Problems on Monte Carlo methods

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. For any statistical test, the p-value obtained for a random sample drawn under H_0 follows a uniform distribution in the unit interval. Simulate $MC = 1000$ samples of $n = 100$ observations of a $\mathrm{U}(0, 1)$ distribution and obtain the p-value of the (two-sided) KS test whose null hypothesis establishes that the sample follows a $\mathrm{U}(0, 1)$ distribution. Check (using the KS test again) that the 1000 resulting p-values are distributed as $\mathrm{U}(0, 1)$

2. Consider the sepal length of the 50 *iris setosa* plants of dataset `iris`.

   a. Run the Shapiro-Wilk normality test (`shapiro.test`) on them in order to check that they are normally distributed.
   b. Run the KS test with H_0 estalishing that the data are normally distributed with mean and variance equal to those estimated from the sample and keep the test statistic.
   c. Simulate $MC = 1000$ samples of $n = 50$ observations of a normal distribution (use function `rnorm`). For each of them, run the KS test with H_0 estalishing that the data are normally distributed with mean and variance equal to those estimated from the corresponding simulated sample and keep the test statistic.
   d. Determine the fraction of test statistics of the simulated samples that are greater that the test statistic obtained for the real dataset in b. It approximates the p-value of the normality KS test (with unespecifed parameters).

3. An insurance company has clients classified as high, medium, and low risk. These clients have a probability of claiming equal to 0.02, 0.01, and 0.0025 respectively. If the probability of being a client of high risk is 0.1, 0.2 of medium risk, and 0.7 of low risk. Approximate the probability that a claim selected at random comes from a hight risk client by selecting clients at random until $MC = 10000$ claims are found and compute then the ratio of the claims that come from high risk clients. Obtain a 99% CI on the probability.

4. Build a 99% CI on $\int_0^1 \exp\{e^x\}\,\mathrm{d}x$ whose width is 0.05 units. Simulate first $MC = 1000$ observations to estimate the standard deviation of $\exp\{e^U\}$ and compute then the number of observations needed.

5. $\int_{-2}^{2} e^{x+x^2}\,\mathrm{d}x$

6. $\int_0^1 \int_0^1 e^{(x+y)^2}\,\mathrm{d}y\,\mathrm{d}x$

7. $\int_0^\infty \int_0^x e^{-(x+y)}\,\mathrm{d}y\,\mathrm{d}x$