Paper Code

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 **Hilary Term 2019**
Year 3

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) Define the terms “sample space” and “random event”. Give an example of each.

[5 marks]

(ii) You perform the following experiment: you take a six-sided die, and roll it. If the number that comes up is six, you stop; otherwise you repeat.

(a) What is the probability that you roll the die once and then the experiment stops? What is the probability that you roll the die 5 times before the experiment stops?

[5 marks]

(b) Suppose now that you roll three dice and stop when one or more of the dice come up six. What is the probability that you roll the dice 5 times before the experiment stops ?

[5 marks]

(iii) Define the conditional probability of an event and state Bayes Rule. [5 marks]

(iv) Suppose two websites A and B take hotel bookings. Site A takes 60% of all bookings and site B takes 40%. However, only 75% of the bookings made on site A result in positive reviews after the hotel stay, while on site B it is 90%. Given that a booking received a positive review, what is the probability that booking was made on site B ? Hint: use Bayes Rule. [5 marks]

Model Solution

(i) The sample space is the set of all possible outcomes of an experiment. A random event is a subset of the sample space. For example, when flipping a coin the sample space is {heads, tails} and a random event might be {heads}.

(ii) (a) Prob that stops after one roll is 1/6. Prob that stops after 5 rolls is $(1-1/6)^4 1/6$. (b) Probability that none of the three dice come up six is $(1-1/6)^3$. Probability that one or more comes up six is $1-(1/6)^3$. So probability roll five times is $((1-1/6)^3)^4(1-(1/6)^3)$.

(iii) For random events E and F the conditional probability $P(E|F)=P(E \text{ and } F)/P(F)$. Bayes Rule states that $P(E|F)=P(F|E)P(E)/P(F)$.

(iv) Let E be event that booking was on site B and F the event that a review was positive. $P(F|E)=0.9$, $P(E)=0.4$ and $P(F)=P(F|E)P(E)+P(F|E^c)P(E^c)=0.9x0.4+0.75x0.6=0.81$. So by Bayes Rule $P(E|F)=0.9x0.4/0.81=0.44$.

2. (i) Define the term “random variable” and give an example. What is the probability mass function of a discrete random variable ? Give an example. [5 marks]

(ii) Define what it means for two random variables to be independent. [5 marks]

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 =2X+2Y.

© Trinity College Dublin, The University of Dublin 2016

(a) Are V and X independent? Explain using the definition of independence given above. [10 marks]

(b) Compute the PMF of V. [5 marks]

Model Solution

(i) A random variable is a mapping from the sample space to a real number. For example when tossing a coin twice a random variable X might be the number of times that heads comes up (the sample space is {(H,H),(H,T),(T,H),(T,T)} and X takes value 2 on outcome (H,H), value 1 on outcomes (H,T),(T,H) value 0 on outcome (T,T). For an RV X which takes values $x_1,\ldots,x_n$ the PMF is the probabilities $P(X\text{-}x_i)$, i=1,…,n. For the coin toss example the PMF is P(X=2)=1/4, P(X=1)=1/2, P(X=0)=1/4.

(ii) Two RVs X and Y are independent when P(X=x and Y=y)=P(X=x)P(Y=y).

(a) No, they are not independent. For example, P(X=3)=1/3, P(V=12)=P(X=3 and Y=3)=1/3x1/3 and P(X=3 and V=12)=$(1/3)^2$ which is not equal to P(X=3)P(V=12)=$(1/3)^3$.

(b) V can take values 4,6,8,10,12. P(V=4)=P(X=1 and Y=1)=$(1/3)^2$, P(V=6)=P(X=1 and Y=2)+P(X=2 and Y=1)=2$(1/3)^2$, P(V=8)= 3$(1/3)^2$,P(V=10)= 2$(1/3)^2$,P(V=12)=P(X=3 and Y=3)=$(1/3)^2$

3. A web server is connected to the internet via two links. The server is required to serve an object consisting of n packets to a client and has to choose which link to send the n packets across. The time taken to send the n packets across link i is $T_i = \sum_{j=1}^{n} X_j^{(i)}$ where $X_j^{(i)}$ is a random variable whose value is the time that would be taken to send packet j across link i.

(i) Suppose the transmit times $X_j^{(1)}, j = 1,2,\ldots$ on link 1 are independent and identically distributed with mean 0.1 and variance 0.05.

(a) What is the expected value of T_1 ? Hint: use the linearity of the expectation i.e. that E[X+Y]=E[X]+E[Y]. [5 marks]

(b) What is the variance of T_1 ? Hint: use the linearity of the variance when variables are independent i.e. that var(X+Y)=var(X)+var(Y). [5 marks]

(ii) Using Chebyshev's inequality give an upper bound on the probability that T_1 is greater than or equal to 0.5n. Recall that for random variable X Chebyshev's inequality is: $P(|X - \mu| \geq k) \leq E[(X- \mu)^2]/k^2$ where $\mu=E[X]$. [5 marks]

(iii)

(a) Calculate the expected value and variance of T_1/n. [5 marks]

© Trinity College Dublin, The University of Dublin 2016

(b) As the number of packets n goes to infinity, explain what Chebyshev's inequality tells us about the probability that T_1/n is greater than its expected value ? Hint: recall the weak law of large numbers. [5 marks]

Model Solution

(i) (a) $E[T_1] = 0.1n$, (b) $Var(T_1)=0.05n$

(ii) By Chebyshev's $P(|T_1 - 0.4n| \geq 0.1n) \leq 0.05n/(0.1n)^2=5/n$. That is, $P(T_1 \geq 0.5n$ or $T_1 \leq 0.3n) = P(T_1 \geq 0.5n) + P(T_1 \leq 0.3n) \leq 5/n$ and so $P(T_1 \geq 0.5n) \leq 5/n$ since $P(T_1 \leq 0.3n) \geq 0$.

(iii) (a) $E[T_1/n]=E[T_1]/n=0.1$, $Var(T_1/n)=Var(T_1)/n^2=0.05/n$. (b) By Chebyshev's $P(|T_1/n - 0.1| \geq k) \leq 0.05/(nk^2)$. For any $k>0$ the RHS goes to zero as n goes to infinity and so the probability that T_1/n is larger than its expected value of 0.1 goes to 0.

4. Suppose that we observe data (x_i,y_i), i=1,2,…,n for n people sampled from the population, where x_i is the distance of the person's home from Dublin city centre, and $y_i=1$ if the i'th person has visited TCD in the last year and 0 otherwise. We model y_i as an observation of random variable Y_i and x_i as an observation of random variable X_i. Recall that a logistic regression classifier uses the model

$$P(Y_i=y \mid X_i=x_i) = 1/(1+\exp(-yz_i))$$

with $z_i=\theta x_i$, θ a parameter and y either 0 or 1.

(i) Assume the Y_i's are independent and that parameter θ is known. Using the above model write an expression for the likelihood $P(\{Y_i=y_i, i=1,2,\ldots,n\} \mid \{X_i=x_i, i=1,2,\ldots,n\}, \theta)$ i.e. for the probability that data y_i, i=1,2,…,n is observed under this model. [5 marks]

(ii) Now suppose parameter θ is not known. Define the maximum likelihood estimate of parameter θ. [5 marks]

(iii) Suppose that in addition to the distance of their home from Dublin we also know for person i the number w_i of their children who currently study at TCD. Describe how we can modify the above logistic regression classifier to include this information. Hint: we now have two inputs/features x_i an w_i. [5 marks]

(iv) This type of logistic model is often said to be most effective for linearly separable data. Explain what linearly separable means, giving an example. What can happen if the data is not linearly separable ? [5 marks]

(v) Suppose the data is not linearly separable. How might we change the inputs/features used in the model to accommodate this ? [5 marks]

Model Solution

(i) $P(\{Y_i=y_i, i=1,2,\ldots,n\} \mid \{X_i=x_i, i=1,2,\ldots,n\}, \theta) = 1/(1+\exp(-y_1\theta x_1))x \ldots x\ 1/(1+\exp(-y_n\theta x_n))$

(ii) The max likelihood estimate is the value of θ that maximises $P(\{Y_i=y_i, i=1,2,\ldots,n\} \mid \{X_i=x_i, i=1,2,\ldots,n\}, \theta)$

© Trinity College Dublin, The University of Dublin 2016

(iii) Extend the model to be $P(Y_i=y \mid X_i=x_i\ W_i=w_i) = 1/(1+\exp(-y(\theta_1 x_{i+}\ \theta_2 w_i))$ i.e. change z_i to be $(\theta_1 x_{i+}\ \theta_2 w_i)$ where θ_1 and θ_2 are unknown parameters.

(iv) We can transform the features to try to make the data linearly separable. For example is y actually depends on x^2 rather than x then we can use x^2 as a feature.

© Trinity College Dublin, The University of Dublin 2016