STATISTICS DEPARTMENT
M.S. EXAMINATION

PART I
CLOSED BOOK

Friday, May 25, 2001

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

School of Education Conference Room, AE 123

Instructions: Complete *all five* problems. Each problem counts 20 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.

1. (a) Suppose that we have 2 different groups of patients with a certain disease, 101 patients in each group. In one group, they take diagnostic test #1 and 95 patients test positive for the disease; in the 2^{nd} group, they take diagnostic test #2 and 88 test positive for the disease. Assume that the total of 202 patients are randomly chosen from some population and are randomly assigned to the two groups. Can one conclude that the probability of a testing positive differs between the two diagnostic tests? Show your work.

   (b) Suppose in another scenario that 101 patients randomly chosen from some population took both diagnostic tests. 82 tested positive for both tests, 13 tested positive for diagnostic test #1 but not for #2, a (+, −) result, and 6 tested positive for #2 but not for #1, a (−, +) result.

      (i) Suppose we focus our attention on only those 19 patients where there was a lack of agreement between the two tests. Can one conclude that the probability of a (+, −) result (versus a (−, +) result) is not equal to .5? Show your work.

      (ii) Explain why a "YES" answer to (i) would allow one to establish that it is not the case that the probability of a positive result is the same for both diagnostic tests.

---

2. A professional gambler expects to have a 'winning day' 4 out of 5 days when he plays a certain gambling game. He wins \$5000 on his winning days; on his losing days, he loses \$10,000. Assume independence between different plays of the game.

(a) What is the probability that his 5^{th} loss occurs on the 15^{th} play of the game?

(b) Assuming that his 5^{th} loss occurs on the 15^{th} play, how much will his gain be after 15 plays?

(c) Given that the gambler has just suffered his 5^{th} loss, how many total games would we have expected him to play?

(d) Given that the gambler has just suffered his 5^{th} loss, what is his expected gain?

(e) Suppose X = number of plays that it takes to achieve the gambler's w^{th} loss. Explain why for w, a large integer, X is approximately normally distributed.

3. Sam installs TV cables. He is about to install a cable in an apartment and a cable in a house. Let X be the time that he requires to install a cable in the apartment and let Y be the time that he requires to install a cable in the house. Suppose that $0 < a < b < \infty$ and that X has a uniform distribution on the interval (a, b). Suppose that $0 < \tau < \infty$ and that Y has the probability density function f_Y defined by

$$f_Y(y) = \begin{cases} 2\tau^2 y^{-3} & \text{for } \tau < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

5 (a) Find $P(Y > b)$.

5 (b) Find $E(Y - X)$.

(c) Suppose that X and Y are independent and suppose that $0 < \tau < a$.

5 (i) Find $E(\frac{1}{X})$.

5 (ii) Let $U = \frac{Y}{X}$. Find the probability density function of U.

4. Let $X_1, X_2, .., X_n$ be independent and identically distributed two-parameter exponential random variables with p.d.f.

$$f(x|\theta,\eta) = \frac{1}{\theta}e^{-(x-\eta)/\theta} \tag{1}$$

where $x > \eta$ and $\theta > 0$.

(a) Calculate the method of moments estimators of θ and η. (Hint: Recall the exponential and gamma densities when computing $E[X]$. Also recall that a location parameter does not change the variance of a random variable. Therefore, the variance of the two-parameter exponential is the same as the one-parameter exponential.)

(b) Calculate the maximum likelihood estimator, $\hat{\theta}$, of θ.

(c) Calculate the maximum likelihood estimator, $\hat{\eta}$, of η.

(d) Find the sufficient statistics for θ and η.

(e) Explain why the maximum likelihood estimators are functions of the sufficient statistics.

For the rest of the problem, assume that $\eta = 0$.

(f) Compute the Cramer-Rao Lower Bound for θ. Show that the MLE of θ, $\hat{\theta}$, achieves the CRLB.

(g) Give an asymptotic $100(1-\alpha)\%$ confidence interval for θ.

(h) Suppose the sample size is $n = 5$. Explain why the confidence interval given above may not be accurate in terms of its coverage probability. As an alternative method of calculating a confidence interval for θ, describe how you would implement the parametric bootstrap to produce the bootstrap estimate of θ, $\bar{\hat{\theta}}^*$, a bootstrap estimate of the standard error, $s_{\hat{\theta}}$, and a 95% Bootstrap confidence interval for θ.

5. Suppose that n is a positive integer and that $0 < p < 1$. Let N be a binomial random variable with parameters n and p. For each positive integer i, let X_i be a random variable whose probability density function is

$$f_X(x) = \begin{cases} \frac{1}{2}x^2 \exp(-x) & \text{if } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Suppose that N, X_1, X_2, $\cdots$ are independent. Let

$$Y_k = \begin{cases} 1 & \text{if } k = 0 \\ \prod_{i=1}^{k} X_i & \text{if } k = 1, 2, \cdots \end{cases}$$

7 (a) Find $P(Y_1 > 2)$.

7 (b) Find $E(Y_N)$ in terms of $\mu = E(X_1)$, n, and p. You need not determine the value of μ.

6 (c) For $k = 1, 2, \cdots$ let $Z_k = \frac{Y_k}{E(Y_k)}$. Show that Z_k, $k = 1, 2, \cdots$ is a Martingale.

CUMULATIVE DISTRIBUTION OF CHI-SQUARE*

Degrees of Freedom	Probability of a Greater Value												
	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	....	....	...	...	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17

Areas for the Normal Curve

02	.03	.04	.05	.06	.07	.08	.09
920	.4880	4840	.4801	.4761	.4721	4681	.4641
522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
129	.4090	.4052	4013	.3974	.3936	.3897	.3859
745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
676	.2643	2611	.2578	.2546	.2514	2483	.2451
358	.2327	2296	2266	.2236	.2206	.2177	.2148
061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
539	.1515	.1492	1469	.1446	.1423	.1401	.1379
314	.1292	1271	.1251	.1230	.1210	.1190	.1170
112	.1093	.1075	1056	.1038	.1020	.1003	.0985
)934	0918	.0901	0885	.0869	.0853	.0838	.0823
)778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
)643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
)526	.0516	.0505	0495	.0485	.0475	.0465	.0455
)427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
)	.0336	.0329	0322	.0314	.0307	.0301	.0294
)2	.0268	.0262	.0256	.0250	.0244	.0239	.0233
)217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
0132	0129	.0125	.0122	.0119	.0116	.0113	.0110
0102	.0099	.0096	.0094	0091	.0089	.0087	.0084
0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
0044	.0043	.0041	.0040	.0039	.0038	0037	.0036
0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

–
3
7
0
9

Percentage Points of the *t* Distribution

df	a = .1	a = .05	a = .025	a = .01	a =
1	3.078	6.314	12 706	31.821	63
2	1 886	2.920	4.303	6.965	9
3	1 638	2.353	3.182	4.541	5
4	1 533	2.132	2.776	3.747	4.
5	1.476	2.015	2.571	3.365	4.
6	1.440	1.943	2.447	3.143	3.
7	1.415	1.895	2.365	2.998	3.
8	1.397	1.860	2.306	2.896	3.
9	1.383	1.833	2.262	2.821	3.
10	1.372	1.812	2 228	2.764	3.
11	1.363	1.796	2.201	2.718	3
12	1.356	1.782	2.179	2.681	3.
13	1 350	1.771	2.160	2.650	3.
14	1.345	1.761	2.145	2.624	2.
15	1.341	1.753	2.131	2.602	2.
16	1.337	1.746	2.120	2.583	2.
17	1.333	1.740	2.110	2.567	2.
18	1.330	1.734	2 101	2.552	2.
19	1.328	1.729	2.093	2.539	2.
20	1.325	1.725	2.086	2.528	2.
21	1.323	1.721	2 080	2.518	2.
22	1.321	1.717	2 074	2.508	2.
23	1.319	1.714	2.069	2.500	2.
24	1.318	1.711	2.064	2.492	2.
25	1.316	1.708	2.060	2.485	2.
26	1.315	1.706	2.056	2.479	2.
27	1.314	1.703	2.052	2.473	2.
28	1.313	1.701	2.048	2.467	2.
29	1.311	1.699	2.045	2.462	2.
30	1.310	1.697	2.042	2.457	2.
40	1.303	1.684	2.021	2.423	2.
60	1.296	1.671	2.000	2.390	2.6
120	1.289	1.658	1.980	2.358	2.6
240	1.285	1.651	1.970	2.342	2.5
[illegible]	1.282	1.645	1.960	2.326	2.5

closed book #1

Answer

(a) Suppose that we have 2 different groups of patients with a certain disease, 101 patients in each group. In one group, they take diagnostic test #1 and 95 patients test positive for the disease and in the 2^{nd} group, they take diagnostic test #2 and 88 test positive for the disease. Assume that the total of 202 patients are randomly chosen from some population and are randomly assigned to the two groups. Can one conclude that it is not the case that the probability of a positive result is the same for both diagnostic tests? Show your work.

(b) Suppose in another scenario that 101 patients randomly chosen from some population took both diagnostic tests. 82 of them tested positive for both tests, 13 of them tested positive for diagnostic test #1 but not for #2, a (+, –), result, and 6 of them tested positive for #2 but not for #1, a (–, +) result.

(i) Supposing we focused our attention on only those 19 patients where there was a lack of agreement between the two tests. Can one conclude that the probability of a (+, –) result (versus a (–, +) result) is not equal to .5? Show your work.

(ii) Explain why if your answer to (i) is "YES", then you have established that it is not the case that the probability of a positive result is the same for both diagnostic tests.

```
>>> DA

! Datafile: C:\backup\E Drive\MSSOLUT\msp2s012nd.cy3

! Table   1 of   1
```

Table1	+	-	TOTAL
test #1	95 (91.5)	6 (9.5)	101
test #2	88 (91.5)	13 (9.5)	101
TOTAL	183	19	202

Pearson's

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2\left\{\frac{(3.5)^2}{91.5} + \frac{(3.5)^2}{9.5}\right\} = 2.847$$

```
>>> RC CH/EX

Datafile: C:\backup\E Drive\MSSOLUT\msp2s012nd.cy3

PEARSON CHI-SQUARE TEST

Statistic based on the observed 2 by 2 table(x) :
 CH(X): Pearson Chi-Square Statistic =      2.847(      2.091 with Yates CC)

Asymptotic p-value: (based on Chi-Square distribution with 1 df )
   Two-sided : Pr { CH(X) .GE.       2.847 } =      0.0916
   One-sided : 0.5 * Two-sided                =      0.0458

Exact p-value and point probabilities :
   Two-sided : Pr { CH(X) .GE.       2.847 } =      0.1466
               Pr { CH(X) .EQ.       2.847 } =      0.0952
   One-sided : Let y be the value in Row 1 and Column 1
       y =95 min(Y) =82 max(Y) =101 mean(Y) =      91.50 std(Y) =      2.080

          Pr { Y .GE. 95 } =      0.0733
          Pr { Y .EQ. 95 } =      0.0476
```

p-value = .0916
Thus at $\alpha = .05$ one cannot conclude that $P(+)$ is different for the 2 tests.

```
>>> DA

! Datafile: C:\backup\E Drive\MSSOLUT\msp2s01.cy3

! Table   1 of   1
```

diagnostic test 2

Table1		+	-	TOTAL
diagnostic test 1	+	82	13	95
	-	6	0	6
TOTAL		88	13	101

```
>>> PS MC/EX

Datafile: C:\backup\E Drive\MSSOLUT\msp2s01.cy3

MCNEMAR'S TEST
```

```
Statistic based on the observed 2 by 2 table(x) :

        Min           Max          Mean        Std-dev     Observed Standardized
      -19.00         19.00        0.0000        4.359        7.000        1.606

 Asymptotic Inference:
     One-sided p-value: Pr { Test Statistic .GE. Observed }  =       0.0541
     Two-sided p-value: 2 * One-sided                        =       0.1083

 Exact Inference:
     One-sided p-value: Pr {  Test Statistic .GE. Observed } =       0.0835
                        Pr {  Test Statistic .EQ. Observed } =       0.0518
     Two-sided p-value: 2*One-Sided                          =       0.1671
```

asymptotic

Newan $\chi^2 = \frac{(13-6)^2}{13+6} = \frac{49}{19} = (1.606)^2$; We can also use the chi-square goodness of fit test

p-value $= .1083$ (2-sided)

Cannot conclude at $\alpha = .05$

exact

$X = \#$ of $(+,-) = 13$

$n = 19$

$= 2P(X \geq 13) = 2(.0835) = .1671$; Cannot conclude at $\alpha = .05$

(from binomial table, $n=19$, $X=13$)

ii) Suppose you conclude that $P((+,-)) > P((-,+))$

then $P(+, \text{test } 1) = P((+,+)) + P((+,-))$

$> P((+,+)) + P((-,+)) = P(+, \text{test } 2)$.

CSX I
Open book #2 Answer

A professional gambler figures to have 'winning day' 4 out of 5 days that he plays a certain gambling game. He wins $5000 on his winning days; otherwise, on his losing days, he loses $10,000. Assume independence between different plays of the game.

(a) What is the probability that his 5^{th} loss occurs on the 15^{th} play of the game?
(b) Assuming that his 5^{th} loss occurs on the 15^{th} play, how much will be his gain after 15 plays?
(c) After the gambler suffers his 5^{th} loss, how many total games would we expect to have been played by then?
(d) What is the gambler's expected gain after suffering his 5^{th} loss?
(e) Suppose X = number of plays that it takes to achieve the gambler's w^{th} loss. Explain why for w, a large integer, X is approximately normally distributed.

(a) X = # of games until (and including) the 5^{th} loss is hypergeometric.

$$P(X=15) = \binom{14}{4}(.8)^{10}(.2)^5$$

(b) $10(5000) - 5(10000) = 0$ (even, steven)

~~(c) $\frac{5}{\left(\frac{1}{5}\right)} = 25 = E(X)$~~

(d) $G = (X-5)\,5000 - 5(10{,}000)$

$$E(G) = (20)\,5000 - 5(10{,}000) = 50{,}000$$

(e) Let Y_1 = # of plays thru 1^{st} loss

$i \geq 1$; Y_{i+1} = # of plays after i^{th} loss thru $(i+1)^{st}$ loss.

$Y_1, Y_2, \ldots$ are iid geometrics

$X = Y_1 + Y_2 + \cdots + Y_w$ is Normal

by the central limit theorem.

~~Open~~book #3

3. Sam installs TV cables. He is about to install a cable in an apartment and a cable in a house. Let X be the time that he requires to install a cable in the apartment and let Y be the time that he requires to install a cable in the house. Suppose that $0 < a < b < \infty$ and that X has a uniform distribution on the interval (a,b). Suppose that $0 < \tau < \infty$ and that Y has the probability density function f_Y defined by

$$f_Y(y) = \begin{cases} 2\tau^2 y^{-3} & \text{for } \tau < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

5 # (a) Find $P(Y > b)$. – $P[Y>b] = \begin{cases} \int_b^\infty 2\tau^2 y^{-3}\,dy = \dfrac{2\tau^2 y^{-2}}{-2}\Big|_{y=b}^{y=\infty} = \dfrac{\tau^2}{b^2} & \text{if } b \geq \tau \\ 1 & \text{if } b < \tau \end{cases}$

5 # (b) Find $E(Y - X)$. $E(Y-X) = E(Y) - E(X)$, $E(X) = \frac{a+b}{2}$, $E(Y) = \int_\tau^\infty y\, 2\tau^2 y^{-3}\,dy = \dfrac{2\tau^2 y^{-1}}{-1}\Big|_{y=\tau}^{\infty} = 2\tau$. Hence $E(Y) - E(X) = 2\tau - \frac{a+b}{2}$

(c) Suppose that X and Y are independent and suppose that $0 < \tau < a$.

~~4 (i) Find $P(Y > X)$.~~ $P[Y>X] = \int_a^b \dfrac{P[Y>x]\,dx}{b-a} = \dfrac{\int_a^b \frac{\tau^2}{x^2}dx}{b-a} = \dfrac{\tau^2\left(\frac{1}{-x}\right)\Big|_{x=a}^{x=b}}{b-a} = \dfrac{\tau^2\left(\frac{1}{a}-\frac{1}{b}\right)}{b-a} = \dfrac{\tau^2}{ab}$

5 # (i) ~~(ii)~~ Find $E(\frac{Y}{X})$. $E(\frac{Y}{X}) = E(Y)\,E(\frac{1}{X}) = 2\tau \int_a^b \frac{1}{x}dx/(b-a) = \frac{2\tau}{b-a}(\ln(b) - \ln(a))$

5 # ~~(iii)~~ (ii) Let $U = \frac{Y}{X}$. ~~Find the probability density function of U~~ $P[U>1]$

~~$P[U>u] = P[Y>uX]$~~

$= \dfrac{\int_a^b P[Y>ux]\,dx}{b-a}$

note U always exceeds $\frac{\tau}{b}$:

$U = \frac{Y}{X} > \frac{Y}{b} > \frac{\tau}{b}$.

$= \dfrac{\int_a^b \frac{\tau^2}{(ux)^2}dx}{b-a} = \frac{1}{u^2}\frac{\tau^2}{ab}$ if $ua \geq \tau$

for $\frac{\tau}{b} \leq u < \frac{\tau}{a}$,

$$P[U>u] = \frac{\int_a^b P[Y>ux]\,dx}{b-a}$$

$= \dfrac{\int_a^b I_{(-\infty,\tau]}(ux)\,dx}{b-a} + \dfrac{\int_a^b \frac{\tau^2}{(ux)^2} I_{(\tau,\infty)}(ux)\,dx}{b-a}$

$\dfrac{\int_a^{\frac{\tau}{u}} dx}{b-a} + \dfrac{\int_{\frac{\tau}{u}}^b \left(\frac{\tau}{u}\right)^2 \frac{1}{x^2}dx}{b-a} = \dfrac{\frac{\tau}{u} - a + \left(\frac{\tau}{u}\right)^2 \left(\frac{-1}{x}\Big|_{x=\frac{\tau}{u}}^{x=b}\right)}{b-a}$

$\frac{\tau}{u} - a + \left(\frac{\tau}{u}\right)^2\left(\frac{u}{\tau} - \frac{1}{b}\right)$ $\quad \frac{2\tau}{u} - a - \left(\frac{\tau}{u}\right)^2 \frac{1}{b}$ $\quad \therefore f_U(u) = \frac{2}{u^3}\frac{\tau^2}{ab} I_{[\frac{\tau}{a},\infty)}(u)$

$2\tau \quad 2\tau^2$

$u = \frac{Y}{X}, \quad v = Y$

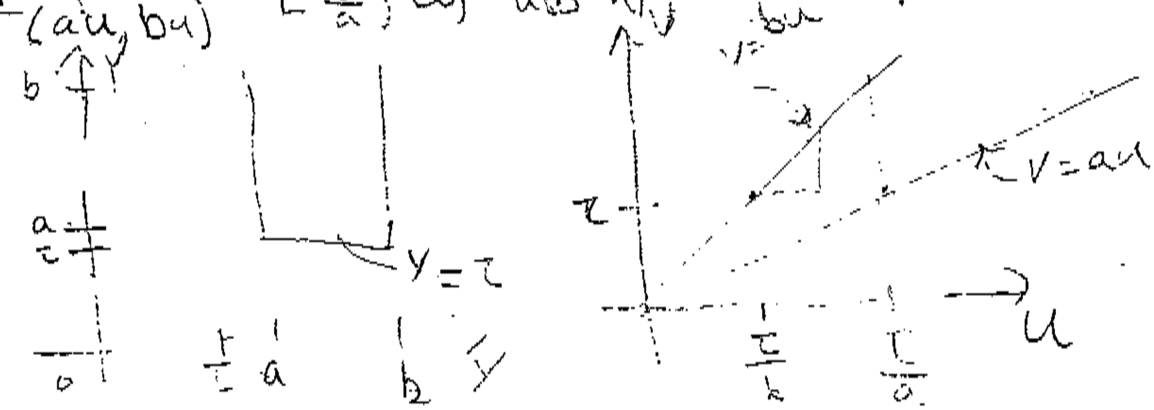
$x = \frac{v}{u}, \quad Y = V$

$$J = \begin{vmatrix} -\frac{v}{u^2} & \frac{1}{u} \\ 0 & 1 \end{vmatrix} = -\frac{v}{u^2}$$

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))\,|J|$$

$$= \frac{1}{b-a} I_{(a,b)}\left(\frac{v}{u}\right) 2\tau^2 v^{-3} I_{(\tau,\infty)}(v) \frac{v}{u^2}$$

$$\cancel{\frac{1}{u^2(b-a)} 2\tau^2 v^{-2} I_{(au,bu)}(v) I_{[\frac{\tau}{a},\infty)}(u) + \frac{1}{u^2(b-a)} 2\tau^2 v^{-2} I_{(\tau,bu)}(v) I_{(\frac{\tau}{b},\frac{\tau}{a})}(u)}$$



$\cdot - 0 \Rightarrow \frac{v}{u} = a, \quad v = au$

$v = bu$

$$f_U(u) = \frac{2\tau^2}{u^2(b-a)}\left[\int_{au}^{bu} v^{-2}\,dv\, I_{[\frac{\tau}{a},\infty)} + \int_{\tau}^{bu} v^{-2}\,dv\, I_{(\frac{\tau}{b},\frac{\tau}{a})}(u)\right]$$

$$= \frac{2\tau^2}{u^2(b-a)}\left[-\frac{1}{v}\Big|_{v=au}^{v=bu} I_{[\frac{\tau}{a},\infty)}(u) + -\frac{1}{v}\Big|_{v=\tau}^{v=bu} I_{(\frac{\tau}{b},\frac{\tau}{a})}(u)\right]$$

$$2\tau^2 \left[\frac{1}{\cdot} - \frac{1}{\cdot}\right] I \quad \cdots \quad + \left(\frac{1}{\tau} - \frac{1}{bu}\right) I_{(\frac{\tau}{b},\frac{\tau}{a})}(u)\Big]$$

$$E(U) = \frac{2\tau^2}{ab}\int_{\frac{\tau}{a}}^{\infty} u^{-2}\,du + \frac{2\tau}{b-a}\int_{\frac{\tau}{b}}^{\frac{\tau}{a}} u^{-1} - \frac{\tau}{b}u^{-2}\,du$$

$$= \frac{2\tau^2}{ab}\left[-u^{-1}\Big|_{u=\frac{\tau}{a}}^{u=\infty}\right] + \frac{2\tau}{b-a}\left[\ln(u) + \frac{\tau}{b}u^{-1}\right]\Bigg|_{u=\frac{\tau}{b}}^{u=\frac{\tau}{a}}$$

$$= \frac{2\tau^2}{ab}\left[\frac{a}{\tau}\right] + \frac{2\tau}{b-a}\left[\ln\left(\frac{\tau}{a}\right) - \ln\left(\frac{\tau}{b}\right) + \frac{\tau}{b}\left(\frac{a}{\tau} - \frac{b}{\tau}\right)\right]$$

$$= \frac{2\tau}{b} + \frac{2\tau}{b-a}\left[\ln(b) - \ln(a)\right] - \frac{2\tau}{b}$$

— closed book # 4

a) $E[X] = \int_\eta^\infty x \frac{1}{\theta} e^{-\frac{x-\eta}{\theta}} dx = \frac{1}{\theta}\int_\eta^\infty x\, e^{-\frac{x-\eta}{\theta}} dx$

$= \frac{1}{\theta}\int_\eta^\infty (x-\eta+\eta)\, e^{-\frac{x-\eta}{\theta}} dx$

$= \frac{1}{\theta}\int_\eta^\infty (x-\eta)\, e^{-\frac{x-\eta}{\theta}} dx + \frac{\eta}{\theta}\int_\eta^\infty e^{-\frac{x-\eta}{\theta}} dx$

let $y = x-\eta$, $dy = dx$ let $z = x-\eta$, $dz = dx$

$= \frac{1}{\theta}\int_0^\infty y\, e^{-\frac{y}{\theta}} dy + \frac{\eta}{\theta}\int_0^\infty e^{-\frac{z}{\theta}} dz$

$= \theta\Gamma(2)\underbrace{\int_0^\infty \frac{1}{\theta^2\Gamma(2)} y^{2-1} e^{-\frac{y}{\theta}} dy}_{\text{Gamma}(\theta, 2)} + \eta\underbrace{\int_0^\infty \frac{1}{\theta} e^{-\frac{z}{\theta}} dz}_{\text{Exp}(\theta)}$

$\Gamma(2) = 1$

$= \theta + \eta$

Using the hint.

$E[X^2] = \text{Var}(X) + (E[X])^2 = \theta^2 + (\theta+\eta)^2$

To find the MME's solve

$$E[X] = \theta + \eta = \frac{1}{n}\sum x_i$$

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

① $\tilde{\eta} = \frac{1}{n}\sum x_i - \tilde{\theta} = \bar{x} - \tilde{\theta}.$

② $\theta^2 + (\theta+\eta)^2 = \frac{1}{n}\sum x_i^2$

$$2\theta^2 + 2\theta\eta + \eta^2 = \frac{1}{n}\sum x_i^2$$

$$2\theta^2 + 2\theta(\bar{x}-\theta) + (\bar{x}-\theta)^2 = \frac{1}{n}\sum x_i^2$$

$$\cancel{2\theta^2} + \cancel{2\theta\bar{x}} - \cancel{2\theta^2} + \bar{x}^2 - \cancel{2\bar{x}\theta} + \theta^2 = \frac{1}{n}\sum x_i^2$$

$$\theta^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

$$\theta^2 = \frac{\sum x_i^2 - \frac{(\sum x)^2}{n}}{n}$$

$$\tilde{\theta} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n}}$$

b) $L(\theta, \eta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i - \eta}{\theta}} \quad x_1 \cdots x_n > \eta$

$= \theta^{-n} e^{-\frac{1}{\theta}\Sigma(x_i - \eta)} \quad x_{(1)} > \eta.$

$l(\theta, \eta) = -n \log \theta - \frac{1}{\theta} \Sigma(x_i - \eta)$

$\frac{\partial}{\partial\theta} l(\theta, \eta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\Sigma(x_i - \eta) = 0$

$\frac{n}{\theta} = \frac{1}{\theta^2}\Sigma(x_i - \eta)$

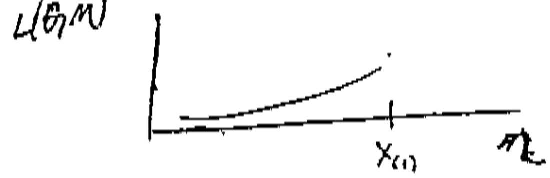
$\hat{\theta} = \frac{\Sigma(x_i - \hat{\eta})}{n}.$ //

---

c) irregular estimation

$L(\theta, \eta) = \theta^{-n} e^{-\frac{1}{\theta}\Sigma x_i} e^{+\frac{\eta}{\theta}}$

$L(\theta, \eta) \propto e^{+\eta} \quad \eta < X_{(1)}$



$\hat{\eta} = X_{(1)}$

d) $f(x_1, \dots, x_n | \theta, \eta) = \theta^{-n} e^{-\frac{1}{\theta}\Sigma(x_i - \eta)} I_{(\eta,\infty)}(x_{(1)})$

$= \theta^{-n} e^{+\frac{n\eta}{\theta}} e^{-\frac{\Sigma x_i}{\theta}} I_{(\eta,\infty)}(x_{(1)})$

$\therefore (\Sigma X_i, X_{(1)})$ are suff. for (θ, η)

e) From the factorization Theorem, the likelihood is $g(T, \theta)\, h(\underset{\sim}{X})$, which depends on θ only through T. Therefore, to maximize this quantity ~~we need only maximize~~ $g(T, \theta)$.

---

f) $f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $\log f(x|\theta) = -\log\theta - \frac{x}{\theta}$

$\frac{d}{d\theta} \log f(x|\theta) = -\frac{1}{\theta} - \frac{x}{\theta}$.

$\frac{d^2}{d\theta^2} \log f(x|\theta) = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$

$I(\theta) = -E\left[\frac{d^2}{d\theta^2} \log f(x|\theta)\right] = -E\left[\frac{1}{\theta^2} - \frac{2x}{\theta^3}\right]$

$= -\frac{1}{\theta^2} + \frac{2}{\theta^3} E[X] = -\frac{1}{\theta^2} + \frac{2}{\theta^3}\theta = \frac{1}{\theta^2}$

$CRLB = \frac{1}{nI(\theta)} = \frac{\theta^2}{n}$.

For the $Exp(\theta)$, the MLE is $\hat{\theta} = \overline{X}$.

Since $Var(\hat{\theta}) = Var(\overline{X}) = \frac{\theta^2}{n}$, $\hat{\theta}$ acheives the CRLB

g) $\hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta})}} \Rightarrow \hat{\theta} \pm z_{\alpha/2} \frac{\hat{\theta}}{\sqrt{n}}$ where $\hat{\theta} = \bar{X}$

e) The asymptotic CI computed in g may not be accurate since it requires a large sample. The coverage probability may be low since the CI would be too narrow as a result of the small sample size.

To implement the parametric bootstrap:

① Using the MLE $\hat{\theta} = \bar{X}$ generate B samples from $Exp(\hat{\theta})$. ② Compute $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ as the sample mean of each bootstrap sample. ③ The bootstrap estimate of θ is $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}_i^*$ ④ The bootstrap estimate of $S_{\hat{\theta}}$ is $S_{\hat{\theta}} = \sqrt{\frac{\Sigma(\hat{\theta}_i^* - \bar{\hat{\theta}}^*)^2}{B-1}}$

⑤ An approximate 95% bootstrap CI can be obtained from the 0.025 and 0.975 percentiles of the $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

closed book #3

5 Suppose that n is a positive integer and that $0 < p < 1$. Let N be a binomial random variable with parameters n and p. For each positive integer i, let X_i be a random variable whose probability density function is

$$f_X(x) = \begin{cases} \frac{1}{2}x^2\exp(-x) & \text{if } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Suppose that N, X_1, X_2, $\cdots$ are independent. Let

$$Y_k = \begin{cases} 1 & \text{if } k = 0 \\ \prod_{i=1}^{k} X_i & \text{if } k = 1, 2, \cdots \end{cases}$$

(a) Find $P(Y_1 > 2)$. $P[Y_1 \geq 2] = P[X_1 \geq 2] = \int_2^\infty \frac{1}{2}x^2 e^{-x}dx = -\frac{1}{2}x^2 e^{-x}\Big|_2^\infty + \int_2^\infty x e^{-x}dx$
$= 2e^{-2} - xe^{-x}\Big|_2^\infty + \int_2^\infty e^{-x}dx = 2e^{-2} + 2e^{-2} - e^{-x}\Big|_2^\infty = 5e^{-2}$ or $P[Y > 2] = P[N(2) < 3] = e^{-2} + 2e^{-2} + \frac{2^2 e^{-2}}{2!}$

~~(b) Find $E(Y_N | N = 10)$ in terms of $\mu = E(X_1)$. You need not determine the value of μ.~~
~~$E(Y_N | N = 10) = E(\prod_{i=1}^{10} X_i) = \prod_{i=1}^{10} E(X_i) = \prod_{i=1}^{10} \mu = \mu^{10}$~~

(b) Find $E(Y_N)$ in terms of $\mu = E(X_1)$, n, and p. You need not determine the value of μ. $E(Y_N) = \sum_{k=0}^{n} E(Y_k | N = k)\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n} \mu^k \binom{n}{k}p^k(1-p)^{n-k} = (1-p+\mu p)^n$

and $\mu = 3$

(c) For $k = 1, 2, \cdots$ let $Z_k = \frac{Y_k}{E(Y_k)}$ Show that Z_k, $k = 1, 2, \cdots$ is a Martingale.

$$E(Z_{k+1} | Z_1 \cdots Z_k) = E\left(Z_k \frac{X_{k+1}}{E(X_{k+1})} \Big| Z_1 \cdots Z_k\right)$$
$$= Z_k E\left(\frac{X_{k+1}}{E(X_{k+1})} \Big| Z_1, \cdots, Z_k\right)$$
$$= Z_k E\left(\frac{X_{k+1}}{E(X_{k+1})}\right)$$
$$= Z_k$$