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 Trinity Term 2016
Year 3 Annual Examinations

ST3009: Statistical Analysis

Tuesday 17th May 2016 Drawing Office 14.00 - 16.00

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. A bag contains 4 red balls and 4 white balls. One ball is drawn from the bag and put to one side. A second ball is now drawn from the bag. What is the probability that the first ball is red and the second ball is white? Are the events of drawing are a red ball and then a white ball independent? [15 marks]

2. We transmit a bit of information which is "0" with probability 1-p and "1" with probability p. Because of noise on the channel, each transmitted bit is received correctly with probability 1-q where q<1/2.

a) Suppose we observe a "1" at the receiver. What is the probability that the transmitted bit was a "1"? [15 marks]

b) Suppose we transmit each information bit n times over the channel e.g. if n=3 and the information bit is a "1" then we send "111". What is the probability that the information bit is "1" given that you have observed n "1"s at the receiver. What happens when n becomes large? [15 marks]

c) For the setup in part (b), what is the probability that the information bit is "1" given that you have observed m "1"s (and n-m "0"s) at the receiver, m≤n. [10 marks]

3. A server has 32GB of memory. We are interested in the probability that the server is overloaded, meaning the memory usage by all of the running jobs exceeds 32GB. Suppose the memory usage of a job is 0.5GB with probability 0.5 and 1GB with probability 0.5, and that the memory usage of different jobs are independent.

a) Suppose exactly 32 jobs are running. Using Markov's inequality, compute an upper bound on the probability that the server is overloaded. [10 marks]

b) Suppose now that a random number N of jobs are running, with $P(N=n)=p(1-p)^{(n-1)}$, where p is a parameter. Using Markov's inequality, compute an upper bound on the probability that the server is overloaded. What value of p should we choose to ensure that the probability of overload is less than 0.1 (based on Markov's inequality). Useful fact:
$$\sum_{n=0}^{\infty} np(1-p)^{(n-1)} = \frac{1}{p}$$
[15 marks]

4. Consider a linear regression model in which random variable Y is the sum of a deterministic linear function of input x plus random noise M. That is, Y=θx+M, where θ is a parameter we would like to estimate. Suppose noise M normally distributed with mean 0 and variance 1.

a) Write down an expression for the probability distribution function of Y given θ and x.

[5 marks]

b) You are given n independent and identically distributed training examples d={(x_1,y_1), ..., (x_n,y_n)}. Write an expression for the likelihood of this training data.   [5 marks]

c) Now write an expression for the log-posterior probability density function for θ assuming a Gaussian prior over θ with mean 0 and standard deviation σ. How is this used to obtain a MAP estimate for θ?   [10 marks]