Université Paris-Dauphine Année 2012-2013
Département de Mathématique

Partiel NOISE, sujet A

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 verify the following limit theorem :

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

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 $\mathcal{N}(0, 300)$, *i.e.* a normal distribution with mean 0 and variance $\sigma^2 = 300$, and hence with density

$$f(x) = \frac{1}{\sqrt{600\pi}} \exp(-\frac{x^2}{600}).$$

Take $\alpha = 0.11$.

a. Simulate $N = 10^3$ samples of size $n = 10^2$ from distribution $\mathcal{N}(0, 300)$, 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 2
Two integers are said to be *coprime* iff the only positive integer that divides both of them is 1. Let I and J be two independent random variables with uniform distribution over the set of integers $\{1, 2, \ldots, N\}$. Then a famous number theory result states that

$$P(I \text{ and } J \text{ are coprime}) \underset{N\to\infty}{\longrightarrow} \frac{6}{\pi^2}.$$

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

1. Write a function `coprime(i,j)` which returns `TRUE` if i and j are coprime, and `FALSE` otherwise. Recall that i and j are coprime iff the only divisor they share is 1, and that `a %% b` calculates $a \mod b$.

2. Take $N = 100$ and produce an estimate of π using $n = 1000$ realizations of I and J.
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 3
Let f be the density corresponding to a mixture of normal distributions

$$p\,\mathcal{N}(0,1) + (1-p)\,\mathcal{N}(3,1)$$

We can get a sample Z from the distribution f of size `nsample` in the following manner :

```
p=0.6
nsample=1000
Z=replicate(nsample, rnorm(1, sample(c(0,3),prob=c(p,1-p)),1))
```

We are interested in the quantity $P = \mathbb{P}(X > 4)$.

1. Compute the exact value of P :
   – by numerical integration ;
   – by using the fact that f is a mixture of normal distributions.

   We now wish to estimate the quantity $P = \mathbb{P}(X > 4)$ using Monte-Carlo Methods :
2. Sample $N = 500$ realizations, each of size `nsample`, following the density f. Derive from those simulations a 95% confidence interval for P.
3. By using a single sample of size `nsample` following the density f, and remembering that P is an estimation of the cumulative distribution function of f, give a 95% confidence interval for P.
4. We would like to use a sample W from a distribution g that differs from f. What distribution do you suggest to estimate P ? Name the method you use and give a 95% confidence interval on P for a sample size of $n = 1000$.

Exercice 4
We consider the random variable X having density

$$f_X(x) = C\left[1 - 3\left(x - \frac{1}{2}\right)^2\right]\mathbb{I}_{[0,1](x)}$$

where C is an unknown normalizing constant.

1. Compute the expected value and the variance of $X \sim f_X$.
2. Propose a method for sampling from $X \sim f_X$ that is based on the Accept-Reject algorithm.
3. Represent the histogram of the realizations and compare it with the exact density f_X for $n = 10^4$.
4. Give an estimation of both the expected value and the variance of X and compute in addition a 95% confidence interval on both quantities.
5. Compute a numerical approximation of the normalizing constant C such that

$$\int_0^1 f_X(x)dx = 1.$$

Exercice 5
We consider the Gamma distribution $\mathcal{G}(k,\theta)$ with support $\mathbb{R}^+$ and density

$$f(x;k,\theta) = \frac{\theta^k}{(k-1)!} x^{k-1} e^{-\theta x}$$

for $k \in \mathbb{N}^*$ and $\theta > 0$.

1. We recall a property of the Gamma distribution : **if** $X_1, X_2, \ldots, X_k$ are independent realizations of the Exponential law $\mathcal{E}(\theta)$ with cumulative distribution function $F(x) = 1 - e^{-\theta x}$ and **if** $Z = X_1 + X_2 + \cdots + X_k$, **then** $Z \sim \mathcal{G}(k,\theta)$. Deduce from this property an algorithm which simulates a sample of size n from a Gamma distribution $\mathcal{G}(k,\theta)$, when k is a (strictly) positive integer.
(Note that the use of the R function `rexp` is allowed)
2. Illustrate graphically the validity of this algorithm using $k = 3$, $\theta = 0.1$, $n = 10000$.
3. Based on the output of the above algorithm, give a Monte Carlo estimate of both the expectation and variance of the distribution $\mathcal{G}(3, 0.1)$. Compare them with the theoretical values k/θ and k/θ^2 respectively.
4. Give a 95% confidence interval for the expectation of the distribution $\mathcal{G}(3, 0.1)$.

Exercice 6
Take $a \in \mathbb{R}$ and $k < -1$. The cdf of the Pareto distribution with parameters (a, k) is given by

$$F(x) = \begin{cases} 0 & \text{when } x \leq a \\ 1 - \left(\frac{x}{a}\right)^k & \text{otherwise} \end{cases}$$

1. Write an R function called `rpareto` with arguments n, a positive integer, and the parameters a and k, which returns n independent simulations from the Pareto(a, k) distribution. This function must use exactly n independent simulations from the uniform $\mathcal{U}_{[0,1]}$ distribution.
2. For a pair of parameters (a, k) equal to $(-3, -3)$, illustrate by a graphical representation the adequacy of this simulation method.
3. Using a sample of size 10^5 simulated via `rpareto`, give a confidence interval at confidence level asymptotically equal to 75% for the mean of a Pareto$(5, -4)$ distribution.

Exercice 7
Create an R code that finds all pairs of so-called "friendly squares", namely the pairs of perfect squares $10^4 > x^2 > 10$ and $10^4 > y^2 > 10$, x and y being integers, such that (a) they have the same number of digits and (b) the translation of all digits of x^2 by a given integer τ (modulo 10) leads to y^2. *Example :* 121 and 676 are friendly squares.