SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES

F21SA
Statistical Modelling and Analysis

Semester 1 - 2020/2021

Duration: 24 hours

Question	Marks
1	9
2	5
3	9
4	9
5	8
Total Marks	40

Attempt ALL FIVE questions.
To receive full credit you must show your work and explain your answers.

Excerpts from Cambridge Statistical Tables are provided on pages 7–9

1. **(a)** Two companies, A and B, produce phones in country X. The number of phones produced by A is twice the number of phones produced by B. We know that 9% of phones produced by A and 10% of phones produced by B are faulty. Given that a randomly selected phone from X is faulty, find the probability that it was produced by B. [3 marks]

(b) The waiting time for processing of a refund by company B has an exponential distribution with mean 100 days. Given that a customer has been already waiting for 120 days for their refund, find the average remaining waiting time. [2 marks]

(c) Use the Central Limit Theorem to approximate the probability that in a batch of randomly selected 1000 phones from X, no more than 80 are faulty. [4 marks]

[$\mathrm{Total}\ 9$ marks]

[PLEASE TURN OVER]

2. The number of a certain kind of bacteria in k litres of water from a local reservoir is believed to have a Poisson distribution with mean $k\lambda$ for any $k \geq 1$, where $\lambda > 0$ is an unknown parameter. A testing method allows researchers to tell whether a sample of water contains the bacteria, but not their exact number. Among 50 samples of 1 litre each, the bacteria was present in 12 samples, and then in additional 5 samples of 2 litre each, no bacteria presence was registered. Based on the available data, prove that the maximum likelihood estimate of the average number of the bacteria in 100 litres of water from the reservoir is $100 \log(5/4)$.

[$\mathrm{Total}\ 5$ marks]

[PLEASE TURN OVER]

3. A group of researchers collected 50 specimens of a certain plant and measured their height. The sample mean was $\bar{x} = 20.68$ cm and the sample standard deviation $s = 2.59$ cm. Based on this experiment, the researchers proposed the hypothesis H_0 that the mean height of this plant is 20 cm, with the alternative hypothesis being that the mean height is greater than 20 cm.

(a) Is there enough evidence to reject H_0 at significance level 1%? [3 marks]

(b) State the p-value for the test in (a). Would you reject H_0 at significance level 5%? [2 marks]

(c) The experiment was subsequently reviewed by a different group of researchers who discovered that the two specimens with extreme heights of 9.02 cm and 29.95 cm, respectively, were incorrectly classified and in fact belong to a different species. Hence their heights should be removed from the sample. With the modified data, is there enough evidence to reject H_0 at significance level 1%? [4 marks]

[$\text{Total } 9$ marks]

4. A factory produces components that are supposed to withstand high temperatures. In an experiment, 10 components were tested and the times until their melting were recorded. The following table summarises the data:

Temperature in $°C$ (x)	180	185	190	195	200	205	210	215	220	225
Melting time in min (y)	67	64	62	57	51	49	46	41	35	33

For these data

$$\sum x_i = 2025, \ \sum x_i^2 = 412125, \ \sum y_i = 505, \ \sum y_i^2 = 26791, \ \sum x_i y_i = 100640$$

(a) Calculate S_{xx}, S_{yy}, and S_{xy}. [2 marks]

(b) Calculate the fitted linear regression equation of decrease in y on x. [3 marks]

(c) In practice, the component will need to be able to work in $200°C$ for at least 50 minutes. Find a suitable one-sided 99% confidence interval for the melting time for a temperature of $200°C$. Based on the result, comment on whether the data fits the model well. [4 marks]

[$\text{Total } 9$ marks]

[PLEASE TURN OVER]

5. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ denote a realised sample from a geometric distribution with unknown parameter $\theta \in (0,1)$. The likelihood of a single observation x_k is

$$L(\theta; x_k) = \theta(1-\theta)^{x_k}\,.$$

Assume that observations are i.i.d. given the value of θ.

(a) Let α, $\beta > 0$. Assuming a $\mathrm{Beta}(\alpha, \beta)$ prior for θ, derive the posterior distribution for θ given the observed data $\boldsymbol{x}$. [3 marks]

(b) Report the posterior mean and variance for θ. [2 marks]

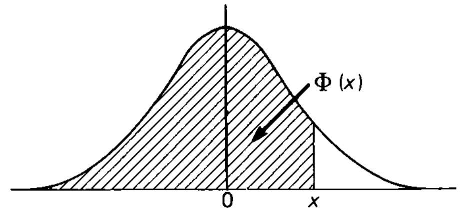
(c) Consider a prior $\pi(\theta) \propto \frac{1}{\theta(1-\theta)}$. Using the law of large numbers, check if the posterior mean is a consistent estimator of θ. [3 marks]

[$\text{Total } 8$ marks]

[END OF PAPER]

TABLE 4. THE NORMAL DISTRIBUTION FUNCTION

The function tabulated is $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$. $\Phi(x)$ is the probability that a random variable, normally distributed with zero mean and unit variance, will be less than or equal to x. When $x < 0$ use $\Phi(x) = 1 - \Phi(-x)$, as the normal distribution with zero mean and unit variance is symmetric about zero.



x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0·00	0·5000	0·40	0·6554	0·80	0·7881	1·20	0·8849	1·60	0·9452	2·00	0·97725
·01	·5040	·41	·6591	·81	·7910	·21	·8869	·61	·9463	·01	·97778
·02	·5080	·42	·6628	·82	·7939	·22	·8888	·62	·9474	·02	·97831
·03	·5120	·43	·6664	·83	·7967	·23	·8907	·63	·9484	·03	·97882
·04	·5160	·44	·6700	·84	·7995	·24	·8925	·64	·9495	·04	·97932
0·05	0·5199	0·45	0·6736	0·85	0·8023	1·25	0·8944	1·65	0·9505	2·05	0·97982
·06	·5239	·46	·6772	·86	·8051	·26	·8962	·66	·9515	·06	·98030
·07	·5279	·47	·6808	·87	·8078	·27	·8980	·67	·9525	·07	·98077
·08	·5319	·48	·6844	·88	·8106	·28	·8997	·68	·9535	·08	·98124
·09	·5359	·49	·6879	·89	·8133	·29	·9015	·69	·9545	·09	·98169
0·10	0·5398	0·50	0·6915	0·90	0·8159	1·30	0·9032	1·70	0·9554	2·10	0·98214
·11	·5438	·51	·6950	·91	·8186	·31	·9049	·71	·9564	·11	·98257
·12	·5478	·52	·6985	·92	·8212	·32	·9066	·72	·9573	·12	·98300
·13	·5517	·53	·7019	·93	·8238	·33	·9082	·73	·9582	·13	·98341
·14	·5557	·54	·7054	·94	·8264	·34	·9099	·74	·9591	·14	·98382
0·15	0·5596	0·55	0·7088	0·95	0·8289	1·35	0·9115	1·75	0·9599	2·15	0·98422
·16	·5636	·56	·7123	·96	·8315	·36	·9131	·76	·9608	·16	·98461
·17	·5675	·57	·7157	·97	·8340	·37	·9147	·77	·9616	·17	·98500
·18	·5714	·58	·7190	·98	·8365	·38	·9162	·78	·9625	·18	·98537
·19	·5753	·59	·7224	·99	·8389	·39	·9177	·79	·9633	·19	·98574
0·20	0·5793	0·60	0·7257	1·00	0·8413	1·40	0·9192	1·80	0·9641	2·20	0·98610
·21	·5832	·61	·7291	·01	·8438	·41	·9207	·81	·9649	·21	·98645
·22	·5871	·62	·7324	·02	·8461	·42	·9222	·82	·9656	·22	·98679
·23	·5910	·63	·7357	·03	·8485	·43	·9236	·83	·9664	·23	·98713
·24	·5948	·64	·7389	·04	·8508	·44	·9251	·84	·9671	·24	·98745
0·25	0·5987	0·65	0·7422	1·05	0·8531	1·45	0·9265	1·85	0·9678	2·25	0·98778
·26	·6026	·66	·7454	·06	·8554	·46	·9279	·86	·9686	·26	·98809
·27	·6064	·67	·7486	·07	·8577	·47	·9292	·87	·9693	·27	·98840
·28	·6103	·68	·7517	·08	·8599	·48	·9306	·88	·9699	·28	·98870
·29	·6141	·69	·7549	·09	·8621	·49	·9319	·89	·9706	·29	·98899
0·30	0·6179	0·70	0·7580	1·10	0·8643	1·50	0·9332	1·90	0·9713	2·30	0·98928
·31	·6217	·71	·7611	·11	·8665	·51	·9345	·91	·9719	·31	·98956
·32	·6255	·72	·7642	·12	·8686	·52	·9357	·92	·9726	·32	·98983
·33	·6293	·73	·7673	·13	·8708	·53	·9370	·93	·9732	·33	·99010
·34	·6331	·74	·7704	·14	·8729	·54	·9382	·94	·9738	·34	·99036
0·35	0·6368	0·75	0·7734	1·15	0·8749	1·55	0·9394	1·95	0·9744	2·35	0·99061
·36	·6406	·76	·7764	·16	·8770	·56	·9406	·96	·9750	·36	·99086
·37	·6443	·77	·7794	·17	·8790	·57	·9418	·97	·9756	·37	·99111
·38	·6480	·78	·7823	·18	·8810	·58	·9429	·98	·9761	·38	·99134
·39	·6517	·79	·7852	·19	·8830	·59	·9441	·99	·9767	·39	·99158
0·40	0·6554	0·80	0·7881	1·20	0·8849	1·60	0·9452	2·00	0·9772	2·40	0·99180

TABLE 4. THE NORMAL DISTRIBUTION FUNCTION

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2·40	0·99180	2·55	0·99461	2·70	0·99653	2·85	0·99781	3·00	0·99865	3·15	0·99918
·41	·99202	·56	·99477	·71	·99664	·86	·99788	·01	·99869	·16	·99921
·42	·99224	·57	·99492	·72	·99674	·87	·99795	·02	·99874	·17	·99924
·43	·99245	·58	·99506	·73	·99683	·88	·99801	·03	·99878	·18	·99926
·44	·99266	·59	·99520	·74	·99693	·89	·99807	·04	·99882	·19	·99929
2·45	0·99286	2·60	0·99534	2·75	0·99702	2·90	0·99813	3·05	0·99886	3·20	0·99931
·46	·99305	·61	·99547	·76	·99711	·91	·99819	·06	·99889	·21	·99934
·47	·99324	·62	·99560	·77	·99720	·92	·99825	·07	·99893	·22	·99936
·48	·99343	·63	·99573	·78	·99728	·93	·99831	·08	·99896	·23	·99938
·49	·99361	·64	·99585	·79	·99736	·94	·99836	·09	·99900	·24	·99940
2·50	0·99379	2·65	0·99598	2·80	0·99744	2·95	0·99841	3·10	0·99903	3·25	0·99942
·51	·99396	·66	·99609	·81	·99752	·96	·99846	·11	·99906	·26	·99944
·52	·99413	·67	·99621	·82	·99760	·97	·99851	·12	·99910	·27	·99946
·53	·99430	·68	·99632	·83	·99767	·98	·99856	·13	·99913	·28	·99948
·54	·99446	·69	·99643	·84	·99774	·99	·99861	·14	·99916	·29	·99950
2·55	0·99461	2·70	0·99653	2·85	0·99781	3·00	0·99865	3·15	0·99918	3·30	0·99952

The critical table below gives on the left the range of values of x for which $\Phi(x)$ takes the value on the right, correct to the last figure given; in critical cases, take the upper of the two values of $\Phi(x)$ indicated.

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
			0·9994		0·99990		0·99995
3·075		3·263		3·731		3·916	
	0·9990		0·9995		0·99991		0·99996
3·105		3·320		3·759		3·976	
	0·9991		0·9996		0·99992		0·99997
3·138		3·389		3·791		4·055	
	0·9992		0·9997		0·99993		0·99998
3·174		3·480		3·826		4·173	
	0·9993		0·9998		0·99994		0·99999
3·215		3·615		3·867		4·417	
	0·9994		0·9999		0·99995		1·00000

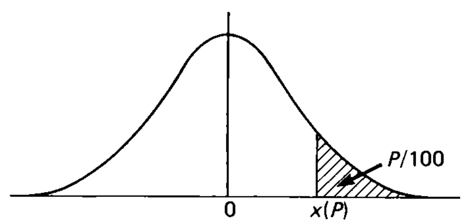
When $x > 3{\cdot}3$ the formula $1-\Phi(x) \doteqdot \frac{e^{-\frac{1}{2}x^2}}{x\sqrt{2\pi}}\left[1-\frac{1}{x^2}+\frac{3}{x^4}-\frac{15}{x^6}+\frac{105}{x^8}\right]$ is very accurate, with relative error less than $945/x^{10}$.

TABLE 5. PERCENTAGE POINTS OF THE NORMAL DISTRIBUTION

This table gives percentage points $x(P)$ defined by the equation

$$\frac{P}{100} = \frac{1}{\sqrt{2\pi}}\int_{x(P)}^{\infty} e^{-\frac{1}{2}t^2}\,dt.$$

If X is a variable, normally distributed with zero mean and unit variance, $P/100$ is the probability that $X \geqslant x(P)$. The lower P per cent points are given by symmetry as $-x(P)$, and the probability that $|X| \geqslant x(P)$ is $2P/100$.



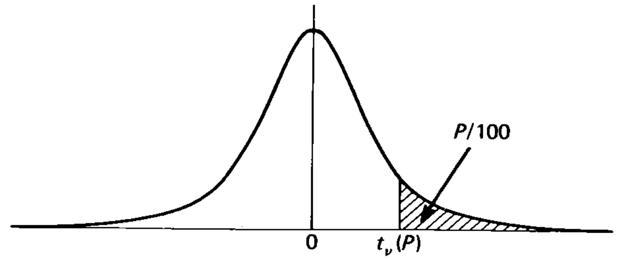
P	$x(P)$	P	$x(P)$	P	$x(P)$	P	$x(P)$	P	$x(P)$	P	$x(P)$
50	0·0000	5·0	1·6449	3·0	1·8808	2·0	2·0537	1·0	2·3263	0·10	3·0902
45	0·1257	4·8	1·6646	2·9	1·8957	1·9	2·0749	0·9	2·3656	0·09	3·1214
40	0·2533	4·6	1·6849	2·8	1·9110	1·8	2·0969	0·8	2·4089	0·08	3·1559
35	0·3853	4·4	1·7060	2·7	1·9268	1·7	2·1201	0·7	2·4573	0·07	3·1947
30	0·5244	4·2	1·7279	2·6	1·9431	1·6	2·1444	0·6	2·5121	0·06	3·2389
25	0·6745	4·0	1·7507	2·5	1·9600	1·5	2·1701	0·5	2·5758	0·05	3·2905
20	0·8416	3·8	1·7744	2·4	1·9774	1·4	2·1973	0·4	2·6521	0·01	3·7190
15	1·0364	3·6	1·7991	2·3	1·9954	1·3	2·2262	0·3	2·7478	0·005	3·8906
10	1·2816	3·4	1·8250	2·2	2·0141	1·2	2·2571	0·2	2·8782	0·001	4·2649
5	1·6449	3·2	1·8522	2·1	2·0335	1·1	2·2904	0·1	3·0902	0·0005	4·4172

TABLE 10. PERCENTAGE POINTS OF THE t-DISTRIBUTION

This table gives percentage points $t_\nu(P)$ defined by the equation

$$\frac{P}{100} = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}{\Gamma(\frac{1}{2}\nu)} \int_{t_\nu(P)}^{\infty} \frac{dt}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}}.$$

Let X_1 and X_2 be independent random variables having a normal distribution with zero mean and unit variance and a χ^2-distribution with ν degrees of freedom respectively; then $t = X_1/\sqrt{X_2/\nu}$ has Student's t-distribution with ν degrees of freedom, and the probability that $t \geqslant t_\nu(P)$ is $P/100$. The lower percentage points are given by symmetry as $-t_\nu(P)$, and the probability that $|t| \geqslant t_\nu(P)$ is $2P/100$.



The limiting distribution of t as ν tends to infinity is the normal distribution with zero mean and unit variance. When ν is large interpolation in ν should be harmonic.

P(%)	40	30	25	20	15	10	5%	2.5	1%	0.5	0.1	0.05
ν = 1	0.3249	0.7265	1.0000	1.3764	1.963	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	0.2887	0.6172	0.8165	1.0607	1.386	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	0.2767	0.5844	0.7649	0.9785	1.250	1.638	2.353	3.182	4.541	5.841	10.21	12.92
4	0.2707	0.5686	0.7407	0.9410	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.2672	0.5594	0.7267	0.9195	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.2648	0.5534	0.7176	0.9057	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.2632	0.5491	0.7111	0.8960	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.2619	0.5459	0.7064	0.8889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.2610	0.5435	0.7027	0.8834	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.2602	0.5415	0.6998	0.8791	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.2596	0.5399	0.6974	0.8755	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.2590	0.5386	0.6955	0.8726	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.2586	0.5375	0.6938	0.8702	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.2582	0.5366	0.6924	0.8681	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.2579	0.5357	0.6912	0.8662	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.2576	0.5350	0.6901	0.8647	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.2573	0.5344	0.6892	0.8633	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.2571	0.5338	0.6884	0.8620	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.2569	0.5333	0.6876	0.8610	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.2567	0.5329	0.6870	0.8600	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.2566	0.5325	0.6864	0.8591	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.2564	0.5321	0.6858	0.8583	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.2563	0.5317	0.6853	0.8575	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.2562	0.5314	0.6848	0.8569	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.2561	0.5312	0.6844	0.8562	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.2560	0.5309	0.6840	0.8557	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.2559	0.5306	0.6837	0.8551	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.2558	0.5304	0.6834	0.8546	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.2557	0.5302	0.6830	0.8542	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.2556	0.5300	0.6828	0.8538	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	0.2555	0.5297	0.6822	0.8530	1.054	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	0.2553	0.5294	0.6818	0.8523	1.052	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	0.2552	0.5291	0.6814	0.8517	1.052	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	0.2551	0.5288	0.6810	0.8512	1.051	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	0.2550	0.5286	0.6807	0.8507	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.2547	0.5278	0.6794	0.8489	1.047	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	0.2545	0.5272	0.6786	0.8477	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.2539	0.5258	0.6765	0.8446	1.041	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.2533	0.5244	0.6745	0.8416	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291