BE in Electronic Engineering with Communications
BE in Electronic Engineering with Computers
BE in Electronic Engineering
BSc in Robotics and Intelligent Devices

Year 3

SEMESTER 1
2018-2019

EE304
Probability and Statistics

Dr. J.K. Slingerland

Time allowed: 2 hours

Question 1 is **compulsory**

Answer question 1 and two other questions

Question 1 carries **50 marks**

All other questions carry **25 marks**

Pages 7 and 8 give mathematical formulae which may be used without derivation.

QUESTION 1 – This question is compulsory

(a) E and F are events, with probabilities $P(E) = 0.55$ and $P(F) = 0.45$. (5 marks)
We are also given that $P(E \cup F) = 0.75$

(i) Calculate $P(E \cap F)$

(ii) Calculate $P(E^c|F)$

(b) A random variable X has a probability density function given by (5 marks)

$$f(x) = \begin{cases} 0\,, & x < 0 \\ 12\left(x^2 - x^3\right)\,, & 0 \leq x < 1 \\ 0\,, & 1 \leq x \end{cases}$$

(i) Calculate the probability that $0.2 < x < 0.4$

(ii) Calculate the cumulative distribution function $F(x)$.

(c) A fair 8-sided die, with sides numbered 1 to 8 and a fair 6-sided die with sides numbered 1 to 6 are each thrown once. What is the probability that the product of the outcomes is (5 marks)

(i) greater than or equal to 32?

(ii) greater than or equal to 32, if we are given that the sum of the scores is an odd number?

(d) How many different permutations of the letters in the word "STATISTICS" exist? If a box contains the letters of the word statistics and you take them all out in a random order, what is the probability that during the process, you draw three consecutive Ts? (5 marks)

(e) In the 2018 midterm elections in the USA, an aggregate of all polls predicted that the Democratic Party would take 49.7% of the vote. During the week before the elections, one particular poll predicted instead that the Democratic Party would take 45%. This poll used a sample of 2500 potential voters. Based on this poll, construct a 99% confidence interval for the proportion of voters that planned to vote for the Democratic Party, assuming that the poll was random and unbiased. (5 marks)

(f) The lifetime of a certain type of mobile phone headset is described by a Weibull distribution with scale parameter 7 years and shape parameter 3. (5 marks)
Headsets of this type are used in a 24 hour call centre that receives calls at a rate of 12 per headset per hour.

(i) After 5 years of use, what is the probability that a headset still functions?

(ii) Calculate the average and the standard deviation of the number of calls that are answered using a given headset in 5 years. Assume that the headset is in use at all times and the calls are uncorrelated random events.

(g) A cheese monger claims that a wheel of cheese they have for sale weighs on average at least 5 kg. To test this claim, a random sample of 12 wheels of cheese is purchased and it is found that the mean weight is 4.9 kg. The standard deviation in the weight of these cheese wheels is known to be 150 g. (5 marks)
Perform a hypothesis test of the cheese monger's claim at a 5% level of significance.

(h) The eruptions of a certain active volcano follow a Poisson distribution with a mean of 1 eruption per 8 years. (5 marks)

(i) How long do we have to wait to be 90% sure of witnessing an eruption?

(ii) If we wait until we have observed 5 eruptions, what is the probility that there was exactly one pair of eruptions that occurred within 3 years of each other?

(i) In 2007, a study found that the average height of Irish men was 1.77 m, while for Irish women it was 1.63 m. The standard deviation of women's heights is approximately 0.06 m. Assuming that the women's heights are normally distributed, calculate the probability for an Irish woman to be taller than the average for an Irish man. (5 marks)

(j) A survey of vegetable gardens has found that on average they host 35 snails per square metre, with a standard deviation of 10 snails. Use Chebyshev's inequality to get an upper bound for the probability of finding more than 70 snails on a given square metre. (5 marks)

QUESTION 2

(a) A gas supplier wants to estimate the mean volume of gas used per household per month. A random sample of 7 households is selected and monitored and the monthly gas usage (in cubic metres) for these households is given below. (12 marks)

$$160, 151, 111, 186, 188, 142, 203$$

(i) Estimate the mean and standard deviation of the monthly gas usage of a household from these samples. Explain briefly how you arrive at your answers.

(ii) Using your estimates, and assuming that the gas usage of a household is normally distributed, find a 99% confidence interval for the mean gas usage per household per month.

(b) Before sampling, the gas company had hypothesized that the monthly gas usage per household is normally distributed with a mean of $170\,\mathrm{m}^3$ and a standard deviation of $25\,\mathrm{m}^3$. In order to test this hypothesis they use the sample data in part (a) and sort it into 3 bins, described by the intervals $(-\infty, 165)$, $[165, 185)$ and $[185, \infty)$. (13 marks)

(i) Construct the expected numbers of samples in each bin, according to the gas supplier's hypothesis.

(ii) Calculate χ^2 for the sample.

(iii) Can you reject the hypothesis on the distribution with 95% confidence?

QUESTION 3

(a) Research shows that about 55% of all email sent is spam. A spam filter is meant to detect this and filter it before it reaches the user. The filter is able to identify a spam message as spam with 99% probability. However there is also a 0.3% probability that it will misidentify a genuine email message as spam. Based on these data, (10 marks)

(i) What percentage of all messages will be identified as spam?

(ii) What percentage of the messages identified as spam are actually not spam?

(iii) What percentage of the messages identified as genuine are actually spam?

(b) Despite the presence of a spam filter, a given user still receives the occasional spam message into her inbox. On average she receives 2 spam messages per day. She thinks that the arrivals of these messages are most likely random, uncorrelated events. If this is true (15 marks)

(i) What is the probability that she will receive at least 3 spam messages on a given day?

(ii) If she has just received a spam message, what is the probility that she will not receive another one for 2.5 days?

(iii) What is the probability that she receives at least 230 spam messages during a period of 100 days?

QUESTION 4

(12 marks)

(a) Two events A and B have probabilities $P(A) = 0.65$, and $P(B) = 0.75$. A third event C, with $P(C) > 0$, is mutually exclusive with $A \cap B$. C is also independent of A. We are also given that $P(A \cup B) = 0.9$

(i) Calculate $P(A \setminus B)$

(ii) Show that $P(C) \leq 0.5$

(iii) Show that C is not independent of $A \setminus B$

(b) Two random variables, X and Y have ranges $\mathcal{R}_X = \{0, 1, 2\}$ and $\mathcal{R}_Y = \{0, 1\}$. (13 marks)
X and Y have joint probability mass function given by the following table:

X \ Y	0	1
0	0.4	0.04
1	0.2	0.1
2	0.05	0.21

(i) Find the marginal probability mass functions f_X and f_Y.

(ii) Show that X and Y are not independent.

(iii) Find the range and the probability mass function of the random variable $X+Y$

(iv) Calculate the variance of $X + Y$

Probability

- Basic Probability

$$0 \leq P(E) \leq 1 \qquad \forall E$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- If $F_1, F_2, ..., F_n$ is a collection of mutually exclusive and exhaustive events, then;

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + ... + P(E|F_n)P(F_n)$$

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + ... + P(E|F_n)P(F_n)}$$

- Binomial distribution with parameters n and p:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Negative binomial distribution with parameters k and r:

$$P(X_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

- Poisson distribution with mean λt:

$$P(X = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

- Cumulative distribution of exponential distribution with parameters λ:

$$F(x) = 1 - e^{-\lambda x}$$

- Cumulative distribution of Weibull distribution with shape parameter β and scale parameter δ:

$$F(x) = 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}$$

Statistics

$\sim \mathcal{N}(0,1)$ has normal distribution with mean 0 and standard deviation 1.
$\sim t_p$ has t-distribution with p degrees of freedom.
$\sim \chi^2_p$ has chi-squared distribution with p degrees of freedom.

- Estimation of mean (large sample):

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

- Estimation of mean (small sample):

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

- Estimation of proportion (large sample):

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1)$$

- Chi-squared goodness to fit:

$$\chi^2 = \sum_{i=1}^{n} \frac{(E_i - O_i)^2}{E_i} \sim \chi^2_{n-p-1}$$

- Chi-squared contingency table:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(E_{ij} - O_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$