Stellenbosch University Iyunivesithi Universiteit
Department of Mathematical Sciences
Division of Applied Mathematics

AIMS | Physics for Machine Learning

Problems for Week 1: Random variables and sampling

Reading

- ML822 Monte Carlo notes: Chap. 1. We will cover some sections in class.
- AM783 Applied Markov Processes notes: Chap 1. There is some overlap between that chapter and the one before.

Practice

Q1. Common random variables. Calculate the mean, variance, and characteristic function of the following random variables:

(a) Bernoulli p: $P(X = 1) = p$, $P(X = 0) = 1 - p$.

(b) Binomial with parameters (n, p).

(c) Gaussian $X \sim \mathcal{N}(\mu, \sigma^2)$.

(d) Exponential: $p(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

(e) Uniform over $[0, L]$.

Theoretical

Q2. Gaussian sums. Show that the sum of two IID Gaussian random variables with mean μ and variance σ^2 is Gaussian-distributed. Generalise to a sum of n IID Gaussian random variables.

Q3. Log-normal random variable. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and define $Y = e^X$. Find the probability density of Y.

Numerical

Q4. Histogram function. Construct a function called `myhistogram(v,a,b,n)` that constructs a histogram of the values contained in the vector `v` in n bins spread uniformly in the interval $[a, b]$. The output of your function is the vector of histogram counts. Show that your function works correctly by comparing it with the similar histogram function available in Matlab, R or Python.

Q5. Non-uniform variates. Use the transformation method to construct random number generators for the following probability distributions and test them with large-enough samples by plotting the sample histogram (properly normalised) with the corresponding theoretical distribution.

(a) Random choice in a list $[1, 2, \ldots, n]$ of n values with probability given in the list $[p_1, p_2, \ldots, p_n]$.

(b) Uniform over $[-1, 1]$.

(c) Exponential with parameter λ.

Q6. Box–Muller method.

(a) Let X, Y be two independent Gaussian RVs with mean 0 and variance 1. Show numerically that $\theta = \arctan(Y/X)$ is uniform over $[-\pi/2, \pi/2]$ and $R = \sqrt{X^2 + Y^2}$ is distributed according to the Rayleigh distribution:

$$p(r) = re^{-r^2/2}, \quad r \geq 0.$$

(b) Code the Box–Muller method and show that it works numerically.

Q7. **Monte Carlo estimation of π.** Choose a point (x, y) at random in the square

$$S = \{(x, y) : x \in [-1, 1] \text{ and } y \in [-1, 1]\}.$$

The probability that the point lies in the circle

$$C = \{(x, y) : x^2 + y^2 = 1\}$$

is equal to $\pi/4$. Turn this result into a program that gives an approximation of π. Show that it works by performing statistical tests, showing convergence with error bars, as in the demo on statistical estimation.

Q8. **Two-state Markov chain.** Write a program that simulates trajectories of the two-state Markov chain with specific values of α and β, and verify with your program that the long-term occupations of the 0 and 1 states are $\beta/(\alpha + \beta)$ and $\alpha/(\alpha + \beta)$, respectively.

Q9. **Random walk on graphs.** Generate a connected graph, say with 10 or more states, and write a program that simulates a trajectory of the unbiased random walk on that graph. Plot a trajectory up to 1000 time steps and use it to verify that the fraction of time a node i is visited is proportional to its degree k_i, so that nodes with higher degree are most often visited. The Python package `networkx` is useful for this question.