



**Coláiste na Tríonóide, Baile Átha Cliath**  
**Trinity College Dublin**  
Ollscoil Átha Cliath | The University of Dublin

**Faculty of Engineering, Mathematics and Science**  
**School of Computer Science and Statistics**

**ST3009: Statistical Methods for Computer Science**  
**Mock Final Exam**

**Instructions to Candidates:**

Attempt **all** questions.

You may not start this examination until you are instructed to do so by the invigilator.

**Materials Permitted for this examination:**

Non-programmable calculators are permitted for this examination please indicate the make and model of your calculator on each answer book used.

**Question 1.** Let  $X$  and  $Y$  be independent random variables that take values in the set  $\{1, 2, 3\}$ . Let  $V = 2X + 2Y$ , and  $W = X - Y$ .

- (a) Assume that  $P(X = x)$  and  $P(Y = x)$  are non-zero for any  $x, y \in \{1, 2, 3\}$ . Can  $V$  and  $W$  be independent? Explain. [10 marks]

For the remaining parts of this problem, assume that  $X$  and  $Y$  are uniformly distributed on  $\{1, 2, 3\}$  i.e. the probability of each value occurring is the same.

- (b) Compute the PMF of  $V$ ,  $E[V]$  and  $var(V)$ . [10 marks]
- (c) Compute the joint PMF of  $V$  and  $W$ . [5 marks]
- (d) Compute  $E[V|W > 0]$ . [10 marks]

**Question 2.** We want to perform a survey of  $n$  people, asking their answer to a sensitive question. We use the randomised response approach whereby each respondent tosses a fair coin and if it comes up heads they respond with “yes” and if tails they respond by truthfully answering “yes” or “no” to the sensitive question being asked. Associate with the  $i$ 'th respondent a random variable  $X_i$  which takes value 1 when the person's truthful answer to the sensitive question is “yes” and value 0 otherwise. Assume that the  $X_i$ ,  $i = 1, 2, \dots, n$  are independent and identically distributed with  $P(X_i = 1) = p$ .

- (a) Let random variable  $Y_i$  take value 1 when the  $i$ 'th respondent answers “yes” and 0 otherwise. Give an expression for  $E[Y_i]$  in terms of  $E[X_i]$ , and using this then rearrange to get  $E[X_i]$  in terms of  $E[Y_i]$ . Hint: use the linearity of expectation. [10 marks]
- (b) Let random variable  $Z = \sum_{i=1}^n Y_i$  be the number of respondents who answer “yes”. Write an expression for the probability that out of  $n$  respondents  $z$  respond “yes”. [10 marks]
- (c) Suppose we use  $Z/n$  as an estimate of  $q = E[Y_i]$ , where  $n$  is the total number of respondents. Recalling that for two random variables  $W$  and  $V$  we have  $E[W + V] = E[W] + E[V]$ , show that  $E[Z/n] = q$ . Using Chebyshev's inequality explain the weak law of large numbers and the behaviour of  $|\frac{Z}{n} - q|$  as  $n$  becomes large. [10 marks]
- (d) Using  $Z/n$  as an estimate of  $q$  explain how you could use bootstrapping to estimate a confidence interval for the accuracy of this estimate. [5 marks]

**Question 3.** Diana the Daredevil is trying to break the world land-speed record using her rocket-powered motorcycle. In order to do so, she needs to cover more than 500 metres in 10 seconds. If her motorcycle has  $Z$  pounds of rocket fuel to start, then it travels a distance of  $X = 50\sqrt{Z}$  metres in 10 seconds.

- (a) Suppose that the amount of rocket fuel  $Z$  is a random variable uniformly distributed over  $[50, 150]$ . Compute the CDF of the distance  $X$ . [10 marks]
- (b) Compute the probability that Diana breaks the world record on any given trial. [10 marks]
- (c) Now suppose that Diana is allowed to take a total of  $n$  trials in order to try and break the world record. (Here  $n$  is some fixed positive integer: i.e.,  $n \in \{1, 2, 3, \dots\}$ ). The amount of rocket fuel at the start is independent from trial to trial. How large does  $n$  have to be in order for Diana to have a better than 90% chance of breaking the world record over the  $n$  trials? [10 marks]