Modelling of Complex Systems

Department of Physics, University of Aveiro.

Project 1.

Probability and probability density.

The central limiting theorem.

Binomial, Poisson and Gaussian distributions.

Part 1 — Probability

Task 1.1 A random number x can take the values $1,2,3, \ldots M$ uniformly at random, i.e., with the same probability $1/M$ each value. Prove **analytically** that the averaged value $\langle x \rangle$ of the random number x is

$$\langle x \rangle = \frac{M+1}{2}, \qquad (1)$$

and the variance σ^2 is

$$\sigma^2 = \langle \delta x^2 \rangle = \frac{M^2-1}{12}, \qquad (2)$$

where $\delta x = x - \langle x \rangle$ are the fluctuations. Show **analytically** that, for $M = 100$, the probability that x is not greater than 60 is $p = 0.6$.

Task 1.2 Check Eqs. (1) and (2), and $p = 0.6$ **numerically**.

Algorithm.
Generate N random integer numbers x_i (with $i = 1,2, \ldots N$), where each $x_i = 1$, or 2, or 3, ... or M, uniformly at random. Find the average value and the variance:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i ,$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2 ,$$

Repeat for the numbers of trials $N = 10^2, 10^4, 10^6$. Compare the numerical results with the theoretical predictions of Eqs. (1) and (2). Show that the mean value $\langle x \rangle$ and, the variance σ^2, and the probability p (the probability that x is smaller than or equal to 60) tends to the theoretical predictions as N increases.

Task 1.3 Show that, if two random numbers $x = \mathrm{rand}(1)$ and $y = \mathrm{rand}(1)$ are uncorrelated random variables, then the mean value of a random variable $z = xy$ is $\langle z \rangle = \langle x \rangle \langle y \rangle$.

Algorithm. Generate N pairs of random numbers $x = \mathrm{rand}(1)$and $y = \mathrm{rand}(1)$. Find their product, $z_i = x_i y_i$, $i = 1,2, \ldots N$. Then find the mean values $\langle x \rangle$, $\langle y \rangle$, and $\langle z \rangle$:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i , \quad \langle y \rangle = \frac{1}{N} \sum_{i=1}^{N} y_i , \text{ and } \langle z \rangle = \frac{1}{N} \sum_{i=1}^{N} z_i .$$

Compare the numerical results with $\langle z \rangle = \langle x \rangle \langle y \rangle$ for numbers of trials $N = 10^2, 10^4, 10^6$.

Part 2 — Probability density

Task. 2.1. Find numerically the probability density distribution function $P(x)$ of a random variable $x = \mathrm{rand}(1)$, i.e. x is a random number in the interval $0 \le x < 1$. Plot $P(x)$.
Task. 2.2. Find numerically the mean value $\langle x \rangle$, and the variance $\sigma^2 = \int_0^1 (x - \langle x \rangle)^2 P(x) dx$. Compare with the expected values $\langle x \rangle = 1/2$, $\sigma^2 = 1/12$.
Task. 2.3. Find the probability density $Q(y)$ of a random variable y defined as $y = \sqrt{\mathrm{rand}(1)}$. Plot $Q(y)$ versus y. Compare with the theoretical result

$$Q(y) = 2y.$$

Use the binning procedure of steps 3 and 4 of the algorithm given below, in Part 3.

Parameters: Numbers of iterations $N = 10^2, 10^4, 10^6$. Width of the bins $\Delta x = 0.005$.

Part 3 — The Central Limit Theorem

Task. 3.1. Generate n random numbers x_i with a mean value $\langle x \rangle$ and the variance σ^2. For example, you can use the uniform random number generator $x = \mathrm{rand}(1)$. The random number Y is defined as a sum of n random numbers x_i,

$$Y = \frac{1}{n} \sum_{i=1}^{n} x_i \,.$$

Find the probability density distribution function $P(Y)$ of the random numbers Y. Plot $P(Y)$. Show that the mean value of Y is equal to $\langle x \rangle$. Calculate the variance of Y,

$$\Lambda^2 = \langle (Y - \langle Y \rangle)^2 \rangle \,,$$

and show that Λ^2 tends to ${\sigma^2}/{n}$ when the number of samples of Y goes to infinity.
Algorithm.

1. Using $x = \mathrm{rand}(1)$, generate n random numbers x_i and calculate Y.
2. Repeat step 1 N times. You will get N random numbers Y_m, $m = 1,2, \dots N$.
3. Divide the interval $[0,1]$ into bins of width Δy, i.e., intervals $[k\Delta y, (k+1)\Delta y]$ where $k = 0,1, \dots k_{\max} = \frac{1}{\Delta y} - 1$. Get the numbers of Y_m in each bin. Let this number in the bin k be $M(k)$. Calculate the distribution function $P(y_k)$ (the probability density),
$$P(y_k) = \frac{M(k)}{N \Delta y}.$$
The parameter $y_k = (k + 0.5)\Delta y$ is the centre of bin k. Check the normalization condition $\sum_{k=0}^{k_{\max}} P(y_k) \Delta y = 1$.
4. Calculate the mean value and the variance of Y_m,
$$\langle Y \rangle = \sum_{k=0}^{k_{\max}} y_k P(y_k) \Delta y \,,$$
$$\Lambda^2 = \sum_{k=0}^{k_{\max}} (y_k - \langle Y \rangle)^2 P(y_k) \Delta y \,.$$
5. Calculate mean value $\langle x \rangle$ and the variance σ^2, averaging over all trails:
$$\langle x \rangle = \frac{1}{N} \sum_{m=1}^{N} \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) ,$$
$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \frac{1}{N} \sum_{m=1}^{N} \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \langle x \rangle)^2 \right) ,$$

Parameters: $n = 10,100,1000$. $\Delta y = 0.005$, the number of trials $N = 10^6$.

Part 4 — Throwing balls

Task. 4.1

There are M boxes and N balls. We throw the balls into the boxes. The N balls are distributed among the boxes uniformly at random. Find the probability to find n balls in a chosen box (for example, a box number 3). Parameters: $M = 9$, $N = 21$, the number of attempts $K = 10^2, 10^4, 10^6$.
[Note that this task is completely equivalent to gambling and Russian roulette, which are discussed in lecture 2. To see this assume that the appearance of a ball in the box 3 corresponds to appearance of the number 777 in the slot machine and the total number N of balls is the number of coins that you have. Then the probability to have n balls in the box 3 after N attempts is the probability to win n times using N coins. To see the equivalence to Russian roulete, assume that appearance of a ball in the box 3 corresponds to a deadly shot. Then the number of balls in the box 3 is the number of deaths after N shots. Chose $M = 6$ (a cylinder of a revolver with 6 rounds). In your report, if you wish you can use any of these models that you like more.]
Algorithm.

1. Let us measure the number of balls **in the box with the number 3**. Generate at uniformly at random N numbers x_i, with $i = 1,2, \dots N$, from the set of integers $1,2, \dots M$.
2. Calculate how many times the number 3 appear among the N integers x_i (i.e., x_1, $x_2, \dots x_N$). Let that number be denoted n.
3. Repeat steps 1 and 2 K times (K trials, or rounds). We will get K random numbers n_j, with $j = 1,2, \dots K$.
4. Calculate the number of attempts $N_{\mathrm{tr}}(n)$ which gave n balls in the box 3. Calculate the probability to find n balls in box 3 as follows:
$$P(n) = \frac{N_{\mathrm{tr}}(n)}{K}$$
5. Plot $P(n)$ versus n, for $n = 0,1,2, .. N$. On the same figure, plot the binomial, Poisson and Gaussian distributions:
$$B(n, N) = C_n^N p^n (1-p)^{N-n} ,$$
$$P_n(Np) = \frac{(Np)^n}{n!} e^{-Np} ,$$
$$G(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\langle n \rangle)^2}{2\sigma^2}} ,$$

Here, $p = 1/M$, $\langle n \rangle = \sigma^2 = Np$. Compare between the $P(n)$ obtained numerically and the theoretical distributions. What distribution, Binomial, Poisson or Gaussian, describes better the numerical results given by $P(n)$ (just by eye)?