G

Probability Final Exam

2:20-4:50 pm Thursday, 22 June 2006

Total: 6 pages, 100 points.

(1) A calculator is NOT allowed. You may not consult the textbook or neighbor.
(2)Please turn off your cell phones before the exam.
(3)Please show your work for partial credit, and underline your answers. Points are awarded for solutions, not answers, so correct answers without justification will not receive full credit.
(4)Please consult the tables in the last two pages to find accurate answers if necessary. Answers without simplifying will not receive full credit.

PROBLEM 1: ***Rolling Die*** **(5%)**

Let X be the smallest number of obtained in rolling a balanced die n times. Calculate the probability distribution function and the probability mass function of X.

PROBLEM 2: ***Sum of Random Variables*** **(5%)**

Let $\{X_1, X_2, X_3, \dots\}$ be a sequence of independent and identically distributed exponential random variables with parameter λ. Let N be a geometric random variable with parameter p independent of $\{X_1, X_2, X_3, \dots\}$. Find the distribution function of $\sum_{i=1}^{N} X_i$.

PROBLEM 3: ***Joint PDF I*** **(9%)**

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed (iid) Gaussian random variables with zero mean and unit variance. Also, let $Y_1, Y_2, \dots, Y_n$ be iid binary random variables with the common probability mass function $p_Y(y) = \frac{1}{2}$ if $y = +1$ or $y = -1$, and $p_Y(y) = 0$ otherwise. In addition, X_k's, Y_k's, $k = 1, \dots, n$, are mutually independent. Now, define a new continuous-valued random variable $Z = \sum_{k=1}^{n} X_k Y_K$. Derive the probability density function of Z.

PROBLEM 4: ***Joint PDF II*** **(5%)**

Let A and φ be two independent continuous-valued random variables where A has the probability density function $f_A(a) = a\exp\{-\frac{a^2}{2}\}$ for $a \ge 0$ and $f_A(a) = 0$ for $a < 0$, and φ is uniformly distributed in $[0, 2\pi)$. Now, define two new random variables

$$X = A\cos(\varphi), \quad Y = A\sin(\varphi).$$

Find the marginal and joint probability density functions of X and Y, i.e., $f_X(x), f_Y(y), f_{X,Y}(x,y)$. What do you find?

GO ON TO THE NEXT PAGE.

PROBLEM 5: ***Limit Theorems* (18%)**

(1) Describe and derive Markov's inequality. (4 %)

(2) Describe and derive Chebyshev's inequality from (1). (4 %)

(3) Describe and derive Weak Law of Large Numbers from (2). (4 %)

(4) Suppose that it is known that the number of items produced in a factory during a week is a random variable with *mean 100*. (6 %)

A. What can be said about the probability that this week's production will exceed 150?

B. If the *variance* of a week's production is known to equal *50*, then what can be said about the probability that this week's production will be between 80 and 120?

PROBLEM 6: ***Central Limit Theorem Applications* (7%)**

The time between events in a certain random experiment are *iid* exponential random variables with m second. Find the probability that the 10000[th] event occurs in the time interval $(10000 \pm 20)m$ second.

PROBLEM 7: ***Unbiased Estimator* (5%)**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Prove that the expected value of the sample variance is equal to the population variance σ^2, i.e., the sample variance is an unbiased estimator of σ^2.

PROBLEM 8: ***Sample Estimate* (10%)**

Consider a finite population of size N. Let p be the proportion of the population that has a certain characteristic of interest. For i=1,…,n_1+n_2, let

$$X_i = \begin{cases} 1 & \text{if the ith member of the sample has the characteristic;} \\ 0 & \text{otherwise.} \end{cases}$$

be a simple random sample from the population without replacement and $\overline{Y}$ be the sample mean. Let a simple random sub-sample without replacement of size n_1 be drawn from the n_1+n_2 samples with sample mean $\overline{Y}_1$.

(a) Show that $E(\overline{Y}_1) = E(\overline{Y}) = p$. (5%)

(b) Let s^2 be the population variance. Derive Var($\overline{Y}$) =? (in terms of s^2 and n_1 and n_2)(5%)

GO ON TO THE NEXT PAGE.

PROBLEM 9: ***Point Estimator*** **(15%)**

Karl takes a sample $X_1, X_2, \ldots, X_{2n}$ from a normal population with unknown mean and unknown variance. Karl wants to estimate the mean and the variance of this population, but he does not have much time. So he gives Bill $X_1, X_2, \ldots, X_n$; he also gives John $X_{n+1}, X_{n+2}, \ldots, X_{2n}$. Bill computes the sample mean and sample variance of $X_1, X_2, \ldots, X_n$. He calls them $\bar{x}_{Bill}$ and s^2_{Bill}, respectively. Similarly, John computes the sample mean and sample variance of $X_{n+1}, X_{n+2}, \ldots, X_{2n}$. He calls them $\bar{x}_{John}$ and s^2_{John}, respectively.

Before you proceed: in this problem, when comparing two estimators, you should check first if they are identical. If they are not, then the one with the smaller variance is the *better* estimator. Hint: the variance of a chi-square random variable is $Var(\chi^2_n) = \frac{n/2}{(1/2)^2}$.

(a) Karl forms an estimate of the population mean as $\frac{1}{2}(\bar{x}_{Bill} + \bar{x}_{John})$. The question is that, if Karl were to compute the sample mean from $X_1, X_2, \ldots, X_{2n}$ directly himself, would he have obtained a better estimate than $\frac{1}{2}(\bar{x}_{Bill} + \bar{x}_{John})$? (5%)

(b) Karl forms an estimate of the population variance as $\frac{1}{2}(s^2_{Bill} + s^2_{John})$. The question is that, if Karl were to compute the sample variance from $X_1, X_2, \ldots, X_{2n}$ directly himself, would he have obtained a better estimate than $\frac{1}{2}(s^2_{Bill} + s^2_{John})$? (10%)

PROBLEM 10 ***Confidence Interval and Hypothesis Testing*** **(15%)**

From past observations, a particular drug delivers an average pain killing duration, μ_0, of 6 hours. However, depending on from which factories a drug is made, the effectiveness of the drug could vary. A researcher is assigned to affirm whether the drugs manufactured at a newly added factory can deliver this expected pain killing duration.

The researcher tracked 25 patients who took the drugs manufactured at that factory. The reaction of the patients can be assumed to be Gaussian. The reported mean pain killing duration is 5 hours and 45 minutes. The sample standard deviation of the pain killing duration is 20 minutes.

(a) There are two hypotheses of interest: $\mu \geq 6$ and $\mu < 6$. To this researcher, which one should be considered the null hypothesis? (Hint: the null hypothesis is the one that the researcher must accept if he or she did not test any patient.) (5%)

(b) What is the lower one-sided 90% confidence interval for the duration of pain killing duration? (5%)

(c) Can the null hypothesis be rejected at 0.1 level? (5%)

GO ON TO THE NEXT PAGE.

PROBLEM 11: ***Maximum Likelihood Estimator*** **(5%)**

A random number generator is known to generate random numbers uniformly in [L, L + 5]. Belinda used this generator to generate the following 8 numbers:

2.5, 3.6, 5.4, ~~4, 8,~~ 4.8 1.7, 2.3, 3.9, 5.2

(a) Find the likelihood of this sample given L. What are the maximal likelihood estimates for L? (5%)

STOP

PLEASE RECHECK YOUR ANSWERS! THANK YOU!

Reference Table I

TABLE AI *Standard Normal Distribution Function:* $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Reference Table II

TABLE A3 *Values of* $t_{\alpha,n}$

n	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.474	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

Other t Probabilities:
$P\{T_8 < 2.541\} = .9825$ $P\{T_8 < 2.7\} = .9864$ $P\{T_{11} < .7635\} = .77$ $P\{T_{11} < .934\} = .81$ $P\{T_{11} < 1.66\} = .94$ $P\{T_{12} < 2.8\} = .984$.

$E(\bar{X}^2) = \frac{1}{n^2} E[(\sum x_i)^2]$