Problems 19: Probability Distributions

Exercise 1. Let X be a random variable. Show that $b < c$ implies that $\mathrm{P}(X \leq b) \leq \mathrm{P}(X \leq c)$.

Exercise 2. A random variable X has cumulative distribution function F. What is the cumulative distribution function of $Y = aX + b$, where a and b are real constants?

Exercise 3. A fish searches a river bed for food, covering an area of 30 m^2 in 15 minutes. Suppose there are 2 pieces of food (A and B) randomly positioned within this area, and that consuming food takes negligible time.

(a) What is the cumulative probability function that the fish finds piece A after t seconds? What about for piece B?

(b) What is the expected time to find food A?

(c) What is the cumulative probability that the fish finds *both* pieces of food within t seconds? What is the expected time taken to find both pieces?

(d) Repeat part (c) for finding either piece of food.

Exercise 4. Suppose the diameter D of a tumour spheroid is measured, and due to measurement error $\log D$ is normally distributed with mean $300\,\mu$m and variance $10\,\mu$m. What are the mean and variance of the log of the volume V of the tumour?

Exercise 5. (Assessed) A species of bacteria has a probability of 0.003 of acquiring a point mutation on each cell division.

(a) What distribution models the number of generations until the next mutation?

(b) How many generations do we expect to elapse between subsequent mutations?

(c) What is the probability that in a population of 1000 bacteria we would see more than 5 mutations appear in one generation?

(d) What is the probabilty that no new mutations would appear?

Exercise 6. A new drug has been developed for a disease, and is undergoing a clinical trial to assess its effectiveness compared to placebo. Patients are randomly allocated between 2 groups: those receiving the drug and those receiving a placebo. On the assumption that the drug is no better than placebo, what is the probability that at most 38 of the 88 patients cured during the trial were in the drug group?

Exercise 7. (Extension) Using Python simulate drawing 1000 random numbers from each of the probability distributions introduced during the lecture. For each one, plot a (normalised) histogram of the result, and compare it to the pdf for the distribution. Experiment with changing the distribution parameters to visualise the effects.

Exercise 8. (Extension) Suppose (simplistically) that the height of a shrub is controlled solely by genetics, and that it has been determined that 3 genes contribute equally to the height. All genes exhibit incomplete dominance, with the dominant allele contributing 4 inches in height, while the recessive allele contributes 2 inches, so that (for instance) a plant heterozygous in all 3 genes will be (4+4+4+2+2+2=) 18 inches high.
Using PYTHON, simulate the heights observed in a field of 10,000 shrubs, where the shrubs are the second generation after cross-breeding for each gene. What do you notice about the distribution of the heights on a histogram? What would happen if 4 genes contributed to the height?