Maynooth University
National University of Ireland Maynooth

SEMESTER 1
2022-2023

EE304FZ
Probability and Statistics

Dr. Lo Fook-Loong, Prof. Ronan Farrell, Dr. Martin 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

Instructions:
You have 2 hours to complete this examination paper. You then have a further 15 minutes on top of that, to scan and upload your work to the dropbox provided on Moodle for EE304FZ. This additional time should allow for any technical difficulties that may arise.

Certification of Authenticity: Please be aware on submission of your exam work, you are committing to your lecturer that all the information on your exam script is your own work with no assistance. The Lecturer reserves the right to interview the student if deemed appropriate.

Lecturer's email address : lo.fook-loong@mu.ie

QUESTION 1

(a) A bag holds 3 red, 3 blue and 3 green balls.

(i) Three balls are taken out at random. What is the probability that all three balls are of the same colour? (3 marks)

(ii) What is the probability that 3 random balls taken out one by one are in the order red, green and blue? (2 marks)

(b) Three balls, colored red, green and blue, are to be placed in a row of 10 cups. What is the total number of ways that the balls can be arranged in the cups? (5 marks)

(c) Two proof-readers are given the task of finding typos in a new book. They work independently and one finds A typos, the other B typos. The number of common typos found by both is C. (5 marks)
Assume that the total number of typos T in the book is large, and that finding a typo can be considered a Bernoulli trial with proportion of success for one reader being p, and for the other q. Estimate the number of undiscovered typos U in terms of A, B and C.

(d) Students in a college will be listed on the Dean's List if they obtain yearly average marks in the top 5%. In a particular year, the students' average marks, which can be described as approximately gaussian, has a mean of 58 and a variance of 12. Determine what minimum marks a student must have to be on the Dean's List. (5 marks)

(e) A fast-food restaurant has 3 servers at the counter each of whom can serve a customer in 2 minutes. Customers arrive at a rate of 4 per minute in a Poisson process. What is the probability that a new customer entering the restaurant can be served immediately? (5 marks)

(f) Time to failure of a machine part is given by the Weibull distribution with scale parameter $\delta = 12$ months and parameter $\beta = 2$. The part has to be replaced before the probability of failure reaches 70% to prevent shutdown of the machine. How often does the part need to be changed? Give your answer in months. (5 marks)

(g) (i) To obtain the 96% confidence interval that the population mean lies within a range centred at the sample mean when it is obtained from a large number of samples, what is the $z_{\alpha/2}$ value that we should use? (3 marks)

(ii) To obtain the 95% confidence interval that the population mean lies within a range centred at the sample mean when it is obtained from 15 samples, what is the $t_{n-1,\alpha/2}$ value that we should use? (2 marks)

(h) A variable y is related to the independent variable x by

$$y = \beta_1 x + \beta_0.$$

When y is measured, because of errors in measurement, a random variable Y is obtained that is y with an error that is gaussian distributed.

(i) A very simple way to estimate β_0 and β_1 is to draw the best straight line by eye passing through the measured sample points (y_i, x_i). How do you obtain the estimates of β_0 and β_1 from the line? (3 marks)

(ii) A more accurate method is to minimize the error sum of squares. Write out the mathematical equation defining the error sum of squares for this case. (2 marks)

QUESTION 2

A digital communication system experiences a bit error rate of 10^{-3}. This means that on average, one bit out of one thousand is received in error. An error occurs when a transmitted one is received as a zero, or vice-versa.
The bits are sent out in blocks of 100, some of which are check bits that allow the receiver to know if any errors have occurred during transmission. If errors are detected, the receiver will request the transmitter to resend the whole block of 100 bits.

(a) What is the probability that the whole block of 100 bits is received without any error? (6 marks)

(b) What is the probability that a block has exactly one bit error? (4 marks)

(c) What is the block error rate, that is, the percentage of blocks received with one or more bit errors? (4 marks)

(d) If the error detection scheme used is upgraded to an error correction scheme, where single bit errors in a block can be corrected, what will be the new block error rate? (6 marks)

QUESTION 3

Two friends, Pascal and Fermat, like to play games of chance and bet small amounts of money. They will put $10 each into the pot, and whoever wins a game will collect the $20 pot.

(a) One of the games they play is to draw one card each from 13 of the same suit. The order of the cards is King, Queen, Jack, 10, 9, down to 1. Whoever draws a higher card in a round wins a point. Whoever gets 4 points wins the pot. (5 marks)
During a game when Pascal obtained 2 points and Fermat 1, they were forced to stop. How should the pot of $20 be divided fairly between them according to the probability of each winning should the game carry on?

(b) The second game is each taking turns to roll a die. Whoever gets a higher number wins a point. What is the probability that after 3 turns, both Pascal and Fermat have 1 point each? (5 marks)

(c) The third game is the rolling of 4 dice. If there is at least 1 six, the person rolling will win the pot. If Pascal is the one rolling the die, what is his probability of winning? (5 marks)

(d) The fourth game is rolling a pair of dice 24 times. The person will win if there is at least one double six. If Fermat is rolling, what is the probability of him winning? (5 marks)

QUESTION 4

A factory bottles energy drinks, filling each bottle with 100ml. The manager does not want to fill more as that will eat into the profit of the company, and she cannot fill less as consumers will complain.
You are asked to check whether the bottles contain 100ml of drink with 2% significance level. A sample of 100 bottles are taken and the mean volume is 99.5 ml. The variance is known to be 1.5ml.

(a) Formulate the null and alternative hypothesis. (4 marks)

(b) Write out the test statistics Z_0. (4 marks)

(c) With the given significance level of 2%, look up the value of $Z_{\alpha/2}$. (4 marks)

(d) Determine the critical regions. (4 marks)

(e) Explain whether it is appropriate to reject the Hypothesis H_0. (4 marks)

QUESTION 5

A medical researcher wishes to know whether obesity is related to the community the person is staying in. If obesity is independent of community, then the probability of finding an obese person will be just a Bernoulli trial, with success meaning that the person is obese.
A survey is taken in 100 communities, where 10 persons in each community have their heights and weights recorded. A BMI (body mass index) of 30 or more indicates obesity. The national average of obese people in the country is known to be 15%.

No. of obese person	0	1	2	3	4	>4
No. of communities	18	35	26	15	6	0

From the table, it can be seen, for example, that there were 35 communities which have 1 person in the sample of 10 who is obese.

(a) Give a formula for the test statistic needed for the χ^2 test. (4 marks)

(b) Find the frequencies of the number of obese people in a community if obesity is random. (6 marks)

(c) What is v, the number of degrees of freedom to use, when looking up the χ^2 table? Obtain the required χ^2 value for a significance level of 0.01. (4 marks)

(d) Calculate the value of the test statistic. (4 marks)

(e) Explain whether the evidence shows that obesity is a trait of people independent of the community they live in. (2 marks)

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:

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

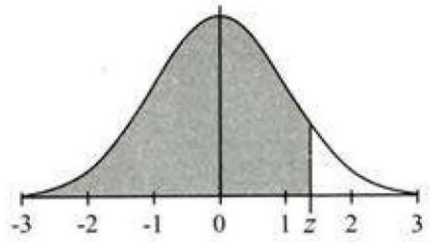
- Chi-squared contingency table:

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

Dóchúlachtaí don dáileadh normalach caighdeánach

I gcás z a thugtar, faightear ón tábla

$$P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$



Probabilities for the standard normal distribution

For a given z, the table gives

$$P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	0.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	0.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	0.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	0.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	0.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	0.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	0.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	0.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	0.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	0.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621

Clár / Contents

an dáileadh normalach (ar lean) — normal distribution (continued)

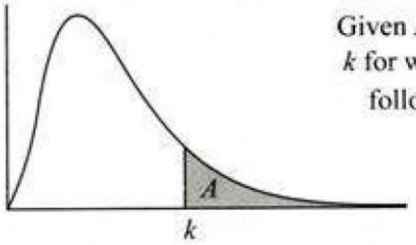
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.1	0.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	0.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	0.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	0.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	0.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	0.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	0.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	0.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	0.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	0.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	0.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	0.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	0.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	0.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	0.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	0.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	0.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	0.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	0.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	0.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Clár Staitisticí / Statistics Contents

Clár / Contents

Dáileadh chí-chearnaithe
luachanna criticiúla tástála aonfhoircní

Nuair a thugtar A, faightear ón tábla an luach ar k mar a bhfuil $P(X > k) = A$, áit a leanann X dáileadh chí-chearnaithe a bhfuil v céim saoirse aige.



Chi-squared distribution
one-tailed critical values

Given A, the table gives the value of k for which $P(X > k) = A$, where X follows a chi-squared distribution with v degrees of freedom.

v \ A	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.0000	0.0002	0.0010	0.0039	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	5.9915	7.3778	9.2103	10.597
3	0.0717	0.1148	0.2158	0.3518	7.8147	9.3484	11.345	12.838
4	0.2070	0.2971	0.4844	0.7107	9.4877	11.143	13.277	14.860
5	0.4117	0.5543	0.8312	1.1455	11.070	12.833	15.086	16.750
6	0.6757	0.8721	1.2373	1.6354	12.592	14.449	16.812	18.548
7	0.9893	1.2390	1.6899	2.1673	14.067	16.013	18.475	20.278
8	1.3444	1.6465	2.1797	2.7326	15.507	17.535	20.090	21.955
9	1.7349	2.0879	2.7004	3.3251	16.919	19.023	21.666	23.589
10	2.1559	2.5582	3.2470	3.9403	18.307	20.483	23.209	25.188
11	2.6032	3.0535	3.8157	4.5748	19.675	21.920	24.725	26.757
12	3.0738	3.5706	4.4038	5.2260	21.026	23.337	26.217	28.300
13	3.5650	4.1069	5.0088	5.8919	22.362	24.736	27.688	29.819
14	4.0747	4.6604	5.6287	6.5706	23.685	26.119	29.141	31.319

Clár / Contents

dáileadh chí-chearnaithe (ar lean)

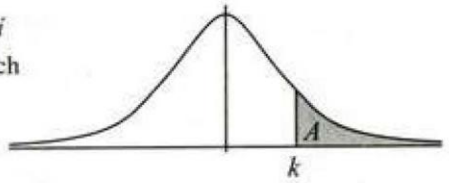
chi-squared distribution (continued)

v \ A	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
15	4.6009	5.2293	6.2621	7.2609	24.996	27.488	30.578	32.801
16	5.1422	5.8122	6.9077	7.9616	26.296	28.845	32.000	34.267
17	5.6972	6.4078	7.5642	8.6718	27.587	30.191	33.409	35.718
18	6.2648	7.0149	8.2307	9.3905	28.869	31.526	34.805	37.156
19	6.8440	7.6327	8.9065	10.117	30.144	32.852	36.191	38.582
20	7.4338	8.2604	9.5908	10.851	31.410	34.170	37.566	39.997
21	8.0337	8.8972	10.283	11.591	32.671	35.479	38.932	41.401
22	8.6427	9.5425	10.982	12.338	33.924	36.781	40.289	42.796
23	9.2604	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.8862	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.43	104.21
80	51.172	53.540	57.153	60.391	101.88	106.63	112.33	116.32
90	59.196	61.754	65.647	69.126	113.15	118.14	124.12	128.30
100	67.328	70.065	74.222	77.929	124.34	129.56	135.81	140.17

Clár Staitisticí / Statistics Contents

Clár / Contents

t-dháileadh Student
luachanna criticiúla tástála aonfhoircnì
Nuair a thugtar A, faightear ón tábla an luach ar k mar a bhfuil $P(T > k) = A$, áit a leanann T, t-dháileadh a bhfuil ν céim saoirse aige.



Student's t-distribution
one-tailed critical values
Given A, the table gives the value of k for which $P(T > k) = A$, where T follows a t-distribution with ν degrees of freedom.

ν \ A	0.1	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	0.00005
1	3.078	6.314	12.71	31.82	63.66	318.3	636.6	3183	6366
2	1.886	2.920	4.303	6.965	9.925	22.33	31.60	70.70	99.99
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92	22.20	28.00
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	13.03	15.54
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	9.678	11.18
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	8.025	9.082
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7.063	7.885
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	6.442	7.120
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	6.010	6.594
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	5.694	6.211
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	5.453	5.921
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	5.263	5.694
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	5.111	5.513
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	4.985	5.363

Clár / Contents

t-dháileadh Student (ar lean) — Student's *t*-distribution (continued)

ν \ A	0.1	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	0.00005
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	4.880	5.239
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	4.790	5.134
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	4.715	5.043
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	4.648	4.966
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	4.590	4.899
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	4.539	4.838
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	4.492	4.785
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	4.452	4.736
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	4.416	4.694
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	4.382	4.654
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	4.352	4.619
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	4.324	4.587
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689	4.299	4.556
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	4.276	4.531
29	1.311	1.699	2.045	2.462	2.756	3.396	3.660	4.254	4.505
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	4.234	4.482
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551	4.094	4.321
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496	4.014	4.228
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460	3.962	4.169
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416	3.899	4.095
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390	3.861	4.054
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.290	3.719	3.891

Clár Staitisticí / Statistics Contents

Clár / Contents