



**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 & Statistics**

**Integrated Computer Science Programme**  
**Year 3**

**Hilary Term 2017**

**ST3009: Statistical Methods for Computer Science**

**DD MMM YYYY**

**Venue**

**00.00 – 00.00**

**Doug Leith**

**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.

1. (i) A bag contains 10 balls, of which 5 are red and the other 5 black.
- (a) Suppose you take out 5 balls from this bag, with replacement. What is the probability that among the 5 balls in this sample exactly 2 are red and 3 are black? [5 marks]
- (b) Now suppose that the balls are taken out of the bag without replacement. What is the probability that out of 5 balls exactly 2 are red and 3 are black? [10 marks]
- (iii) Three people get into an elevator at the ground floor of a hotel which has four upper floors. Assuming each person gets off at a floor independently and is equally likely to choose each of these four floors, what is the probability that no two people get off at the same floor? [10 marks]
- (b) (i) Define the terms “random event” and “random variable” and give an example of each. [5 marks]
- (ii) For a random variable  $X$ , define  $E[X]$  and  $\text{var}(X)$ . [5 marks]
- (iii) A random variable  $X$  has  $P(X=1)=0.2$ ,  $P(X=2)=0.3$ ,  $P(X=3)=0.5$  and  $P(X=x)=0$  for all values of  $x$  other than 1,2 or 3. What is the mean and variance of  $X$ ? [5 marks]
- (iv) Define what it means for two random variables to be independent. [5 marks]
- (v) Let  $X$  and  $Y$  be independent random variables that take values in the set  $\{1,2,3\}$ . Assume that  $X$  and  $Y$  are uniformly distributed on  $\{1, 2, 3\}$  i.e. the probability of each value occurring is the same. Let  $V = XY$ . Are  $V$  and  $X$  independent? Explain. [5 marks]
- (c) (i) Write down expressions for  $E[X]$  and  $E[X/n]$  for random variable  $X$  and  $n \neq 0$ . Show that  $E[X/n]=E[X]/n$ . [5 marks]
- (ii) Give a proof that the expected value is linear i.e.  $E[X+Y]=E[X]+E[Y]$  for random variables  $X$  and  $Y$ . [5 marks]
- A sequence of  $n$  bits is sent across a wireless link. Let random variable  $Y_i$  take value 1 when the  $i$ 'th bit is received without error and 0 otherwise. Suppose the random variables  $Y_i$   $i=1,2,\dots,n$  are independent and identically distributed with  $E[Y_i]=\mu$ .
- (iii) Let random variable  $Z = \sum_{i=1}^n Y_i$  be the number of bits received without error. Show that  $E[Z/n] = \mu$ . Hint: use the linearity of the expected value. [5 marks]
- (iv) Using Chebyshev's inequality explain the weak law of large numbers and the behaviour of  $|Z/n - \mu|$  as  $n$  becomes large. Recall that for random variable  $X$  Chebyshev's inequality is:  $P(|X - \mu| \geq k) \leq E[(X - \mu)^2]/k^2$  for an  $k$  and  $\mu$ . [5 marks]

(v) Explain what a confidence interval is, using  $Z/n$  as an estimate of  $\mu$  as an example. Describe how to use bootstrapping to estimate a confidence interval.

[5 marks]

(d) (i) With reference to Bayes Rule explain what is meant by the likelihood, prior and posterior.

[5 marks]

(ii) Explain how the maximum a posteriori (MAP) estimate of a parameter differs from the maximum likelihood estimate.

[5 marks]

(iii) We observe data  $(x_i, y_i)$ ,  $i=1,2,\dots,n$  from  $n$  people, where  $x_i$  is the person's height and  $y_i$  is the person's weight.

(a) Explain how to construct a linear regression model for this data. [10 marks]

(b) Suppose we suspect that the weight of a person is not linearly related to their height but rather is related to the square root of their height. Explain how to modify the linear regression model to accommodate this. [5 marks]