Maynooth University
National University of Ireland Maynooth

SEMESTER 1
2021-2022
SOLUTIONS

EE304FZ
Probability and Statistics

Dr. Lo Fook Loong, Prof. R. Farrell, Dr M. Hayes

Time allowed: 2 hours

Question 1 is **compulsory** and is worth 40 marks.

Answer ***three*** questions from the remaining four. Each is worth 20 marks.

Instructions

	Yes	No
Log Books Allowed	ϒ	X
Formula Tables Allowed	X	ϒ
Other Allowed (*enter details)*	X	ϒ

General (*enter details)*

Formula table is attached to the end of the exam paper.
Non-programmable calculators are allowed.

QUESTION 1

(a)	A password can be formed from capital letters ‘A’ to ‘Z’, small letters ‘a’ to ‘z’ and digits ‘0’ to ‘9’.		
	(i)	If the password must be 8 characters long and contain at least one capital letter, one small letter and one digit, how many possible passwords are there?	(3 marks)
		We first choose a capital letter, a small letter and a digit. No. of ways of doing so is 26 x 26 x 10. We can arrange them in 3! ways. Now we place them in 8 positions. First character has 8 choices, second has 7 choices and 3^{rd} has 6 choices, making 8 x 7 x 6 ways. Finally we fill in the 5 empty positions with any characters. No. of ways is 62^5. Hence total no. of possible passwords is, multiplying them together, 1.2485×10^{16}.	
	(ii)	If the password must be 9 characters long, must contain at least one small letter, one capital letter, one digit and just one of the following special characters ‘@’, ‘#’, ‘$’ and ‘%’, how many possible passwords are there?	(2 marks)
		First choose a special character. There are 4 ways. Then squeeze it in front, at the back or in between the 8 character password. There are 9 ways. Hence no. of possible passwords is 4.4946×10^{17}.	
(b)	When a player in a poker game has the 5 cards Ten, Jack, Queen, King and Ace of any suit, then he is said to be holding a royal flush. What is the probability of getting a royal flush if 5 random cards are taken from a well shuffled pack of 52 cards?		(5 marks)
	For a particular suit, probability is $\frac{5}{52} \times \frac{4}{51} \times \frac{3}{50} \times \frac{2}{49} \times \frac{1}{48} = 3.8477 \times 10^{-7}$. Since royal flush can come from any suit, the probability is 4 times that, or 1.5391×10^{-6}.		
(c)	The continuous random variable X has a pdf given below: $f(x) = \begin{cases} 0 & x < 0 \\ kx & 0 \le x < 5 \\ 0 & x > 5 \end{cases}$		
	(i)	Find the constant k.	(4 marks)
		$\int_0^5 kxdx = 1$; Solving, $k = 2/25$	
	(ii)	Calculate the expected value of X, E(X).	(3 marks)
		$E(X) = \int_0^5 xkx\,dx$; Solving, $E(X) = \mu = 10/3$	
	(iii)	Find the variance of X, Var(X). $Var(X) = E(X^2) - \mu^2$; Solving, $Var(X) = \frac{25}{2} - \frac{100}{9} = 1.3889$	(3 marks)
(d)	A bottling factory fills each bottle with juice, and it is known that the volume in the bottles has a normal distribution. The mean volume is 200 ml, and standard deviation is 10 ml. What is the probability that a bottle has less than 185 ml of juice?		(5 marks)

	Converting to a standard normal random variable, $Y = \frac{X - 200}{10}$ For X < 185, we find the probability that Y is less than -15/10. Looking up the standard gaussian distribution table, P(Y< -15/10) = 6.68%		
(e)	An electricity supply company is interested to know how fast it can respond to reported power outages. The company wants the mean time from report to arrival at the area of power outage to be within 20 minutes. It recorded the response time from 40 outages and found it to have a mean of 18 mins. The standard deviation is known to be 3 mins. Can the company be 95% sure that the response time is acceptable?		(5 marks)
	We are only interested to know whether the response time is greater than 20 mins, and not if it is much shorter than 20mins. For the one-tailed probability, 95% confidence requires the critical value to be 1.65. Hence, we want to check if $\mu \leq \bar{x} + \frac{z_\alpha \sigma}{\sqrt{n}}$ We have $18 + \frac{1.65\times3}{\sqrt{40}} = 18.78$. Thus the response time is acceptable.		
(f)	The mean weight of a bag of flour is supposed to be μ = 500g. It is known that the standard deviation is σ = 10g. The null hypothesis is μ = 500, and the alternative hypothesis is μ ≠ 500. We will reject the null hypothesis if the sampled means of 50 bags $\bar{x}$ is such that $\bar{x} > 502$ or $\bar{x} < 498$.		
	(i)	What is the test statistic Y that is to be used?	(2 marks)
		$Y = \frac{\bar{X} - 500}{10/\sqrt{50}}$	
	(ii)	With the help of the standard normal table, obtain the probability of a Type I error.	(3 marks)
		Type I error occurs if $\bar{X} < 498$ or $\bar{X} > 502$ $P(\bar{X} < 495) + P(\bar{X} > 505)$ $= P\left(Z < \frac{498-500}{10/\sqrt{50}}\right) + P\left(Z > \frac{502-500}{10/\sqrt{50}}\right)$ $= P(Z < -1.4142) + P(Z > 1.4142)$ = 2 x (1 - 0.9207) = 0.1586	
(g)	The simplest relationship between a dependent variable Y and a variable x is $Y = \beta_0 + \beta_1 x + \epsilon$.		
	(i)	An important hypothesis to test for linear regression analysis is whether β_1= 0. Briefly explain why.	(3 marks)
		β_1 gives the slope of the straight line. If it is zero, then a linear relationship does not hold between the variable x, and the dependent random variable Y.	
	(ii)	Which distributions can be used to test the hypothesis above.	(2 marks)
		Chi square distribution & t-distribution.	

大样本，正态分布
单边（不超过）

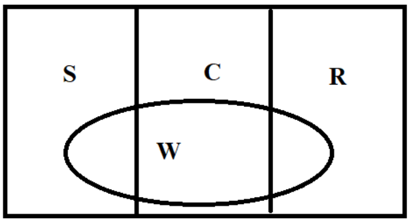
$\bar{x}=18$
$s=3$

$\bar{x} + \frac{1.65\times3}{\sqrt{40}} = 18.78$

大样本，
双边，正
态分布

H0对，但不接受

QUESTION 2

Over a long period of time in a town, it was found that the proportion of sunny days is 0.6, cloudy days is 0.3 and rainy days is 0.1. It was also found that cloudy and rainy days are twice as likely to be windy compared to sunny days. The record shows that 20% of the days are windy.		
(a)	Draw the Venn diagram showing the events sunny (S), cloudy (C), rainy (R) and windy (W).	(5 marks)
		
(b)	What is the probability that the town has windy conditions given that it is sunny?	(5 marks)
	P(W) = P(W|S)P(S) + P(W|C)P(C) + P(W|R)P(R) 0.2 = k0.6 + 2k0.3 + 2k0.1 = k(0.6+0.6+0.2) k = 0.2/1.4 = 0.1429	
(c)	What is the probability that the town experience a sunny day given that it is windy?	(5 marks)
	P(S|W) = P(S,W)/P(W) = P(W|S)P(S)/P(W) = 0.1429*0.6/0.2 = 0.4287	
(d)	What is the probability that a day is not rainy but windy, P(~R,W)?	(5 marks)
	P(~R,W) = P(S,W) + P(C,W) = P(W|S)P(S) +P(W|C)P(C) = 0.1429 x 0.6 + 0.2858 x 0.3 = 0.1714	

QUESTION 3

Consider the 18-letter string “TO BE OR NOT TO BE”.		
(a)	If the 18 cards each of which has one of the letters above written on it are placed in a bag, what is the probability of picking out the 5 cards with spaces on them without replacement?	(5 marks)
	The probability is 5/18 x 4/17 x 3/16 x 2/15 x 1/14 = 1.1671 x 10^{-4}	
(b)	How many different arrangements can be made from these 18 characters?	(5 marks)
	Number of different arrangements is 18!/(5!4!3!2!2!) = 9.2627 x 10^{10}	
(c)	A keyboard with 27 keys, from ‘A’ to ‘Z’, and space (‘ ’) is connected to a monitor screen. What is the number of different 18-letter strings that can be produced from this keyboard?	(5 marks)
	Number of 18 letter strings is 27^{18} = 5.815e25	

(d)	If a monkey pressed the keys on the keyboard randomly at one keypress per second, what is the expected amount of time before the string above will appear on the monitor screen?	(5 marks)
	Consider these attempts as Bernoulli trials. The probability one particular character typed is the first of the whole 18 in order is 1/5.815e25 and can be considered a success. Let Y be the geometric random variable that records the trials until the first success at the k^{th} character typed. We know E(Y) = 1/p = 5.815e25 Hence on average, the time needed to get the whole string out correctly is 5.815e25 seconds since each trial takes 1 second. Translated to years, it is 1.844e18 years.	

QUESTION 4 Confidence Intervals

A factory making expensive pieces of a delicate sensor needs to check how often each sensor can take a shock before it fails. The results on tests on 10 pieces of sensor is shown below:

10	12	11	13	12	15	13	11	14	13

A Student t-distribution is used to construct a 98% confidence interval for the mean number of shocks before failure.

小样本，
t分布，
双边

(a)	What is the sample mean for the number of shocks the sensor can take before it fails?	(4 marks)
	Sample mean is 12.40	
(b)	What is the sample standard deviation for the number of shocks?	(6 marks)
	Sample std dev is 1.5055	
(c)	From the t-distribution table, obtain the critical values for a 98% confidence interval for the mean for this case.	(4 marks)
	From table, critical value is 2.821 for 98% two-tailed confidence interval with 9 degrees of freedom.	
(d)	Calculate the range of the estimated mean number of shocks within that 98% confidence interval.	(6 marks)
	The test statistic is $t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$. The 98% confidence interval is therefore $\bar{x} - t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$ Plugging in values, we get $11.0570 \le \mu \le 13.7430$	

卡方检测 柏松分布的模型

Chi-Square Tests

QUESTION 5

The number of customers arriving at a bank is thought to have a Poisson distribution. The number of customers walking into the bank was recorded for a total of 150 randomly selected 1-minute intervals. The result is shown below:

No. of customers	0	1	2	3	>4
Frequencies	27	46	44	33	0

From the table, it can be seen, for example that there were 27 1-minute intervals where no customer stepped into the bank.

(a) What is the estimated parameter λ? (5 marks)

Estimated $\lambda = \frac{\sum_{i=0}^{4} i n_i}{150} = 1.5533$

(b) What is the test statistic to be used in a χ^2 test? (3 marks)

$\chi_0^2 = \sum_{i=1}^{4} \frac{(E_i - O_i)^2}{E_i}$ brackets essential here

(c) The null hypothesis H_0 is that the arrivals of customers follow a Poisson process. What value must χ_0^2 be greater than to reject H_0? (4 marks)

From the table, $\chi^2_{2,0.05} > 5.99$

(d) Calculate the test statistic value. (5 marks)

0.1723

$\chi^2_{2,0.05} = 0.6968 + 0.2209 + 0.8483 + 0.2444 = 2.3594$

(e) Explain briefly if we can reject the null hypothesis for significance level 0.05. (3 marks)

Since what we got is within the limit, we cannot reject the hypothesis that the arrival of the customers follows a Poisson process with λ = 1.5533.

0	1	2	3	>4	
27	46	44	33	0	=150
31.7	493	38.3	17.8	10.9	=150
			30.7		

$$\boxed{150 \times \frac{e^{-\lambda}(\lambda)^k}{k!}}$$