Problem sheet 1 solutions for the Lecture course "Algorithms and Computations in Physics" 2024

Ylias Sadki, Werner Krauth

University of Oxford

(Dated: 25 January 2024)

1. **direct-pi** Implement the Algorithm **direct-pi** in Python and confirm that the output approaches $\pi/4$. Modify the program, so that it runs many times, for a given number or trials, say $N = 1000$. Plot a histogram of the output and characterize it completely (use that the variance of a Bernoulli random variable with parameter θ is $\theta(1-\theta)$).

   Each sample is drawn from the Bernoulli distribution

   $$B = \begin{cases} 1 & \pi/4 \\ 0 & 1-\pi/4 \end{cases}$$

   Each "run" is composed of N pebble throws:

   $$R = \frac{1}{N}\sum_{i=1}^{N} B_i \tag{1}$$

   For each run, R will be sampled from it's own distribution, with moments:

   $$\begin{aligned} \langle R \rangle &= \frac{1}{N}\langle \sum B_i \rangle = \frac{1}{N} N \langle B \rangle = \frac{\pi}{4} \\ \mathrm{Var}(R) = \langle R^2 \rangle - \langle R \rangle^2 &= \frac{1}{N^2}\mathrm{Var}(\sum B_i) = \frac{1}{N^2} N \mathrm{Var}(B) = \frac{1}{N}(\pi/4 - (\pi/4)^2) \end{aligned}$$

   Since the Bernoulli distribution has finite mean $\langle B \rangle$ and variance $\mathrm{Var}(B)$, the central limit theorem holds for R, so that as N is large, each R is drawn from a Gaussian distribution with mean and variance given above:

2. **walker** In Lecture 1, we discussed Walker's algorithm, a prodigious sampling method of the distribution $\{p_1, \dots, p_K\}$ in $\mathcal{O}(1)$ per sample as $K \to \infty$.

   - Implement the table construction for this algorithm and explain why at each stage of this construction, one starts with one small stick and one tall stick. What would happen if the construction started with the tall sticks?

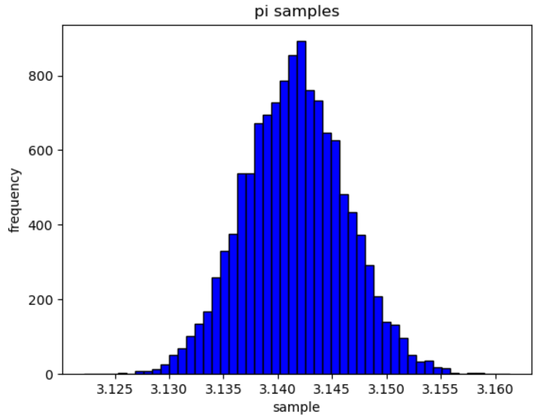

FIG. 1: Samples of direct-pi

Short sticks are placed first, then tall sticks on top, so there is a guarantee of a maximum of two sticks in each column. If we started by placing tall sticks, you would keep placing until you only had short sticks. Then you would stack short sticks into (perhaps multiple) towers. Sampling the algorithm might then go as log(N), as you effectively have tower sampling.

- Implement the sample stage for this algorithm and explain why it is $\mathcal{O}(1)$ per event.

  The algorithm is $\mathcal{O}(1)$ per event, as each sample you throw one random variable, read 3 numbers, then throw another random number. As there are only two sticks in each column, you do not have to tower search, so it scales as $\mathcal{O}(1)$.

- Test your implementation in a "Saturday-night" problem of your choice and make sure that it actually picks any element i with probability p_i, to good precision.

3. **Convergence to cumulative distribution** (optional) In Lecture 1, we discussed the Alg. **direct-gamma**, obtained from samples $\texttt{ran}(0,1)^\gamma$. We learned that the random variable corresponding to this algorithm, for $\gamma < 1/2$ is very tricky.

   - Take $\gamma = -0.75$. What is the probability distribution $\pi(\mathcal{O})$ of $\mathcal{O} = \texttt{ran}(0,1)^\gamma$?

     As indicated in eq. (19), $\mathcal{O}$ has for $\gamma < 0$ the distribution

$$\pi(\mathcal{O}) = \frac{-1}{\gamma}\mathcal{O}^{-1+1/\gamma} \overset{\gamma=-3/4}{=} \frac{4}{3}\mathcal{O}^{-7/3} \quad \textit{for } 1 \le \mathcal{O} < \infty, \tag{2}$$

which is normalized. The expectation (mean value) of $\mathcal{O}$ can be computed in two ways: On the sample space $[0,1]$, *it is given by*

$$\langle \mathcal{O} \rangle = \frac{\int_0^1 \mathrm{d}x \; x^{-0.75}}{\int_0^1 \mathrm{d}x} = 4, \tag{3}$$

mirroring what is written in eq.(4). But we can also compute the mean value as

$$\langle \mathcal{O} \rangle = \frac{4}{3} \int_1^\infty \mathrm{d}\mathcal{O} \; \mathcal{O}\mathcal{O}^{-7/3} = 4. \tag{4}$$

The second moment of $\mathcal{O}$ is infinite, whether we compute it on the sample space,

$$\frac{\int_0^1 \mathrm{d}x \; [x^{-0.75}]^2}{\int_0^1 \mathrm{d}x} = \infty, \tag{5}$$

or determine it as

$$\frac{4}{3} \int_1^\infty \mathrm{d}\mathcal{O} \; \mathcal{O}^2\mathcal{O}^{-7/3} = \infty \tag{6}$$

NB: The cumulative distribution function of the distribution of $\mathcal{O}$, as $\gamma = -0.75 < 0$, is

$$F(y) = \begin{cases} 0 & \textit{for } \mathcal{O} \le 1 \\ \frac{4}{3} \int_1^y \mathrm{d}\mathcal{O} \; \mathcal{O}^{-7/3} = 1 - \frac{1}{y^{4/3}} & \textit{for } 1 < \mathcal{O} < \infty \end{cases}, \tag{7}$$

and it corresponds to the cumulative histogram of $x^{-0.75}$ on the sample space $[0,1]$.

- Write a Python program to test explicitly what you just found out.

  See `AlgebraicHistogram.py` *and* `AlgebraicCumulativeHistogram.py`. *and the output in Fig. 3.*

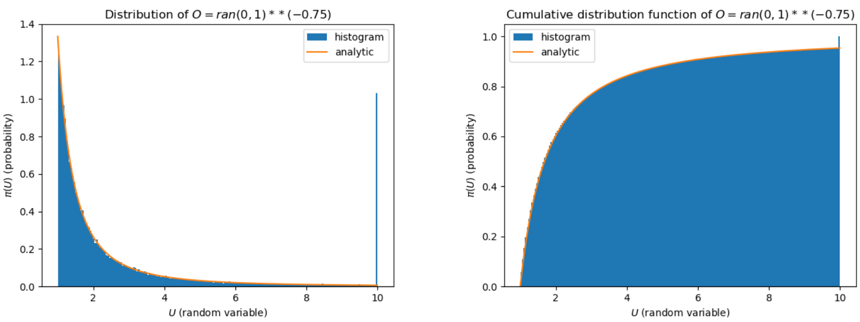

FIG. 2: Comparison of theory and experiment for $\mathtt{ran}(0,1)^{-0.75}$. (a): Distribution density. (b): Cumulative distribution function.

- Modify your program, so that it generates N samples of $\texttt{ran}(0,1)^\gamma$. Sort these samples so that $x_1 < x_2 < \cdots < x_N$ and compute the empirical cumulative distribution function $\hat{F}(x)$, the function that is zero for $x < x_1$, that is $1/N$ for $x_1 < x < x_2$, etc. Compare $\hat{F}$ to the true cumulative distribution function $F(x)$ that you obtained analytically.
- Explicitly check the Glivenko–Cantelli Theorem, stating that

$$\sup_x |\hat{F}_N(x) - F(x)| \to 0 \quad \text{for } N \to \infty.$$

For added value, we now consider the distribution that corresponds to $\texttt{ran}(0,1)^\gamma$ *with* $\gamma > 0$*, and take the example* $\gamma = 0.75$*. We now have that the random variable (the observable)* $\mathcal{O}$ *is between* 0 *and* 1*, with the distribution:*

$$\pi(O) = \frac{1}{\gamma}\mathcal{O}^{1/\gamma-1} \overset{\gamma=+3/4}{=} \frac{4}{3}\mathcal{O}^{1/3} \tag{8}$$

The integrated distribution function, for the case $\gamma > 0$*, is given by:*

$$F(x) = \int_0^x \frac{1}{\gamma}\mathcal{O}^{1/\gamma-1} = x^{1/gamma} \overset{\gamma=+3/4}{=} x^{4/3} \quad \textit{for } 0 < x < 1 \tag{9}$$

Outside this window, the distribution is zero $(x < 0)$ *or one* $(x > 1)$*.*

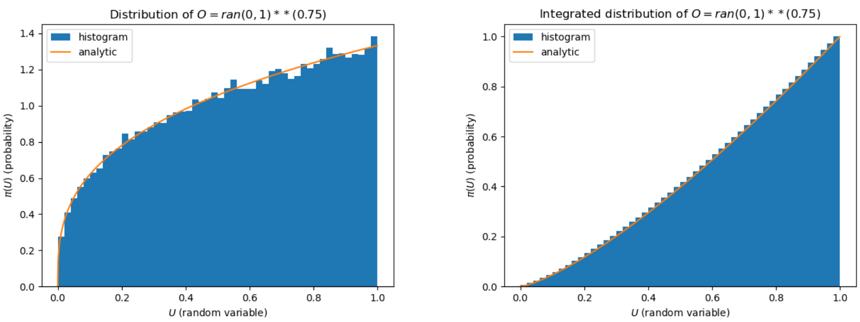

FIG. 3: Comparison of theory and experiment for $\texttt{ran}(0,1)^{+0.75}$. (a): Distribution density. (b): Cumulative distribution function.

The solution for the cumulative histogram, proposed by the plotting routine, consists in equal-width bars, but this is non-ideal. To compare how $N = 20$ *(or so) samples of a distribution approach the latter, it is much better to sort the data and use the empirical cumulated distribution. In our example, we write a program to sample* N *data:*

$$\mathcal{O}_1 < \mathcal{O}_2 < \cdots < \mathcal{O}_N \tag{10}$$

and we write the empirical cumulative distribution function

$$\hat{F}(\mathcal{O}) = \begin{cases} 0 & \mathcal{O} < \mathcal{O}_1 \\ 1/N & \mathcal{O}_1 \leq \mathcal{O} < \mathcal{O}_2 \\ etc & \end{cases} \tag{11}$$

The sup mentioned in the theorem is illustrated in Fig. 4.

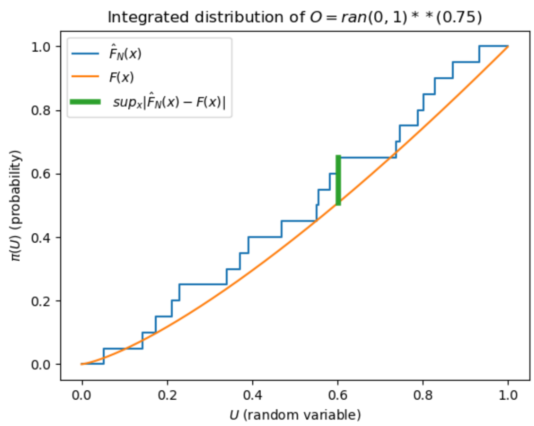

FIG. 4: Maximum distance for $N = 20$, as used in the GlivenkoCantelli theorem. We should now study how this approaches 0 as $N \to \infty$

- If you have more time, check the Dvoretzky-Kiefer-Wolfowitz inequality, that states:

$$\mathbb{P}\left(\sup_x |\hat{F}_N(x) - F(x)| > \epsilon\right) \leq 2\mathrm{e}^{-2N\epsilon^2}$$

  This inequality, miraculously, does not depend on the distribution you're sampling, as long as F is continuous (see Wasserman, Theorem 7.5).