STA 114: Statistics

HW 3

Due Wed Sep 21 2011

1. A machine goes through 4 hazard levels θ, coded 0 through 3 (from low hazard to high hazard) with use over time. The hazard level can be measured by frequency of hazardous incidents X, again coded 0 through 3 (low frequency to high frequency). Suppose X is modeled with pmfs $f(x|\theta)$, $\theta \in \Theta = \{0, 1, 2, 3\}$ as given by the rows of the following table.

θ	$f(0\|\theta)$	$f(1\|\theta)$	$f(2\|\theta)$	$f(3\|\theta)$
0	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
1	0	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
2	0	0	$\frac{2}{3}$	$\frac{1}{3}$
3	0	0	0	1

   (a) Give a simple expression for the MLE $\hat{\theta}_{\text{MLE}}(x)$ in terms of $x \in \{0, 1, 2, 3\}$.

   (b) Consider the ML set $A_{1/2}(x) = \{\theta \in \Theta : L_x(\theta) \geq \frac{1}{2} L_x(\hat{\theta}_{\text{MLE}}(x))\}$. Find $A_{1/2}(x)$ (list all elements of the set) for each of $x = 0, 1, 2, 3$.

   (c) Find the coverage probability $\gamma(A_{1/2}; \theta_0)$ for each of $\theta_0 = 0, 1, 2, 3$.

2. Consider the statistical model $X_i \overset{\text{IID}}{\sim} \mathsf{Logis}(\mu, \sigma)$, $\mu \in (-\infty, \infty)$, σ fixed, where $\mathsf{Logis}(\mu, \sigma)$ denotes the logistic distribution with pdf

$$g(y|\mu, \sigma) = \frac{e^{-(y-\mu)/\sigma}}{\sigma(1 + e^{-(y-\mu)/\sigma})^2}, \quad y \in (-\infty, \infty).$$

   (a) For large n, which one is a better estimator of μ: $\bar{X}$ or X_{med}? Explain. [Hint: If $Y \sim \mathsf{Logis}(\mu, \sigma)$ then $\text{Var}(Y) = \pi^2\sigma^2/3$.]

   (b) Does your answer to part (a) change if σ was not fixed, but included as a model parameter ranging in $(0, \infty)$? Explain.

   (c) For the fixed σ model, give the expression of an approximately 95% confidence interval for μ based on the estimator X_{med} (give precise formulas for the two end-points of the interval).

3. Consider measuring the overall lactic acid concentration of a cheese slab from n randomly sampled pieces of it. The sample measurements are modeled $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\sigma = 1/3$. Here μ is the overall concentration of the slab.

   (a) Suppose the actual overall concentration of the slab is 1. How many samples should one collect so that the estimate $\bar{X}$ of μ is within 0.01 distance from the true value with at least 95% probability? Explain.

(b) Would the answer to part (a) change if the true concentration was 2? Explain.

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. Find a 95% confidence interval for θ based on the observations $(10.4, 19.6, 12.8, 14.8, 1.3, 0.7, 5.8, 6.9, 8.9, 9.4)$.

5. For the opinion poll data $X \sim \mathsf{Binomial}(n, p)$, $p \in [0, 1]$, $n = 500$, a researcher found $B_{1.96} = [0.357, 0.443]$ based on observation $X = 200$. She was interested to know if $p = 0.5$ is tenable. She concludes that "either p is not equal to 0.5 or a coincidence has occurred that does not occur more than once in twenty trials" – which is a valid statement in light of the fact that $\gamma(B_{1.96}, 0.5) \approx 95\%$. Would she be also right to conclude that there is a 19/20 chance that p is not 0.5? Explain.

6. (Bonus question) A pdf $f(x)$ over $(-\infty, \infty)$ is said to be symmetric around a point a if $f(a + h) = f(a - h)$ for every real number h. Prove the following results

   (a) The pdf $f(x)$ of a scalar random variable X is symmetric around a if and only if the pdfs of $Y^+ = X - a$ and $Y^- = a - X$ are identical. [Hint: what's the pdf of Y^+?]

   (b) If the pdf $f(x)$ of a scalar random variable X is symmetric around a then $\mathrm{E}X = a$. [Hint: from part (a), $\mathrm{E}Y^+ = \mathrm{E}Y^-$.]

   (c) Suppose $X_1, \cdots, X_n \overset{\text{IID}}{\sim} f(x)$ where $f(x)$ is symmetric around a. Then the pdf of X_{med} is symmetric around a. [Hint: look at $Y_i^+ = X_i - a$ and $Y_i^- = a - X_i$, $i = 1, \cdots, n$ and the corresponding Y_{med}^+ and Y_{med}^-. From part (a), these two random variables have identical pdfs.]