MATH 244 (L2) Applied Statistics

Final Examination December 16, 2000

Time allowed : 3 hours
Answer all five questions.

1. (*25 marks*) Consider the following poker game. You need to pay \$10 for **each** trial of this game. Based on the outcome of 5 cards randomly drawn from a deck of 52 poker cards, you can obtain a certain amount of money (denote this as X) according to the following table :

Outcome	**Probability**	**Payoff (X)**
Royal flush	3.848×10^{-7}	\$10,000
Straight flush	1.501×10^{-5}	\$5,000
Four of a kind	2.401×10^{-4}	\$1,000
Full house	1.441×10^{-3}	\$400
Flush	1.965×10^{-3}	\$100
Straight	3.925×10^{-3}	\$50
Three of a kind	0.02113	\$30
Two pairs	0.04754	\$20
One pair	0.4226	\$10
Otherwise (Lose)	p	\$0

For example, if the 5 cards form a full house, then you can get back \$400, i.e. win \$390.

(a) Find $E(X)$ and $Var(X)$. (8 marks)

(b) Let Y be the net profit you can make in one trial of the game. Express Y in term of X and hence find $E(Y)$ and $Var(Y)$. (4 marks)

(c) Is this a fair game ? Why ? (2 marks)

(d) Find p , the probability that you will lose money in one trial of the game. (2 marks)

(e) Suppose you play the game for 100 times independently. Let W be the number of games that you will lose money. What is the distribution of W ? (2 marks)

(f) Find the probability that you will lose in more than 50 games if you played the game for 100 times. (3 marks)

(g) Find the probability that you will lose money after you played the game for 100 times, i.e. the overall gain after 100 games is negative. (4 marks)

2. (*20 marks*) Let X be a number randomly drawn from { 1, 2, 3, 4, …, K } with equal probabilities, i.e.

$$\Pr(X=1)=\Pr(X=2)=\cdots\Pr(X=K)=\frac{1}{K}$$

where K is unknown.

(a) Find $E(X)$ and $Var(X)$ in terms of K. (6 marks)

(b) Let $\{X_1, X_2, ..., X_n\}$ be n numbers drawn from this distribution with replacement, i.e. $\{X_1, X_2, ..., X_n\}$ is a random sample from this distribution. Find the method of moment estimator of K. (3 marks)

Suppose we observed the following sample :

2	4	7	8	8	10	13	15	18	19
21	21	26	27	28	32	34	34	37	39
41	41	43	47	47					

(c) Use the estimator in (b) to estimate K. (2 marks)

(d) Construct a stem-and-leaf plot for this sample. (4 marks)

(e) Find the lower quartile, median and upper quartile and hence sketch the box-plot for this sample. (5 marks)

(HINT : $1+2+3+\cdots+k=\dfrac{k(k+1)}{2}$, $1^2+2^2+3^2+\cdots+k^2=\dfrac{k(k+1)(2k+1)}{6}$)

3. (*10 marks*) Cars pass a point on the highway at a Poisson rate of one per minute.

(a) What is the probability that exactly 6 cars pass the point within 5 minutes ? (3 marks)

(b) What is the probability that at least 70 cars pass the point within one **hour** ? (3 marks)

(c) What is the probability that we need to wait longer than one **hour** to observe 70 cars passing the point ? (4 marks)

(HINT : You may need to use normal approximation.)

4. (*25 marks*) The power of supercomputers derives from the idea of parallel processing. Engineers at Cray Research are interested in determining whether one of two parallel processing designs produces faster average computing time, or whether the two designs are equally fast. The following are the results, in seconds, of independent random computation times using the two designs.

Design 1 :	2.1	2.2	1.9	2.0	1.8	2.4	2.0	1.7	2.3	2.8
	1.9	3.0	2.5	1.8	2.2	2.6				
Design 2 :	2.6	2.5	2.0	2.1	2.6	3.0	2.3	2.0	2.4	2.8
	3.1	2.7	2.6							

Assume that the two populations of computing time are normally distributed and that the two population variances are equal.

(a) Suppose you are asked to do the test based on these data. Write down the hypotheses and explain why you construct your hypotheses in this way. (4 marks)

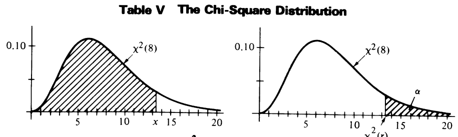
(b) Test your hypotheses in (a) at $\alpha = 0.05$. What can be concluded ? (6 marks)

(c) Test the equal variance assumption at $\alpha = 0.1$. (5 marks)

(d) For **each** design, construct a 95% confidence interval for the average computing time. (6 marks)

(e) Construct a 90% confidence interval for the difference in the average computing times of the two designs. (4 marks)

Table V The Chi-Square Distribution



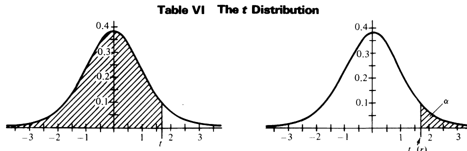
$$P(X \le x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2}\, dw$$

	$P(X \le x)$ 0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21
11	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72
12	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22
13	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69
14	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14
15	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58
16	5.812	6.908	7.962	9.312	23.54	26.30	28.84	32.00
17	6.408	7.564	8.672	10.08	24.77	27.59	30.19	33.41
18	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.80
19	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19
20	8.260	9.591	10.85	12.44	28.41	31.41	34.17	37.57
21	8.897	10.28	11.59	13.24	29.62	32.67	35.48	38.93
22	9.542	10.98	12.34	14.04	30.81	33.92	36.78	40.29
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31
26	12.20	13.84	15.38	17.29	35.56	38.88	41.92	45.64
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
40	22.16	24.43	26.51	29.05	51.80	55.76	59.34	63.69
50	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
60	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
70	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4
80	53.34	57.15	60.39	64.28	96.58	101.9	106.6	112.3

This table is abridged and adapted from Table III in *Biometrika Tables for Statisticians*, edited by E. S. Pearson and H. O. Hartley. It is published here with the kind permission of the *Biometrika* Trustees.

614

Table VI The *t* Distribution



$$P(T \le t) = \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1 + w^2/r)^{(r+1)/2}}\, dw$$
$$[P(T \le -t) = 1 - P(T \le t)]$$

	$P(T \le t)$ 0.60	0.75	0.90	0.95	0.975	0.99	0.995
r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0 765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.997
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576

This table is taken from Table III of Fisher and Yates: *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver and Boyd, Edinburgh), by permission of the authors and publishers.

5. (*20 marks*) Last year a large organization is being investigated to determine if their recruitment is sex biased. Two random samples of male and female applicants were drawn from the applicants of secretarial and sales positions. Following table shows the classification of the applicants for secretarial and sales positions according to sex and result of interview.

Secretarial and sales positions

	Offered	Denied	Total
Male	180	120	300
Female	125	175	300

(a) Test $H_0 : p_M = p_F$ against $H_1 : p_M \neq p_F$ where p_M is the proportion of male applicant who was offered the position and p_F is the proportion of female applicant who was offered the position. Use $\alpha = 0.05$. (5 marks)

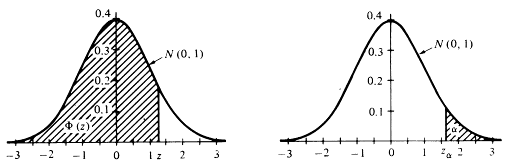
(b) What is the *p-value* of the test in (a) ? (4 marks)

(c) Construct a 95% confidence interval for $p_M - p_F$. What can be concluded from this confidence interval ? (5 marks)

(d) Suppose this year you are asked to do the survey again. If you want both the 95% confidence intervals for p_M and p_F to have lengths less than 0.1, what are the minimum numbers of male and female applicants you should draw as you samples? (6 marks)

<END>

Table IV The Normal Distribution



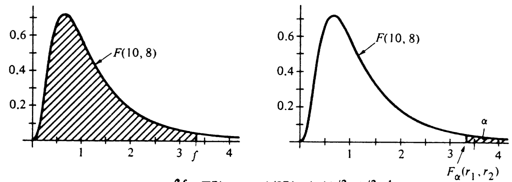
$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw$$

$$[\Phi(-z) = 1 - \Phi(z)]$$

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	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

α	0.400	0.300	0.200	0.100	0.050	0.025	0.010	0.005	0.001
z_α	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090
$z_{\alpha/2}$	0.842	1.036	1.282	1.645	1.960	2.240	2.576	2.807	3.291

Table VII The *F* Distribution



$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1 + r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2 - 1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1 + r_1 w/r_2)^{(r_1+r_2)/2}}\, dw$$

α	$P(P \le f)$	Den. d.f. r_2	Numerator Degrees of Freedom, r_1: 12	15	20	24	30	40	60	120	∞
0.05	0.95	12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
0.025	0.975		3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
0.01	0.99		4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
0.05	0.95	15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
0.025	0.975		2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
0.01	0.99		3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
0.05	0.95	20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
0.025	0.975		2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
0.01	0.99		3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
0.05	0.95	24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
0.025	0.975		2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
0.01	0.99		3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
0.05	0.95	30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
0.025	0.975		2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
0.01	0.99		2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
0.05	0.95	40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
0.025	0.975		2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
0.01	0.99		2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
0.05	0.95	60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
0.025	0.975		2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
0.01	0.99		2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
0.05	0.95	120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
0.025	0.975		2.05	1.95	1.82	1.76	1.69	1.61	1.53	1.43	1.31
0.01	0.99		2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
0.05	0.95	∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00
0.025	0.975		1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00
0.01	0.99		2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00