NOT TO BE TAKEN AWAY

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT2602 PROBABILITY AND STATISTICS II

13 May, 2023 Time: 2:30 p.m. - 4:30 p.m.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Answer **ALL FIVE** questions. Marks are shown in square brackets.

1. The following dataset shows the temperatures recorded at two monitoring stations on sixteen randomly selected summer afternoons. The figures are given in degrees Celsius.

Mongkok	32.0	34.2	34.8	33.1	32.4	31.6	32.2	33.0
Yuen Long	27.6	29.5	28.2	32.1	33.0	26.9	27.2	26.2

Suppose that the two sets of temperatures, recorded at Mongkok and Yuen Long, are independent random samples from the normal distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ respectively, where $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ are unknown parameters.

(a) Test at the 2% significance level the null hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ against the alternative hypothesis $H_0 : \sigma_x^2 \neq \sigma_y^2$.
[Note: $F_{0.01,7,7} = 6.99$ and $F_{0.99,7,7} = 0.143$, where $P(F_{n_1,n_2} \geq F_{\alpha,n_1,n_2}) = \alpha$ and F_{n_1,n_2} follows an F distribution with degree of freedom (n_1, n_2).]

[5 marks]

(b) Assuming that $\sigma_x^2 = \sigma_y^2 = \sigma^2$ but that σ^2 is unknown, construct a 95% two-sided equal tailed confidence interval for the difference $\mu_x - \mu_y$.

[5 marks]

(c) Using (b), or otherwise, determine whether you would reject the null hypothesis $H_0 : \mu_x = \mu_y$ against the alternative $H_1 : \mu_x \neq \mu_y$ at the 5% significance level. **[5 marks]**

[Total: 15 marks]

2. Let $X_1, X_2, \ldots, X_n$ be independent random samples from the uniform distribution $U(0, \theta)$. Let $M_n = \max(X_1, X_2, \ldots, X_n)$. Denote the test with rejection region $R_c = \{M_n \geq c\}$ as δ_c, where c is a constant.

(a) Find the power function for the test δ_c and show that it is a monotone increasing function of θ.

[6 marks]

(b) To test $H_0 : \theta \leq \frac{1}{2}$ against $H_1 : \theta > \frac{1}{2}$, what choice of c would make the test δ_c have size exactly equal to 0.05?

[5 marks]

(c) How large should n be so that the test specified in (b) has power 0.98 for $\theta = \frac{3}{4}$? **[5 marks]**

[Total: 16 marks]

3. A journal reported that, in a bag of m&m's chocolate peanut candies, there are 25% brown, 35% yellow, 20% blue, 15% red and 5% green candies. Suppose you purchase a bag of m&m's chocolate peanut candies at a store and find 17 brown, 20 yellow, 10 blue, 10 red, and 3 green candies, for a total of 60 candies. At the 0.05 level of significance, does the bag purchased agree with the distribution suggested by the journal?

[Note: $\chi^2_{0.05,3} = 7.81$, $\chi^2_{0.05,4} = 9.49$ and $\chi^2_{0.05,5} = 11.07$, where $P(\chi^2_n \geq \chi^2_{\alpha,n}) = \alpha$ and χ^2_n follows a χ^2 distribution with degree of freedom n.]

[Total: 6 marks]

4. Let $X_1, \ldots, X_n$ be a random sample from the below density

$$f(x) = \begin{cases} (\theta+1)x^{\theta} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > -1$ is an unknown parameter.

(a) Estimate θ by the method of moments.

[5 marks]

(b) Show that the method of moments estimator is consistent.

[5 marks]

(c) Is the method of moments estimator found in (a) the uniformly minimum variance unbiased estimator (UMVUE)? Give your reasons.

[5 marks]

(d) Find the maximum likelihood estimator of θ.

[5 marks]

(e) Is the maximum likelihood estimator an unbiased estimator for θ? If it is not, suggest an unbiased modification to it.

[Hint: Find the distribution of $\sum_{i=1}^{n}(-\log X_i)$. Also if $Z_1, \ldots, Z_m$ is a random sample from an exponential density $g(z) = \lambda e^{-\lambda z}, z > 0, \lambda > 0, T = \sum Z_i$ has a gamma distribution $Ga(\alpha, \beta)$ with $\alpha = m, \beta = 1/\lambda$, and probability density function

$$h(t) = \frac{t^{\alpha-1}e^{-t/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad t > 0.$$

And $E(T) = \alpha\beta$, $var(T) = \alpha\beta^2$.]

[6 marks]

[Total: 26 marks]

5. Let $X_1, X_2, \ldots, X_n$ be independent identical Poisson random variables with mean $\lambda > 0$, and probability function

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad \text{for } x = 0, 1, 2, 3, \ldots.$$

(a) Show that the likelihood function based on $X_1, X_2, \ldots, X_n$ is

$$L(\lambda) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}.$$

[5 marks]

(b) Let $0 < \lambda_0 \leq 1$ and $\lambda_1 > 1$ be chosen arbitrarily. Show that $\frac{L(\lambda_0)}{L(\lambda_1)}$ is a decreasing function of $\sum_{i=1}^n X_i$.

[5 marks]

(c) Suppose that we wish to test the hypotheses

$$H_0 : \lambda = \lambda_0 \text{ against } H_1 : \lambda = \lambda_1,$$

with $0 < \lambda_0 \leq 1$ and $\lambda_1 > 1$. Deduce the rejection region of the uniformly most powerful test.

[You are NOT required to derive the critical value explicitly. Hint: You may need to use the result in (b).]

[6 marks]

(d) Assume $\lambda_0 = 1$ and $\lambda_1 > 1$. Suppose $n = 4$ and the X_i's are observed as follows: $X_1 = 2, X_2 = 0, X_3 = 0, X_4 = 0$.

(i) Show that the test in (c) yields a p-value

$$p = P(Y \geq 2),$$

where Y is a Poisson random variable with mean 4.

[Hint: If X and Y are independent Poisson random variables with means λ_X and λ_Y respectively, then $X + Y$ is a Poisson random variable with mean $\lambda_X + \lambda_Y$.]

[5 marks]

(ii) Compute p.

[5 marks]

(iii) Determine whether you would reject H_0 at the 5% significance level.

[5 marks]

(e) Suppose that we wish to test

$$H_0 : \lambda = \lambda^* \text{ against } H_1 : \lambda > \lambda^*,$$

for a given constant $\lambda^* > 0$. Find the generalized likelihood ratio test for testing H_0 against H_1.

[You are not required to derive the critical value explicitly.] [6 marks]

[Total: 37 marks]

********** **END OF PAPER** **********

A LIST OF STATISTICAL FORMULAE

1. $M_X(t) = \mathrm{E}(e^{tX})$. $\quad \mathrm{E}(X^r) = \left(\dfrac{\mathrm{d}^r}{\mathrm{d}t^r} M_X(t) \right) \Bigg|_{t=0}$.

2. $\overline{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$. $\quad S^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$.

3. CLT: $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathrm{N}(0,1)$ for large n.

4. Normal population $\Longrightarrow \dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$, $\dfrac{\overline{X} - \mu}{S/\sqrt{n-1}} \sim t_{n-1}$.

5. $\chi^2_\nu \sim \sum_{i=1}^{\nu} Z_i^2$. $\quad t_\nu \sim \dfrac{Z}{\sqrt{\chi^2_\nu/\nu}}$. $\quad F_{\nu_1,\nu_2} \sim \dfrac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$.

6. $\mathrm{Bias}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$. $\quad \mathrm{E}\left[(\hat{\theta} - \theta)^2 \right] = \mathrm{Var}(\hat{\theta}) + \left[\mathrm{Bias}(\hat{\theta}) \right]^2$.

7. $I(\theta) = \mathrm{E}\left[\left(\dfrac{\partial \log f(X;\theta)}{\partial \theta} \right)^2 \right] = \mathrm{E}\left[-\dfrac{\partial^2 \log f(X;\theta)}{\partial \theta^2} \right]$. $\quad \mathrm{Var}(\hat{\theta}) \geq \dfrac{1}{nI(\theta)}$.

8. $\mathbf{f}(x_1, x_2, \ldots, x_n; \theta) = g(u(x_1, x_2, \ldots, x_n), \theta) \cdot h(x_1, x_2, \ldots, x_n)$.

9. $\dfrac{\overline{X} - \mu_0}{\sigma_0/\sqrt{n}}$. $\quad \dfrac{\overline{X} - \mu_0}{S/\sqrt{n-1}}$.

10. $\dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$. $\quad \dfrac{\overline{X} - \overline{Y} - \delta}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $S_p^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$.

11. $\overline{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$. $\quad \overline{x} \pm t_{\alpha/2,n-1} \dfrac{s}{\sqrt{n-1}}$. $\quad \left(\dfrac{ns^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{ns^2}{\chi^2_{1-\alpha/2,n-1}} \right)$.

12. $\overline{x} - \overline{y} \pm z_{\alpha/2} \sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}$. $\quad \overline{x} - \overline{y} \pm t_{\alpha/2,n_x+n_y-2} \sqrt{\dfrac{n_x s_x^2 + n_y s_y^2}{n_x + n_y - 2} \left(\dfrac{1}{n_x} + \dfrac{1}{n_y} \right)}$.

$$\left(\frac{n_x(n_y-1)s_x^2}{n_y(n_x-1)s_y^2} \cdot \frac{1}{F_{\alpha/2,n_x-1,n_y-1}}, \frac{n_x(n_y-1)s_x^2}{n_y(n_x-1)s_y^2} \cdot F_{\alpha/2,n_y-1,n_x-1} \right).$$

13. $F_{1-\alpha,m,n} = \dfrac{1}{F_{\alpha,n,m}}$.

14. $-2 \ln \Lambda \approx \sum_i \dfrac{(O_i - E_i)^2}{E_i} = \sum_i \dfrac{O_i^2}{E_i} - n$. $\quad -2 \ln \Lambda \approx n \left(\sum_i \sum_j \dfrac{O_{i,j}^2}{n_{i.} n_{.j}} - 1 \right)$.

Cumulative distribution function $\Phi(x)$ for $N(0,1)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Upper percentile for the student's t distribution t_ν

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090