STA 114: Statistics

HW 2

Due Wed Sep 14 2011

1. A supermarket wants to research the popularity of corn based breakfast cereals over those made of wheat. It records the number X of purchases of corn based cereals out of a total n purchases of cereals made from corn or wheat. If a fraction p of all customers of the supermarket prefer corn based cereals, then we can describe this data as $X \sim \mathsf{Binomial}(n, p)$, $p \in [0, 1]$.

   (a) Evaluate (approximately) the ML interval $B_{1.96}(x)$ of p for observed data $x = 58$ with $n = 70$.

   (b) Does the observed data lend support to the theory that corn based cereals are more popular than the ones made from wheat? Explain.

2. Consider modeling annual hurricane counts (in the north Atlantic) $X_1, X_2, \cdots, X_n$ for $n = 10$ years as $X_i \overset{\text{IID}}{\sim} \mathsf{Poisson}(\mu)$.

   (a) Find the expressions for $\hat{\mu}_{\text{MLE}}(x)$ and I_x for data $x = (x_1, \cdots, x_n)$.

   (b) Find an approximate expression for the ML interval $B_c(x)$.

   (c) Evaluate this (approximate) interval for observed data $x = (12, 14, 15, 12, 16, 14, 27, 10, 14, 16)$ and $c = 1.96$.

3. Suppose the time interval (in minutes) between two successive eruptions of a geyser follows an $\mathsf{Exponential}(\lambda)$ distribution. The data available consists of n observed intervals $X_1, X_2, \cdots, X_n$.

   (a) Find the expressions for $\hat{\lambda}_{\text{MLE}}(x)$ and I_x for data $x = (x_1, \cdots, x_n)$.

   (b) Find an approximate expression for the ML interval $B_c(x)$.

   (c) Evaluate this (approximate) interval for observed data $x = (79, 54, 74, 62, 85, 55, 88, 85, 51, 85)$ for $c = 1.96$.

4. Suppose the duration (in seconds) of a smile of a certain eight week old baby follows a $\mathsf{Uniform}(0, \theta)$ distribution with $\theta > 0$ unknown. The data available consists of n observed durations $X_1, X_2, \cdots, X_n$ from the baby.

   (a) For data $X_i = x_i$, $i = 1, 2, \cdots, n$ with each $x_i > 0$, show that the likelihood function is

   $$L_x(\theta) = \begin{cases} \theta^{-n} & \text{for } \theta \geq x_{\max} \\ 0 & \text{for } \theta < x_{\max} \end{cases}$$

   where $x_{\max} = \max(x_1, x_2, \cdots, x_n)$.

   (b) Argue that $\hat{\theta}_{\text{MLE}}(x) = x_{\max}$ (making a plot $L_x(\theta)$ against θ might help).

(c) Find (directly from the expression of $L_x(\theta)$) the expression for the interval $A_k(x) = \{\theta : L_x(\theta) \geq kL_x(\hat{\theta}_{\text{MLE}}(x))\}$ where k is a positive fraction.

(d) Evaluate this interval for observed data $x = (10.4, 19.6, 12.8, 14.8, 1.3, 0.7, 5.8, 6.9, 8.9, 9.4)$ and $k = 0.05$.

(e) Under the model $X_i \sim \mathsf{Uniform}(0, \theta)$, the average duration is $\mu = \theta/2$. What interval would you report for μ?

5. Now consider modeling the smile durations as $X_i \sim \mathsf{Normal}(\mu, \sigma^2)$, $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$. Evaluate the approximate ML interval $B_c(x) \approx \bar{x} \mp cs_x/\sqrt{n}$ for μ with $c = 1.96$ for the observed data in question 4(d).