Marius Hofert, Riley Metzger University of Waterloo 2015-02-06

STAT 340: Computer Simulation of Complex Systems

Assignment 2 (due on 2015-02-27)

Exercise 10 (Unbiasedness and updating of the sample variance) (3+4=7 points)
Let $X_1, \dots, X_n$ be independent with mean $\mu = \mathbb{E}X_i$ and variance $\sigma^2 = \operatorname{Var} X_i$, $i \in \{1, \dots, n\}$.

a) Show that the sample variance is unbiased, i.e.,

$$\mathbb{E}[S_n^2] = \sigma^2,$$

where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is the *sample variance*.

b) Show that the sample variance admits the recursive formula

$$S_n^2 = \frac{n-2}{n-1} S_{n-1}^2 + \frac{(X_n - \bar{X}_{n-1})^2}{n}.$$

Exercise 11 (2 points)
Show that there is no discrete uniform distribution on $\mathbb{N}$.

Exercise 12 (Middle-square method) (4 points)
The *middle-square method* (von Neumann (1949)) is the first known method for generating pseudo-random numbers (with 4 digits). Starting from a 4-digit integer seed, the number is squared, producing an 8-digit number (if the result is less than 8 digits, leading zeroes are added). The middle 4 digits of the result are then taken as the next number in the sequence, and so forth. To obtain pseudo-random numbers in $[0, 1]$, divide the obtained numbers by 10 000. Implement the middle-square method in R and compute the first twenty numbers based on the seeds 7182 and 1009. Based on these numbers, assess the quality of this pseudo-random number generator.

Exercise 13 (Linear congruential generators) (3+2=5 points)

a) Consider the linear congruential generator (LCG) defined by

$$x_n = (5x_{n-1} + 3) \bmod 16, \quad n \in \mathbb{N},\ x_0 = 7.$$

   i) Determine the period of this LCG.
   ii) Compute x_n, $n \in \{1, \dots, 19\}$, by hand.
   iii) Implement this LCG in R and compute the first 19 numbers with your implementation. Compare your results with ii).

b) Consider the multiple-recursive generator (MRG) defined by

$$x_n = (3x_{n-1} + 5x_{n-2}) \bmod 100, \quad n \geq 2,\ x_0 = 23,\ x_1 = 66.$$

Compute x_n, $n \in \{2, \dots, 15\}$.

Exercise 14 (The pseudo-random number generator RANDU) (5 points)
The multiplicative congruential generator (MCG) "RANDU" appeared in the 1960s. It is given by

$$x_n = (2^{16} + 3)x_{n-1} \bmod 2^{31}, \quad n \in \mathbb{N},$$

and pseudo-random numbers can be obtained by setting $u_n = x_n/2^{31}$, $n \in \mathbb{N}$. A graphical way of checking u_n, $n \in \mathbb{N}$, for uniformity and independence is to consider, for various $d \in \mathbb{N}$, $(u_n, u_{n+1}, \dots, u_{n+d-1})$, $n \in \mathbb{N}$. Starting from $x_0 = 123\,456\,789$, generate $n = 2000$ random numbers and plot, for $d \in \{1, 2, 3\}$, the vectors $(u_k, u_{k+1}, \dots, u_{k+d-1})$, $k \in \{1, \dots, n-d+1\}$. Describe what you see and assess the quality of this pseudo-random number generator.

Exercise 15 (Sampling a generic discrete distribution function) (5 points)
Use R's `findInterval()` to write a pseudo-random number generator for any discrete distribution with distinct jumps $x_1, \dots, x_n$ of corresponding heights $p_1, \dots, p_n \in (0, 1]$. Test your implementation by sampling $n = 10^5$ samples from the discrete distribution which puts mass 1/6, 1/3, 1/2 at the integers 1, 2, 3, respectively, and test whether, roughly, 1/6 of your samples lead to 1, 1/3 lead to 2 and 1/2 lead to 3.
Hint. Useful functions are: `order()`, `cumsum()`, `runif()`, `table()`, `all.equal()`.

Exercise 16 (Sampling methods) (2+2+2+2=8 points)
Give sampling algorithms for the following situations.

a) Use the composition method to sample the distribution function $F(x) = \frac{x+x^3+x^5}{3}$, $x \in [0, 1]$.

b) Use the inversion method to sample the density $f(x) = \begin{cases} e^{2x}, & \text{if } x \in (-\infty, 0), \\ e^{-2x}, & \text{if } x \in [0, \infty). \end{cases}$

c) Use the rejection method to sample the density $f(x) = \frac{1+x}{2}e^{-x}$, $x \in (0, \infty)$.

d) Use a method of your choice to sample the density $f(x) = \frac{1}{4} + 2x^3 + \frac{5}{4}x^4$, $x \in [0, 1]$.

Exercise 17 (Multivariate distribution functions, copulas and sampling) (2+3+1+2=8 points)
Note. This exercise will *not* be graded. Feel free to work on it anyways.
Let $Z_1, Z_2, Z_3 \overset{\text{iid}}{\sim} F(x) = \exp(-1/x)$, $x > 0$. Furthermore, let

$$X_1 = \max\{Z_1, Z_2\},$$
$$X_2 = \max\{Z_1, Z_3\}.$$

a) Calculate the joint distribution function H of $\boldsymbol{X} = (X_1, X_2)$.

b) Calculate the copula C which corresponds to H by Sklar's Theorem. Show that it indeed has $\mathrm{U}[0, 1]$ marginal distribution functions.
Hint. Recall that the marginal distribution functions of a bivariate distribution function H can be computed via $F_1(x) = \lim_{x_2 \to \infty} H(x_1, x_2)$ and $F_2(x) = \lim_{x_1 \to \infty} H(x_1, x_2)$.

c) Give a sampling algorithm for H.
Hint. You may use the above stochastic representation; F can easily be sampled with the inversion method.

d) Give a sampling algorithm for C.
Hint. Componentwise apply the probability integral transformation.