Probability and Statistics Reminders

National University of Singapore
A. Thiery

1 Uniform Distribution

A uniform distributino on the unit interval $[0, 1]$ is a distribution with density $f(x) = 1$ for $x \in [0, 1]$ and 0 otherwise. Detail the computation of the expectation and variance of a random variable X with a uniform distribution on $[0, 1]$.

2 Larger than two standard deviations

Use simulation to estimate the probability that a random variable X with a Gaussian distribution $\mathrm{N}(0, 1)$ is larger than two standard deviations, i.e. $\mathrm{P}(X > 2)$. Compare the empirical probability to the exact value.

3 Computation of variance

Consider the distribution on $[-1, 1]$ with density given by

$$f(x) \propto x^2.$$

Compute the expectation and variance of a random variable X with this distribution.

4 Sample average

Consider $N \geq 1$ independent random variables $X_1, \ldots, X_N$ with expectation μ and variance σ^2. Define the sample average as

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Compute the expectation and variance of $\bar{X}_N$. Detail the computation.

5 Simulating an aproximate Gaussian distribution

Consider $N = 6$ independent random variables $X_1, \ldots, X_N$ with a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. Define the random variable

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i.$$

The random variable Z is often used to approximate a Gaussian distribution.

1. Compute the expectation and variance of Z.
2. Simulate a large number of samples of Z and plot the histogram of the samples. Superimpose the density of a Gaussian distribution with the same expectation and variance.

6 Running average and LLN

Consider a sequence of independent random variables $X_1, X_2, \ldots$ that are uniformly distributed on $[-1, 1]$. Figure 1 shows the evolution of the running average. Reproduce this figure by simulating the running average of a large number of samples.

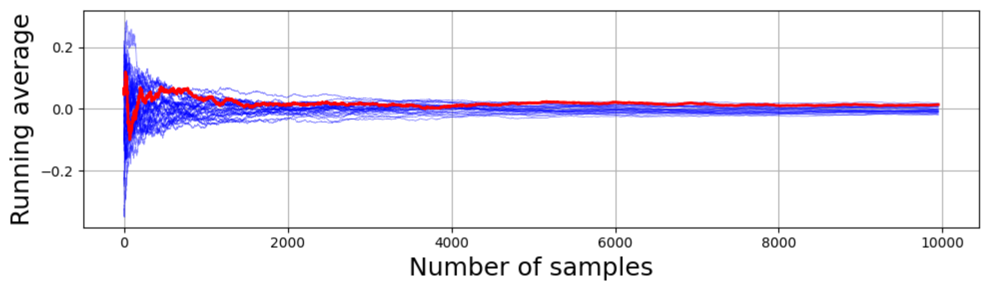

Figure 1: Running average of a sequence of random variables

7 Sample average with correlated random variables

Consider $N \geq 1$ random variables $X_1, \ldots, X_N$ with expectation μ and variance σ^2. Assume that the random variables are correlated with $\mathrm{Cov}(X_i, X_j) = \rho$ for $i \neq j$. Define the sample average as

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Compute the expectation and variance of $\bar{X}_N$. Detail the computation.

8 Galton board

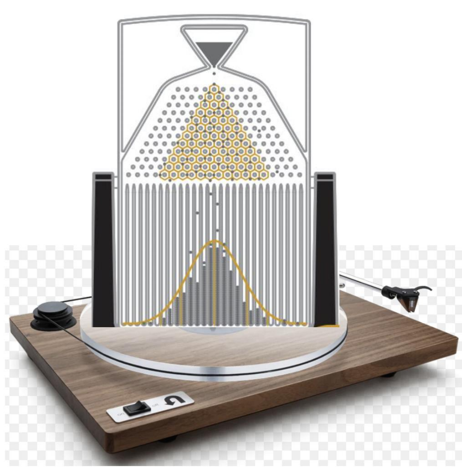

Figure 2: Galton board

Firgure 2 shows a Galton board. A ball is dropped from the top and bounces left or right at each level. Use simulation to estimate the distribution of the final position of the ball. Compare the empirical distribution to the Gaussian distribution with the same expectation and variance.

9 Multivariate Gaussian distribution

Consider a random vector $X \in \mathbb{R}^D$ with a Gaussian distribution with expectation $\mu \in \mathbb{R}^D$ and covariance matrix $\Sigma \in \mathbb{R}^{D \times D}$. Consider an invertible matrix $A \in \mathbb{R}^{D \times D}$ and a vector $b \in \mathbb{R}^D$. Define the random vector $Y = AX + b$. Compute the expectation and covariance matrix of Y.

10 Ratio of two Gaussian random variables

Consider two independent random variables $X \sim \mathrm{N}(0, 1)$ and $Y \sim \mathrm{N}(0, 1)$. Show that the ratio $Z = X/Y$ is a Cauchy random variable with density

$$f(z) = \frac{1}{\pi} \frac{1}{1 + z^2}.$$

For this purpose, take a test function $g(z)$ and compute the expectation of $g(Z)$ using the density of Z:

$$\mathbb{E}[g(Z)] = \int_{x=-\infty}^{x=+\infty} \int_{x=-\infty}^{x=+\infty} g\left(\frac{x}{y}\right) f(x)\, f(y)\, dx dy.$$

A possible approach is then to use the change of variable $u = x/y$.

11 Detecting Non Randomness

Write a python function that takes a sequence of zeros and ones and tries to detect whether it is a sequence of flips of a fair coin, or a sequence of zeros and ones generated by a human. As a matter of fact, humans are typically bad at generating random sequences and tend to generate sequences with too many alternations, and other artifacts. Here is a sequence of 0 and 1 generated by a human (me):

$0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1$

The function should return "random" or "non-random" depending on the result of the test. There is no unique way to detect non-randomness, but it is a good exercise to try to come up with a simple test that works well in practice. Indeed, you are encouraged to use simulations. Can your approach detect the non-randomness of the sequence above?