Statistics 613  
T. E. Wehrly  

Homework 3  
Due October 18, 2002

1. Estimation for the Gamma Distribution

(a.) Problem 27(g), p. 185, Shao.

(b.) Prove that the expression in Shao (p. 136, 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.

(d.) For \(n = 50\), record the values of the inverse Hessian matrix. Is the average close to the inverse of the information matrix? How variable are these values?

2. Estimation for the Weibull Distribution

Consider the three-parameter Weibull distribution with p.d.f. given by

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

where \(\theta, \alpha > 0, \text{ and } \lambda > 0\) are the parameters.

(a.) Suppose that \(\theta=0\). Answer Problem 92, p. 273, Shao. Also, find the asymptotic covariance matrix of the mles.

(b.) 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 in my file “twehrly/stuff/break.dat”.

<table>
<thead>
<tr>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
<th>Value 7</th>
<th>Value 8</th>
<th>Value 9</th>
<th>Value 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.37</td>
<td>0.40</td>
<td>0.70</td>
<td>0.75</td>
<td>0.80</td>
<td>0.81</td>
<td>0.83</td>
<td>0.86</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<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>1.35</td>
<td>1.35</td>
<td>1.37</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>1.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We first need to check goodness of fit of the proposed model to the data. To assess the fit visually, one can plot the quantiles in a reasonable way. Obtain the quantile function for the Weibull distribution. 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 \( (i - .44)/(n + .12) \). Set \( \theta \) equal to 0, -3, and -5, and determine whether any of the plots are linear. Which value of \( \theta \) seems most reasonable?

(c.) We will proceed to estimate the parameters for the three-parameter Weibull distribution. You might wish to estimate the parameters \( (\theta, \alpha, \log(\lambda)) \) as did the authors of the paper cited above. Fit an appropriate regression line for the “most linear” plot in (a.) to obtain starting values for an iterative procedure. Then use your results in (b.) to compute (numerically) the maximum likelihood estimate for these data. If you have time, you might also want to try to maximize the modified likelihood of Cheng and Traylor:

\[
M(\theta, \alpha, \lambda) = \log\left\{ F(y_1 + h) - F(y_1) \right\} + \sum_{i=2}^{n} \log f(y_i|\theta, \alpha, \lambda).
\]

3. Problems on \( \rho \) from the bivariate normal distribution.

(a.) Problem 91, p. 273, Shao.

(b.) Problem 114, p. 275, Shao.

4. (Generalized Linear Models) Suppose that \( Y_1, \ldots, Y_n \) are independent random variables where \( Y_i \) has density of the exponential form:

\[
f(y_i; \theta_i) = c(\theta_i) \exp\{y_i b(\theta_i)\} h(y_i).
\]

Let \( \beta = (\beta_1, \ldots, \beta_k)^T \) be a set of parameters and \( x_1, \ldots, x_n \) be explanatory variables. Let \( g \) be the link function where \( g(\mu_i) = x_i^T \beta = \eta_i \) and \( \mu_i = E(Y_i) \).

(a.) Show that the likelihood equations can be expressed as

\[
\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) = 0, \quad j = 1, \ldots, k.
\]

(b.) Show that the information matrix \( I(\beta) \) has entries of the form

\[
I_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\text{Var}(Y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2.
\]
(c.) Problem 4.2, Silvey. On the Aegean island of Kalythos, the inhabitants suffer from a congenital eye disease whose effects become more marked with increasing age. Samples of fifty people were taken at five different ages and the number of blind people counted:

<table>
<thead>
<tr>
<th>Age</th>
<th>20</th>
<th>35</th>
<th>45</th>
<th>55</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of blind</td>
<td>6</td>
<td>17</td>
<td>26</td>
<td>37</td>
<td>44</td>
</tr>
</tbody>
</table>

It is conjectured that the probability of blindness at age \( x \), \( \pi(x) \), can be expressed in the form:

\[
\pi(x) = \left\{ 1 + e^{-(\alpha + \beta x)} \right\}^{-1}.
\]

Comment on whether this hypothesis is reasonable, by constructing a suitable graph. Estimate \( \alpha \) and \( \beta \) from the graph, and then obtain maximum-likelihood estimates. Also, obtain the estimated asymptotic covariance matrix for these estimates. Estimate also the age at which it is just more likely than not that an islander is blind.

5. Problem 100, p. 274, Shao.

6. Problem 7.29, parts a,b,c, p. 360, Casella and Berger.