1. Estimation for the Gamma Distribution
(a.) Problem 32(g), p. 221, Shao.

(b.) Prove that the expression in Shao (p. 170, bottom) for the Fisher information for a reparameterization holds. Then use this expression to find the Fisher information for the gamma distribution with parameters $(\psi_1, \psi_2) = (\log(\alpha), \log(\gamma))$.

(c.) Use the applet on maximum likelihood estimation to verify the theoretical results. For each of sample sizes $n = 10, 50, \text{ and } 200$, find the m.l.e. and record its value. Repeat this 25 times. Find the means, variances, and covariances of the 25 sample estimates. Compare these to the asymptotic values. Discuss your results. Plot histograms and normal probability plots to investigate the shapes of the distributions. Discuss your results.

2. Estimation for the Gumbel Distribution
(a.) The values in the table below were measurements of the breaking strength of glass fibers that were 15 cm long. These data were fit to the Weibull distribution in the article, “A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution,” which appeared in Applied Statistics, 1987, pp. 358-369. The data are available on the 613 web page.

<table>
<thead>
<tr>
<th>0.37</th>
<th>0.40</th>
<th>0.70</th>
<th>0.75</th>
<th>0.80</th>
<th>0.81</th>
<th>0.83</th>
<th>0.86</th>
<th>0.92</th>
<th>0.92</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.94</td>
<td>0.95</td>
<td>0.98</td>
<td>1.03</td>
<td>1.06</td>
<td>1.06</td>
<td>1.08</td>
<td>1.09</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>1.13</td>
<td>1.14</td>
<td>1.15</td>
<td>1.17</td>
<td>1.20</td>
<td>1.20</td>
<td>1.21</td>
<td>1.22</td>
<td>1.25</td>
<td>1.28</td>
</tr>
<tr>
<td>1.28</td>
<td>1.29</td>
<td>1.29</td>
<td>1.30</td>
<td>1.30</td>
<td>1.35</td>
<td>1.35</td>
<td>1.37</td>
<td>1.38</td>
<td>1.40</td>
</tr>
<tr>
<td>1.40</td>
<td>1.42</td>
<td>1.43</td>
<td>1.51</td>
<td>1.53</td>
<td>1.53</td>
<td>1.61</td>
<td>1.61</td>
<td>1.61</td>
<td>1.61</td>
</tr>
</tbody>
</table>

An alternative to the analysis published there would be to model the data using the extreme value distribution with p.d.f.

$$f(x|\alpha, \mu) = \alpha \exp\{-\alpha(x - \mu) - \exp[-\alpha(x - \mu)]\}, \quad -\infty < x < \infty.$$ 

We first need to check goodness of fit of the proposed model to the distribution. To assess the fit visually, one can plot the quantiles in a reasonable way. First obtain the distribution function for $X$ and use it to obtain the quantile function of $X$. Use this to motivate plotting the order statistics $x_{(i)}$ versus $-\log(-\log(F_i))$ where the plotting position $F_i$ is taken to be $(i - .44)/(n + .12)$. 


Next consider the distribution of $Y$ where $X = -Y$ has the extreme value distribution and obtain the quantile function of $Y$. Use this to motivate plotting the order statistics $x_{(i)}$ versus $-\log(-\log(1 - F_i))$ where the plotting position $F_i$ is taken to be $(i - .44)/(n + .12)$.

We can also form a plot to check the assumption of the Weibull distribution with density

$$f(x|\theta, \alpha, \lambda) = \lambda \alpha (x - \theta)^{\alpha - 1} \exp\{-\lambda(x - \theta)\alpha\}, \quad x > \theta,$$

where $\theta, \alpha > 0$, and $\lambda > 0$ are the parameters. Obtain the quantile function for the Weibull distribution where $\theta$ is set equal to $-3.5$, the m.l.e. found in the article. Use this to motivate plotting $\log(x_{(i)} - \theta)$ versus $\log(-\log(1 - F_i))$ where the plotting position $F_i$ is taken to be $(i - .44)/(n + .12)$.

Plot the order statistics for all three distributions and comment on the linearity of the plots. Which distribution appears most reasonable for these data?

(b.) You should have found that the data were fit better by either the distribution of $Y$ or by the Weibull distribution. Let’s examine the distribution of $Y = -X$ more fully. Show that the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\lambda}$ are given by

$$e^{-\hat{\alpha}\hat{\lambda}} = \frac{1}{n} \sum e^{\hat{\alpha}y_i} \quad \text{and} \quad \frac{1}{\hat{\alpha}} = -\bar{y} + \frac{\sum y_i e^{\hat{\alpha}y_i}}{\sum e^{\hat{\alpha}y_i}}.$$

Also, show that in large samples that

$$n \ Var(\hat{\alpha}) = \frac{\alpha^2}{(\pi^2/6)},$$

$$n \ Var(\hat{\lambda}) = \frac{1}{\alpha^2} \left[1 + \frac{(1 - \gamma)^2}{\pi^2/6}\right],$$

$$n \ Cov(\hat{\alpha}, \hat{\lambda}) = -(1 - \gamma)/(\pi^2/6),$$

where $\gamma$ is Euler’s constant $0.5772\ldots$.

(c.) We will proceed to estimate the parameters for the distribution of $Y$ (the negative of the Gumbel (extreme-value) distribution). Fit an appropriate regression line to your plot for the distribution of $Y$ in (a.) to obtain starting values for an iterative procedure. Then use your results in (b.) to compute the maximum likelihood estimate for these data.
3. Problem 114, p. 314, Shao. Answer part (b) for the case when $\theta$ is known.

4. Problem 115, p. 314, Shao, only for the distribution from Problem 40, Chapter 2.


6. Problem 7.30, p. 360, Casella and Berger. Add the following part:

(d) Consider again the breaking strength data in problem 3.2. Suppose that you wish to model the data as coming from a mixture of normal distributions, where one is $N(0.4, 0.02^2)$ and the other is $N(1.2, 0.2^2)$. Use the E-M algorithm to estimate the mixture proportions. Start with $p_0 = 0.5$.

7. Three-parameter Lognormal Distribution. Suppose that $\log(Y_i - \gamma)$ is $N(\mu, \sigma^2)$, where $\gamma < Y_i < \infty$, and the three parameters $(\gamma, \mu, \sigma^2)$ are to be estimated. Defining

$$
\hat{\mu}(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \log(Y_i - \gamma), \quad \hat{\sigma}^2(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \{\log(Y_i - \gamma) - \hat{\mu}(\gamma)\}^2,
$$

show that the profile likelihood, $L^{**}(\gamma) = \sup_{\mu, \sigma^2} L(\gamma, \mu, \sigma^2, y)$ satisfies

$$
L^{**}(\gamma) \propto \{\hat{\sigma}(\gamma)\}^{-n} \prod_{i=1}^{n} (Y_i - \gamma)^{-1}
$$

and that if $y(1)$ is the smallest observed value of $Y$,

$$
\lim_{\gamma \to y(1)} L^{**}(\gamma) = +\infty,
$$

so that the mle of $(\gamma, \mu, \sigma^2)$ for this three-parameter lognormal distribution is always $(Y(1), \infty, +\infty)$. 
