Part A Simulation and Statistical Programming HT19

Problem Sheet 3 – due Thursday 10am in week 6

Please hand in the solutions at 24-29 St Giles, and email the R code, in a single well-commented R-script, to sinan.shi@stats.ox.ac.uk

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

   (a) The arrival rate Λ for packets at an internet switch has a log-normal distribution LogNormal(μ, σ) with parameters μ and σ. The LogNormal(μ, σ) 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(μ, σ).

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

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

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

   (d) Suppose now we have m iid samples

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

   for one pair of (μ,σ)-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 μ', σ' values can the $\Lambda^{(j)}$-realisation be safely 're-cycled' in this way?

2. Let $X = (X_0, X_1 \ldots)$ be a homogeneous Markov chain taking values in a discrete state space Ω, 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 $Y_k \sim \mathrm{U}\{1, \ldots, m\}$ and $U_k \sim \mathrm{U}[0, 1]$ 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 R's 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 $n = 10^3, 10^4, 10^5$?