Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

FACULTY OF SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS

Electronic & Electrical Engineering

Engineering
Junior Sophister
Annual Examinations

Semester 2, 2021

Probability and Statistics (3E3)

Date: 15th May 2021 **Venue: ONLINE, REAL-TIME** **Time: 12:00 – 14:00**

Anthony Quinn

Instructions to Candidates:

ANSWER QUESTION 1, and any THREE of the remaining five questions.

Question 1 is worth 30 marks *in toto*.
All remaining questions are worth equal marks (*i.e.* 70/3 marks each). The percentage division of marks within each question is indicated on the paper.
All symbols have their usual meaning.

Materials Permitted for this Examination:

This is an OPEN-BOOK examination, subject to declarations which must be signed by all candidates.

Computational supports, such as Matlab, are permitted. Their use should be noted in any answer supported in this way. In such cases, only the material appearing in your answer book will be assessed, and so you should fully specify in the answer book the method you have employed. Code will not be assessed in this examination.

Q.1 [Compulsory]

Answer ALL the following FIVE questions.

(a) Consider three propositions, S_1, S_2 and S_3, where $S_1 \Rightarrow S_2$, and where $\Pr[S_1] = 0.4$ and $\Pr[S_3] = 0.5$. Imposing no other logical constraints, but completing your probability model in any consistent way, evaluate $\Pr[\overline{S_3}|\overline{S_1}]$.

[20 %]

(b) Gamer A has a winning rate of 0.8 in a particular (repetitive) game, while gamer B's rate is 0.5. Given a run of 4 wins, what is the probability that it is gamer A who is playing?

[20 %]

(c) During the first fortnight of April 2021, the 14-day cumulative incidence rate of confirmed cases of COVID-19 per 100,000 was 192 in Dublin and 99 in Wicklow. The populations of these counties are 1.3 million and 140,000 respectively. Consider a community of 1000 people, located in one of these two counties. If no COVID-19 cases were reported there, what is the probability that it a Wicklow community?

[20 %]

[Q.1 CONTINUES OVERLEAF]

Q.1 [Continued]

(d) As of 13th April 2021, vaccine administrations in Ireland were as shown in Figure Q.1.

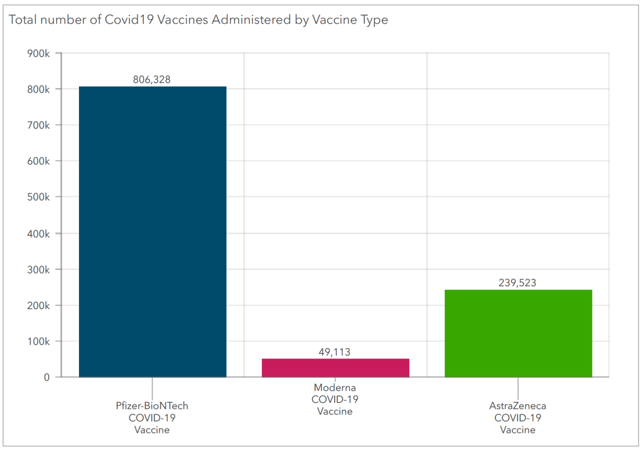

Figure Q.1

©Rialtas na hÉireann 2021

The population of Ireland is about 4.8 million. Consider a random sample of 7 people from the general population (vaccinated and unvaccinated). Write down an expression—which you do not need to evaluate—for the probability that the majority of these have been vaccinated with Pfizer-BioNTech, and equal numbers with Moderna and AstraZeneca.

[20 %]

(e) A device manufacturer tests the time to failure, t_i, $i = 1, \ldots, 10$, of ten of its devices, as follows:

t_i (years)	4.3	4.9	2.6	2.1	1.7	7.7	2.3	2.5	10.9	12.8

Estimate—using *both* a nonparametric and a parametric technique—the probability that the failure time of an untested device will be between 2 and 5 years.

[20 %]

Q.2

(a) Consider the probabilistic graphical model in Figure Q.2, which involves four Boolean (i.e. 2-state) nodes. Assume that I is sufficient for D. What is the maximum number of probability parameters required to complete the model, and under what conditions is this maximum number required? Define these probability parameters, and briefly specify statistical experiments for their estimation.

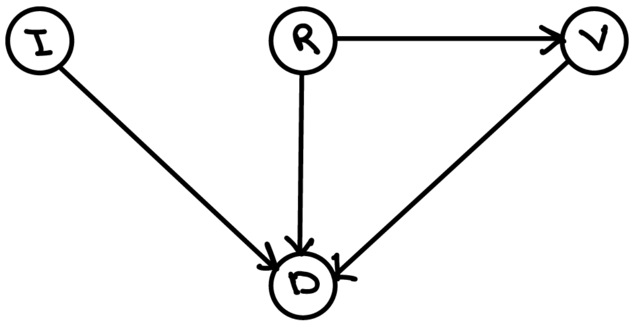

Figure Q.2

[40 %]

(b) A particular population is studied for the efficacy of a vaccine against COVID-19. In this population, 30% are considered to be at high-risk (R) from this disease. The vaccine was administered (V) to 95% of these high-risk subjects, but to only 10% of low-risk subjects. Infection (I) with the SARS-CoV-2 virus was subsequently confirmed in 15% of this (entire) population. Among those who had been vaccinated in the high-risk group, 15% went on to develop symptoms of COVID-19 (D), compared to 10% who had been vaccinated in the low-risk group. Meanwhile, among those who had *not* been vaccinated, these statistics were 35% and 15% in the high-risk and low-risk groups, respectively.

Use these statistics to parameterize the graphical model in Figure Q.2, and hence compute the following:

(i) the probability that an infected subject went on to develop symptoms of COVID-19;

[30 %]

(ii) the probability that a subject who develops symptoms of COVID-19 had been vaccinated.

[30 %]

Q.3

The DNA of every gene is formed as a particular sequence of four possible bases, labelled A, C, G and T, respectively. In a particular gene sub-sequence of length 1572, the following cooccurrence matrix was recorded:

$$\mathbf{C} = \begin{bmatrix} 185 & 101 & 69 & 161 \\ 74 & 41 & 45 & 103 \\ 86 & 6 & 34 & 100 \\ 171 & 115 & 78 & 202 \end{bmatrix}.$$

Here, the (i, j)th count, $c_{i,j}$, $1 \leq i, j \leq 4$, is the number of transitions from base j to base i (with the bases ordered as listed above) that occur at adjacent locations along the sub-sequence.

(a) A machine learning algorithm estimates the parameters of a homogeneous Markov chain (HMC) model of the gene DNA by consistently processing $\mathbf{C}$. Using these estimates, address the following:

(i) if all four bases are equiprobable at a particular position along the DNA, what is the probability that the base at the 3rd-next position is equal to the base at the 7th-next position?

[25 %]

(ii) if base A is found at a position along the DNA in the long-run, what is the probability that either base A or D occurs four positions earlier?

[25 %]

(b) A different machine learning algorithm assumes that the bases are independently and identically distributed (iid) at every position along the DNA. Once again consistently processing $\mathbf{C}$, compute the following inferences:

(i) the probability that the next base A will be observed between 5 and 9 positions from the current position.

[25 %]

(ii) the most probable numbers of times that the four bases will occur in a length-9 sequence, and the probability that these are the actual numbers that will occur.

[25 %]

Q.4

At a certain time of day, emails arrive at a server at an average rate of 2.5 per minute, while emails are deleted off the server at an average rate of 2.1 per minute. Stating your assumptions, address each of the following:

(a) Deduce the probability that the time to the second-next *change* in the number of emails on the server—due either to arrivals or deletions—is between 20 and 40 seconds.

[30 %]

(b) If the time to arrival of the second-next email is more than 40 seconds from now, deduce the probability that the time to arrival of the third-next email is at most one minute from now.

[35 %]

(c) Deduce an expression for the probability that the total number of emails on the server increases by more than five, in an interval of 10 minutes. You do not need to evaluate this expression.

[35 %]

Q.5

The target area of an annular β-radiation detector is illustrated in Figure Q.5.

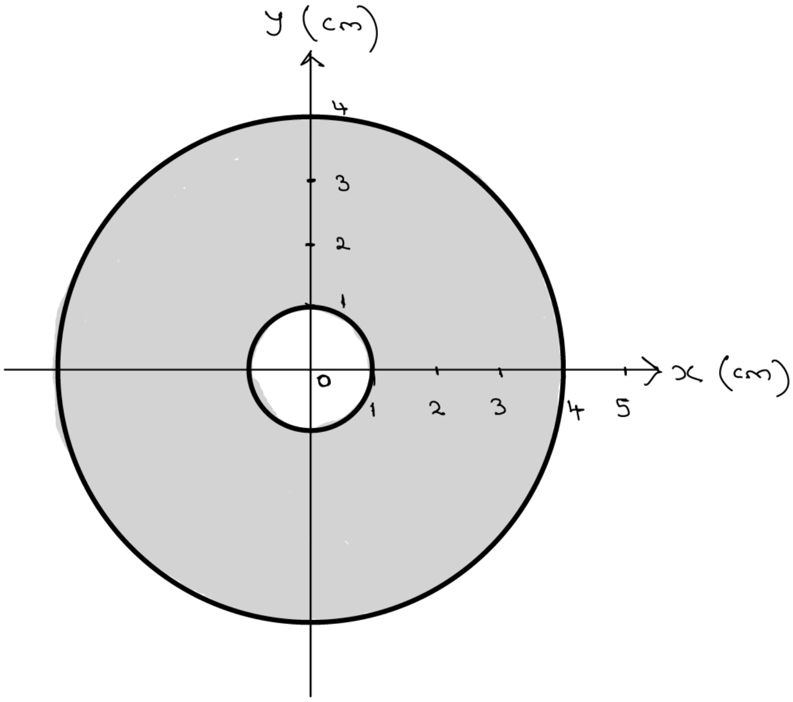

Figure Q.5

Let (X, Y) be the (Cartesian) x- and y-coordinates of a β-particle impinging on the detector, and let (D, Θ) be its polar coordinates. Stating all assumptions that you make, address each of the following:

(a) Provide a calibrated graph of the probability density function (PDF) of D. Hence, evaluate its standard interval and the probability of this interval.

[40 %]

(b) Deduce the regression function of D, regressing onto $X = x$. Plot this function appropriately, along with its its standard interval.

[40 %]

(c) Choose two of the four positional random variables above which are independent, and also two that are uncorrelated, justifying your choice in each case.

[20 %]

Q.6

(a) An electronic device manufacturer records the voltages, $V_{1,i}$ and $V_{2,i}$ (each in mV, relative to earth), at two distinct points in each of n test devices $i = 1, \ldots, n$.

(i) State the assumptions under which these data may be processed into a bivariate Gaussian model for a future voltage pair, (V_1, V_2), and specify the processing involved.

[25 %]

(ii) Assume that the Gaussian model parameters in (i) are estimated as follows:

$$\mathbf{m} = \begin{bmatrix} +1.5 \\ +1.0 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} +1.2 & -1.1 \\ -1.1 & +2.3 \end{bmatrix}$$

Deduce and sketch the regression line of V_1, regressing onto $V_2 = v_2$. Hence, or otherwise, find the range of values of v_2 for which the conditional probability that V_1 is negative is at most 0.1.

[35%]

(b) The variable strain induced by stresses at a particular location on a pylon is sequentially transduced to a voltage (all positive), v_i (mV), $i = 1, 2, \ldots$. The pylon can be in one of two states, $s_i \in \{1, 2\}$, at each measurement. $n = 15$ such strain-state measurements are recorded as follows:

v_i	0.6	1.2	1.4	1.5	0.3	0.7	0.5	1.0	1.2	0.9	0.2	0.6	1.1	1.0	.8
s_i	1	1	2	2	1	1	1	1	2	2	1	1	2	2	1

Two observers, O_1 and O_2, process these data into Gaussian models for prediction of future transduced strains. O_1 believes that the strains, V_i, are dependent on the state, S_i, while O_2 assumes they are independent. Estimate the parameters of an appropriate model for each observer. Hence, deduce the probability—for each observer—that the magnitude of a future strain voltage will be at most 2 mV. If this event occurs, what is O_1's probability that the pylon is in state 1?

[40%]

A note on the bivariate normal distribution:

If

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}), \quad \text{where} \quad \mathbf{m} = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix},$$

with

[CONTINUED OVERLEAF]

$$\sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y,$$

then

$$f(x|Y = y) = \mathcal{N}(m_{X|y}, \sigma^2_{X|y}),$$

where

$$\begin{aligned} m_{X|y} &= m_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - m_Y), \\ \sigma^2_{X|y} &= \sigma^2_X\left(1 - \rho^2_{XY}\right). \end{aligned}$$

Error Function Table

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

x	Hundredths digit of x									
	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	0.01128	0.02256	0.03384	0.04511	0.05637	0.06762	0.07886	0.09008	0.10128
0.1	0.11246	0.12362	0.13476	0.14587	0.15695	0.16800	0.17901	0.18999	0.20094	0.21184
0.2	0.22270	0.23352	0.24430	0.25502	0.26570	0.27633	0.28690	0.29742	0.30788	0.31828
0.3	0.32863	0.33891	0.34913	0.35928	0.36936	0.37938	0.38933	0.39921	0.40901	0.41874
0.4	0.42839	0.43797	0.44747	0.45689	0.46623	0.47548	0.48466	0.49375	0.50275	0.51167
0.5	0.52050	0.52924	0.53790	0.54646	0.55494	0.56332	0.57162	0.57982	0.58792	0.59594
0.6	0.60386	0.61168	0.61941	0.62705	0.63459	0.64203	0.64938	0.65663	0.66378	0.67084
0.7	0.67780	0.68467	0.69143	0.69810	0.70468	0.71116	0.71754	0.72382	0.73001	0.73610
0.8	0.74210	0.74800	0.75381	0.75952	0.76514	0.77067	0.77610	0.78144	0.78669	0.79184
0.9	0.79691	0.80188	0.80677	0.81156	0.81627	0.82089	0.82542	0.82987	0.83423	0.83851
1.0	0.84270	0.84681	0.85084	0.85478	0.85865	0.86244	0.86614	0.86977	0.87333	0.87680
1.1	0.88021	0.88353	0.88679	0.88997	0.89308	0.89612	0.89910	0.90200	0.90484	0.90761
1.2	0.91031	0.91296	0.91553	0.91805	0.92051	0.92290	0.92524	0.92751	0.92973	0.93190
1.3	0.93401	0.93606	0.93807	0.94002	0.94191	0.94376	0.94556	0.94731	0.94902	0.95067
1.4	0.95229	0.95385	0.95538	0.95686	0.95830	0.95970	0.96105	0.96237	0.96365	0.96490
1.5	0.96611	0.96728	0.96841	0.96952	0.97059	0.97162	0.97263	0.97360	0.97455	0.97546
1.6	0.97635	0.97721	0.97804	0.97884	0.97962	0.98038	0.98110	0.98181	0.98249	0.98315
1.7	0.98379	0.98441	0.98500	0.98558	0.98613	0.98667	0.98719	0.98769	0.98817	0.98864
1.8	0.98909	0.98952	0.98994	0.99035	0.99074	0.99111	0.99147	0.99182	0.99216	0.99248
1.9	0.99279	0.99309	0.99338	0.99366	0.99392	0.99418	0.99443	0.99466	0.99489	0.99511
2.0	0.99532	0.99552	0.99572	0.99591	0.99609	0.99626	0.99642	0.99658	0.99673	0.99688
2.1	0.99702	0.99715	0.99728	0.99741	0.99753	0.99764	0.99775	0.99785	0.99795	0.99805
2.2	0.99814	0.99822	0.99831	0.99839	0.99846	0.99854	0.99861	0.99867	0.99874	0.99880
2.3	0.99886	0.99891	0.99897	0.99902	0.99906	0.99911	0.99915	0.99920	0.99924	0.99928
2.4	0.99931	0.99935	0.99938	0.99941	0.99944	0.99947	0.99950	0.99952	0.99955	0.99957
2.5	0.99959	0.99961	0.99963	0.99965	0.99967	0.99969	0.99971	0.99972	0.99974	0.99975
2.6	0.99976	0.99978	0.99979	0.99980	0.99981	0.99982	0.99983	0.99984	0.99985	0.99986
2.7	0.99987	0.99987	0.99988	0.99989	0.99989	0.99990	0.99991	0.99991	0.99992	0.99992
2.8	0.99992	0.99993	0.99993	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995	0.99996
2.9	0.99996	0.99996	0.99996	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998
3.0	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999	0.99999
3.1	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
3.2	0.99999	0.99999	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000