Part IA: Mathematics for Natural Sciences B
Examples Sheet 9: Discrete and continuous random variables, and functions of random variables

Please send all comments and corrections to jmm232@cam.ac.uk.

Questions marked with a $(*)$ are difficult and should not be attempted at the expense of the other questions.

Discrete random variables, expectation, and variance

1. A complex experiment consists of two independent stages, each of which involves the random generation of a number. In the first stage, the possible numbers are $0, 1, 2$, with probabilities $1/2, 1/4, 1/4$ respectively. In the second stage, the possible numbers are $2, 3$, with probabilities $2/3, 1/3$ respectively.

   Let X_1, X_2 denote the results of the first and second stages of the experiment, respectively. Let $X = 2X_1 - X_2$, that is, let X be the random variable given by taking twice the result of the first stage, and then subtracting the result of the second stage.

   (a) Find and sketch the probability mass function of X.

   (b) Find and sketch the cumulative distribution function of X.

   (c) Compute the probability $\mathbb{P}(0 \leq X \leq 2)$ by: (i) summing values of the probability mass function; (ii) taking the difference of two values of the cumulative distribution function.

   (d) Define the *expectation* (or *mean*) $\mathbb{E}[X]$ of a discrete random variable X, taking values $x_1, ..., x_n$. Find $\mathbb{E}[X]$ for the variable X defined above.

   (e) Give two expressions for the *variance* $\mathrm{Var}[X]$ of a discrete random variable X, and prove that they are equivalent. Find $\mathrm{Var}[X]$ for the random variable X defined above.

2. Let X be the result of a roll of a biased die, which displays one with probability $p/2$, two, three, four or five with probability p, and six with probability $2p$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

3. (a) Show that if X_1, X_2 are discrete random variables, and a, b are constants, we have:

   $$\mathbb{E}[aX_1 + bX_2] = a\mathbb{E}[X_1] + b\mathbb{E}[X_2].$$

   (b) Show that if X_1, X_2 are discrete *independent* random variables, we have

   $$\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2].$$

   (c) Now, let X, X_1 and X_2 be the random variables defined in Question 1. Verify that both $\mathbb{E}[X] = 2\mathbb{E}[X_1] - \mathbb{E}[X_2]$ and $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$ hold.

4. Show that if X_1, X_2 are discrete *independent* random variables, and a, b are constants, we have:

   $$\mathrm{Var}[aX_1 + bX_2] = a^2\mathrm{Var}[X_1] + b^2\mathrm{Var}[X_2].$$

   Now, let X, X_1 and X_2 be the random variables defined in Question 1. Verify that $\mathrm{Var}[X] = 4\mathrm{Var}[X_1] + \mathrm{Var}[X_2]$ holds.

5. Three standard 6-sided dice are tossed onto a table. Calculate the mean and variance of:

   (a) the sum of the values shown by the dice;

   (b) the sum of the squares of the values shown by the dice.

6. (**The law of the unconscious statistician**) Let X be a discrete random variable taking values $x_1, ..., x_n$, and let g be a real function defined on these values. Using the definition of expectation, show that:

$$\mathbb{E}[g(X)] = \sum_{k=1}^{n} g(x_k)\mathbb{P}(X = x_k),$$

and give an intuitive explanation of this result.

The geometric distribution

7. Explain what it means to say that a discrete random variable X has a *geometric distribution*, $X \sim \mathrm{Geo}(p)$. Prove that:

$$\mathbb{E}[X] = 1/p, \qquad \mathrm{Var}[X] = (1-p)/p^2.$$

8. After a late night out in Cambridge, you attempt to open the door to your college room with the three keys in your pocket, only one of which is the correct door key. Assuming that you are equally likely to select any of the keys in your pocket, compute the probability mass functions, the expectations, and variances of:
   (a) the random variable X, which is the number of attempts required to open the door if once you try a key from your pocket, you discard it on the ground if it doesn't work;
   (b) the random variable Y, which is the number of attempts required to open the door if once you try a key from your pocket, you place it back in your pocket again.

The binomial and Poisson distributions

9. (a) Explain what it means to say that a random variable X has a *binomial distribution*, $X \sim B(n, p)$. Prove that:

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}[X] = np(1-p).$$

   (b) An opaque bag contains 10 green counters and 20 red. One counter is selected at random and then replaced: green scores one and red scores zero. Five draws are made. If X is the total score, determine its expectation and variance.

10. (a) Explain what it means to say that a random variable X has a *Poisson distribution*, $X \sim \mathrm{Po}(\lambda)$. Prove that:

$$\mathbb{E}[X] = \lambda, \qquad \mathrm{Var}[X] = \lambda.$$

   Prove also that the limit of a binomial distribution $B(n, p)$ as $n \to \infty, p \to 0$ with $np = \lambda$ fixed, is a Poisson distribution $\mathrm{Po}(\lambda)$.

   (b) The probability of seeing a shooting star in any given hour is 0.44. Explaining all assumptions you make, estimate the probabilty of seeing a shooting star in any given half hour.

Continuous random variables

11. Let X be a continuous random variable with probability density function:

$$f_X(x) = \begin{cases} 10dx^2, & 0 \leq x \leq \frac{3}{5}, \\ 9d(1-x), & \frac{3}{5} \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Find the value of d, and hence sketch the function f_X.
   (b) Find and sketch the cumulative distribution function of X. Show also that the derivative of the cumulative distribution function is equal to the probability density function f_X - is this an accident?
   (c) Compute the probability $\mathbb{P}(0 \leq X \leq 2/5)$ by: (i) integrating the probability density function; (ii) taking the difference of two values of the cumulative distribution function.
   (d) Compute the mode, median, mean, and variance of the variable X.

12. A continuous random variable X, taking values in the interval $[0, 1]$, has cumulative distribution function:

$$F_X(x) = A\left(\frac{x^3}{3} - \frac{x^4}{4}\right),$$

for values in the range $0 \leq x \leq 1$.

   (a) Find the value of the constant A, and hence sketch the function F_X for all values of x.
   (b) Find and sketch the probability density function of X.
   (c) Calculate the mode, median, mean, and variance of the random variable X.
   (d) Calculate the probability that X lies in the interval $[\mathbb{E}[X] - \sigma, \mathbb{E}[X] + \sigma]$, where σ is the standard deviation of the random variable X.

13. A continuous random variable X, taking values in the interval $[-2, 2]$, has probability density function:

$$f_X(x) = \frac{1 + e^{-|x|}}{N},$$

for values in the range $-2 \leq x \leq 2$.

   (a) Find the value of the constant N, and sketch the function f_X for all values of x.
   (b) Find and sketch the cumulative distribution function of X.
   (c) Calculate the mode, median, mean, and variance of the random variable X.

   Another continuous random variable Y, taking values in the interval $[-2, 2]$, has probability density function:

$$f_Y(y) = \begin{cases} 0, & y < -2, \\ \mathbb{P}(X \leq y)/M, & -2 \leq y \leq 2, \\ 0, & y > 2. \end{cases}$$

   (d) Calculate the constant M.

The normal distribution and the central limit theorem

14. Explain what it means to say that a random variable X is *normally distributed*, $X \sim N(\mu, \sigma^2)$. Prove that:

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}[X] = \sigma^2.$$

15. (a) State the *central limit theorem*. [*The proof is a bit too hard for this course!*]
   (b) Using the central limit theorem, show that the limit of a binomial distribution $B(n, p)$ as $n \to \infty$, with $\mu = np, \sigma = np(1-p)$ fixed, is a normal distribution $N(\mu, \sigma^2)$. This is called the *normal approximation* to a binomial distribution.
   (c) Using the normal approximation, estimate the probability that a binomial random variable $X \sim B(100, 1/5)$ lies within one standard deviation of its mean. [*You will need to look up some values of the cumulative distribution function of the normal distribution online.*]

Functions of random variables

[*This topic is not really lectured, but is sometimes examined!*]

16. Let X be a continuous random variable with probability density function f_X. Let $U = X + c$ be a new random variable, where c is a constant. By considering $\mathbb{P}(u \leq U \leq u + du)$, for small du, find the probability density function for U in terms of f_X.

17. Let X be a continuous random variable, let f_X be its probability density, and let F_X be its cumulative distribution function. Suppose that $Y = g(X)$ is another random variable, where g is an arbitrary function. Show that if g is invertible, differentiable, and increasing, then the probability density function of Y is given by:

$$f_Y(y) = \frac{dg^{-1}}{dy}(y) \cdot f_X(g^{-1}(y)).$$

[*Hint: consider the equality of cumulative distribution functions* $\mathbb{P}(y \leq Y) = \mathbb{P}(y \leq g(X)) = \mathbb{P}(g^{-1}(y) \leq X)$.] How does this formula change if g^{-1} is invertible, differentiable, and *decreasing*?

18. A cannon fixed at the origin in the xy plane fires cannonballs at a screen placed $x = x_0$. The cannonballs are inclined at a random angle Θ to the x-axis, where Θ is uniformly distributed on the interval $[-\pi/2, \pi/2]$. Let Y be the random y-coordinate of the point of collision with the screen.

    (a) Using the result of the previous question, show that the probability density function of Y is given by:

    $$f_Y(y) = \frac{x_0}{\pi(x_0^2 + y^2)}.$$

    This distribution is called a *Cauchy distribution*.

    (b) Hence, show that the mean and standard deviation of Y are not well-defined.

Functions of multiple random variables

[*Similarly, this topic is not really lectured, but is sometimes examined!*]

19. (a) If X, Y are independent discrete random variables, explain why the distribution of their sum $Z = X + Y$ is given by the convolutional sum:

    $$\mathbb{P}(Z = z) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = z - k).$$

    (b) Hence, show that if $X \sim \mathrm{Po}(\lambda)$, $Y \sim \mathrm{Po}(\mu)$ are independent Poisson variables, their sum is also a Poisson variable, $X + Y \sim \mathrm{Po}(\lambda + \mu)$.

    (c) The number of patients arriving at Addenbrooke's hospital in a morning is on average 42, whilst the number of patients arriving at Addenbrooke's hospital in the afternoon is on average 64. Estimate the probability that the number of patients arriving on a single day will exceed 100, explaining any modelling assumptions you require. What is the variance of the number of patients arriving on a single day, under your model?

20. By considering the probability $\mathbb{P}(X_1 + X_2 = r)$, where $X_1 \sim B(m, p)$, $X_2 \sim B(n, p)$ are independent binomial variables, prove *Vandermonde's identity* for the binomial coefficients:

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}.$$

21. (a) If X, Y are independent continuous random variables with densities $f_X(x), f_Y(y)$ respectively, explain why the distribution of their sum $Z = X + Y$ has density $f_Z(z)$ given by the convolutional integral:

    $$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx.$$

    (b) Hence, show that if $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent normal variables, their sum is also a normal variable, $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

22. Let X be a continuous random variable with probability density function $f(x) = \pi^{-1}(1+x^2)^{-1}$, for $-\infty < x < \infty$, and let Y be a continuous random variable with uniform density on the interval $[-1/2, 1/2]$. Let $Z = X + Y$.
    (a) Find the probability density function of Z, and sketch it.
    (b) Find the mode and mean values of Z.

23. A pedestrian arrives at a crossing, where they press the button and wait for the lights to change. The time taken before the lights change colour is a uniform random variable T_1, taking values between 0 and 1 minutes.
    (a) Sketch the probability density function of T_1, and find its mean and variance.

    After the lights change, the pedestrian crosses and walks to another independent set of lights, which operate under the same conditions, changing at a time T_2 after the button is pressed. The pedestrian 5 minutes in total to walk to the second set of lights, after the first set of lights has changed. Let U denote the difference between the total time taken for the journey, from the moment that the first button is pressed until the second set of lights changes, and 5 minutes.

    (a) Find an expression for U in terms of T_1, T_2, and hence obtain its probability density.
    (b) By direct integration of the probability density, find the mean and variance of U, and show that this is consistent with the general properties of expectation and variance from Questions 3 and 4.

24. A coffee machine produces random amounts of liquid for each cup, always in the range 270 to 330ml, with the distribution function:

$$f(v) = \frac{\pi}{120} \cos\left(\frac{\pi(v - 300\,\text{ml})}{60\,\text{ml}}\right)(\text{ml})^{-1}$$

    Each coffee is made independently of all other coffees.

    (a) Sketch the distribution function, and find both the mean amount of liquid per cup and the variance in the amount of liquid per cup.
    (b) Suppose that the machine is used to make two coffees. Calculate the probability that the total amount of liquid produced by the machine for these two cups exceeds 630ml.
    (c) Find the mean and standard deviation in the amount of liquid produced by the machine when making two coffees by: (i) using standard properties of expectations and variances of independent random variables; (ii) direct calculation of the expectation and variance from the distribution function of the amount of liquid produced by the machine when making two coffees.

25. Let $X, Y \sim \text{Unif}(0, 1)$ be independent uniform random variables on the interval $[0, 1]$. Define $Z = X/Y$.
    (a) Show that the probability density function of Z is given by:

$$f_Z(z) = \begin{cases} \frac{1}{2}, & \text{if } 0 < z < 1, \\ \frac{1}{2z^2}, & \text{if } z > 1, \\ 0, & \text{otherwise.} \end{cases}$$

    (b) Hence, show that the probability that the closest integer to Z is even is given by:

$$\frac{5 - \pi}{4}.$$

    [*Hint: Consider the Taylor series of* $\arctan(x)$.]
    (c) $(*)$ Show that $\mathbb{P}(\lfloor Z \rfloor \text{ is even}) = 1 - \log(2)/2$.