E}\left[(a-\mathbb{E}[a])^2\right] = \mathbb{E}\left[a^2\right]-(\mathbb{E}[a])^2 \\
\text{Cov}[a,b] &\equiv \mathbb{E}\Big[(a-\mathbb{E}[a])\,(b-\mathbb{E}[b])\Big]
\end{aligned}$$

Review of elementary results

If a, b are independent random variables then

$$\mathbb{E}[f(a)\; g(b)] = \mathbb{E}[f(a)]\; \mathbb{E}[g(b)]$$

Hence, Cov$[a, b] = 0$ and therefore $\mathbb{V}[a+b] = \mathbb{V}[a] + \mathbb{V}[b]$

Extending this to a set of N iid (independent identically distributed) r.v.'s x_n, we have

$$\mathbb{V}\left[\sum_{n=1}^{N} x_n\right] = \sum_{n=1}^{N} \mathbb{V}[x_n] = N\, \mathbb{V}[x]$$

and so

$$\mathbb{V}\left[N^{-1} \sum_{n=1}^{N} x_n\right] = N^{-1}\, \mathbb{V}[x]$$

Antithetic variables

The simple estimator from the last lecture has the form

$$N^{-1} \sum_i f(W^{(i)})$$

where $W^{(i)}$ is the value of the random Weiner variable $W(T)$ at maturity.

$W(T)$ has a symmetric probability distribution so $-W(T)$ is just as likely.

Antithetic variables

Antithetic estimator replaces $f(W^{(i)})$ by

$$\overline{f}^{(i)} = \tfrac{1}{2}\left(f(W^{(i)}) + f(-W^{(i)}) \right)$$

Clearly still unbiased since

$$\mathbb{E}[\overline{f}] = \tfrac{1}{2}\left(\mathbb{E}[f(W)] + \mathbb{E}[f(-W)] \right) = \mathbb{E}[f(W)]$$

The variance is given by

$$\begin{aligned}
\mathbb{V}[\overline{f}] &= \tfrac{1}{4}\left(\mathbb{V}[f(W)] + 2\,\mathsf{Cov}[f(W), f(-W)] + \mathbb{V}[f(-W)] \right) \\
&= \tfrac{1}{2}\left(\mathbb{V}[f(W)] + \mathsf{Cov}[f(W), f(-W)] \right)
\end{aligned}$$

Antithetic variables

The variance is always reduced, but the cost is almost doubled, so net benefit only if Cov$[f(W), f(-W)] < 0$.

Two extremes:

- A linear payoff, $f = a + b\,W$, is integrated exactly since $\overline{f} = a$ and Cov$[f(W), f(-W)] = -\mathbb{V}[f]$
- A symmetric payoff $f(W) = f(-W)$ is the worst case since Cov$[f(W), f(-W)] = \mathbb{V}[f]$

General assessment – usually not very helpful, but can be good in particular cases where the payoff is nearly linear

Control Variates

Suppose we want to approximate $\mathbb{E}[f]$ using a simple Monte Carlo average $\overline{f}$.

If there is another payoff g for which we know $\mathbb{E}[g]$, can use $\overline{g} - \mathbb{E}[g]$ to reduce error in $\overline{f} - \mathbb{E}[f]$.

How? By defining a new estimator

$$\widehat{f} = \overline{f} - \lambda\,(\overline{g} - \mathbb{E}[g])$$

Again unbiased since $\mathbb{E}[\widehat{f}] = \mathbb{E}[\overline{f}] = \mathbb{E}[f]$

Control Variates

For a single sample,

$$\mathbb{V}[f - \lambda\,(g - \mathbb{E}[g])] = \mathbb{V}[f] - 2\,\lambda\,\textsf{Cov}[f,g] + \lambda^2\,\mathbb{V}[g]$$

For an average of N samples,

$$\mathbb{V}[\overline{f} - \lambda\,(\overline{g} - \mathbb{E}[g])] = N^{-1}\left(\mathbb{V}[f] - 2\,\lambda\,\textsf{Cov}[f,g] + \lambda^2\,\mathbb{V}[g]\right)$$

To minimise this, the optimum value for λ is

$$\lambda = \frac{\textsf{Cov}[f,g]}{\mathbb{V}[g]}$$

Control Variates

The resulting variance is

$$N^{-1}\,\mathbb{V}[f]\left(1-\frac{(\textsf{Cov}[f,g])^2}{\mathbb{V}[f]\,\mathbb{V}[g]}\right)=N^{-1}\,\mathbb{V}[f]\left(1-\rho^2\right)$$

where ρ is the correlation between f and g.

The challenge is to choose a good g which is well correlated with f – the covariance, and hence the optimal λ, can be estimated from the data.

Control Variates

Possible choices:

- for European call option (ignoring its known value) could use $g = S$ since $\exp(-rt)\, S(t)$ is a martingale:

$$\mathbb{E}[S(T)] = \exp(rT)\; S(0)$$

- for a general European payoff $f(S)$ could use a combination of put and call options

More opportunities for path-dependent options – will discuss next term. The idea can also be taken further using multiple control variates.

General assessment – can be very effective, depending on the application

Application

MATLAB code, part 1 – estimating optimal λ:

```
r=0.05; sig=0.2; T=1; S0=110; K=100;


N = 1000;
U = rand(1,N);              % uniform random variable
Y = norminv(U);             % inverts Normal cum. fn.
S = S0*exp((r-sig^2/2)*T + sig*sqrt(T)*Y);
F = exp(-r*T)*max(0,S-K);
C = exp(-r*T)*S;
Fave = sum(F)/N;
Cave = sum(C)/N;
lam = sum((F-Fave).*(C-Cave)) / sum((C-Cave).^2);
```

Application

MATLAB code, part 2 – control variate estimation:

```
N  = 1e5;
U  = rand(1,N);             % uniform random variable
Y  = norminv(U);            % inverts Normal cum. fn.
S  = S0*exp((r-sig^2/2)*T + sig*sqrt(T)*Y);
F  = exp(-r*T)*max(0,S-K);
C  = exp(-r*T)*S;
F2 = F - lam*(C-S0);

Fave   = sum(F)/N;
F2ave  = sum(F2)/N;
sd     = sqrt( sum((F -Fave ).^2)/(N*(N-1)) )
sd2    = sqrt( sum((F2-F2ave).^2)/(N*(N-1)) )
```

Application

Results:

```
>>> lec3

sd =
    0.0599



sd2 =
    0.0151

est. price (no  CV)  = 17.676995 +/- 0.179683
est. price (with CV) = 17.659708 +/- 0.045310
exact price          = 17.662954
```

Importance Sampling

Importance sampling involves a change of probability measure. Instead of taking X from a distribution with p.d.f. $p_1(X)$, we instead take it from a different distribution with p.d.f. $p_2(X)$.

$$\begin{aligned}\mathbb{E}_1[f(X)] &= \int f(X)\, p_1(X)\, \mathrm{d}X \\ &= \int f(X)\, \frac{p_1(X)}{p_2(X)}\, p_2(X)\, \mathrm{d}X \\ &= \mathbb{E}_2[f(X)\; R(X)]\end{aligned}$$

where $R(X) = p_1(X)/p_2(X)$ is the Radon-Nikodym derivative.

Importance Sampling

We want the new variance $\mathbb{V}_2[f(X)\ R(X)]$ to be smaller than the old variance $\mathbb{V}_1[f(X)]$.

How do we achieve this? Ideal is to make $f(X)R(X)$ constant, so its variance is zero.

More practically, make $R(X)$ small where $f(X)$ is large, and make $R(X)$ large where $f(X)$ is small.

Small $R(X) \Longleftrightarrow$ large $p_2(X)$ relative to $p_1(X)$, so more random samples in region where $f(X)$ is large.

Particularly important for rare event simulation where $f(X)$ is zero almost everywhere.

Importance Sampling

Really simple rare event example: suppose random variable X takes value 1 with probability $\delta \ll 1$ and is otherwise 0.

$$\mathbb{E}[X] = \delta$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \delta - \delta^2$$

Hence,

$$\frac{\sqrt{\mathbb{V}[X]}}{\mathbb{E}[X]} = \sqrt{\frac{1-\delta}{\delta}} \approx \sqrt{\frac{1}{\delta}}$$

If we want the relative error to be less than ε, the number of samples required is $O(\varepsilon^{-2}\delta^{-1})$.

Importance Sampling

Digital put option:

$$P = \exp(-rT)\; H(K - S(T)) = \exp(-rT)\; H(\log K - \log S(T))$$

where

$$X = \log S(T) = \log S(0) + (r-\tfrac{1}{2}\sigma^2)\, T + \sigma\, W(T)$$

is Normally distributed with p.d.f.

$$\phi_1(X) = \frac{1}{\sqrt{2\pi\sigma^2 T}}\; \exp\left(\frac{-(x-\mu)^2}{2\sigma^2 T}\right)$$

with $\mu = \log S(0) + (r-\tfrac{1}{2}\sigma^2)\, T$.

Importance Sampling

A digital put option with very low strike (e.g. $K = 0.4\ S(0)$) is sometimes used as a hedge for credit derivatives.

If the stock price falls that much, there is a strong possibility of credit default.

Problem: this is a rare event. The probability that $S(T) < K$ can be very low, maybe less than 1%, leading to a very high r.m.s. error relative to the true price.

Solution: importance sampling, adjusting either mean or volatility

Importance Sampling

Approach 1: change the mean from μ_1 to $\mu_2 < \mu_1$ by using

$$X = \mu_2 + \sigma\, W(T)$$

The Radon-Nikodym derivative is

$$\begin{aligned} R(X) &= \exp\left(\frac{-(X-\mu_1)^2}{2\sigma^2 T}\right) \Big/ \exp\left(\frac{-(X-\mu_2)^2}{2\sigma^2 T}\right) \\ &= \exp\left(\frac{(X-\frac{1}{2}(\mu_1+\mu_2))(\mu_1-\mu_2)}{\sigma^2 T}\right) \\ &> 1 \text{ for } X > \tfrac{1}{2}(\mu_1+\mu_2) \\ &< 1 \text{ for } X < \tfrac{1}{2}(\mu_1+\mu_2) \end{aligned}$$

Choosing $\mu_2 = \log K$ means half of samples are below $\log K$ with very small $R(X) \Longrightarrow$ large variance reduction

Importance Sampling

Approach 2: change the volatility from σ_1 to $\sigma_2 > \sigma_1$ by using

$$X = \mu + \sigma_2 \, W(T)$$

The Radon-Nikodym derivative is

$$\begin{aligned} R(X) &= \sigma_1^{-1} \exp\left(\frac{-(X\!-\!\mu)^2}{2\sigma_1^2 T} \right) / \sigma_2^{-1} \exp\left(\frac{-(X\!-\!\mu)^2}{2\sigma_2^2 T} \right) \\ &= \frac{\sigma_2}{\sigma_1} \exp\left(\frac{-(X\!-\!\mu)^2 (\sigma_2^2 \!-\! \sigma_1^2)}{2\sigma_1^2 \sigma_2^2 T} \right) \\ &> 1 \text{ for small } |X\!-\!\mu| \\ &\ll 1 \text{ for large } |X\!-\!\mu| \end{aligned}$$

This is good for applications where both tails are important
– not as good in this application.

Final Words

- antithetic variables – generic and easy to implement but limited effectiveness
- control variates – easy to implement and can be very effective but requires careful choice of control variate in each case
- importance sampling – very useful for applications with rare events, but needs to be fine-tuned for each application

Overall, a tradeoff between simplicity and generality on one hand, and efficiency and programming effort on the other.