Project 1
Computational Numerical Statistics
DM, FCT-UNL
Group: G6

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Deadline: 28/10/19

Random variable generation

In the resolution of the following problems **do not** use any of the `R` random variable generation built-in functions, nor any `R` function referring to the densities of random variables.

Fix your `R` random seed to 123 in all simulations.

1. Let $X \sim Laplace(\mu, \lambda)$, which has probability density function (p.d.f.)

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}, \qquad \mu \in \mathbb{R}, \qquad \lambda > 0, \qquad x \in \mathbb{R}.$$

   (a) Show that $Y = |X - \mu| \sim Exp(\lambda)$.

   (b) It can be shown that the random variable X may defined as

$$X = \begin{cases} \mu + Y & \text{with prob.}0.5 \\ \mu - Y & \text{with prob.}0.5 \end{cases}.$$

   Use this result to implement in `R` a routine `sim.lap()` for generating a sample from the Laplace distribution. Let routine `sim.lap()` receive as input a generic sample size m as well as the Laplace parameters μ and λ. Provide both algorithm and **R** code.

   (c) Use the `sim.lap()` routine to generate a sample of size $m = 10000$ of f.

   (d) Plot the histogram referring to (b) with the true p.m.f. superimposed.

2. Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} 1 + x & \text{if } -1 \le x \le 0 \\ 1 - x & \text{if } \quad 0 \le x \le 1 \end{cases}.$$

   (a) Derive the cumulative distribution function (c.d.f.) F of X.

   (b) Describe and implement the **inverse-transform** method in `R` for generating a sample from f. Call your routine `sim.f.itm()` and let it receive as input a generic sample size m. Provide both algorithm and **R** code.

   (c) Describe and implement the **acceptance-rejection** method in `R` for generating a sample from f. Call your routine `sim.f.arm()` and let it receive as input a generic sample size m. Provide both algorithm and **R** code.

(d) Use routines `sim.f.itm()` and `sim.f.arm()` to generate a sample of size $m = 10000$ of f.

(e) Plot the histograms referring to (d) with true p.d.f. superimposed.

(f) Display the the hit-and-miss plot referring to `sim.f.arm(10)`.

Monte Carlo integration

In the resolution of the problems in this section you may already use the `R` random variable generation built-in functions as well as any `R` function referring to the densities of random variables.

Fix your `R` random seed to 456 in all simulations.

3. Let $\mathcal{I} = \displaystyle\int_0^1 \frac{e^{-x}}{1+x^2} \mathrm{d}x$.

   (a) Use the `R` function `integrate()` to compute the value of $\mathcal{I}$.

   (b) Describe and implement in `R` the Monte Carlo method of size $m = 10000$ for estimating $\mathcal{I}$. Report an estimate of the variance of the Monte Carlo estimator $\hat{\mathcal{I}}_{MC}$ of $\mathcal{I}$.

   (c) Describe and implement in `R` the Monte Carlo method of size $m = 10000$ based on **control variables** for estimating $\mathcal{I}$. Report an estimate of the variance of the Monte Carlo estimator $\hat{\mathcal{I}}_C$ of $\mathcal{I}$.

   Report an estimate of the variance of the Monte Carlo estimator $\hat{\mathcal{I}}_C$ of $\mathcal{I}$.

   (d) What's the percentage of variance reduction that is achieved when using $\hat{\mathcal{I}}_C$ instead of $\hat{\mathcal{I}}_{MC}$?

Hypothesis testing

In the resolution of the problems in this section you may already use the `R` random variable generation built-in functions as well as any `R` function referring to the densities of random variables.

Fix your `R` random seed to 789 in all simulations.

4. Let $X_1, \ldots, X_n$ be a random sample from the population $X \sim N(4, 2)$.

   (a) Validate via a Monte Carlo simulation study that $T = \dfrac{\overline{X} - 4}{S/\sqrt{n}} \sim t_{n-1}$. Consider in your study the number of simulations $m = 1000$ and a sample size of $n = 25$. For validation purposes provide:

   - the histogram of the simulated $T_1, \ldots, T_{1000}$ with the theoretical density superimposed;
   - the plot of the empirical cumulative distribution (ecdf) of $T_1, \ldots, T_{1000}$ with the theoretical cumulative probability function (pdf) superimposed (use the `R` built-in `ecdf()` function to create the sample ecdf);

- perform a Kolmogorov-Smirnov two-sided test to test your hypothesis (use the `R` built-in `ks.test()` function);

- Compute the theoretical quantiles $0.90,\ 0.95, 0.975$ with the empirical quantiles (use the `R` built-in `qchisq()` and `quantile()` functions, respectively).

(b) Assume one wants to test the hypothesis

$$H_0 : \mu = 4 \qquad\qquad H_1 : \mu \neq 4$$

at the significance level $\alpha = 0.05$. Compute the empirical p-values for this test using the previous Monte Carlo simulation. Compute the empirical significance level $\hat{\alpha}$ and perform the binomial test to assess if $\hat{\alpha}$ departs significantly from α.

(c) For the hypothesis test above and the m simulations, construct a power plot for the alternative values $\mu = 5, \ldots, 49, 50$. In face of these results, how far from H_0 does one need to be so that the power of the test gets higher than 80%?