EXAMPLES SHEET 2

Probability

Christopher Yau

November 27, 2020

These are some exercises designed to practice your knowledge and ability to apply some of the concepts of probability you learnt in lectures.

Answers are given at the end but try to attempt the questions first before looking at the solutions.

1. (a) Bayes Theorem states that, for two events A and B, the conditional probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Given that $P(A) = 0.1$ and the conditional probability of B given A is 0.5 and A given B is 0.2.

What is the conditional probability $P(B|\overline{A})$?

(b) In a randomized control trial, breast cancer patients are randomly assigned with probability 0.6 and 0.4 to one of two groups who will receive a standard treatment and a novel experimental agent respectively.

i. Historical data indicates that the standard treatment would clear the disease in 70% of patients. However, in preclinical trials, 85% of those treated with the experimental agent showed no signs of disease.

A. What is the probability that a cancer patient recruited to the trial will be cleared of disease?

B. What is the probability that a patient received the standard treatment, given that they were given an all-clear after treatment?

ii. The trial involves patients recruited from a number of hospitals. In one local regional hospital, 4 patients were recruited.

A. What is the probability that at least one of these 4 patients would get an all-clear after treatment?

B. The local regional hospital only has limited capacity for supporting recurrent disease. What is the probability that 2 or more patients would get disease recurrence?

The following expression for the probability mass function of the Binomial distribution for n trials and success probability p maybe useful:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

and the binomial coefficient is given by:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

2. A GP practice receives patients from two local villages. Each village has 500 and 2,000 residents respectively. All residents use the GP practice as it is the only one located in close proximity.

   (a) Historical averages, suggest that 1 in 5 people from the smaller village will use the GP practice in any single year whilst the rate is 2 in 5 from the larger village.

      i. What is the expected number of people who will use the GP practice in any one year?

      ii. A patient attends a GP appointment. What is the probability that they come from the small village?

   (b) The GP practice randomly selects 10 individuals from the area it serves to take part in a trial of a new online appointment booking service.

      i. What is the probability that at least one individual will make use of the service during the year?

      ii. What is the probability that there is at least one individual from each village included in the trial?

3. (a) A discrete random variable X can take on the values 0, 1, 2, 3 and 4. Its probability mass function has the form:

$$p(X = x) = \frac{1}{Z} \exp(-x)$$

Compute the value of Z then find the expectation $E[X]$ and variance $V[X]$.

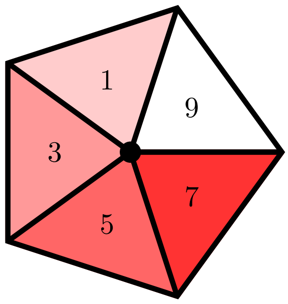
(b) Two identical six-sided die are rolled. What is the expected value and variance of the difference in the two values?

(c) A random variable Y has the following pmf:

y	0	1	2
$P(Y = y)$	0.5	0.3	0.2

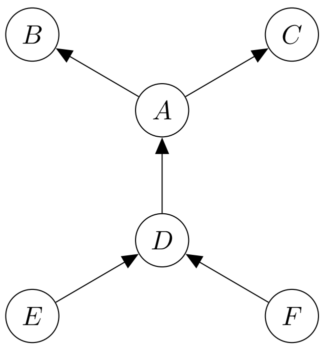
Compute the expected value of $2Y - 1$ and its variance.

(d) A five sided spinner is spun twice:

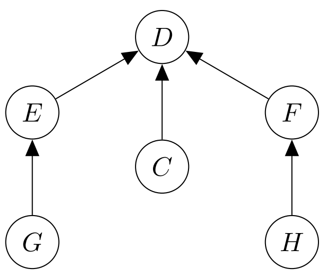


What is the probability that the sum of the scores is greater than 10?

(e) Given the following directed acyclic graph (DAG), factorise the joint distribution $P(A, B, C, D, E, F)$:



(f) Given the following directed acyclic graph (DAG), factorise the joint distribution $P(C, D, E, F, G, H)$.:



4. (a) A function is given by:

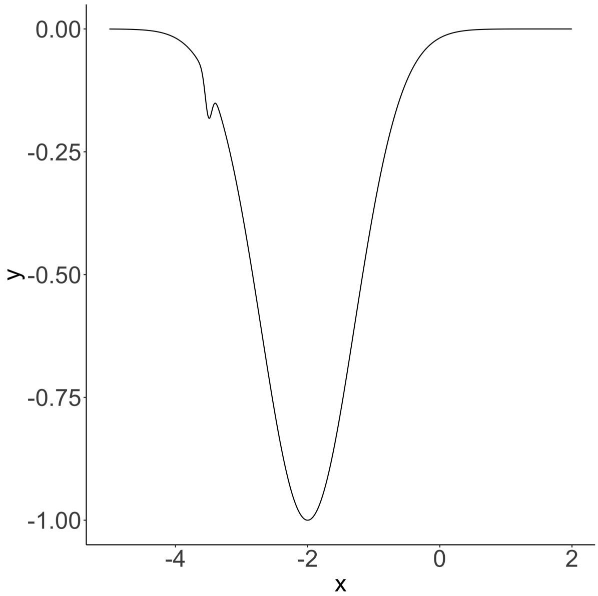
$$h = 2x^2 + 6x + y^2 - 4y + 5$$

i. Find the partial derivatives of h with respect to x and y.

ii. We apply gradient descent to find the values of (x, y) values that minimise h. Show the form of the gradient descent updates.

iii. Determine the values of x and y at convergence and the value of h at its minimal value.

(b) The following shows a function $y = f(x)$:



Explain why stochastic gradient descent might be preferable to gradient descent when attempting to find the value of x that minimises this function.

5. In a recycling plant, the proportion of refuse which is Class A or Class B recyclable material occurs with probability θ and $1-\theta$ respectively. The probability θ will be used to determine how many workers to assign to process each type of item.

You are given a data set which consists of a random sample of 20 independent refuse items that were manually classified and contained 14 Class As and 6 Class Bs.

(a) If $X_1, X_2, \ldots, X_{20}$ are binary random variables which represent the events that each of the 20 objects belong to Class A (1) or not (0).

Show that the likelihood $p_\theta(X_1, \ldots, X_{20})$ of observing the sample data given θ is given by:

$$p_\theta(X_1, \ldots, X_{20}) = \theta^{14}(1-\theta)^6$$

Hint: You may use the fact that the probability mass function for a random variable Z that follows a Bernoulli distribution with parameter h is given by:

$$p(Z = z) = h^z(1-h)^{1-z}$$

(b) Design a gradient descent approach to find the value of θ that maximises the likelihood. What is the form of the gradient descent update expression?

(c) Show that the algorithm converges when the value of the parameter θ reaches 0.7.

(d) At a certain point in the gradient descent algorithm, $\theta = 0.65$. What is the value of θ after the next update with step length 0.1? What does this suggest about the step length?

6. The personal salary of online shoppers is known to determine the probability that shoppers will actually make a purchase after browsing an online store.

You are given the following data from six independent shoppers which includes their relative (to the population average) salary and whether they made a purchase on their last visit to an online store:

Data Item	1	2	3	4	5	6
Relative Salary (x)	0.376	0.918	-0.527	0.557	0.4667	-1.791
Purchase (y)	1	1	0	1	0	0

You are asked to devise a binary classification algorithm to estimate the probability that a shopper will make a purchase given their salary (x) as an input.

(a) Suppose you decide to approach this problem using logistic regression. Show that the likelihood of the data is given by:

$$\prod_{i=1,2,4} \frac{1}{1+\exp(-z_i)} \prod_{i=3,5,6} \frac{\exp(-z_i)}{1+\exp(-z_i)}$$

where $z_i = b_0 + b_1 x_i$ and (b_0, b_1) are regression parameters.

(b) Show that the update expressions for an online stochastic gradient descent algorithm to estimate the parameters b_0 and b_1 using the given dataset are given by:

$$b_0' = b_0 - \lambda(p(y_i) - y_i),$$
$$b_1' = b_1 - \lambda x_i(p(y_i) - y_i)$$

(c) You decide to compare the performance of the logistic regression algorithm against a deterministic perceptron algorithm. Given initial weights $(w_0, w_1) = (1, 1)$, where w_0 is the bias term. Determine the classification of each of the data items.

(d) Now, apply one iteration of the perceptron update, compute the new weights and compute the updated classification.

(e) Explain why the perceptron algorithm will not be able to classify all items correctly.