STA 250: Statistics

HW 7

Due Wed Oct 30 2013

1. Time intervals (in minutes) $X_1, \cdots, X_n$ between successive eruptions of a geyser are modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Exponential}(\lambda)$, $\lambda > 0$. We want to report a plausible range for the median interval time. The median of the $\mathsf{Exponential}(\lambda)$ pdf equals $\log 2/\lambda = 0.693/\lambda$.

   (a) Using the fact that the exponential model is a regular model, give the expression for a (asymptotically) 95% confidence interval for λ.

   (b) Calculate a 95% confidence interval for the median interval time. Justify why the rule you used to construct this interval has 95% confidence.

2. Eruption durations (in minutes) $X_1, \cdots, X_n$ of a geyser are modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Uniform}(0, \theta)$, $\theta > 0$.

   (a) Give a simplified expression for the ML interval $A_b = \{\theta : L_x(\theta) \ge b \cdot L_x(\hat{\theta}_{\text{MLE}}(x))\}$ for a $b \in (0, 1)$.

   (b) What should we fix as b to get a 95% ML confidence interval?

   (c) Evaluate the 95% ML confidence interval for θ with $n = 272$ and $X_{\max} = 5.1$.

3. In 2012, a supermarket sold corn and wheat cereals in a 3:1 ratio. It would like to know whether customer preference of corn cereal to wheat cereal has gone up this year, based on $n = 70$ purchases of the two kinds it has recorded in the first week on January 2013. Letting X denote the number of corn cereal purchases among these n cereal purchases, we could model $X \sim \mathsf{Binomial}(n, p)$, where $p \in [0, 1]$ denotes current customer preference. The research question can be formulated as testing between $H_0 : p \le \frac{3}{4}$ and $H_1 : p > \frac{3}{4}$. If p is ascertained to be larger than 3/4, the supermarket will allocate an additional 50% aisle space to corn cereal, otherwise it will continue with current allocation. If demand remains same or goes down, then allocating extra space will result in some extra cost. On the other hand, if demand goes up but no extra space is allocated, then the supermarket will suffer a large "opportunity cost". Suppose a market analyst asserts the following loss table (in ten thousand dollars):

		Follow-up result	
		$p \le \frac{3}{4}$	$p > \frac{3}{4}$
Decision	Same space allocation	0	5
	50% extra allocation	1	0

(a) The observed data shows $X = 44$. Under the Jeffreys' prior on p, what is posterior probability of $p \leq \frac{3}{4}$?

(b) Given the above loss table, should the supermarket allocate extra space to corn cereal under the Jeffreys' analysis?

(c) Will the decision in part (b) change if we labeled $H_0 : p > \frac{3}{4}$ and $H_1 : p \leq \frac{3}{4}$? Justify and elaborate on how your answer compares with the classical testing approach.

4. From the NSW study we have data $X_1, \cdots, X_n$ on the increase in earning of $n = 185$ individuals who enrolled in the program. Consider the model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ where we fix $\sigma = 9000$ and the only model parameter is $\mu \in (-\infty, \infty)$. We would look at formal Bayes testing $H_0 : \mu = 2000$ against $H_1 : \mu \neq 2000$ with the prior

$$\xi(\mu) = p_0 \cdot \delta_{2000}(\mu) + (1 - p_0) \cdot \xi_1(\mu)$$

where $p_0 = 0.5$ and $\xi_1(\mu) = \mathsf{Normal}(2000, 10000^2)$ [so, under the prior, $\mu = 2000$ with probability 1/2 and is distributed according to $\mathsf{Normal}(2000, 10000^2)$ with the remaining half probability]. The posterior pdf equals

$$\xi(\mu|x) = p_0(x) \cdot \delta_0(\mu) + (1 - p_0(x)) \cdot \xi_1(\mu|x)$$

where

$$\xi_1(\mu|x) = \frac{L_x(\mu)\xi_1(\mu)}{c_1}, \quad \text{with} \quad c_1 = \int L_x(\mu)\xi_1(\mu)d\mu$$

and

$$p_0(x) = \frac{1}{1 + \frac{1-p_0}{p_0}\frac{c_1}{L_x(2000)}}.$$

Observed data has $n = 185$ and $\bar{X} = 4253.57$.

(a) Identify $\xi_1(\mu|x)$ for the observed data (give name and parameters).

(b) Argue that $c_1/L_x(2000) = \xi_1(2000)/\xi_1(2000|x)$.

(c) Calculate $P(H_0|x)$ for the observed data.

5. For the NSW study and model given above, also consider using the simple prior $\xi(\mu) = \mathsf{Normal}(2000, 10000^2)$. What is the Bayes tail area probability for testing $H_0 : \mu = 2000$ against $H_1 : \mu \neq 2000$?