Numerical Methods II
M. Giles

Problem sheet 1: solutions

1. Putting
$$r = \sqrt{-2\,\log y_1}, \qquad \theta = 2\pi\, y_2,$$
then
$$\frac{\partial r}{\partial y_1} = -\frac{1}{y_1 r} = -\frac{1}{r} e^{\frac{1}{2}r^2},$$
and hence
$$\det \frac{\partial(r,\theta)}{\partial(y_1,y_2)} = \det \left(\begin{array}{cc} -\frac{1}{r} e^{\frac{1}{2}r^2} & 0 \\ 0 & 2\pi \end{array} \right),$$
and so the probability density for r,θ is
$$p_{r,\theta} = \frac{r}{2\pi} e^{-\frac{1}{2}r^2}.$$
Next, putting
$$x_1 = r\cos\theta, \quad x_2 = r\sin\theta,$$
then
$$\det \frac{\partial(x_1,x_2)}{\partial(r,\theta)} = \det \left(\begin{array}{cc} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{array} \right) = r,$$
and so the joint probability density for
$$p_{x_1,x_2} = \frac{1}{2\pi}\, e^{-\frac{1}{2}r^2} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x_1^2} \;\; \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x_2^2}.$$
Because it has factored into a product of two unit Normal distributions, this completes the proof that x_1 and x_2 are independently distributed with a Normal distribution with zero mean and unit variance.

2. Because the different X_n are independent, we have
$$\mathbb{E}[X_m\, X_n] = \left\{ \begin{array}{ll} 0, & m \neq n \\ \mu_2, & m = n \end{array} \right.$$
where $\mu_2 = \mathbb{E}[X^2] = \mathbb{V}[X]$. Looking at the expression for $\widehat{\sigma}^2$, the only terms with a non-zero expectation are those involving products $X_m X_n$ for $m\!=\!n$. This gives
$$\mathbb{E}[\widehat{\sigma}^2] = \left(\frac{N}{N\!-\!1}\right) \left(N^{-1} N\, \mu_2 - N^{-2} N\, \mu_2\right) = \mu_2,$$
proving that $\widehat{\sigma}^2$ is an unbiased estimator.

(a) When X is a $N(0,1)$ random variable, $\mu_2 = 1$ and $\mu_4 = 3$ so

$$\mathbb{V}[\widehat{\sigma}^2] = 2N^{-1}.$$

If we assume that for large N, $\widehat{\sigma}^2$ has an approximately Normal distribution, then using a ± 3 standard deviations confidence interval, we want

$$3\sqrt{2N^{-1}} \;=\; 0.1 \quad\Longrightarrow\quad N = 1800.$$

(b) In this case, $\mu_2 = 2p$ and $\mu_4 = 2p$ so

$$\mathbb{V}[\widehat{\sigma}^2] \approx 2N^{-1}p.$$

Using a ± 3 standard deviations confidence interval, we want

$$3\sqrt{2N^{-1}p} \;\approx\; 0.2\,p, \quad\Longrightarrow\quad N \approx 450\,p^{-1}.$$

Note that since $p \ll 1$, it takes a very large number of samples to get an accurate estimate of the variance.

3. The variance of a single sample is

$$\begin{aligned} & \mathbb{V}\left[f - \lambda\,(g - \mathbb{E}[g]) - \gamma\,(h - \mathbb{E}[h])\right] \\ = \; & \mathbb{V}\left[f - \lambda g - \gamma h\right] \\ = \; & \mathbb{V}[f] + \lambda^2 \mathbb{V}[g] + \gamma^2 \mathbb{V}[h] - 2\lambda\;\mathrm{cov}(f,g) - 2\gamma\;\mathrm{cov}(f,h) + 2\lambda\gamma\;\mathrm{cov}(g,h) \end{aligned}$$

This quadratic function is a minimum when the derivatives with respect to each of λ and γ are zero, which requires that

$$\left(\begin{array}{cc} \mathbb{V}[g] & \mathrm{cov}(g,h) \\ \mathrm{cov}(g,h) & \mathbb{V}[h] \end{array}\right) \left(\begin{array}{c} \lambda \\ \gamma \end{array}\right) = \left(\begin{array}{c} \mathrm{cov}(f,g) \\ \mathrm{cov}(f,h) \end{array}\right).$$

The matrix above is singular when g and h are perfectly correlated; in this case h is proportional to g and so there is nothing to be gained by including h as a second control variate.

4. (a) Let $X = \varepsilon\, x$ so x is a standard zero mean unit variance Normal. Then

$$f(X) \approx f_0 + \varepsilon f_1 x + \tfrac{1}{2}\varepsilon^2 f_2\, x^2$$

where $f_0 = f(0)$, $\;f_1 = f'(0)$, $\;f_2 = f''(0)$, and hence the standard estimate

$$N^{-1} \sum_n f(X^{(n)})$$

has variance equal to

$$N^{-1}\mathbb{V}[f] \approx N^{-1}\varepsilon^2 f_1^2.$$

On the other hand,

$$\tfrac{1}{2}\left(f(X) + f(-X)\right) \approx f_0 + \tfrac{1}{2}\varepsilon^2 f_2 x^2$$

and so the antithetic variable estimate

$$N^{-1} \sum_n \tfrac{1}{2} \left(f(X^{(n)}) + f(-X^{(n)}) \right)$$

has variance equal to

$$N^{-1} \mathbb{V} \left[\tfrac{1}{2} \left(f(X) + f(-X) \right) \right] \approx \tfrac{1}{4} N^{-1} \varepsilon^4 f_2 \, \mathbb{V}[x^2].$$

(b) We start by noting that

$$\mathbb{E}[f(X)] = \mathbb{E}[g(X)] + f_0$$

Next, we observe that

$$g(X) \approx \tfrac{1}{2} \varepsilon^2 f_2 \, x^2$$

and so the main contribution to its expectation comes from the extreme values of x.

This suggests the use of importance sampling whereby instead of sampling x from $N(0,1)$ we instead take x from $N(0,\sigma^2)$ with $\sigma > 1$.

Introducing the Radon-Nikodym derivative

$$r(x) = \phi(x)/\phi_\sigma(x)$$

which is the ratio of the original and new probability density functions, the desired expectation is

$$\mathbb{E}[g(X)] = \mathbb{E}_\sigma[r(X/\varepsilon)\; g(X)].$$

The optimal choice for σ is the one which minimises

$$\mathbb{V}_\sigma[r(x)\; x^2].$$

It can be shown that this also minimises

$$\mathbb{E}[r(x)\; x^4].$$

5. (a) This payoff is one-sided, large when $X \gg 0$ and small when $X \ll 0$, so in this case it is best to change the mean to move the whole distribution towards larger X to better sample that tail of the distribution.

The Radon-Nikodym derivative is

$$r(x) = \exp\left(\tfrac{1}{2}(x-\mu)^2 - \tfrac{1}{2}x^2 \right) = \exp\left(-\mu x + \tfrac{1}{2}\mu^2 \right).$$

The variance of the new estimator is

$$\mathbb{V}_2[r(x) e^x] = \mathbb{E}_2[r^2(x) e^{2x}] - \left(\mathbb{E}_2[r(x) e^x] \right)^2$$

where the subscript on the expectation $\mathbb{E}_2$ denotes that it is with respect to the new, shifted distribution.

Since

$$\mathbb{E}_2[r(x)\, e^x] = \mathbb{E}_1[e^x]$$

changing μ does not affect this. Looking instead at the first term,

$$\begin{aligned}
\mathbb{E}_2[r^2(x)\, e^{2x}] &= \mathbb{E}_1[r(x)\, e^{2x}] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left((2-\mu)x + \tfrac{1}{2}\mu^2\right)\ \exp(-\tfrac{1}{2}x^2)\ \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}(x-(2-\mu))^2 + \tfrac{1}{2}\mu^2 + \tfrac{1}{2}(2-\mu)^2\right)\ \mathrm{d}x \\
&= \exp\left(\tfrac{1}{2}\mu^2 + \tfrac{1}{2}(2-\mu)^2\right) \\
&= \exp\left(\mu^2 - 2\mu + 2\right)
\end{aligned}$$

This is clearly a minimum when $\mu\!=\!1$.
In this case, $r(x) = \exp(\frac{1}{2} - x)$ and so $r(x)\, e^x = \exp(\frac{1}{2})$ which is constant.
Hence, the variance is actually zero in this highly unusual case.

(b) This payoff is two-sided, large in the tails where $|X|$ is large, so in this case it is best to change the increase the variance to get more samples in both tails of the distribution.

The Radon-Nikodym derivative is

$$r(x) = \sigma\ \exp\left(-\tfrac{1}{2}x^2 + \tfrac{1}{2\sigma^2}\, x^2\right).$$

The variance of the new estimator is

$$\mathbb{V}_2[r(x)x^4] = \mathbb{E}_2[r^2(x)x^8] - \left(\mathbb{E}_2[r(x)x^4]\right)^2.$$

Looking again at the first term,

$$\begin{aligned}
\mathbb{E}_2[r^2(x)\, x^8] &= \mathbb{E}_1[r(x)\, x^8] \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^8\ \exp\left(-x^2 + \tfrac{1}{2\sigma^2}\, x^2\right)\ \mathrm{d}x \\
&= \frac{\sigma}{\lambda^9}\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^8 \exp(-\tfrac{1}{2}y^2)\ \mathrm{d}y
\end{aligned}$$

using the substitution $y = \lambda x$ with $\lambda^2 = 2 - \sigma^{-2}$.

Using integration by parts, it is easily proved that for a unit Normal random variable y,

$$\mathbb{E}[y^{m+2}] = (m\!+\!1)\ \mathbb{E}[y^m]$$

and hence

$$\mathbb{E}_2[r^2(x)\, x^8] = \frac{105\,\sigma}{\lambda^9} = \frac{105\ \sigma}{(2\!-\!\sigma^{-2})^{9/2}}.$$

Differentiating this, it is found that the minimum is at $\sigma^2\!=\!5$, and the variance for this value is approximately 7.67, compared to the original value of $105 - 9 = 96$ without importance sampling, so in this case importance sampling reduces the variance, and hence the computational cost of Monte Carlo sampling, by a factor of approximately 12.5.

6. In this case we have for the j^{th} stratum,

$$f(U) = f(U_j) + f'(U_j)(U{-}U_j) + \tfrac{1}{2}f''(U_j)(U{-}U_j)^2 + \tfrac{1}{6}f'''(U_j)(U{-}U_j)^3 + O((U{-}U_j)^4).$$

Using the antithetic pair the linear and cubic terms cancel and we get

$$\tfrac{1}{2}(f(U) + f(U_{anti})) = f(U_j) + \tfrac{1}{2}f''(U_j)(U-U_j)^2 + O((U-U_j)^4).$$

and so

$$\mathbb{V}[\tfrac{1}{2}(f(U) + f(U_{anti}))] \;\approx\; \tfrac{1}{4}\left(f''(U_j)\right)^2 \mathbb{V}[(U-U_j)^2] \;=\; \frac{1}{720\,N^4}\left(f''(U_j)\right)^2$$

Summing over all of the strata, and dividing by N^2 due to averaging, the variance of the average is approximately

$$\frac{1}{720\,N^5}\int_0^1 \left(f''(U)\right)^2\;\mathrm{d}U.$$