# Software Development Principles for Statistical Modelling
# Principles of Probability with Python/R

Mick Cooney
mickcooney@gmail.com

January 31, 2017

`https://www.github.com/kaybenleroll/training_courses`.

Code is available in the `sdpsm_intro` directory.

Content in this workshop is based on the 'Software Carpentry' course: `http://software-carpentry.org/`

## 1. Introduction to Probability

**Exercise 1.1** Generate 100 random uniform numbers between 0 and 1.

**Exercise 1.2** Generate 100 random numbers distributed as a normal distribution with mean 0 and standard deviation 1.

**Exercise 1.3** Generate 100 normally distributed numbers with mean 50 and standard deviation 10.

**Exercise 1.4** Transform the standard normal numbers from the previous question and convert it to a distribution with mean 50, sd 10 and compare to the data generated directly.

## 2. Principles of Probability

**Exercise 2.1** Using simulation, calculate the value of $\pi$. Estimate the error in your calculated value.

**Exercise 2.2** Starting with a sample size of 10 and scaling in whatever way you wish, illustrate the *Law of Large Numbers* by plotting the mean of the $(0, 1)$ uniform distribution of data by sample size.

**Exercise 2.3** Repeat the above exercise with the standard normal distribution.

**Exercise 2.4** Repeat the above exercise with a Pareto distribution with $\alpha = 1.5$

**Exercise 2.5** Repeat the above exercise with a Pareto distribution with $\alpha = 2$

**Exercise 2.6** Repeat for the Cauchy$(0, 1)$ distribution.

**Exercise 2.7** Using the above distributions, illustrate the *Central Limit Theorem*.

# 3. Exercises in Probability

**Exercise 3.1** The incidence rate of a disease in the general population is 1 in a thousand people. A medical test for this disease has a 99% hit rate, i.e. if the patient has the disease, the test returns a positive result 99 times out of 100. The false alarm rate is 5% meaning that a uninfected patient tests positive 5 times out of 100. Using simulation, calculate the probability of a person being infected if they test positive?

**Exercise 3.2** Suppose there exists a second test for the disease that is independent of the first test but has the same reliability metrics. If a person tests positive on both tests, what is the probability of infection?

**Exercise 3.3** Create plots of the probability of being infected as a function of hit rate and of the false alarm rate for both one and two tests.
For the remainder of this workshop, we are going to use the book *Introduction to Probability* by Grinstead and Snell, it is available in various formats at the following site:

`https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html`

**Exercise 3.4** Answer Q4 on page 13 — the question on winning at racquetball.

**Exercise 3.5** Create plots of the probability of victory as a function of serve success and of return success.

**Exercise 3.6** How does the game length affect the probability of victory?

**Exercise 3.7** Choose another exercise from that group and solve it as you wish. We suggest you look at Q6, Q10, Q12 or Q13 but choose one that is interesting as a group.