STATISTICS DEPARTMENT
M.S. EXAMINATION

PART I
CLOSED BOOK

Friday, May 16, 2003

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

Biella Room (Library, First Floor)

Instructions: Complete _four of the five_ 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.

Tables of some distributions are provided. Use them as appropriate.

1. Let $X_{[1]} \leq X_{[2]} \leq X_{[3]} \leq \ldots \leq X_{[41]}$ be 41 ordered observations from the variable X = number of movies seen per month for a random sample of 41 people. The data are given below. Let θ be the population 70[th] percentile for this variable.

Note the data and the cumulative binomial probabilities given below.

(a) Explain why $P(X > \theta) = 0.3$.
(b) For θ_0, some specific value of θ, what type of random variable is Y = the number of the 41 observations that are greater than θ_0?
(c) Suppose we test $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$, accepting H_0 if $9 \leq Y \leq 18$, and rejecting H_0 otherwise. What is the value of α = P(Type I error) for this test?
(d) If we test $H_0: \theta = 4.0$ versus $H_1: \theta \neq 4.0$, what is the conclusion?
(Notice that 4.0 is the 23[rd] observation.)
If we test $H_0: \theta = 6.0$ versus $H_1: \theta \neq 6.0$, what is the conclusion?
(Notice that 6.0 is the 33[rd] observation.)
(e) Explain why $\{\theta: 4.0 \leq \theta < 6.0\}$ is a $100(1-\alpha)\%$ confidence interval for θ.
<u>**Use the value of α obtained in part (c).**</u>

Data

1.0	1.0	1.5	1.5	1.5	1.5	1.5	1.5	2.0	2.0
2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
3.0	4.0	4.0	4.0	4.5	4.5	4.5	4.5	5.0	5.0
5.0	5.0	6.0	6.0	6.0	6.0	8.0	10.0	11.0	12.5
15.0									

Cumulative binomial probabilities: n = 41, p = 0.3

k	P(Y<=k)
0	0.00000
1	0.00001
2	0.00008
3	0.00045
4	0.00197
5	0.00680
6	0.01922
7	0.04583
8	0.09429
9	0.17045
10	0.27490
11	0.40105
12	0.53621
13	0.66543
14	0.77619
15	0.86163
16	0.92114
17	0.95864
18	0.98007
19	0.99119
20	0.99643
21	0.99868
22	0.99955
23	0.99986
24	0.99996
25	0.99999
26	1.00000

(Part I)

Solutions #1 CB

1) Let $X_{[1]} \le X_{[2]} \le X_{[3]} \ldots \le X_{[41]}$ be 41 ordered observations from the variable X = number of movies seen per month for a random sample of 41 people. The data is given below. **Note also the cumulative binomial probabilities given below.** Let θ = the population 70th percentile for this variable.

(a) Explain why $P(X > \theta) = .3$.
(b) For θ_0, some specific value of θ, what type of random variable is Y = the number of the 41 observations that are greater than θ_0.
(c) Suppose we test H_0: $\theta = \theta_0$ versus H_1: $\theta \ne \theta_0$, accepting H_0 if $9 \le Y \le 18$, and rejecting H_0 otherwise. What is the value of α = P(type I error) for this test?
(d) Suppose we test H_0: $\theta = 4.0$ versus H_1: $\theta \ne 4.0$, what is the conclusion (note that 4.0 is the 23rd observation)? Suppose we test H_0: $\theta = 6.0$ versus H_1: $\theta \ne 6.0$, what is the conclusion (note that 6.0 is the 33rd observation)?
(e) Explain why $(4.0 \le \theta < 6.0)$ is a $100(1-\alpha)\%$ confidence interval for θ.

nomovie

1.0	1.0	1.5	1.5	1.5	1.5	1.5	1.5	2.0	2.0
2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
3.0	4.0	4.0	4.0	4.5	4.5	4.5	4.5	5.0	5.0
5.0	5.0	6.0	6.0	6.0	6.0	8.0	10.0	11.0	12.5
15.0									

Data Display (binomial, n = sample size = 41, p = probability of successs = .3, cumulative probabilities)

k	P(Y<=k)
0	0.00000
1	0.00001
2	0.00008
3	0.00045
4	0.00197
5	0.00680
6	0.01922
7	0.04583
8	0.09429
9	0.17045
10	0.27490
11	0.40105
12	0.53621
13	0.66543
14	0.77619
15	0.86163
16	0.92114
17	0.95864
18	0.98007
19	0.99119
20	0.99643
21	0.99868
22	0.99955
23	0.99986
24	0.99996
25	0.99999
26	1.00000

(a) Since by definition of 70th percenti[le] $P(X \le \theta) = 0.7 \Rightarrow P(X > \theta) = 1 - P(X \le \theta) = 1 - 0.3$ $= 0.3$

(b) binomial $n = 41$, $p = .3$.

(c) $(1-\alpha) = P(9 \le Y \le 18) = P(Y \le 18) - P(Y \le 8)$
$= .98007 - .09429 \cong .886$
Thus $\alpha \cong .114$

(d) $H_0: \theta = 4.0$, $9 \le Y = 17 \le 18$
Thus accept H_0
(For $\theta_0 < 4.0$, $Y > 18$ and we would reject H_0

$H_0: \theta = 6.0$, $Y = 5$ and we would reject H_0; also for $\theta_0 > 6.0$.
For $4.0 \le \theta_0 < 6.0$, $Y \ge 9$ and $Y \le 18$, and we would accept H_0.

(e) see answer to (d)

2. Consider the following display for data from an incidence study for a disease.

	Disease status		
Risk factor status	*Disease*	*No disease*	*Total*
Exposed	a	b	$a+b$
Not exposed	c	d	$c+d$
Total	$a+c$	$b+d$	n

Where $n = a+b+c+d$. We define the *risk* of the disease in the sample as

$$risk = r = \text{number of cases of disease / number of people at risk} = (a+c)/n$$

which is used to estimate the *population risk*, ϕ. We define the *exposure-specific risks*, for those with the risk factor as $a/(a+b)$ and for those without the risk factor as $c/(c+d)$. We also define the *relative risk* for those with the risk factor, compared to those without the risk factor as

$$\lambda = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)} \,. \tag{1}$$

(a) Let $X = a+c$. What is the distribution of the number X of cases of disease in the sample of size n, assuming constant risk over the risk factor status? Give the formula for the likelihood function of the *risk* ϕ. Show that the maximum likelihood estimate (m.l.e.) of the *risk* ϕ is $\hat{\phi} = (a+c)/n$, based on observing the number of cases of disease in the sample $X = (a+c)$.

(b) What is the large sampling distribution of the m.l.e. $\hat{\phi}$ of ϕ? Give a large sample confidence interval for the *population risk* ϕ.

The derivation of the large sample confidence interval for the *population relative risk* λ is slightly more difficult to derive. The large sample distribution of the *sample relative risk* $\hat{\lambda}$ is skewed and a *log* transformation is used to achieve approximate normality. On the log scale it can be shown that

$$\hat{se}(log(\hat{\lambda})) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}} \,. \tag{2}$$

Therefore, a 95% confidence interval for $log(\lambda)$ is

$$\log(\hat{\lambda}) \pm 1.96\hat{se}(\log\hat{\lambda}) \tag{3}$$

with lower and upper confidence limits of

$$L_{log} = \log(\hat{\lambda}) - 1.96\hat{se}(\log\hat{\lambda}) \tag{4}$$

$$U_{log} = \log(\hat{\lambda}) + 1.96\hat{se}(\log\hat{\lambda})\ . \tag{5}$$

Since we want a 95% confidence interval for λ itself, we can obtain the two limits by raising (L_{log}, U_{log}) to the power of exp. That is

$$L = \exp(L_{log}) \tag{6}$$

$$U = \exp(U_{log}) \tag{7}$$

to give a 95% confidence interval for λ, the *population relative risk*.

(c) Explain why $\hat{\lambda}$, the *sample relative risk* statistic, might have a skewed distribution? In which direction would the distribution be skewed? Why might the log transformation help make the sampling distribution of the *sample relative risk* more normal? What property of m.l.e.s is being used when the confidence interval for $\log(\lambda)$ is transformed using exp? Explain why the results will be valid.

Suppose risk factors for coronary heart disease are being studied in men. The following table gives the smoking status of men entering the study and whether or not a coronary event occurred during the 10 years the study was conducted.

	Coronary event?		
Smoker entering the study	*Yes*	*No*	*Total*
Yes	166	1176	1342
No	50	513	563
Total	216	1689	1905

(d) Compute the estimated *population risk* $\hat{\phi}$ of coronary disease in the study. Compute a 95% confidence interval for the *population risk*, ϕ.

(e) Compute the estimated *relative risk* of coronary disease of smokers to nonsmokers. Compute a 95% confidence interval for the $\log(\lambda)$. Compute a 95% confidence interval for λ. Conduct a hypothesis test of $H_0 : \lambda = 1$ versus $H_1 : \lambda \neq 1$ using the final confidence interval computed in part (d) above. Is there statistically significant evidence that smoking is a risk factor for coronary heart disease?

3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with the density function $f(x) = (1/\theta^2)\, x\, e^{-x/\theta}$, for $x > 0$. The parameter θ is unknown.

(a) Identify the distribution family of this population, specifying the value of any known parameter of the population distribution. Derive $E(X_i)$ in terms of θ. State $V(X_i)$ if you know it, otherwise derive it.

(b) Find the method of moments estimators of θ and of $\tau = \tau(\theta) = 1/\theta$.

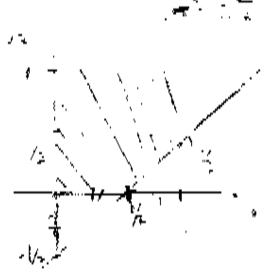
(c) Find the maximum likelihood estimator $\hat{\theta}$ of θ. State the maximum likelihood estimator of τ and the name of the principle by which you found it.

(d) Find the Cramér-Rao bound on unbiased estimators of θ. Can this result be used to determine whether $\hat{\theta}$ is the UMVUE of θ? Why or why not? Can this result be used to find a UMVUE for τ? Why or why not?

(e) Based on $n = 800$ observations from this distribution, suppose that the sample mean is 273.1. Give approximate 95% confidence intervals for θ and τ. Quote appropriate theorem(s) to justify your method.

4. Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables that have the probability density function

$$f(y) = I_{(0,1)}(y) = \begin{cases} 1 \text{ if } 0 < y < 1 \\ 0 \text{ elsewhere.} \end{cases}$$



5 (a) Find $P\left(Y_1 \le Y_2 + \frac{1}{2}\right)$.

5 (b) For $k = 1, 2, \cdots$, let $X_k = -\ln(Y_k)$.

For $n = 1, 2, \cdots$, let $W_n = \sum_{k=1}^{n} X_k$.

5 (i) Find $P(W_2 \le w)$ for $0 < w < \infty$.

5 (ii) Suppose that $0 < t \le 1$. Find $E\left(t^{W_n}\right)$.

10 (iii) For $0 < \theta < 1$, let N be a random variable that has the probability mass function

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

Suppose that N is independent of $X_1, X_2, \cdots$.

Find $E(W_N)$. Also find $E\left(e^{tW_N}\right)$ for $-\infty < t \le 0$.

5. The number of items that arrive at a repair facility by time t is a Poisson process $Y(t)$ for $0 \le t < \infty$. Assume that the items arrive at a rate of λ items per hour.

(a) Let n be a positive integer and suppose that $0 < s < t < \infty$.

5 (i) If $Y(s) = n$, then what is the expected value of $Y(t)$?

5 (ii) If $Y(t) = n$, then what is the variance of $Y(s)$?

(b) Twenty percent of the arriving items require an expensive repair. Assume that the items are independent of each other and that the items independent of $Y(t)$. For $0 \le t < \infty$, let $X(t)$ be the number of items requiring an expensive repair that arrive by time t.

5 (i) For $0 \le t < \infty$, what is the expected value of $X(t)$?

5 (ii) For $0 \le t < \infty$, what is variance of $X(t)$?

5 (iii) For $0 \le t < \infty$ and $k = 0, 1, 2, \cdots$, what is the probability that $X(t)$ is equal to k?

Upper-tail Areas for the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	Area
3.500	.00023263
4.000	.00003167
4.500	.00000340
5.000	.00000029

Percentage Points of the t Distribution

df	α = .1	α = .05	α = .025	α = .01	α = .005	α = .001
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
240	1.285	1.651	1.970	2.342	2.596	3.125
inf.	1.282	1.645	1.960	2.326	2.576	3.090

Solution - #2 CB

2

a) $X \sim Bin(n, \phi)$

$$L(\phi) = f(x \mid \phi) = L(\phi) = \binom{n}{x} \phi^{x} (1-\phi)^{n-x}$$

$$\ell(\phi) = \log \binom{n}{x} + x \log(\phi) + (n-x) \log(1-\phi)$$

$$\ell'(\phi) = \frac{x}{\phi} - \frac{n-x}{1-\phi} = 0$$

$$\frac{x}{\phi} = \frac{n-x}{1-\phi}$$

$$\frac{1}{\phi} - 1 = \frac{1-\phi}{\phi} = \frac{n-x}{x} = \frac{n}{x} - 1$$

$$\hat{\phi} = \frac{x}{n} = \frac{a+c}{n}$$

b) $\ell''(\phi) = -\dfrac{x}{\phi^{2}} - \dfrac{n-x}{(1-\phi)^{2}}$

$$E[-\ell''(\phi)] = E\left[\frac{x}{\phi^2} + \frac{n-x}{(1-\phi)^2}\right]$$

$$= \frac{1}{\phi^2} E[x] + \frac{1}{(1-\phi)^2} E[n-x]$$

$$= \frac{n\phi}{\phi^2} + \frac{n-n\phi}{(1-\phi)^2}$$

$$= \frac{n(1-\phi)^2 + \phi(n-n\phi)}{1-\phi(1-\phi)^2}$$

$$= \frac{n - 2n\phi + n\phi^2 + n\phi - n\phi^2}{\phi(1-\phi)^2}$$

$$= \frac{n - n\phi}{\phi(1-\phi)^2} = \frac{n(1-\phi)}{\phi(1-\phi)^2} = \frac{n}{\phi(1-\phi)}$$

$$AV(\hat{\phi}) = \frac{1}{E[-l''(\phi)]} = \frac{\phi(1-\phi)}{n}$$

$$\hat{\phi} \sim N\left(\phi, \frac{\phi(1-\phi)}{n}\right) \qquad se(\hat{\phi}) = \sqrt{\frac{\phi(1-\phi)}{n}}$$

large sample CI

$$\hat{\phi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\phi}(1-\hat{\phi})}{n}} \qquad \hat{se}(\hat{\phi}) = \sqrt{\frac{\hat{\phi}(1-\hat{\phi})}{n}}$$

or.

$$r \pm z_{\alpha/2}\sqrt{\frac{r(1-r)}{n}}$$

c) Since the relative risk λ can take values from 0 to $+\infty$ and since $\lambda = 1$ when the exposure-specific risks are equal the sampling distribution of λ should be skewed to the right. The log transformation should help by lowering the values of $\lambda > 1$ and spreading out the values below $\lambda < 1$. The invariance property is used, and since the transformation e^x is one-to-one it is valid.

d) $r = \hat{p} = (a+c)/n = 216/1905 = .1134$

CI for p

$$r \pm z_{\alpha/2} \sqrt{\frac{r(1-r)}{n}}$$

$$.1134 \pm 1.96 \sqrt{\frac{(.1134)(.8866)}{1905}}$$

$$.1134 \pm .0142 \qquad (.0992, .1276)$$

c) Smokers $a/(a+b) = 166/1342 = .1237$

nonsmokers $c/(c+d) = 50/563 = .0888$

relative risk $\hat{\lambda} = \frac{a/(a+b)}{c/(c+d)} = \frac{.1237}{.0888} = 1.3930$

$$\hat{se}(\log(\hat{\lambda})) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

$$= \sqrt{\frac{1}{166} - \frac{1}{1342} + \frac{1}{50} - \frac{1}{563}}$$

$$= .1533$$

$$L_{\log} = \log_e(1.3930) - 1.96\,(.1533) = .0310$$

$$U_{\log} = \log(1.3930) + 1.96\,(.1533) = .6319$$

$$L = e^{.0310} = 1.0315$$

$$U = e^{.6319} = 1.8813$$

$\therefore (1.0315, 1.8813)$

$H_0: \lambda = 1 \qquad H_1: \lambda \neq 1$

Reject H_0 since $\lambda = 1$ is not contained in the CI.

#3 CB

Answers

(a) (Using the notation of Bain and Englehardt) gamma with shape parameter $\kappa = 2$ and unknown scale parameter θ. The derivation (shown in many probability and mathematical statistics texts) of $E(X) = \kappa\theta$ based on the fact that a gamma density for $\kappa = 3$ and θ integrates to 1. That $V(X) = \kappa\theta^2$ can be derived similarly by finding $E(X^2)$, but here it is sufficient just to state the result.

(b) The MME $\bar{X}/2$ of θ is found by setting $\mu = 2\theta = \bar{X}$ and solving for θ. By invariance, the MME of τ is $2/\bar{X}$. (The parameter τ is the rate of the underlying Poisson process and it is often called λ.)

(c) $L(\theta) = \Pi f(X_i|\theta) = \theta^{-2n}\, \Pi X_i\, e{-}S/\theta$, where $S = \Sigma X_i$.
Then $\ln L(\theta) = \ell(\theta) = -2n \ln\theta + \Sigma \ln X_i - \theta^{-1}\Sigma X_i$ and $\ell'(\theta) = -2n/\theta + S/\theta^2$.
Solving $\ell'(\theta) = 0$ for θ, we get $\hat{\theta} = \bar{X}/2$ for the MLE (which agrees with the MME). By invariance, the MLE of τ is $2/\bar{X}$.

(d) Fisher's information for a single observation X is

$$I(\theta) = -E[(d^2/d\theta^2)\, f(X|\theta)] = -E[2\theta^{-2} - 2X\theta^{-3}] = 2\theta^{-2}.$$

So CRLB $= \theta^2/2n$. Because $V(\hat{\theta}) = V(\bar{X}/2) = \theta^2/n =$ CLRB, we know that $\hat{\theta}$ is UMVUE for θ. Because τ is a nonlinear function of θ, we know that the variance of an unbiased estimator of τ cannot achieve its CRLB, so this method won't work.

(e) For large n, the MLE of θ is approximately normal with mean θ and standard deviation $\sigma_n = \theta(2n)^{-1/2}$. This is a standard theorem about the asymptitic properties of MLEs. Considering $n = 800$ as large, we have $\sigma_{800} = \theta/40$ and

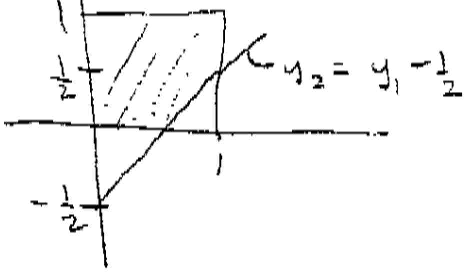
$$P\{\theta - 2\theta/40 < \bar{X}/2 < \theta + 2\theta/40\} = P\{1.9\theta < \bar{X} < 2.1\theta\} = P\{\bar{X}/2.1 < \theta < \bar{X}/1.9\} \approx 0.95$$

(We might have multiplied σ_n by 1.96, but this is only an approximate procedure.) Thus an approximate 95% CI for θ based on $\bar{X} = 273.1$ is (130.0, 143.7). Similarly, a 95% CI for $\tau = 1/\theta$ is $\{1.9/\bar{X} = 0.00696,\ 2.1/\bar{X} = 0.00769\}$. Note that both CIs are based on MLEs and thus on the sufficient statistic $\bar{X}$. [Even though the MLE of τ is not unbiased, it is *asymptotically* unbiased so $2/\bar{X} = 0.00732$ is not a bad point estimate of τ.]

Note to those studying for future MS exams: In (d) you might want to find the constant c that unbiases the MLE of τ; that is, such that $E(2c/\bar{X}) = \tau$. Then find a UMVUE of τ by using theorems of Rao-Blackwell and Lehmann-Scheffé and the ideas of sufficiency, completeness, and standard exponential families. Also, in (e) note that with statistical software you could find CIs based on the exact distribution of $\bar{X}$ (which is what?) rather than using a normal approximation. Note the similarity of this method to the method used to find a CI for the variance (or standard deviation) of a normal population.

Solution (4) CB

1. a.



$P[Y_1 \le Y_2 + \frac{1}{2}] = 1 - \frac{1}{8}$

unshaded area is $\frac{1}{8}$ of the square.

b. (i) Using (ii) W_2 has density $f(w) = \frac{w^{2-1} e^{-w}}{\Gamma(2)} I_{(0,\infty)}(w)$

or using the fact that $X_k \sim f(x) = e^{-x} I_{(0,\infty)}(x)$

$W_1 = X_1$, $W_2 = X_1 + X_2$ so $X_1 = W_1$, $X_2 = W_2 - W_1$

$J = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$ $\quad f_{W_1,W_2}(w_1, w_2) = e^{-w_1} e^{-(w_2 - w_1)} I_{(0,\infty)}(w_2) I_{(0,w_2)}(w_1)$

$f_{W_2}(w) = \int_0^w e^{-w_1} e^{-(w-w_1)} dw_1 = \int_0^w e^{-w} dw_1 = w e^{-w} I_{(0,\infty)}(w)$

$P[W_2 \le w] = \int_0^w y e^{-y} dy = -ye^{-y}\Big|_0^w + \int_0^w e^{-u} du = -we^{-w} + 1 - e^{-w}$

(ii) $E(t^{W_q}) = [E(t^{X_1})]^q = [1 - \ln(t)]^{-q}$.

$$P[X_1 \le x] = P[-\ln(Y_1) \le x] = P[Y_1 \ge e^{-x}] = 1 - e^{-x} \text{ for } x > 0. \therefore f_{X_1}(x) = e^{-x} I_{(0,\infty)}(x)$$

$$E(t^{X_1}) = \int_0^\infty t^x e^{-x} dx = \int_0^\infty e^{-x[1-\ln(t)]} dx = \frac{1}{1 - \ln(t)}$$

(iii) Let $s = e^t$. Then $-\infty < t \le 0 \iff 0 < s \le 1$.

Hence $E(e^{tW_n}) = E(s^{W_n}) = [1 - \ln(s)]^{-n} = [1-t]^{-n}$ for $n = 1, 2, \ldots$

Hence, $E(e^{tW_N}) = \sum_{n=1}^{\infty} [1-t]^{-n} \theta^{n-1}(1-\theta)$

$$= \frac{1-\theta}{1-t} \sum_{n=1}^{\infty} \left[\frac{\theta}{1-t}\right]^{n-1} = \frac{1-\theta}{1-t}\left[\frac{1}{1-\frac{\theta}{1-t}}\right] = \frac{1-\theta}{1-t-\theta} = m(t)$$

or $E(W_N) = E(X_1)E(N) = \frac{1}{1-\theta}$, $N \sim Ge(1-\theta)$ so $E(N) = \frac{1}{1-\theta}$, $X_1 \sim Exp(1)$ so $E(X_1) = 1$.

$m^{(1)}(t) = \frac{1-\theta}{(1-t-\theta)^2}$ Hence $E(W_N) = m^{(1)}(0) = \frac{1}{1-\theta}$.

$E(W_N) = E(X_1)E(N)$ $\quad N \sim Ge(1-\theta)$

Solution #5 CB

(5) The number of items that arrive at a repair facility by time t is a Poisson process $Y(t)$ for $0 \le t < \infty$. Assume that the items arrive at a rate of λ items per hour.

(a) Let n be a positive integer and suppose that $0 < s < t < \infty$.

(i) If $Y(s) = n$, then what is the expected value of $Y(t)$?

$$E(Y(t)|Y(s)=n) = E[Y(t)-Y(s)+Y(s)|Y(s)=n] = E(Y(t)-Y(s)) + n = \lambda(t-s)+n$$

(ii) If $Y(t) = n$, then what is the variance of $Y(s)$?

(b) Twenty percent of the arriving items require an expensive repair. Assume that the items are independent of each other and that the items are independent of $Y(t)$. For $0 \le t < \infty$, let $X(t)$ be the number of items requiring an expensive repair that arrive by time t.

Let $I_i = \begin{cases} 1 & i\text{th item expensive} \\ 0 & \text{o.w.} \end{cases}$

(i) For $0 \le t < \infty$, what is the expected value of $X(t)$?

$$E(X(t)) = E\left(\sum_{i=1}^{Y(t)} I_i\right) = E(Y(t))\,E(I_1) = \lambda t(.2) = .2\lambda t$$

(ii) For $0 \le t < \infty$, what is variance of $X(t)$?

$$Var(X(t)) = Var\left(\sum_{i=1}^{Y(t)} I_i\right) = [E(I_1)]^2 Var(Y(t)) + E(Y(t))Var(I_1) = (.2)^2\lambda t + \lambda t(.2)(1-.2) = .2\lambda t.$$

(iii) For $0 \le t < \infty$ and $k = 0, 1, 2, \cdots$, what is the probability that $X(t)$ is equal to k?

(a) (ii)
$$P[Y(s)=k|Y(t)=n] = \frac{P[Y(t)-Y(s)=n-k,\ Y(s)=k]}{P[Y(t)=n]}$$
$$= \frac{\dfrac{[\lambda(t-s)]^{n-k}e^{-\lambda(t-s)}}{(n-k)!}\,\dfrac{(\lambda s)^k e^{-\lambda s}}{k!}}{\dfrac{(\lambda t)^n e^{-\lambda t}}{n!}} = \binom{n}{k}\left(\frac{s}{t}\right)^k\left(1-\frac{s}{t}\right)^{n-k}$$
$\therefore Y(s)|Y(t)=n \sim Binomial(n, \frac{s}{t})$ $\therefore Var(Y(s)|Y(t)=n) = n\frac{s}{t}\left(1-\frac{s}{t}\right)$.

(b) (iii)
$$P[X(t)=k, Y(t)=n]$$
$$= P[X(t)=k|Y(t)=n]\,P[Y(t)=n]$$
$$= \binom{n}{k}(.2)^k(1-.2)^{n-k}\frac{(\lambda t)^n e^{-\lambda t}}{n!}$$
$$= \frac{(.2\lambda t)^k e^{-\lambda t}}{k!}\,\frac{((1-.2)\lambda t)^{n-k}}{(n-k)!} \quad \text{for } n = k, k+1, \ldots$$
$$\therefore P[X(t)=k] = \frac{(.2\lambda t)^k e^{-\lambda t}}{k!}\sum_{n=k}^{\infty}\frac{((1-.2)\lambda t)^{n-k}}{(n-k)!}$$
$$= \frac{(.2\lambda t)^k e^{-\lambda t}}{k!}\,e^{(1-.2)\lambda t} = \frac{(.2\lambda t)^k e^{-.2\lambda t}}{k!} \quad \text{for } k=0,1,2,\ldots$$
$\sim Poisson(.2\lambda t)$.

$\therefore$ (i) $E(X(t)) = .2\lambda t$

(ii) $Var(X(t)) = .2\lambda t$