STA 114: Statistics

Practice Problems

1. The area X (in mm^2) of a microchip made by a machine is distributed according to the pdf

$$f(x|\theta) = \begin{cases} 1 - |x - \theta| & \text{if } x \in [\theta - 1, \theta + 1] \\ 0 & \text{otherwise} \end{cases}$$

where θ denotes the unknown "target value" as determined by the tuning of the machine.

   (a) Based on the observation $X = 3$, make a plot of the likelihood function in θ.

   (b) Does this observation support the theory that the machine is currently tuned for a target area of 1? Justify.

   (c) Identify the ML interval $A_{0.1}(x) = \{\theta : L_x(\theta) \geq 0.1 \times \max_\theta L_x(\theta)\}$ based on the observation $x = 3$.

   (d) What is the confidence coefficient of $A_{0.1}$?

2. Let $X_1, \cdots, X_n$ denote the first serve success rates of a tennis player from n matches. Consider the model $X_i \overset{\text{IID}}{\sim} g(x_i|\theta)$, $\theta \in (0, \infty)$, where the pdf $g(x_i|\theta) = \theta x_i^{\theta - 1}$ for $0 < x_i < 1$ and is zero elsewhere.

   (a) Find the expression for $\hat{\theta}_{\text{MLE}}(x)$ and I_x based on observations $x = (x_1, \cdots, x_n)$.

   (b) Find a ML, asymptotically, 95%-CI for θ based on observations (0.93, 0.42, 0.88, 0.84, 0.82, 0.90, 0.99, 0.95, 0.70, 0.92).

   (c) The quantity of interest is the average success rate $\eta = E_{[X_1|\theta]} X_1 = \frac{\theta}{\theta+1}$. Give a 95%-CI for η and justify why it is a 95%-CI.

3. Let $X_1, X_2, \cdots, X_n$ denote the numbers of revolutions (in millions) until failure of n ball bearings manufactured by a company. Consider the statistical model: $X_i \overset{\text{IID}}{\sim} g(x_i|\mu, \lambda)$, $\mu \in (0, \infty)$ is the model parameter and λ is a fixed positive number; here $g(x_i|\mu, \lambda)$ is the inverse-Gaussian pdf:

$$g(x_i|\lambda) = \begin{cases} \left(\frac{\lambda}{2\pi x_i^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x_i - \mu)^2}{2\mu^2 x_i}\right\} & \text{if } x_i > 0 \\ 0 & \text{if } x_i \leq 0 \end{cases}.$$

   (a) Show that the log-likelihood function based on data $x = (x_1, x_2, \cdots, x_n)$ can be written as

$$\ell_x(\mu) = \text{const.} + \frac{n\lambda}{\mu} - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n} x_i$$

(b) Show that $\hat{\mu}_{\text{MLE}}(x) = \bar{x}$ and $I_x = n\lambda/\bar{x}^3$.

(c) Give the ML, asymptotically 90%, 95% and 99%-CIs for μ (assume the pdf family is sufficiently regular) for observed data x with

$$n = 23, \quad \bar{x} = 72.26, \quad \bar{x}^3 = 377306.1, \quad s_x = 37.49$$

and for $\lambda = 232$.

4. For the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$, is $T_1(x) = x_1$ a better estimator of μ than $T_2(x) = \bar{x}$? Explain.

5. Lactic acid measurements of a set of n cheese samples are modeled as $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$. Suppose, instead of focusing on μ, we are interested in X_{n+1}, the lactic acid concentration of a hypothetical future sample from the same cheese slab. We can describe $X_{n+1} \sim \mathsf{Normal}(\mu, \sigma^2)$, with same (μ, σ^2) that applies to the first n measurements, and, beyond sharing this common pdf, X_{n+1} is independent of $X_1, \cdots, X_n$.

(a) Based on observation $x = (x_1, \cdots, x_n)$ on the n actual measurements, $X = (X_1, \cdots, X_n)$, we can construct a "predictive interval" $A(x)$ for the hypothetical sample X_{n+1} to quantify a range of values we expect X_{n+1} to take (this is of more importance to customers than an interval for μ). Consider one such interval,

$$A_c(x) = \bar{x} \mp c s_x \sqrt{1 + 1/n}$$

Show that for any (μ_0, σ_0^2),

$$P_{[X, X_{n+1} | (\mu_0, \sigma_0^2)]}(X_{n+1} \in A_c(X)) = 2\Phi_{n-1}(c) - 1$$

and argue that $A_{z_{n-1}(\alpha)}$ has a $100(1-\alpha)\%$ guarantee of containing X_{n+1}.

[Hint: rearrange terms to express $X_{n+1} \in A_c(X)$ as $-c \leq \frac{X_{n+1} - \bar{X}}{s_X \sqrt{1+1/n}} \leq c$ and then argue that $T = \frac{X_{n+1} - \bar{X}}{s_X \sqrt{1+1/n}} \sim t(n-1)$ when $X_1, \cdots, X_n, X_{n+1} \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_0, \sigma_0^2)$. Toward this note that $X_{n+1} \sim \mathsf{Normal}(\mu_0, \sigma_0^2)$, $\bar{X} \sim \mathsf{Normal}(\mu_0, \sigma^2/n)$ and $(n-1)s_X^2/\sigma_0^2 \sim \chi^2(n-1)$ and these are all independent of each other; use the definition of the t-distribution.]

(b) Contrast this with the ML, $100(1-\alpha)\%$ -CI for μ given by $B_{z_{n-1}(\alpha)}(x) = \bar{x} \mp z_{n-1}(\alpha) s_x / \sqrt{n}$. The width of $B_{z_{n-1}(\alpha)}(x)$, relative to s_x, collapses to 0 as n becomes large. On the other hand, the width of $A_{z_{n-1}(\alpha)}(x)$, relative to s_x, converges to $2z(\alpha)$. Should this difference concern you? Explain.