STA 250 (Fall 13): Midterm I

Total time: 1hr 10min

The **three** questions below carry a total of 30 points. Attempt all question parts and show work to guarantee partial/full credit. Make use of the tables and basic probability facts attached at the end. **No other cheat sheets allowed**. You will be provided with sheets of white paper to write your answers on. Please remember to staple them before turning in and write your name on the top. You should be able to attempt any part of any question whether or not you have attempted the parts before it.

Tip for quick table look-up: $\boxed{z_k(2\alpha) = c \iff \alpha = 1 - \Phi_k(c)}$.

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	$\frac{11}{12}$	$\frac{1}{12}$
3	0	0	0	1

For testing $H_0 : \theta \in \{0, 1\}$ vs $H_1 : \theta \in \{2, 3\}$ answer the following. [$5 + 4 = 9$ points]

   (a) Argue that the testing rule "reject H_0 if $X = 3$" is an ML test of size 1/6.

   (b) Now consider the testing rule where one rolls a 6 sided fair die and rejects H_0 if the die shows '5' [so essentially, the information in data X is completely disregarded]. This test too has size 1/6. Is the ML test in part (a) better than this die-rolling test rule? Why or why not?

2. Consider data consisting of $n = 15$ observations $X_1, X_2, \cdots, X_n$. Below are three possible statistical models for this data and a pair of hypotheses for each model. For each model-hypotheses set below indicate whether or not to reject H_0 according to the size-5% ML test if we observe $\bar{X} = 3.33$, $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 55.73$. [$2 \times 4 = 8$ points]

   (a) *Model.* $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ with $-\infty < \mu < \infty$ unknown and $\sigma^2 = 2.5$.
   Hypotheses. $H_0 : \mu \leq 2.5$ against $H_1 : \mu > 2.5$.

   (b) *Model.* $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ with both $-\infty < \mu < \infty$ and $\sigma^2 > 0$ unknown.
   Hypotheses. $H_0 : \mu \leq 2.5$ against $H_1 : \mu > 2.5$.

3. Let $X = (X_1, \cdots, X_n)$ denote the first serve success rates of a tennis player from n tournament 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. Answer the following questions and show your work. [$6 + 4 + 3 = 13$ points]

   (a) Give a simplified expression for the log-likelihood function and argue that

$$\hat{\theta}_{\text{MLE}}(x) = -n / \sum_{i=1}^{n} \log x_i, \quad I_{\text{OBS}}(x) = \frac{(\sum_{i=1}^{n} \log x_i)^2}{n}$$

   (b) What is the p-value of $H_0 : \theta \leq 4$ against $H_1 : \theta > 4$ based on ML tests when observed data shows the following?

$$n = 20, \quad \bar{X} = 0.802, \quad s_X = 0.106, \quad \sum_{i=1}^{n} \log X_i = -4.59, \quad X_{\text{med}} = 0.805.$$

   You may assume that the pdf family is sufficiently smooth.

   (c) The quantity of interest is the average success rate $\eta = E_{[X_1|\theta]} X_1 = \frac{\theta}{\theta+1}$. Could you calculate the ML test based p-value of $H_0 : \eta \leq 80\%$ against $H_1 : \eta > 80\%$ from the p-value you reported in part (b)? If you answer "yes", justify why and please specify what the p-value is. If you answer "no", justify why and indicate how one can proceed to calculate the new p-value.