STA 114: Midterm I

Total time: 1hr 10min

The **five** questions below carry a total of 43 points. Your exam will be graded out of 40 points – your score is either the points you secure or 40, whichever is less. Answer each question to the best of your ability and show work to guarantee partial/full marks. Make use of the tables and basic probability facts attached at the end. You'd be provided with white papers to write your answers. Please write your name on each sheet of paper and remember to staple them before turning in.

1. For each statement below, state "TRUE" or "FALSE" and give a brief justification. [$4 \times 2 = 8$ points]

   (a) Every ML interval contains the MLE.

   (b) Every ML interval contains the MLE as its mid-point.

   (c) For the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$, the interval $\bar{x} \mp z(\alpha) s_x/\sqrt{n}$ for μ has confidence coefficient at least $1 - \alpha$.

   (d) For the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ^2 fixed, the interval $\bar{x} \mp z_{n-1}(\alpha)\sigma/\sqrt{n}$ has confidence coefficient at least $1 - \alpha$.

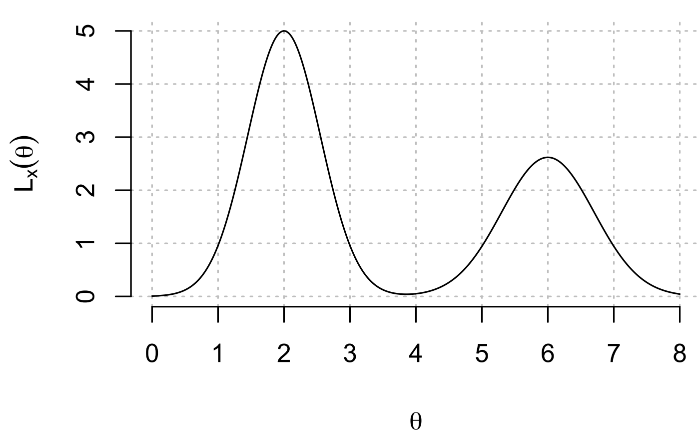
2. For some statistical model $X \sim f(x|\theta)$, $\theta \in [0, 8]$, the likelihood function $L_x(\theta)$, based on some observation x looks like the curve in the following figure. Identify the correct answer for each of the following questions [$2 \times 2 = 4$ points]

   (a) The MLE $\hat{\theta}_{\text{MLE}}(x)$ is

   (I) 2 (II) 6 (III) BOTH 2 AND 6 (IV) UNDEFINED.

   (b) The ML set $A_{0.2}(x) = \{\theta : L_x(\theta) \geq 0.2 \times \max_\theta L_x(\theta)\}$ is (roughly)

   (I) [1, 3] (II) [5, 7] (III) [1, 7] (IV) [1,3] $\cup$ [5,7]



3. A count data X is modeled as $X \sim f(x|\theta)$, $\theta \in \{1/4, 1/2, 3/4\}$, where the pmfs $f(x|\theta)$, defined over $x \in \{0, 1, 2, 3\}$ (i.e., $f(x|\theta) = 0$ for any other x) are as in the table below. Answer the following questions. [$5 + 4 = 9$ points]

θ	$f(0\|\theta)$	$f(1\|\theta)$	$f(2\|\theta)$	$f(3\|\theta)$
$\frac{1}{4}$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$\frac{1}{2}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$\frac{3}{4}$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

(a) Describe $A_{1/2}(x) = \{\theta : L_x(\theta) \geq \frac{1}{2} \times \max_\theta L_x(\theta)\}$ for each of $x = 0, 1, 2, 3$ (list the elements of $A_{1/2}(x)$ in each case).

(b) Calculate the coverage $\gamma(A_{1/2}; \theta_0)$ for each of $\theta_0 = 1/4, 1/2, 3/4$.

4. A code breaking algorithm, when applied to a certain class of cryptic codes, has an unknown failure probability $p \in (0, 1)$. The algorithm is tried on randomly chosen codes until it fails, and the number of trials until failure (counting the failed trial) is recorded as X_1. This process is then repeated a further $n - 1$ times, giving us trial lengths $X_1, X_2, \cdots, X_n$, modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Geometric}(p)$, $p \in (0, 1)$, where $\mathsf{Geometric}(p)$ has pmf $g(y|p) = (1-p)^{y-1}p$, $y = 1, 2, \cdots$ and $g(y|p) = 0$ otherwise. Answer the following questions. [$3 + 4 + 4 = 11$ points]

(a) Show that the log-likelihood function $\ell_x(p)$, based on observation $x = (x_1, \cdots, x_n)$ can be written as $\ell_x(p) = n(\bar{x} - 1)\log(1-p) + n\log p$.

(b) Show that $\hat{p}_{\text{MLE}}(x) = 1/\bar{x}$ and $I_x = n\bar{x}^3/(\bar{x} - 1)$.

(c) Give ML, asymptotically 90%, 95% and 99%-CIs for p based on observation x for which:

$$n = 10, \quad \bar{x} = 67.2, \quad x_{\text{med}} = 35.5, \quad s_x = 77.85$$

5. I am to measure the lactic acid concentrations of two batches of samples from a cheese slab. Let the measurements from the first batch be $X_1, \cdots, X_n$ and those from the second batch be $Y_1, \cdots, Y_m$. We model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $Y_j \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, X_i's and Y_j's are independent, with model parameter $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$. Answer the following questions. [$6 + 2 + 3 = 11$ points]

(a) Show that the interval $B(x, y) = \bar{y} \mp z_{n-1}(0.05)s_x/\sqrt{m}$, based on observations $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_m)$, is a 95%-CI for μ.

(b) Suppose I collect the second batch after recording the observations in the first batch. I find $s_x = 0.31$ with $n = 10$. What is the minimum number of observations m I need to collect in the second batch so that my reported interval $B(x, y)$ has total width less than 0.2?

(c) Does $B(x, y)$ remain a 95%-CI for μ if my choice of m depends on x? Explain.