# Part A Simulation and Statistical Programming HT16

## Problem Sheet 1 – due Week 2 Friday 10am 24-29 St Giles

1. Consider the integral
$$\theta = \int_0^\pi x \cos x dx.$$
   (a) Evaluate $\theta$.
   (b) Give a Monte Carlo estimator $\widehat{\theta}_n$ for numerically approximating $\theta$, using uniform random variables on $[0, \pi]$.
   (c) Calculate the bias and the variance of this estimator.
   (d) Using Chebyshev's inequality, determine how large $n$ needs to be to ensure that the absolute error between $\widehat{\theta}_n$ and $\theta$ is less than $10^{-3}$, with probability exceeding 0.99.
   (e) Same question using the Central Limit Theorem.

2. Consider the family of distributions with probability density function (pdf)
$$f_{\mu,\lambda}(x) = \lambda \exp\left(-2\lambda \left|x - \mu\right|\right), \qquad x \in \mathbb{R},$$
where $\lambda > 0$ and $\mu \in \mathbb{R}$ are parameters.
   (a) Given $U \sim \mathrm{U}[0,1]$, use the inversion method to simulate from $f_{\mu,\lambda}$.
   (b) Let $X$ have pdf $f_{\mu,\lambda}$. Show that $a + bX$ has pdf $f_{\mu',\lambda'}$ for $b \neq 0$. Find the parameters $\mu', \lambda'$.
   (c) Let $Y, Z \sim \mathrm{Exp}(r)$. Show that $Y - Z$ has pdf $f_{\mu',\lambda'}$. Find the parameters $\mu', \lambda'$. Hence, use the transformation method to simulate from $f_{\mu,\lambda}$ for any $\lambda > 0$ and $\mu \in \mathbb{R}$, given $U_1, U_2 \sim \mathrm{U}[0,1]$ independent.

3. (a) Let $Y \sim \mathrm{Exp}(\lambda)$ and fix $a > 0$. Let $X = Y|Y \geq a$. That is, the random variable $X$ is equal to $Y$ conditioned on $Y \geq a$. Calculate $F_X(x)$ and $F_X^{-1}(u)$. Give an algorithm simulating $X$ from $U \sim \mathrm{U}[0,1]$.
   (b) Let $a$ and $b$ be given, with $a < b$. Show that we can simulate $X = Y|a \leq Y \leq b$ from $U \sim \mathrm{U}[0,1]$ using
$$X = F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U),$$
   i.e., show that if $X$ is given by the formula above, then $\Pr(X \leq x) = \Pr(Y \leq x|a \leq Y \leq b)$. Apply the formula to simulate an exponential rv conditioned to be greater than $a$.
   (c) Here is a very simple rejection algorithm simulating $X = Y|Y > a$ for $Y \sim \mathrm{Exp}(\lambda)$:
      1 Let $Y \sim \mathrm{Exp}(\lambda)$. Simulate $Y = y$.
      2 If $Y > a$ then stop and return $X = y$, and otherwise, start again at 1.

   Calculate the expected number of trials to the first acceptance. Why is the inversion method to be preferred over this rejection algorithm for $a \gg 1/\lambda$?

## Part A Simulation and Statistical Programming HT16

## Problem Sheet 2 – due Week 4 Friday 10am 24-29 St Giles

1. Suppose $X$ is a discrete random variable taking values $X \in \{1, 2, ..., m\}$ with probability mass function (pmf) $p(i) = \Pr(X = i)$. Let $q(i) = 1/m$ be the pmf of the uniform distribution on $\{1, 2, ..., m\}$. Give a rejection algorithm simulating $X \sim p$ using proposals $Y$ distributed according to $q$. Calculate the expected number of simulations $Y \sim q$ per returned value of $X$ if $p = (0.5, 0.25, 0.125, 0.125)$.

2. Let $Y \sim q$ with probability density function (pdf) $q(x) \propto \exp(-|x|)$ for $x \in \mathbb{R}$. Let $X \sim \mathrm{N}(0, 1)$ be a standard normal random variable, with pdf $p(x) \propto \exp(-x^2/2)$.

   (a) Find $M$ to bound $p(x)/q(x)$ for all real $x$.

   (b) Give a rejection algorithm simulating $X$ using $q$ as the proposal pdf.

   (c) Can we simulate $Y \sim q$ by rejection using $p$ as the proposal pdf?

3. Consider a discrete random variable $X \in \{1, 2, \ldots\}$ with probability mass function

$$p(x; s) = \frac{1}{\zeta(s)} \frac{1}{x^s}, \qquad \text{for } x = 1, 2, 3, ....$$

   where $s > 1$.

   (a) The normalising constant $\zeta(s)$ is hard to calculate. However, when $s = 2$ we do have $\zeta(2) = \pi^2/6$. Give an algorithm to simulate $Y \sim p(y; 2)$ by inversion.

   (b) Implement your inversion algorithm as an R function. Your function should take as input an integer $n > 0$ and return as output $n$ iid realisations of $Y \sim p(y; 2)$. Say briefly how you checked your code.

   (c) Give a rejection algorithm simulating $X$ with pmf $p(x; s)$ for $s > 2$, using the rejection algorithm and draws from $Y \sim q$ where the proposal is $q(y) = p(y; 2)$. You will need to derive the upper bound $M' \geq \tilde{p}(x; s)/\tilde{q}(x)$ for all $x$.

   (d) Compute the expected number of simulations of $Y \sim q$ for each simulated $X$ in the previous part question, giving your answer in terms of $\zeta(s)$.

   (e) Implement your algorithm as an R function. Your function should take as input $s$ and return as output $X \sim p(x; s)$ and the number of trials $N$ it took to simulate $X$.

4. Suppose $X \sim N(0, \sigma^2)$ is a Gaussian random variable with mean 0 and variance $\sigma^2$. We want to estimate $\mu_\phi = \mathbb{E}(\phi(X))$ for some function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(X)$ has finite mean and variance. Suppose we have iid samples $Y_1, ..., Y_n$ with $Y_i \sim N(0, 1), i = 1, 2, ..., n$. We consider the following two estimators for $\mu_\phi$:

$$\widehat{\theta}_{1,n} = \frac{1}{n} \sum_{i=1}^{n} \phi(\sigma Y_i)$$

   and

$$\widehat{\theta}_{2,n} = \frac{1}{n\sigma} \sum_{i=1}^{n} \exp\left[-Y_i^2 \left(\frac{1}{2\sigma^2} - \frac{1}{2}\right)\right] \phi(Y_i).$$

   (a) Show that $\widehat{\theta}_{1,n}$ and $\widehat{\theta}_{2,n}$ are unbiased and give the expression of their variances.

   (b) What range of values must $\sigma$ be in for $\widehat{\theta}_{2,n}$ to have finite variance? Can you give a weaker condition if it is known that $\int_{-\infty}^{\infty} \phi^2(x)\, dx < \infty$?

   (c) Why might we prefer $\widehat{\theta}_{2,n}$ to $\widehat{\theta}_{1,n}$, for some values of $\sigma^2$ and functions $\phi$? (Hint: consider estimating $\mathbb{P}(X > 1)$ with $\sigma \ll 1$).

# Part A Simulation and Statistical Programming HT16

## Problem Sheet 3 – due Week 6 Friday 10am

Please hand in the solutions at 24-29 St Giles, and email the R code for questions 3 and 4, in a single well-commented R-script, to
- thibaut.lienart@univ.ox.ac.uk (Monday 2pm class) or
- andreas.anastasiou@jesus.ox.ac.uk (Monday 4pm class).

1. We are interested in performing inference about the parameters of an internet traffic model.

   (a) The arrival rate $\Lambda$ for packets at an internet switch has a log-normal distribution LogNormal$(\mu, \sigma)$ with parameters $\mu$ and $\sigma$. The LogNormal$(\mu, \sigma)$ probability density is

   $$p_\Lambda(\lambda; \mu, \sigma) = \frac{1}{\lambda\sqrt{2\pi\sigma^2}} \exp\left(-(\log(\lambda) - \mu)^2/2\sigma^2\right),$$

   Show that if $V \sim N(\mu, \sigma^2)$ and we set $W = \exp(V)$ then $W \sim$ LogNormal$(\mu, \sigma)$.

   (b) Given an arrival rate $\Lambda = \lambda$, the number $N$ of packets which actually arrive has a Poisson distribution, $N \sim$ Poisson$(\lambda)$. Suppose we observe $N = n$. Show that the likelihood $L(\mu, \sigma; n)$ for $\mu$ and $\sigma$ is

   $$L(\mu, \sigma; n) \propto \mathbb{E}(\Lambda^n \exp(-\Lambda)|\mu, \sigma).$$

   (c) Give an algorithm simulating $\Lambda \sim$ LogNormal$(\mu, \sigma)$ using $Y \sim N(0, 1)$ as a base distribution, and explain how you could use simulated $\Lambda$-values to estimate $L(\mu, \sigma; n)$ by simulating values for $\Lambda$.

   (d) Suppose now we have $m$ iid samples

   $$\Lambda^{(j)} \sim \text{LogNormal}(\mu, \sigma), j = 1, 2, ..., m$$

   for one pair of $(\mu,\sigma)$-values. Give an importance sampling estimator for $L(\mu', \sigma'; n)$ at new parameter values $(\mu', \sigma') \neq (\mu, \sigma)$, in terms of the $\Lambda^{(j)}$'s.

   (e) For what range of $\mu', \sigma'$ values can the $\Lambda^{(j)}$-realisation be safely 'recycled' in this way?

2. Let $X = (X_0, X_1 \ldots)$ be a homogeneous Markov chain taking values in a discrete state space $\Omega$, with transition matrix $P = (p_{ij})_{i,j\in\Omega}$.

   (a) Show that if the Markov chain is irreducible, and $p_{ii} > 0$ for some $i \in \Omega$, then the chain is aperiodic.

   (b) Consider the homogeneous Markov chain $(X_0, X_1, \ldots)$ with $X_n \in \{1, \ldots, m\}$ and transition matrix

   $$p_{ij} = \frac{1}{m} \min\left(1, \frac{p(j)}{p(i)}\right)$$

   for $i \neq j$ and

   $$p_{ii} = 1 - \frac{1}{m} \sum_{j\neq i} \min\left(1, \frac{p(j)}{p(i)}\right)$$

   where $p$ is a probability mass function on $\{1, \ldots, m\}$ with $p(i) > 0$ for all $i = 1, 2, \ldots, m$ and $X_0 = 1$.

   (i) Show that the Markov chain is irreducible and aperiodic, and admits $p$ as invariant distribution.

   (ii) Propose an algorithm to simulate the Markov chain $(X_0, X_1, X_2, \ldots)$ using independent random variables $Y_k \sim \text{U}\{1, \ldots, m\}$ and $U_k \sim \text{U}[0, 1]$ for $k = 1, 2, \ldots$

3. Here is an algorithm converting a non-negative number $x \in [0, 1)$ to its binary expansion.

   Let $b$ be the binary representation of $x$. Compute the first $I$ binary places as follows. Let $i = 1$ and $y = 2x$. If $y$ is greater than or equal one set $b_i = 1$ otherwise set $b_i = 0$; let $x = y - b_i$. If $x$ is now zero or $i = I$ then stop (as either there are no more non-zero places, or we have reached the limit of our number of digits), otherwise increase $i$ by one and repeat.

   (a) Write an R function implementing this algorithm. Your function should take as input a single non-negative number $x$ between 0 and 1 and return the corresponding binary representation. Represent the binary number as a vector, so for example decimal 0.125 becomes `c(0,0,1)` in binary.

   (b) At what binary place do Rs numerical values for 0.3 and $0.1 + 0.1 + 0.1$ differ?

   (c) Adapt your function to take two positive integers $0 < p < q$ as input, and return the binary expansion of $p/q$ *exactly*.

4. Consider a sequence of observations $x_1, \ldots, x_n$. Let $m_i$ and $s_i^2$ denote the mean and sample variance of the first $i$ observations $i \leq n$. How many operations (additions, subtractions, multiplications or divisions) are needed to calculate the sequence of means $m_1, \ldots, m_n$, if each mean is calculated separately?

   (a) Derive an expression for $m_{i+1}$ in terms of $m_i$ and $x_{i+1}$ and write an R function that calculates $m_1, \ldots, m_n$ using this sequential formula. How many operations will this function use? *[Hint: it is important for speed to initialise your output vector with the correct length at the start using* `numeric()`*, rather than appending one answer at a time.]*

   (b) Now consider the sequence of sample variances $s_1^2, \ldots, s_n^2$. Find an expression for $s_{i+1}^2$ in terms of $s_i^2$, $m_i$, $m_{i+1}$ and $x_{i+1}$. Write an R function to evaluate the sample variances using a sequential method.

   (c) (Optional.) Write a function to calculate the sample means non-sequentially (using a loop or, for example, `sapply()`). How long does it take to run when $n = 10^3, 10^4, 10^5$?

## Part A Simulation and Statistical Programming HT16

## Problem Sheet 4 – due Week 7 Friday 10am

Please hand in the solutions at 24-29 St Giles, and email the R code, in a single well-commented R-script, to
• thibaut.lienart@univ.ox.ac.uk (Monday 2pm class) or
• andreas.anastasiou@jesus.ox.ac.uk (Monday 4pm class).

1. (a) Give a Metropolis-Hastings algorithm with a stationary Gamma probability density function,

$$\pi(x) \propto x^{\alpha-1}\exp(-\beta x), \quad x > 0$$

with parameters $\alpha, \beta > 0$. Use the proposal distribution $Y \sim \text{Exp}(\beta)$.

(b) Write an R function implementing your MCMC algorithm. Your function should take as input values for $\alpha$ and $\beta$ and a number $n$ of steps and return as output a realization $X_1, X_2, ..., X_n$ of a Markov chain targeting $\pi$. State briefly how you checked your code.

2. MCMC for Bayesian inference (first two parts were an exam Q in 2009)

(a) Let $X \sim \text{Binomial}(n, r)$ be a binomial random variable with $n$ trials and success probability $r$. Let $\pi(x; n, r)$ be the pmf of $X$. Give a Metropolis-Hastings Markov chain Monte Carlo algorithm with stationary pmf $\pi(x; n, r)$.

(b) Suppose the success probability for $X$ is random, with $\Pr(R = r) = p(r)$ given by

$$p(r) = \begin{cases} r & \text{for } r \in \{1/2, 1/4, 1/8, ...\}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

An observed value $X = x$ of the Binomial variable in part (a) is generated by simulating $R \sim p$ to get $R = r^*$ say, and then $X \sim \text{Binomial}(n, r^*)$ as before. Specify a Metropolis-Hastings Markov chain Monte Carlo algorithm simulating a Markov chain, $(R_t)_{t=0,1,2,...}$ with equilibrium probability mass function $R_t \xrightarrow{d} p(r|x)$ where

$$p(r|x) \propto \pi(x; n, r)p(r)$$

is called the posterior distribution for $r$ given data $x$.

(c) Write an R function implementing your MH MCMC algorithm with target distribution $p(r|x)$. Suppose $n = 10$ and we observe $x = 0$. Run your MCMC algorithm and estimate the mode of $p(r|x)$ over values of $r$.

3. Let $X$ be an $n \times p$ matrix of fixed covariates with $n > p$, and suppose that $X$ has full column rank $p$.

(a) Explain why the $p \times p$ matrix $X^T X$ is invertible.

Consider the linear model given by

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$.

(b) Write down the distribution of $Y_i$, and use it to write out the log-likelihood for $\beta = (\beta_1, \dots, \beta_p)$.

(c) Show that the MLE is equivalent to minimising the sum of squares:

$$R(\beta) = \sum_{i=1}^{n}(Y_i - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2.$$

(d) By differentiating and writing the problem as a system of linear equations, show that the MLE is $\hat{\boldsymbol{\beta}} = (X^TX)^{-1}X^TY$.

4. Consider the linear model $Y = X\beta + \epsilon$ where $Y$ is a vector of $n$ observations, $X$ is an $n \times p$ matrix with each column containing a different explanatory variable and $\epsilon$ is a vector of $n$ independent normal random errors with mean zero and unknown variance $\sigma^2$. The maximum likelihood estimator for $\beta$ is

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

The sample variance is

$$s^2 = \frac{1}{n-p}\|X\hat{\beta} - Y\|^2$$

where $p$ is the length of $\beta$. The standard error for $\beta$ is

$$\text{se}(\hat{\beta}_i) = s\sqrt{[(X^TX)^{-1}]_{ii}}$$

(a) The trees data give Girth, Height and Volume measurements for 31 trees. Fit the model

$$Y_i = \beta_1 + x_i^{height}\beta_2 + x_i^{girth}\beta_3 + \epsilon_i$$

using the R commands

```
> data(trees)
> summary(lm(Volume ~ Girth + Height, data=trees))
```

and briefly interpret the output.

(b) Write a function of your own (using `solve()` or your solution to question 3, not `lm()`) to fit a linear model. Your function should take the length 31 vector `trees$Volume` and the $31 \times 3$ matrix `X = cbind(1, trees$Girth, trees$Height)` as input and return estimates of $\beta$, the residual standard error $s$, and the standard errors of each $\beta_i$. Check your output against the corresponding results from the `summary(lm())` output in (a).

5. Here is an algorithm to compute the QR factorisation of an $n \times p$ matrix $A$ with $p \leq n$. That is, it returns an $n \times p$ orthogonal matrix $Q$ and a $p \times p$ upper triangular matrix $R$ sich that $A = QR$.

Let $|v|$ denote the Euclidean norm of a vector $v$. Let $A_{[,a:b]}$ denote the matrix formed from the columns $a, a+1, \ldots, b$ of $A$.

1. Create $n \times p$ matrix $Q$ and $p \times p$ matrix $R$.
2. Set $Q_{[,1]} = A_{[,1]}/|A_{[,1]}|$ and $R_{11} = |A_{[,1]}|$.
3. If $p = 1$ then we are done; return $Q$ and $R$.
4. Otherwise (i.e. if $p > 1$), set $R_{[1,2:p]} = Q_{[,1]}^T A_{[,2:p]}$ and $R_{[2:p,1]} = \mathbf{0}$.
5. Set $A' = A_{[,2:p]} - Q_{[,1]}R_{[1,2:p]}$.
   *[Notice that $Q_{[,1]}R_{[1,2:p]}$ is an outer product of an $n$ component column vector and a $(p-1)$ component row vector, so $A'$ is a new $n \times (p-1)$ matrix. Either make use of the* `outer()` *command or, if you use* `[` *be careful to use the* `drop` *argument when forming these sub-matrices.]*
6. Compute the QR factorisation of $A'$ (so $A' = Q'R'$ say).
7. Set $Q_{[,2:p]} = Q'$ and $R_{[2:p,2:p]} = R'$ and return $Q$ and $R$.

(a) Implement this algorithm as a recursive function in R. Your function should take as input an $n \times p$ matrix $A$ and return two matrices $Q$ and $R$ as a list. State briefly how you checked your function was correct.

(b) Using your QR function, and the R command `backsolve()`, give a least squares solution to the over-determined system

$$X\beta = Y$$

where $X$ and $Y$ take their values from the `trees` data in question 4.