MATH372301

This question paper consists of 5 printed pages, each of which is identified by the reference **MATH3723**.

Only approved basic scientific calculators may be used. Statistical tables are provided on pp. 4-5.

UNIVERSITY OF LEEDS

Examination for the Module MATH3723
(May/June 2004)

Statistical Theory

Time allowed: **3 hours**

Attempt not more than **four** questions.
All questions carry equal marks.

1. (a) Let $X = e^Z$ where $Z \sim \mathcal{N}(\mu, \sigma^2)$. Using the rule for computing the density, $f_X(x) = dP(X \le x)/dx$, show that the density $f_X(x)$ can be represented in the form,

$$f_X(x) = x^{-1}(2\pi\sigma^2)^{-1/2}e^{-(\ln x-\mu)^2/2\sigma^2}, \quad x > 0.$$

[Hint: events $X \le x$ and $Z \le \ln x$ are equal.]
Using this density or otherwise, show that statistic $T(X_1, \ldots, X_n) = \sum_{i=1}^n \ln X_i$ based on observations $X_1, \ldots, X_n$, is sufficient for μ if σ^2 is given.
Suggest an unbiased estimator for μ based on observations $X_1, \ldots, X_n$.

(b) Define Fisher's information and formulate the Cramér-Rao lower bound for unbiased estimators. Compute the Cramér-Rao lower bound for the parameter μ for the parametric family introduced in (a); here σ^2 is given.
Is this bound attained by any unbiased estimator? Justify your answer.

(c) Define what is meant by a best linear estimator and a best linear unbiased estimator in the mean square error sense.
State the definition of normal random variable $\mathcal{N}(\mu, \sigma^2)$ via its density.
Two independent random variables $X_1, X_2 \sim \mathcal{N}(\mu, \sigma^2)$ with known σ^2 are given.
Define all linear estimators of μ based on X_1 and X_2. Show that $(X_1 + X_2)/2$ is the best linear unbiased estimator.
Assuming $-1 \le \mu \le 1$, show that there is a biased linear estimator of μ which is strictly better than the best linear unbiased for any μ from this range, in the sense of the mean square error.
[Hint: consider the mean square error for the statistic $\alpha(X_1 + X_2)/2$ for α close to and slightly less than one.]

2. (a) Define in detail the notions of the Method of Moments Estimator and the Maximum Likelihood Estimator for the 1-parametric statistical model (either for continuous or discrete case, or both).

(b) Find the Estimator of θ for the sample from the uniform distribution $U[-\theta, 0]$ based on Method of Moments. Is it biased? Compute its Mean Square Error using its density or otherwise. Comment on the rate of convergence.

(c) For the sample $X_1, \ldots, X_n$ from the distribution $U[-\theta, 0]$, compute the density of $X_{(1)} = \min(X_1, \ldots, X_n)$.
Compute the Maximal Likelihood Estimator for $\theta \in [0, 1]$. Is this statistic sufficient? Is it biased? Compute the Mean Square Error of this estimator. Is it less than or greater than that of the estimator from part (b) of this question?

3. (a) Define the notion of sufficient statistics.
State the Factorization Criterion for sufficient statistics and the Rao-Blackwell Theorem on improvement of unbiased estimators.

(b) Compute the minimal sufficient statistics for the distribution

$$X_i = \begin{cases} +1, & \text{with probability } \; p, \\ -1, & \text{with probability } \; (1-p), \end{cases}$$

and sample $X_1, \ldots, X_n$, $p \in (0, 1)$. [Neither an exact proof of minimality, nor the definition of minimality is required.]

(c) In the model from (b), suggest an unbiased estimator for parameter $p \in (0, 1)$ based on one observation X_1.
Show its unbiasedness.
Use the Rao-Blackwell Theorem to improve it, given minimal sufficient statistic based on two observables X_1, X_2.
[Hint: show that $E(X_1|X_1 + X_2) = (X_1 + X_2)/2$.]
Compare the Mean Square Error of the improved estimator with the Cramér-Rao lower bound and comment on it.

4. (a) Define in detail the notion of a Bayesian Estimator with the Mean Square Error criterion. Either for a general model or for any example, explain why $E(\theta|X)$ is a Bayesian Estimator, where $X = (X_1, \ldots, X_n)$ is a sample.

(b) Consider a sample $X = (X_1, \ldots, X_n)$ from the Bernoulli distribution, $\theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $0 \le \theta \le 1$. Let the prior for θ be $q(\theta = 1) = q(\theta = 0.1) = 1/2$. Compute the Bayesian Estimator and evaluate its Mean Square Error.

(c) State the definition of the Bayesian decision rule in the hypothesis testing problem. Let $H_0 : \theta = 0.1$, $H_1 : \theta = 1$. Argue why the decision rule "accept H_1 iff $\bar{X} = 1$" is plausible for this hypothesis testing problem.
Is this decision rule Bayesian?
Compute both errors in the classical sense, assuming H_0 is the null hypothesis.
Find a sample size n sufficient for the first type error $e_1 \le 0.05$.
Find a sample size n sufficient for the Bayesian error e_B to be at most 0.05.

5. (a) State the definition of critical region, and give definitions of first and second type errors.
Formulate the Neyman-Pearson Lemma.

(b) Let

$$f_{\theta_0}(x) = \frac{1}{2}e^{-|x|}, \qquad f_{\theta_1}(x) = \frac{1}{4}e^{-\sqrt{|x|}}.$$

Show that both functions are densities.
Show that given the sample $X_1, \ldots, X_n$, the best test for testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ can be represented in the form,

$$\sum_{i=1}^{n}(\sqrt{|X_i|} - 1/2)^2 > c.$$

In the case $n = 1$, show that the test $|X| > 1.96$ is the best with significance level $\alpha = 0.05$.
[Hint: $|a| > 1.4$ is equivalent to $||a| - 1/2| > 0.9$.]

(c) Describe the notion of confidence interval or set for an unknown parameter.
State the theorem on how to construct most accurate confidence intervals (sets) based on the Most Powerful Test for testing $H_0 : \theta = \theta_0$.
Given a sample $X_1, \ldots, X_n$ and $\sum_{i=1}^{n} X_i = S$, from a distribution family $F(x - \theta)$ with unknown location $\theta = E(X_1)$ and variance $\sigma^2 = 1$, construct a two-sided confidence interval with confidence level $1 - \alpha$ for $\theta \in R$.
Let $H_0 : \mu = 0$, and $H_1 : \mu = 1$. Show how to construct a critical region with the first type error at most α, using this confidence interval.
Let $\sigma = 1$, first type error $e_1 \approx 0.05$, $n = 26$. What is approximately the second type error?
Write down the theoretical statement (theorem) which you have used while performing this calculation.

END OF QUESTIONS

Normal distribution (areas)

Area ($\alpha = P(Z > z)$) in the tail of the standardized Normal curve, $Z \sim N(0,1)$, for different values of z. Example: Area beyond $z = 1.96$ (or below $z = -1.96$) is $\alpha = 0.02500$. For Normal curve with $\mu = 10$ and $\sigma = 2$, area beyond $x = 12$, say, is the same as the area beyond $z = \dfrac{x-\mu}{\sigma} = \dfrac{12-10}{2} = 1$, i.e. $\alpha = 0.15866$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
4.0	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

α	0.4	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.001
z_α	0.2533	0.6745	0.8416	1.0361	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

CONTINUED

Percentage Points of the t-Distribution

The table gives the value a, such that $P(T < a) = p$ where T is a random variable from a t-distribution with ν degrees of freedom. Example: $P(T < 2.093) = 0.975$, where T has 19 d.o.f.

p	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
ν 1	1.000	3.078	6.314	12.706	31.821	63.656	636.578
2	0.816	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.689
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.660
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	1.299	1.676	2.009	2.403	2.678	3.496
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
70	0.678	1.294	1.667	1.994	2.381	2.648	3.435
80	0.678	1.292	1.664	1.990	2.374	2.639	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	3.390
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

END