

# Mathematics for Data Science Tutorial 6 (week 12)

Semester 2, 2019

1. Consider the discrete random variable  $X$  described by the following table

$k$	1	3	4	6	8	9
$\Pr(X = k)$	0.37	0.08	$q$	0.06	0.09	0.21

where  $X$  can take on only those values of  $k$  shown.

- (a) What is the value of  $q$ ?
  - (b) Determine  $\mathbb{E}[X]$ .
  - (c) Determine  $\text{Var}(X)$ .
2. Consider a complex simulation which is performed  $n$  times, each instance operating independently and providing a ‘positive’ result with probability  $p$  (and a ‘negative’ result otherwise). Suppose a consensus is determined by taking the result which is the majority of the simulations.
- (a) Consider performing the simulation 3 times, what is the probability of a ‘positive’ consensus?
  - (b) Consider performing the simulation 5 times, what is the probability of a ‘positive’ consensus?
  - (c) For which values of  $p$  is performing the simulation 5 times more likely to produce a ‘positive’ consensus than performing the simulation 3 times?
  - (d) Suppose that the first three simulations gave a negative consensus, but then, after an additional two simulations, a positive consensus is reached. What value of  $p$  maximises the probability of this event?
  - (e) What is the expectation and variance for the number of ‘positive’ results in  $n$  simulations?
3. Consider a random variable  $X$  with probability density function

$$f(x) = \begin{cases} \frac{x}{32} & \text{for } 2 < x < 6 \\ \frac{1}{8} & \text{for } 6 < x < 8 \\ a(100 - x^2) & \text{for } 8 < x < 10 \\ 0 & \text{elsewhere} \end{cases}$$

- (a) What is the value of  $a$ ?

- (b) Determine  $\mathbb{E}[X]$ .
  - (c) Determine  $\text{Var}[X]$ .
4. Suppose an analysis of social media posts on a particular website reveals that the length of time between any two posts can be modelled with a probability distribution function of the form

$$f(t) = c2^{-t}, \quad t \geq 0.$$

(Implicitly take  $f(t) = 0$  for  $t < 0$  in this context, i.e. there cannot be negative time in between posts.)

- (a) What must  $c$  be for this to be a valid probability distribution?
  - (b) What is the mean and variance in the time between posts?
5. Suppose  $X, Y$  are random variables and that  $Y = -1 + 4X - X^2$ .
- (a) If  $\mathbb{E}[X] = 2$  and  $\mathbb{E}[Y] = 1$ , what is  $\text{Var}(X)$ ?
  - (b) Why is it not possible that  $\mathbb{E}[X] \leq 0$  if  $\mathbb{E}[Y] = 1$  (assuming  $X, Y$  are real valued)?