Université Paris-Dauphine Année 2012-2013
Département de Mathématique

Partiel NOISE, sujet B

Résoudre trois et uniquement trois exercices au choix.
Enregistrez très régulièrement votre travail, afin d'éviter toute perte de fichiers en cas de problème informatique. La composition doit s'effectuer en anglais.

Exercice 1

We wish to sample realizations from a Poisson distribution with parameter $\lambda > 0$ starting from realizations of the uniform distribution $\mathcal{U}_{[0,1]}$.

1. Build a function `rexp2` which takes as arguments an integer n and a rate λ, and which samples n independent realizations from the Exponential distribution of parameter λ, with density $\lambda \exp\{-\lambda x\}$ over $\mathbb{R}_+$.
2. Let $Y_1, Y_2, \ldots$ be random variables independent and identically distributed with Exponential distribution of parameter λ. Define

   $$Z_k = Y_1 + Y_2 + \ \ldots \ + Y_k$$

   and X such that

   $$X = k \text{ iff } Z_k < 1 \leq Z_{k+1}$$

   In other words, we sample Exponential random variables with parameter λ and we count the number of simulations needed for the cumulative sum to exceed 1.

   Write a function `rpoiss2` with arguments λ and n, which samples n values from the random variable X with parameter λ.
3. Use the function `rpoiss2` to simulate a sample of size 1000. Propose a graphical method to verify that X follows a Poisson distribution of parameter λ.

Exercice 2

Let us consider the probability density defined on $\mathbb{R}$ by

$$f(x) \propto x^2 \frac{3+\sin^2(x)}{(1+\cos^2(x))^2} \exp\{-x^2\}$$

and let $g(x) = x^2 \exp\{-x^2\}/\Gamma(3/2)$.

1. Using a change of variable, show that g is a probability density. This can be checked using R, by typing

   ```
   > integrate(function(x) x^2*exp(-x^2)/gamma(1.5),-10,10)
   1 with absolute error < 3.2e-05
   ```
2. Build an accept-reject algorithm to generate a random variable with density f (hint : use the `rgamma()` function).
3. Using the above algorithm, write an R code that computes a Monte-Carlo estimate of the constant

   $$\int_{\mathbb{R}} x^2 \frac{3+\sin^2(x)}{(1+\cos^2(x))^2} \exp\{-x^2\} \mathrm{d}x$$

4. Consider the following R output :

```
> integrate(function(x) x^2*(3+sin(x)^2)*exp(-x^2)/
+ ((1+cos(x)^2)^2*gamma(1.5)),-10,10)
2.453387 with absolute error < 6.9e-11
```

Can you deduce from this result the probability of acceptance of the accept-reject algorithm ?

Exercice 3
The goal of this exercise is to evaluate the following integral :

$$\mathcal{I} = \int_0^1 e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

1. Give the R command to give the (almost) exact numerical value of $\mathcal{I}$ under the form `0.8556244 with absolute error < 9.5e-15`
2. Propose a first Monte-Carlo estimate of $\mathcal{I}$, based on the generation of n Gaussian random variables. Give the R code to compute a 95% confidence level of $\mathcal{I}$, for $n = 1\,000$.
3. Propose a second Monte-Carlo estimate of $\mathcal{I}$, based on the generation of a sample of n uniform random variables.
4. How can you compare both methods ?
5. Show that

$$\mathcal{I} = \int_0^1 \exp\left(-\frac{(1-x)^2}{2}\right) \mathrm{d}x = \frac{1}{2}\int_0^1 \left[\exp\left(-\frac{x^2}{2}\right) + \exp\left(-\frac{(1-x)^2}{2}\right)\right] \mathrm{d}x$$

6. Deduce from the above equality a new Monte Carlo estimate for $\mathcal{I}$ based on the generation of n uniform variables.
7. Propose yet another Monte Carlo estimate for $\mathcal{I}$, using simulations from the Beta distribution $B(a, a)$, obtained by `rbeta(n,a,a)`. Describe an algorithm to optimize the choice of a. Give the corresponding R code.

Exercice 4
We wish to verify the following limit theorem :

$$\sqrt{n}(\hat{q}_n - F^{-1}(\alpha)) \to_{\mathcal{D}} \mathcal{N}(0, \sigma^2) \tag{1}$$

where $\hat{q}_n$ is the empirical estimator of quantile $F^{-1}(\alpha)$, *i.e.*

$$\hat{q}_n = X_{(\lfloor \alpha n \rfloor + 1)},$$

with the variance of the asymptotic normal distribution equal to

$$\sigma^2 = \frac{\alpha(1-\alpha)}{[f(F^{-1}(\alpha))]^2}.$$

We suppose that F is the cumulative distribution function of $\mathcal{N}(0, 16)$, *i.e.* a normal distribution with mean 0 and variance $\sigma^2 = 16$, and hence with density

$$f(x) = \frac{1}{4\sqrt{2\pi}} \exp(-\frac{x^2}{32}).$$

Take $\alpha = 0.28$.

a. Simulate $N = 10^3$ samples of size $n = 10^2$ from distribution $\mathcal{N}(0, 16)$, and compute $\hat{q}_n$ for each sample : Use the R function quantile and create an appropriate R function taking (N, n, σ) as input and returning the sample of $\hat{q}_n$'s.

b. Plot the histogram of the $\hat{q}_n$'s thus obtained and check graphically that it fits the asymptotic distribution given in (1).

c. Give a 95% confidence interval for the variance of $\hat{q}_n$.

Exercice 5

Let K be a random variable following the uniform distribution over the set of integers $\{1, 2, \ldots, N\}$. Then

$$P(K \text{ is divisible by a perfect square}) \underset{n\to\infty}{\to} \frac{6}{\pi^2}$$

or more formally

$$P(\exists i \geq 2 : K \mod i^2 = 0) \underset{n\to\infty}{\to} \frac{6}{\pi^2}.$$

We wish to use this property to estimate π in R.

1. Write a function `squaremultiple(k)` which returns `TRUE` if k is divisible by a perfect square, and `FALSE` otherwise. Recall that `a %% b` calculates $a \mod b$.
2. Take $N = 100$ and produce an estimate of π using $n = 1000$ realizations of K.
3. Keep n constant and check that the quality of the estimate increases with N. (Caution : very large values of N may cause your computer to crash.)

Exercice 6

Take two parameters, $\lambda > 0$ and $k > 0$. We consider the associated cdf :

$$F(x) = \begin{cases} 0 & \text{when } x \leq 0 \\ 1 - e^{-(\lambda x)^k} & \text{otherwise} \end{cases}$$

which corresponds to the Weibull(k, λ) distribution.

1. Write an R function called rweibull with arguments n, a positive integer, and the parameters λ and k, which returns n independent simulations from the Weibull(k, λ) distribution. This function must use *exactly* n independent simulations from the uniform $\mathcal{U}_{[0,1]}$ distribution.
2. For a pair of parameters (λ, k) equal to $(2, 3)$, illustrate by a graphical representation the adequacy of this simulation method.
3. Based on a sample of size 5000 simulated via rweibull, give a confidence interval at confidence level asymptotically equal to 5% for the mean of a Weibull$(2, 3)$ distribution.

Exercice 7

Create an R code that solves the following puzzle : Given a lottery with N tickets numbered from 1 to N, all tickets being sold, the winning tickets are such that one of their digits is 1, and another digit *on the right of* 1 is 3. For instance, 123 and 8135 both are winning tickets. Determine the value of $999 < N < 9999$ such that there is a proportion of winning tickets exactly equal to 10%.