Estimation and Confidence Intervals

Math 363 - November 29, 2011

1. The independent random variables \( X_1, X_2, \ldots, X_n \) have common cumulative probability distribution function

\[
F_X(x|\alpha, \beta) = \begin{cases} 
0 & \text{if } x < 0, \\
(x/\beta)^\alpha & \text{if } 0 \leq x < \beta, \\
1 & \text{if } x \geq \beta.
\end{cases}
\]

(a) Find the density of these random variables.
(b) The maximum likelihood estimator \( \hat{\beta} \) for \( \beta \) is the maximum value in the data. Find the maximum likelihood estimators for \( \alpha \).
(c) The length (in millimeters) of cuckoos’ eggs found in hedge sparrow nests can be modeled with this distribution. Give the maximum likelihood estimates for the data:

\[
22.0 \quad 23.9 \quad 20.9 \quad 23.8 \quad 25.0 \quad 24.0 \quad 21.7 \quad 23.8 \quad 22.8 \quad 23.1 \quad 23.1 \quad 23.5 \quad 23.0 \quad 23.0
\]

2. Let \( X_1, X_2, \ldots, X_n \) be a random sample with probability density function

\[
f_X(x|\theta) = \theta x^{\theta-1}, \quad 0 \leq x \leq 1, \quad \theta > 0.
\]

(This is a \( \text{Beta}(\theta, 1) \) distribution.)

(a) Find the maximum likelihood estimator for \( \theta \).
(b) Find the method of moments estimator of \( \theta \).
(c) Use simulation to estimate the standard deviation of these estimators for \( \theta = 1/4 \) and \( \theta = 4 \).
(d) (Extra) Find the Fisher information and use the delta method to estimate the standard deviations of the two estimators.

3. The focal length of a lens in your lab has been mislabeled. You decide to make repeated measurements of the distance to the image and the object in order to estimate the focal length. Here \( r_1 \) is the distance from the lens to the object and \( r_2 \) is the distance from the lens to the real image of the object. Here are the measurements:

\[
> r1<-c(9.63, 9.925,10.67,10.30,9.73,9.60,10.49,9.43,9.69,10.20,9.74,9.44,9.33,10.60, + 10.01,9.78,10.39,10.08,10.04,9.55,10.52,10.24,9.47,9.86,9.68) \\
> r2<-c(9.97,9.90,9.95,9.94,10.07,10.14,10.07,10.16,9.88,10.06,10.17,10.13)
\]

(a) Give a 98% confidence interval for \( r_1 \) and for \( r_2 \).
(b) The focal length $f$ is determined by using the thin lens formula,
\[
\frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{f}.
\]
Give an estimate $\hat{f}$ based on these measurements and the thin lens formula.

(c) Use the delta method to give the standard deviation of $\hat{f}$.

(d) Use this to devise a 98% confidence interval for $f$.

4. In this problem, we will examine the sugar content of several national brands of cereals, here measured as a percentage of weight.

<table>
<thead>
<tr>
<th>children</th>
<th>40.3</th>
<th>55.0</th>
<th>45.7</th>
<th>43.3</th>
<th>50.3</th>
<th>45.9</th>
<th>53.5</th>
<th>43.0</th>
<th>44.2</th>
<th>44.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33.6</td>
<td>55.1</td>
<td>48.8</td>
<td>50.4</td>
<td>37.8</td>
<td>60.3</td>
<td>46.6</td>
<td>47.4</td>
<td>44.0</td>
<td></td>
</tr>
<tr>
<td>adult</td>
<td>20.0</td>
<td>30.2</td>
<td>2.2</td>
<td>7.5</td>
<td>4.4</td>
<td>22.2</td>
<td>16.6</td>
<td>14.5</td>
<td>21.4</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.8</td>
<td>10.6</td>
<td>10.6</td>
<td>16.2</td>
<td>14.5</td>
<td>4.1</td>
<td>15.8</td>
<td>4.1</td>
<td>2.4</td>
</tr>
</tbody>
</table>

|          | 3.3  | 10.0 | 1.0  | 4.4  | 1.3  | 8.1  | 6.6  | 7.8  | 10.6 | 10.6 |
|          | 16.2 | 14.5 | 4.1  | 15.8 | 4.1  | 2.4  | 3.5  | 8.5  | 4.7  | 18.4 |

(a) Give a summary of these two data sets.

(b) Create side-by-side boxplots and interpret what you see.

(c) Use \texttt{t.test} to create a 95% confidence interval for the difference in mean sugar content and explain your result.

5. A company with a fleet of 150 cars found that the emission system of 7 out of the 22 cars tested failed to meet pollution guidelines.

(a) Write a hypothesis to test if more than 20% of the entire fleet might be out of compliance.

(b) Test the hypothesis based on the binomial distribution and report a p-value.

(c) Is the test significant at the 10%, 5%, 1% level?

6. National data in the 1960s showed that about 44% of the adult population had never smoked.

(a) State a null and alternative hypothesis to test that the fraction of the 1995 population of adults that had never smoked had increased.

(b) A national random sample of 891 adults were interviewed and 463 stated that they had never smoked. Perform a z-test of the hypothesis and give an appropriate p-value.

(c) Create a 96% confidence interval for the proportion of adults who had never been smokers.

(d) Give the value of the power function $\pi(p)$ for $p = 0.46, 0.48, 0.50, 0.52$ with the choice of $\alpha = 0.04$ and a “greater than” alternative hypothesis.

(e) Compute the power function for these values if we increase the sample to 1600. Explain why these values increased.

7. The body temperature in degrees Fahrenheit of 52 randomly chosen healthy adults is measured with the following summary of the data:

\[
n = 52, \quad \bar{x} = 98.2846.
\]

Assume a standard deviation of $\sigma = 0.68$. 

(a) Are the necessary conditions for constructing a valid $z$-interval satisfied? Explain.

(b) Find a 98% confidence interval for the mean body temperature and explain its meaning.

(c) Give a hypothesis test for a mean body temperature of 98.6° Fahrenheit versus the alternative of 98.2° Fahrenheit. Use the information above to evaluate a test with significance level $\alpha = 0.02$.

(d) Find the power of the test at the parameter value $\mu = 98.2$ and indicate this value using the cutoff value for the test and drawing the sample distribution for the null and alternative hypothesis.