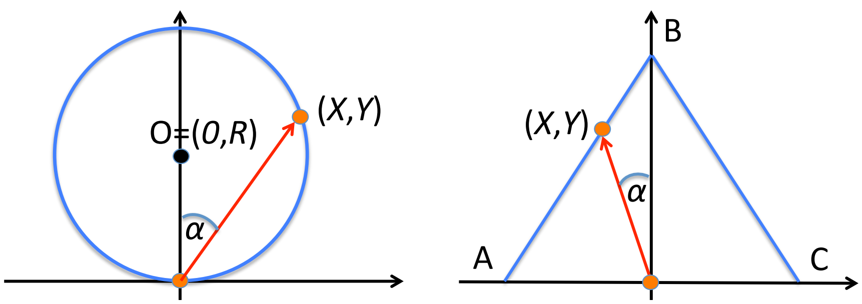
List 1 of exercises (1st semester 2020). Monte Carlo methods.
Prof.A.Iambartsev (yambar@ime.usp.br)

This is an individual work. Choose items summing up 10 points. Report solutions and used codes on an editor of your choice and send me pdf file. You have some doubts or you found some misprinting or mistakes in formulations, or you think that it is mistake, then inform me on our g-meeting on Wednesday, 29/04 or by e-mail.

1. Suppose we want to generate a reflection vector for a particle moving into a bounded domain. There are various methods for simulation of a vectors. Recall that a vector (bi-dimensional) is an directed interval, arrow, that consist on the initial point, in our case $(0,0)$ and the second point, direction, (x,y). Consider the following generator of the vector.



See the right-hand side figure: the direction point (x,y) is chosen uniformly in the intervals $[A,B]\cup[B,C]$.

(a) (1 point) Let $A=(-1,0), B=(0,3), C=(a,0), a>0$. Find the distribution (density) of the angle α for this simulation procedure. And suggest an direct (from the density of α) simulation of the angle α.

See the left-hand side of the figure: the direction point (x,y) is chosen uniformly from the circle with radius R.

(a) (1 point) Find the distribution (density) of the angle α for this simulation procedure. Suggest an direct (from the density of α) simulation of the angle α.

2. [2], (1 point) Construct the algorithm for simulating a random variable which density function is given by the following formula

$$f(u)=\int_{1/2}^{\infty}\frac{e^{-4uv}}{v^4}dv,\quad u>0,$$

and construct the histogram based on 1000 simulated points.

3. [2], A random vector (X,Y) has the following joint distribution

$$f(x,y)=3x^2(1+x)y^xe^{-x^3},\quad x>0, y\in(0,1).$$

(a) (1 point) Construct the algorithm to generate and plot the histogram based on 1000 simulated points.

(b) (1 point) Suggest the numerical (Monte Carlo) calculation of the expectation $\mathbb{E}(XY)$, check the conditions of convergence plotting 95% confidence interval.

4. [2], A random variable has the following density

$$f(x)=\frac{\sin x}{4}+\frac{e^x}{4(e^{\pi/2}-1)}+\frac{1}{\pi},\quad 0<\pi<\frac{\pi}{2}.$$

(a) (1 point) Suggest an efficient method to generate the random variable and plot the histogram based on 1000 simulated points.

(b) (1 point) Suggest the numerical (Monte Carlo) calculation of the expectation $\mathbb{E}(X)$, check the conditions of convergence plotting 95% confidence interval.

5. [3], (1 point) Suggest an efficient method to generate the vector (X,Y) with the distribution

$$\mathbb{P}(X=x,Y=y)=\frac{\theta^{x+y}(1+x)^y}{x!y!}e^{-2\theta-\theta x},\quad x,y\in\{0,1,2,\dots\}.$$

6. [2], (1 point) Suggest an efficient method to generate the random variable with the density

$$f(x) = \frac{5x}{(1+x^2)^2} + \frac{\pi}{4\sqrt{2}x^2} \sin\left(\frac{\pi}{2x}\right), \quad 1 < x < 2,$$

and construct the histogram based on 1000 simulated points.

7. [2], Let X be a discrete random variable with $M, M < \infty$, values $X \in \{x_1, \ldots, x_M\}$ and respective probabilities $p_m = \mathbb{P}(X = x_m)$. Using the generalization of inverse of cumulative distribution function we construct an algorithm for simulating the variable X.

   (a) (1 point) Describe the algorithm. There are two operations used in the algorithm: (a) generating uniform distribution, and (b) comparison operator, $c < d$. Let a be the cost of the first operation (a) and let b be the cost of the comparison operator (b). Find the average cost for generating discrete random variable X.

   (b) (2 points) Prove that the optimal (that minimizes the average cost of generation) version of the algorithms from the previous item should re-arrange the probabilities in the following order:

$$p_{(M)} \geq p_{(M-1)} \geq \cdots \geq p_{(1)}.$$

8. The density $f(x, y)$ of bi-dimensional vector (X, Y) is inversely proportional to the $1 + x^2 + x^2y^2 + y^2$ on the square $x, y \in [-2, 2]$, i.e.

$$f(x, y) = \frac{c}{1 + x^2 + x^2y^2 + y^2}, \quad x, y \in [-2, 2].$$

   (a) (1 point) Construct the algorithm for simulation of the vector (X, Y). Plot the function $f(x, y)$. Simulate 1000 points of (X, Y) and construct the corresponding histogram.

   (b) (1 point) Using Monte Carlo method ("classical") calculate the probability $\mathbb{P}(|X - Y| \leq 0.5)$. Check empirically (plotting the 95% confidence interval) the convergence for the true value.

   (c) (1 point) Using Monte Carlo ("classical") method calculate the integral

$$I = \int_0^2 dx \int_{-2}^1 dy \frac{xy}{1 + x^2 + x^2y^2 + y^2}.$$

   Check empirically the convergence adding the 95% confidence interval.

9. (Inspired by Example 1.1, [1]) Let X be uniformly distributed random variable on the set $\{-1, 0, 1, 2, 3, 4, 5\}$. Define $Y = \lfloor 1 + X/2 + \eta \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part of a number, and η has the normal distribution $N(0, 10)$. The Pearson correlation coefficient, ρ, we will use as the measure of dependence between these two variables. The estimator of ρ based on a sample of size n is

$$r_n = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_x s_y}.$$

   (a) (1 point) Generate large numbers of samples of size $n = 20$ according the model, plot the density distribution for r_{20}: construct the histogram, and then, smooth the histogram using a kernel operators (using Gaussian, for example). Report 2.5%- and 97.5%-quantiles.

   (b) (1 point) Generate a sample with the same sample size. Using large number of (non-parametric) bootstrap replicates plot the histogram for the bootstrapped density for r_{20} together with the previous smoothed histogram, comment. Report 2.5%- and 97.5%-quantiles.

10. Suppose the random variable X has beta distribution with parameters α, β. Fix some values for the parameters $\alpha = \alpha_0, \beta = \beta_0$. Generate 10 values from beta distribution with chosen parameters α_0, β_0. Choose some estimators for the parameters $\hat{\alpha}_n$ and $\hat{\beta}_n$, for example, MLE estimators or using method of moments.

   (a) (1 point) Given a sample size 10, we are interested in the distribution of the estimators $\hat{\alpha}_{10}$ and $\hat{\beta}_{10}$. Generate a large number of samples of size 10 and construct histograms for $\hat{\alpha}_{10}$ and $\hat{\beta}_{10}$ smoothing theirs with a kernel operator.

   (b) (1 point) Generate and fix some sample with the same sample size 10. Generate a large number of bootstrapped values of statistics $\hat{\alpha}_{10}$ and $\hat{\beta}_{10}$. Construct the respective bootstrapped histograms. Compare and comment if there exists a difference between the bootstrapped histogram and the histogram from the previous item.

11. In order to calculate the integral below we use the instrumental random variable with exponential distribution of rate $\lambda = 0.5$ for the importance sampling numerical MC method.

$$I = \int_0^\infty \frac{1}{1+e^x} dx.$$

Find the value of integral I.

   (a) (1 point) Check that the variance of importance sampling estimator $\hat{I}_n$ is finite. Can you find the variance explicitly? Plot the convergence of importance sampling estimator $\hat{I}_n$ to the true value, adding the 95% confidence interval to control the convergence. (Construct the confidence variance with unknown variance.)

   (b) (1 point) Consider another estimator $\hat{I}_n^{(1)}$ based on instrumental random variable with density function $g(x) = e^{-x}$. Is the variance of $\hat{I}_n^{(1)}$ finite? In the constructed plot add the convergence of estimator $\hat{I}_n^{(1)}$ with corresponding 95% confidence interval. (Even if the variance is infinite, calculate formally the confidence interval for unknown variance and plot it.)

12. (1 point) Let U_1, U_2 are independent uniformly distributed random variables. Let us follow the following simulation (of a random variable) algorithm: (1) generate two numbers u_1, u_2 from the variables U_1, U_2; (2) if $u_1 > u^3$ then we accept the value of u_1, and reject otherwise. Find the density and cumulative distribution function of the generated random variable. Provide the distribution of acceptance time: the time we spent up to the first acceptance of the value u_1 as a value of generated random variable.

13. (1 point) The two parameters (P_1, P_2) was measured in 10 patients (in the R command line):

```
p1<-c(19.15, 11.72, 19.89, 8.70, 19.25, 12.64, 13.86, 13.88, 10.64, 11.51)
```

```
p2<-c(14.23, 10.85, 15.83, 14.49, 17.17, 17.19, 22.13, 16.10, 13.53, 9.03)
```

Construct and execute numerically the bootstrap procedure to test the hypothesis about the existence of a dependence between parameters, for example, test the hypothesis, that the correlation coefficient is equal or does not to zero. State the null hypothesis and construct the histogram of bootstrapped values of correlation coefficient under the null hypothesis.

Referências

[1] Robert, C. P., Casella, G., & Casella, G. (2010). Introducing Monte Carlo methods with R (Vol. 18). New York: Springer.

[2] Voitishek, Lecture on Monte Carlo methods. Notes on Novosibirsk University course.

[3] Prof. Luís Esteves, Entrance exams problems.