Part A Simulation and Statistical Programming HT14

Lecturer: Geoff Nicholls

University of Oxford

Notes and Problem sheets are available at

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

"*Monte Carlo Simulation: An analytical technique for solving a problem by performing a large number of trail runs, called simulations, and inferring a solution from the collective results of the trial runs.*", glossary at **www.nasdaq.com**.

"*Anyone who can do solid statistical programming will never miss a meal.*", Prof David Banks, 2008.

Course structure

- 8 Lectures here in SPR1
- 6 2hr sessions in HT weeks 2-6 and 8 in Evenlode room, OUCS, 13 Banbury road
- 4 problem sheets and 4 classes 4-5pm on Tuesdays in weeks 3 and 7 and 10-11am on Fridays in week 5 and 8 in 2 South Parks Road.
- Exam paper A12: 3 questions, with the best two questions counting towards a candidate's total mark for the paper.
- Equivalent to a 16hr lecture course.

Why Simulation and Statistical Programming, and why together?

We fit complex realistic models and analyze large data sets.

Taking expectations is a fundamental operation in Statistics.

Expectations are integrals and integrals are hard.
Use simulation to do hard integrals.

Simulation theory is applied probability.
Doing simulation on a computer is statistical programming.

Taken together Simulation and Statistical Programming empower you to do a large chunck of statistical inference.

What is R?

R is an open-source package for Statistical Computing. The Statistical Programming segment of this course is designed to teach you to do statistical programming in R.

- It is freely available - **http://cran.r-project.org/**
- Millions of users worldwide in Universities and Industry
- The BS1 course in Part B will use R extensively

R is actively supported and updated and has many add-on packages that specialize in specific applications.

I recommend you install it for your own use.

Simulation: motivation

In many settings we wish to estimate expectations.

Suppose $X \in \Omega$ is a random variable (rv) with density $X \sim p(x)$, $f : \Omega \longrightarrow \Re$ is a function and we want to evaluate the expectation

$$E_p(f) = \int_\Omega f(x)p(x)dx.$$

If we have $X_i, i = 1, 2, ..., n$ with $X_i \sim p$ then

$$\overline{f}_n = n^{-1} \sum_{i=1}^{n} f(X_i)$$

is an unbiased estimator for $E_p(f)$.

If $f(X_i)$ are iid with $0 < \mathrm{var}(f(X)) < \infty$ and $|E_p(f)| < \infty$ then

$$\frac{\overline{f}_n - E(f)}{\sqrt{\mathrm{var}(f)/n}} \to N(0,1),$$

by the CLT. If $Z_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(f(X_i) - \overline{f}_n)^2$$

is an unbiased estimator for $\mathrm{var}(f)$ we can report a level-α CI

$$\overline{f}_n \pm Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

for $E_p(f)$.

Example: $X \sim N(0,1)$, what is $a = P(\sin(\pi X) > 0.5)$?

Here $P(\sin(\pi X) > 0.5) = \int_{-\infty}^{\infty} \mathbb{I}_{\sin(\pi x)>0.5}\, p(x)dx$, where

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$$

and

$$\mathbb{I}_{\sin(\pi x)>0.5} = \begin{cases} 1 \text{ if } \sin(\pi x) > 0.5 \\ 0 \text{ otherise.} \end{cases}$$

We fix n and draw $x_i \sim N(0,1)$ then form

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\sin(\pi X_i)>0.5}$$

Simulation: overview. Our task in simulation is to give algorithms for generating the X_i's: given a probability density or mass function $p(x)$ (the target distribution), give an algorithm for simulating $X \sim p$.

We will use

- the inversion and transformation methods,
- rejection sampling,
- importance sampling,
- Markov chain Monte Carlo

You will learn why they work, and what their relative advantages are in different settings.

You will learn to adapt these algorithms to simulation for different target distributions.

Inversion

Say $X \in \Omega$ is a scalar rv with cdf $F(x) = \Pr(X \leq x)$ at $X = x$ and we want to simulate $X \sim F$.

Claim: suppose $F(x)$ is continuous and strictly increasing with x. If $U \sim U(0,1)$ and $X = F^{-1}(U)$ then $X \sim F$.

Proof: $\Pr(X \leq x) = \Pr(F^{-1}(U) \leq x)$ so applying strictly increasing F to both sides of $F^{-1}(U) \leq x$ we have $\Pr(X \leq x) = \Pr(U \leq F(x))$ which is $F(x)$, since U is uniform.

Remark: if we define

$$F^{-1}(u) = \min(z; F(z) \geq u)$$

then $F^{-1}(U) \sim F$ without F continuous or strict increasing.

Example: if we want $X \sim \text{Exp}(r)$, *ie* $X \sim p(x)$ with

$$p(x) = r \exp(-rx),$$

then the CDF is

$$F(x) = 1 - \exp(-rx)$$

and its inverse is

$$F^{-1}(u) = -(1/r) \log(1 - u).$$

The algorithm is

$$U \sim U(0, 1)$$

$$X \leftarrow -\log(U)/r$$

and note I replaced $1 - U$ with U since $U \sim 1 - U$.

Inversion: discrete random variables Suppose $X \sim F$ with X a discrete rv with pmf $p(x), x = 0, 1, 2, \ldots$. Consider the following algorithm

$U \sim U(0, 1)$

set X equal the unique x satisfying $\sum_{i=0}^{x-1} p(i) < u < \sum_{i=0}^{x} p(i)$

with $\sum_{i=0}^{x-1} p(i) \equiv 0$ if $x = 0$.

Proof: It is easy to check that $X \sim p$. (board presentation)

Remark: the cdf here is $F(x) = \sum_{i=0}^{x} p(i)$ and the above algorithm is actually just $x = F^{-1}(u)$ again with the more general definition for F^{-1} above.

Example: If $0 < p < 1$ and $q = 1 - p$, and we want to simulate $X \sim$ Geometric(p) then

$$p(x) = pq^{x-1}$$

and the cdf

$$F(x) = \sum_{i=0}^{x} p(i)$$

is $F(x) = 1 - q^x$ for $x \in \mathbb{N}$.

Smallest x giving $1 - q^x \geq u$ is

$$x = \left\lceil \frac{\log(1-u)}{\log(q)} \right\rceil$$

where $\lceil x \rceil$ rounds up.

Transformation Methods

Say $Y \sim Q$, $Y \in \Omega_Q$ we **can** simulate (for example $Y \sim U(0,1)$)

$X \sim P$, $X \in \Omega_P$ we **want to** simulate (eg $X \sim Exp(1)$).

If we can find a function $f : \Omega_Q \to \Omega_P$ with the property that

$$f(Y) \sim P$$

then we can simulate X by simulating

$$Y \sim Q \qquad \text{and setting} \qquad X = f(Y)$$

(for example, set $f(y) = -\log(y)$ and we know from above that if $X = f(Y)$ then $X \sim Exp(1)$).

Example: Suppose we want to simulate $X \sim Exp(1)$ and we can simulate $Y \sim U(0,1)$. Try setting $X = -\log(Y)$. Does that work?

Ans: recall that if Y has density $q(y)$ and $X = f(Y)$ then the density of X is $p(x) = q(y(x))|dy/dx|$. Now $q(y) = 1$ (the density of $U(0,1)$) and $y(x) = \exp(-x)$, so $p(x) = \exp(-x)$ and so X has the density of an $Exp(1)$ rv.

Example: Inversion is a transformation method: Q is $U(0,1)$; Y is U; and $X = f(Y)$ with $f(y) = F^{-1}(y)$ and F the CDF of the target distribution P.

We can generalize the idea. We can take functions of collections of variables.

Example: Suppose we want to simulate $X \sim \Gamma(a, \beta)$ with $a \in 1, 2, 3, ...$ and we can simulate $Y \sim Exp(1)$ (the $\Gamma(a, \beta)$ density is $p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ for $x > 0$).

Simulate $Y_i \sim Exp(1), i = 1, 2, ..., a$ and set $X = \sum_{i=1}^{a} Y_i/\beta$. Then $X \sim \Gamma(a, \beta)$.

Proof: Use moment generating functions. The MGF of the $Exp(1)$ rv Y is

$$E\left(e^{tY}\right) = (1-t)^{-1}$$

so the MGF of X is

$$E\left(e^{tX}\right) = \prod_{i=1}^{a} E\left(e^{tY_i/\beta}\right) = (1 - t/\beta)^{-a}$$

which is the MGF of a $\Gamma(a, \beta)$ variate.

See you next week

Our next meeting is here in this lecture theater next week.

Our first lecture and practical in statistical programming will be next week on Friday afternoon in Evenlode - the OUCS computer teaching facility.

The first problem sheet is due Monday 9am of Week 3.

Homework: install R and run the code from this lecture.