STA 410S/2102S: Homework #2
Due March 1, 2005

Note: The answers to each question should be written or typed as a report to a non-statistician. The numerical results should be summarized in tables, and the text should provide the interpretation. All computations and code should be included as appendices. Code should be commented. Highlighting selected pieces of R text is sometimes helpful, but is **not** a suitable way to answer the question. You are encouraged to discuss the assignments with each other, but please write up your work on your own.

1. The data in Table 1, which is Table T.1 in Cox & Snell's *Applied Statistics*, gives the intervals in service-hours between failures of the air-conditioning eequipment in 10 Boeing 720 jet aircraft. [1] Since the data are failure times, a natural model to entertain is the two-parameter gamma model

$$\frac{1}{\Gamma(\beta)}\left(\frac{\beta}{\mu}\right)^{\beta} y^{\beta-1}e^{-\beta y/\mu}.$$

   The following possible models are under consideration:

   (a) separate gamma distributions fitted to all aircraft, with 20 parameters;

   (b) separate gamma distributions with a common β, with 11 parameters;

   (c) espearate exponential distributions to all aircraft ($\beta = 1$, separate μ), with 10 parameters;

   (d) common exponential distribution to all aircraft ($\beta = 1$), with 1 parameter.

   Use `glm` to find the maximum likelihood estimates under each of these models, and to choose the best-fitting model of these four. Once you have chosen the model, summarize the conclusions in non-technical language.

2. *STA 2102 (bonus for STA410):* Develop some plots that are useful for illustrating the model choice you made in Question 1.

3. The Cauchy distribution is a location scale model with very long tails. The density for a single observation from the Cauchy is

$$f(y;\mu,\sigma) = \frac{1}{\pi\sigma[1 + \{(y-\mu)/\sigma\}^2]}.$$

   (a) Write an R program to compute the maximum likelihood estimate of μ, for the special case where σ is known and equal to 1. The program should take as input a sample of size n, and a starting value for the estimate of μ, and use Newton-Raphson to compute the maximum likelihood estimate. Test your program on several sets of data, and plot the log-likelihood function for each set.

[1] The data is also on the web page and in /u/reid/aircraft.data on CQUEST.

Table 1: Intervals between failures (operating hours)

Aircraft number									
1	2	3	4	5	6	7	8	9	10
413	90	74	55	23	97	50	359	487	102
14	10	57	320	261	51	44	9	18	209
58	60	48	56	87	11	102	12	100	14
37	186	29	104	7	4	72	270	7	57
100	61	502	220	120	141	22	603	98	54
65	49	12	239	14	18	39	3	5	32
9	14	70	47	62	142	3	104	85	67
169	24	21	246	47	68	15	2	91	59
447	56	29	176	225	77	197	438	43	134
184	20	386	182	71	80	188		230	152
36	79	59	33	246	1	79		3	27
201	84	27	15	21	16	88		130	14
118	44	153	104	42	106	46			230
34	59	26	35	20	206	5			66
31	29	326		5	82	5			61
18	118			12	54	36			34
18	25			120	31	22			
67	156			11	216	139			
57	310			3	46	210			
62	76			14	111	97			
7	26			71	39	30			
22	44			11	63	23			
34	23			14	18	13			
	62			11	191	14			
	130			16	18				
	208			90	163				
	70			1	24				
	101			16					
	208			52					
				95					

(b) Run a simulation of size 100 using your program in part (a), and compare the simulation variance of the maximum likelihood estimate with the asymptotic variance using the second derivative of the log-likelihood function.

(c) *STA 2102 (bonus for STA410):* Extend your program to estimate the mean and variance by maximum likelihood, and run a simulation that compares simulation *covariance* of $\hat{\mu}$ and $\hat{\sigma}$ with that obtained from the observed Fisher information matrix.