STATISTICS DEPARTMENT
M.S. EXAMINATION

PART I
CLOSED BOOK

Friday, May 26, 2000

9:00 a.m. - 1:00 p.m.

School of Science Conference Room, SC N137

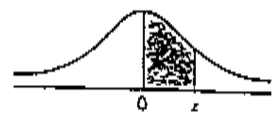
Instructions: **Complete *all four* problems. Each problem counts 25 points. Unless otherwise noted, points are allocated approximately equally to lettered parts of a problem. Spend your time accordingly.**

Begin each problem on a new page. Write the problem number and the page number in the specified locations at the top of each page. Also write your chosen ID code number on every page. Please write only within the black borderlines, leaving at least 1" margins on both sides, top and bottom of each page. Write on one side of the page only.

At the end of this part of the exam you will turn in your answers sheets, but you will keep the question sheets and your scratch paper.

STANDARD NORMAL DISTRIBUTION TABLE

The entries in this table give the areas under the standard normal curve from 0 to z.



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	0120	0160	0199	.0239	.0279	0319	.0359
0.1	.0398	.0438	0478	0517	.0557	.0596	0636	.0675	.0714	.0753
0.2	0793	0832	0871	.0910	.0948	0987	.1026	.1064	1103	.1141
0.3	1179	.1217	.1255	1293	1331	1368	.1406	.1443	.1480	1517
0.4	1554	1591	.1628	1664	1700	1736	.1772	1808	.1844	1879
0.5	1915	1950	1985	2019	.2054	2088	2123	2157	2190	.2224
0.6	.2257	.2291	.2324	2357	2389	2422	.2454	.2486	2517	.2549
0.7	.2580	.2611	2642	2673	2704	.2734	2764	2794	2823	2852
0.8	.2881	.2910	2939	2967	2995	.3023	.3051	.3078	.3106	.3133
0.9	3159	3186	3212	.3238	3264	3289	.3315	.3340	.3365	.3389
1.0	3413	.3438	.3461	3485	3508	3531	3554	.3577	.3599	.3621
1.1	3643	3665	3686	3708	3729	.3749	.3770	3790	.3810	3830
1.2	3849	.3869	3888	3907	3925	.3944	3962	3980	.3997	4015
1.3	.4032	4049	4066	4082	4099	4115	4131	.4147	4162	.4177
1.4	.4192	.4207	4222	4236	4251	4265	4279	4292	4306	.4319
1.5	4332	4345	4357	.4370	4382	4394	4406	.4418	.4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	.4525	4535	.4545
1.7	.4554	4564	4573	4582	.4591	4599	4608	.4616	4625	.4633
1.8	4641	4649	.4656	4664	4671	4678	4686	4693	.4699	4706
1.9	4713	4719	4726	.4732	4738	4744	.4750	.4756	.4761	.4767
2.0	4772	.4778	.4783	.4788	4793	.4798	4803	4808	4812	4817
2.1	.4821	4826	4830	4834	4838	4842	.4846	4850	.4854	4857
2.2	.4861	.4864	.4868	4871	.4875	.4878	4881	4884	.4887	.4890
2.3	4893	.4896	.4898	4901	.4904	4906	.4909	.4911	.4913	4916
2.4	.4918	4920	.4922	.4925	4927	4929	4931	4932	.4934	4936
2.5	.4938	.4940	4941	4943	4945	.4946	4948	4949	.4951	4952
2.6	4953	.4955	.4956	4957	4959	4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	4967	4968	.4969	4970	.4971	.4972	4973	.4974
2.8	4974	.4975	.4976	.4977	.4977	4978	.4979	.4979	.4980	.4981
2.9	4981	4982	4982	.4983	4984	4984	4985	4985	.4986	.4986
3.0	4987	4987	.4987	.4988	4988	.4989	4989	.4989	4990	.4990

1. We wish to show that a certain proposition is favored by a majority (greater than 50%) of a certain population. Use the approximately normal one-sample binomial statistic $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ as the test statistic, where p = population proportion and $\hat{p}$ = sample proportion.

(a) Utilizing a sample of size $n = 100$, and operating at $\alpha = .01$, at least what value would Z have to be for us to prove that a majority of the population favored the proposition? At least what value would $\hat{p}$ have to be?
(b) At $p = .55$, for $n = 100$ and $\alpha = .01$, what would be the power of our test to conclude that a majority of the population favored the proposition?
(c) At $p = .55$, for $\alpha = .01$, how large a sample size would be needed to have a power of .99?

2. Suppose that a job requires the completion of tasks A and B. Let X (in hours) be the completion time of A and let Y (in hours) be the completion time of B. Suppose that X and Y have the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} cx^2y^{-2} & \text{if } 0 < x < 1 \text{ and } x < y < \frac{1}{x} \\ 0 & \text{elsewhere,} \end{cases}$$

where c is a positive constant.

(a) Find the probability density function, $f_X(x)$, of X and find the value of the constant c.

(b) Let E be the event that task A is completed in less than $\frac{1}{2}$ hour and let F be the event that task B is completed in less than $\frac{1}{2}$ hour.

(i) Find $P(E)$.

(ii) Find $P(F)$.

(iii) Find $P(E \cap F)$.

(iv) Are X and Y independent? Prove your answer.

(c) Find $P(XY < \frac{1}{2})$.

3. A particular exponential distribution describes the interarrival times of a Poisson process with a mean rate of 1 every 3 minutes; thus the mean of this exponential distribution is 3 minutes. *In the following parts you may simply state answers that you know (giving brief explanations as appropriate), **unless** you are explicitly asked for a derivation.*

   (a) Write the cumulative distribution function and density function of *this particular* exponential distribution, and use the density function to *derive* the moment generating function.

   (b) A small bank has one teller whose service times have this exponential distribution. A line of customers has formed. One customer is being served and Mary is the third person in line (behind two others waiting to be served). (i) State the distribution [name and parameter(s)], mean, and variance of the length of time until Mary *finishes* being served. (ii) Briefly indicate how this distribution can be derived from what is known about the distribution of service times. (iii) Briefly explain why you do not need to know how long the customer now with the teller has been there in order to answer (i).

   (c) A large bank has four tellers, each of which has exponentially distributed service times with mean 3 minutes. Currently all four of them are busy serving customers. John is first in line and will start being served as soon as the first of the four tellers is available. (i) State the distribution [name and parameter(s)], mean, and variance of the length of time until John *begins* to be served. (ii) *Derive* this distribution (cdf) from the distribution of service times.

   (d) Return to the situation in (c) where all four tellers are currently busy. By any correct method or argument find (i) the mean and (ii) the variance of the length of time until all four of the customers currently with tellers have finished being served. Show your method. [Hint: It *may* be easier for you to argue the value of the mean (and variance) based on ideas in previous parts than to start by finding the distribution of the waiting time. What is the expected wait for the first of these four customer to leave? The second? And so on.]

4. A salesperson interviews customers one at a time. Let X_i be the time (in hours) that the salesperson spends interviewing the ith customer for $i = 1, 2, \cdots$. Suppose that these times are independent and have the probability density function

$$f(x) = \begin{cases} x\exp(-x) & \text{if } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Find the probability that the salesperson spends less that 1 hour interviewing the first customer; that is, find $P(X_1 < 1)$.

(b) Let Y be the time that the salesperson spends interviewing the first two customers; that is, let $Y = X_1 + X_2$. Find the probability density function of Y.

(c) Assume that $0 < p < 1$ and that the salesperson interviews N customers, where N has the probability function

$$p_N(n) = P(N = n) = (1-p)^{n-1} p, \ n = 1, 2, \cdots.$$

Let S_N be the time that the salesperson spends interviewing the N customers; that is, let $S_N = \sum_{i=1}^{N} X_i$. Assume that the random variables $N, X_1, X_2, \cdots$ are independent.

(i) Find $E(S_N)$.

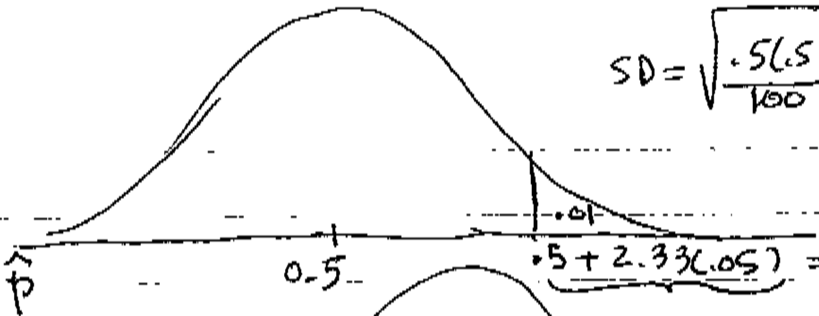
(ii) Find $Var(S_N)$

(iii) Suppose that $t < 1$. Find $E[\exp(tS_N)]$.

(a) $Z \geq 2.33$

or $\hat{p} \geq .6165$

$SD = \sqrt{\frac{.5(.5)}{100}} = \frac{.5}{10} = .05$

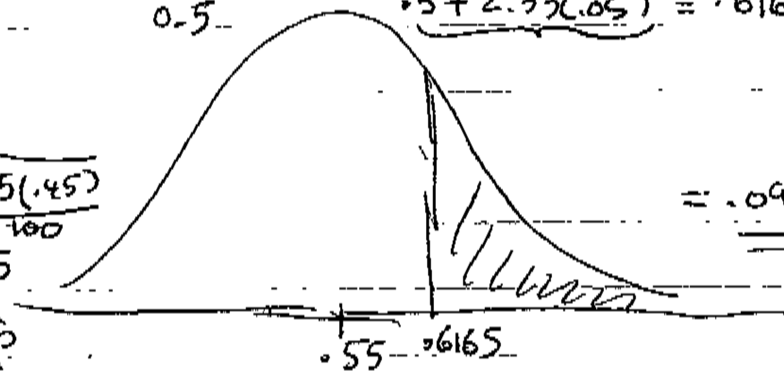


$.5 + 2.33(.05) = .6165$

(b)

$SD = \sqrt{\frac{.55(.45)}{100}}$

$= .04975$

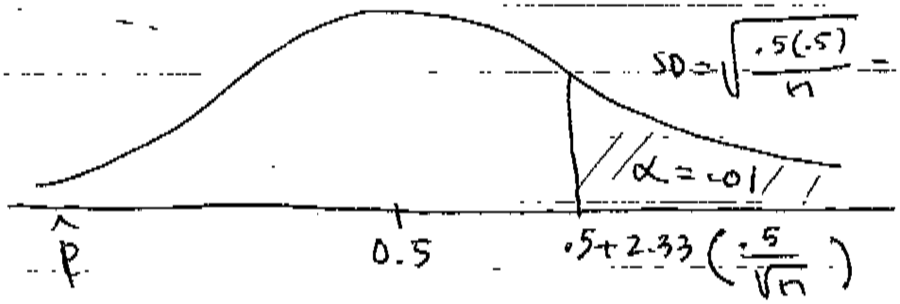


$= .0901$

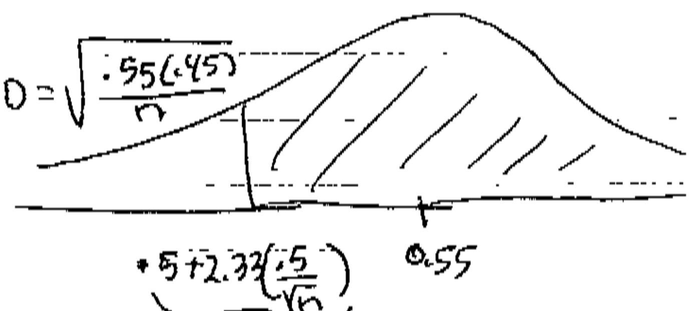
$Z \qquad 1.34$

(c)

$SD = \sqrt{\frac{.5(.5)}{n}} = \frac{.5}{\sqrt{n}}$



$SD = \sqrt{\frac{.55(.45)}{n}}$



$$Z \qquad \frac{-.05 + 2.33\left(\frac{.5}{\sqrt{n}}\right)}{\sqrt{(.55)(.45)}/\sqrt{n}} = \frac{-.05\sqrt{n} + 2.33(.5)}{\sqrt{.55(.45)}}$$

For power = 0.99

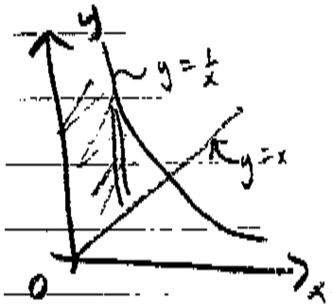
Set $$\frac{-.05\sqrt{n} + 2.33(.5)}{\sqrt{.55(.45)}} = -2.33$$

or $$n = \left(\frac{2.33\left(\sqrt{.55(.45)} + 0.5\right)}{.05}\right)^2 \simeq 2161$$

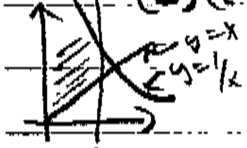
(#2)

(a) $f_X(x) = \int_x^{\frac{1}{x}} c x^2 y^{-2} dy = -cx^2 y^{-1} \Big|_{y=x}^{y=\frac{1}{x}} = cx^2\left(\frac{1}{x} - x\right)$

$= c(x - x^3)$ for $0 < x < 1$ and $f_X(x) = 0$ elsewhere



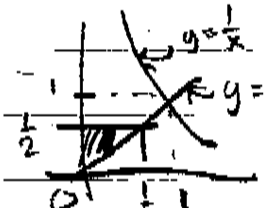
$c = 1 \Big/ \int_0^1 x - x^3 \, dx = 1 \Big/ \left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_0^1 = 1 \Big/ \left(\frac{1}{2} - \frac{1}{4}\right)$



(b)(i) $P[E] = P[X < \frac{1}{2}] = \int_0^{\frac{1}{2}} f_X(x)\,dx$

$= \int_0^{\frac{1}{2}} c\,(x - x^3)\,dx = c\left(\frac{(\frac{1}{2})^2}{2} - \frac{(\frac{1}{2})^4}{4}\right)$

(ii) $P[F] = \int_0^{\frac{1}{2}} \int_x^{\frac{1}{2}} c x^2 y^{-2} \, dy\, dx$



$= \int_0^{\frac{1}{2}} c\, x^2 \left(-y^{-1}\Big|_{y=x}^{y=\frac{1}{2}}\right) dx$

$= \int_0^{\frac{1}{2}} c\, x^2 \left(\frac{1}{x} - \frac{1}{\frac{1}{2}}\right) dx = \int_0^{\frac{1}{2}} c x^2 \left(\frac{1}{x} - 2\right) d$

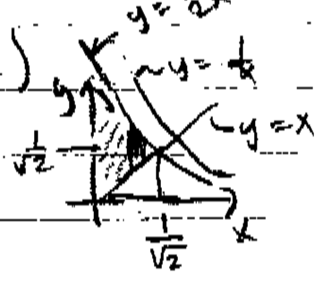
$= \int_0^{\frac{1}{2}} c(x - 2x^2)\,dx = c\left(\frac{x^2}{2} - \frac{2x^3}{3}\right)\Big|_0^{\frac{1}{2}}$

$= c\left(\frac{(\frac{1}{2})^2}{2} - 2\,\frac{(\frac{1}{2})^3}{3}\right)$

(iii) Since $F \subset E$, we have $E \cap F = F$; hence
$P[E \cap F] = c\left(\frac{(\frac{1}{2})^2}{2} - 2\,\frac{(\frac{1}{2})^3}{3}\right)$

(iv) $P[X < \frac{1}{2}, Y < \frac{1}{2}] = P[Y < \frac{1}{2}] > P[X < \frac{1}{2}]\,P[Y < \frac{1}{2}]$
hence X & Y are not independent.

(c)



$x = \frac{1}{2x}$
$x^2 = \frac{1}{2}$
$x = \frac{1}{\sqrt{2}}$

$P[XY < \frac{1}{2}] = \int_0^{1/\sqrt{2}} \int_x^{\frac{1}{2x}} c x^2 y^{-2} \, dy\, dx$

$= \int_0^{1/\sqrt{2}} c\, x^2 \left(-y^{-1}\Big|_x^{\frac{1}{2x}}\right) dx = \int_0^{1/\sqrt{2}} c x^2 \left(\frac{1}{x} - \frac{1}{\frac{1}{2x}}\right) dx$

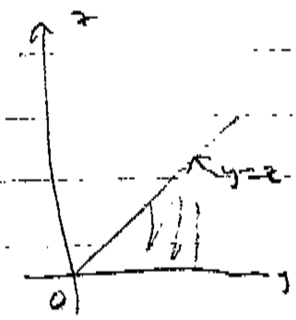
$= \int_0^{1/\sqrt{2}} c\,(x - 2x^3)\,dx = c\left(\frac{x^2}{2} - 2\frac{x^4}{4}\right)\Big|_0^{1/\sqrt{2}} = \frac{c}{2}\left(\frac{1}{2} - \frac{1}{4}\right.$

4. (a) $P[X_1 < 1] = \int_0^1 x e^{-x} dx$
$= -xe^{-x}\Big|_0^1 + \int_0^1 e^{-x} dx = xe^{-x}\Big|_0^1 - e^{-x}\Big|_0^1 = 1 - e^{-1} - e^{-1}$

$P[X(1) \geq 2] = 1 - e^{-1} - e^{-1}$
$X_i \sim Y_1 + Y_2$ where $Y_i \overset{iid}{\sim} exp(1)$

(b) $Y = X_1 + X_2$, $Z = X_1$ $\quad X_1 = Z$, $X_2 = Y - Z$

$J = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$



$f_{Y,Z}(y,z) = f_{X_1}(x_1(y,z))\, f_{X_2}(x_2(y,z))$
$= z e^{-z} (y-z) e^{-(y-z)} I_{(0,\infty)}(z) I_{(0,\infty)}(y-z)$

$f_Y(y) = \int_0^y z e^{-z} (y-z) e^{-(y-z)} dz = \int_0^y z(y-z) e^{-y} dz$
$= e^{-y} \int_0^y z(y-z)\, dz = e^{-y}\left(y\frac{z^2}{2} - \frac{z^3}{3}\right)\Big|_0^y$
$= \frac{1}{6} y^3 e^{-y}$ for $y > 0$, 0 elsewhere

note: $Y = Y_1 + Y_2 + Y_3 + Y_4 \sim \Gamma(4, 1)$ $\therefore f_Y(y) = \frac{y^{4-1}}{\Gamma(4)} e^{-y}$.

(c) (iii) $\psi_{S_N}(t) = E(\exp(t S_N))$
$= E(E(\exp(t S_N | N)))$
$= E\left([E(\exp(tX_1))]^N\right)$

$E(\exp(tX_1)) = \int_0^\infty \exp(tx)\, x e^{-x} dx$
$= \int_0^\infty x e^{-(1-t)x} dx = \frac{1}{(1-t)^2} \int_0^\infty \frac{(1-t)^2 x^{2-1} e^{-(1-t)x}}{\Gamma(2)}$
$= 1/(1-t)^2$

$E\left(\left(\frac{1}{(1-t)^2}\right)^N\right) = \sum_{n=1}^{\infty} \left(\frac{1}{(1-t)^2}\right)^n (1-p)^{n-1} p$
$= \frac{p}{(1-t)^2} \sum_{n=1}^{\infty} \left(\frac{1-p}{(1-t)^2}\right)^{n-1}$
$= \frac{p}{(1-t)^2} \frac{1}{1 - \frac{1-p}{(1-t)^2}} = \frac{p}{(1-t)^2 - (1-p)}$

$\psi^{(1)}_{S_N}(t) = \frac{2(1-t)p}{[(1-t)^2 - (1-p)]^2}$

(i) $\therefore E(S_N) = \psi^{(1)}_{S_N}(0) = \frac{2p}{p^2} = \frac{2}{p}$

$\psi^{(2)}_{S_N}(t) = \frac{-2p[(1-t)^2 - (1-p)]^2 + 2(2(1-t))[(1-t)^2 - (1-p)]\, 2(1-t)p}{[(1-t)^2 - (1-p)]^4}$

(ii) $\therefore Var(S_N) = \psi^{(2)}_{S_N}(0) - (E(S_N))^2 = \frac{-2p^3 + 8p^2}{p^4} - \frac{4}{p^2} = \frac{4 - 2p}{p^2}$