STA 250 (Spring 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. Let $X = (X_1, \cdots, X_n)$ denote log-intensities of all $n = 291$ tropical cyclones from the north Atlantic basin between 1980 and 2006. Log-intensity of a storm system is defined as the logarithm of its maximum sustained windspeed (measured in knots). A tropical cyclone is labeled a category 5 hurricane if its maximum sustained windspeed is 137 knots or more, i.e., if its log-intensity is 4.92 or more. We are interested in the probability $p \in [0, 1]$ of a tropical cyclone attaining category 5 strength. Consider two possible models for the data and answer the related questions. [$6 + 4 = 10$ points]

   (a) Consider the model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ where it is known $\sigma = 0.33$ and $\mu \in (-\infty, \infty)$ is an unknown parameter. Under this model $p = \Phi(\{\mu - 4.92\}/\sigma)$, where $\Phi(x)$ denotes the standard normal CDF. Calculate the p-value of $H_0 : p \le 1\%$ against $H_1 : p > 1\%$ based on ML tests when observed data shows

   $$\bar{X} = 4.20, \quad \sqrt{n} = 17.06, \quad \#\{X_i \ge 4.92\} = 11.$$

   [Hint: $\Phi(z)$ is monotone increasing in z and $\Phi^{-1}(1\%) = -2.33$.]

   (b) Let $Y = \#\{X_i \ge 4.92\}$ denote the number of tropical cyclones that attain category 5 strength. We could directly model $Y \sim \mathsf{Binomial}(n, p)$, $p \in [0, 1]$. Find the p-value of $H_0 : p \le 1\%$ against $H_1 : p > 1\%$ based on ML tests for this model with data summaries as given above.

2. A count data X is modeled as $X \sim f(x|\theta)$, $\theta \in \{0, 1\}$, where the pmfs $f(x|\theta)$, defined over $x \in \{1, 2, 3\}$ (i.e., $f(x|\theta) = 0$ for any other x) are as in the table below.

θ	$f(1\|\theta)$	$f(2\|\theta)$	$f(3\|\theta)$
0	0.9	0.05	0.05
1	0.09	0.055	0.855

   For testing $H_0 : \theta = 0$ vs $H_1 : \theta = 1$ answer the following. [$5 + 3 = 8$ points]

   (a) Argue that the 5% ML test is given by "reject H_0 if $x = 3$".

   (b) Describe one more test that is also of size 5%. Would you prefer using this test over the 5% ML test of part (a)? Justify your answers.

3. 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 $n-1$ IID times, giving us data on n trial lengths $X = (X_1, \cdots, X_n)$, modeled as $X_i \sim \mathsf{Geometric}(p)$, $p \in (0,1)$, 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 and show your work. [$4+4+4 = 12$ points]

   (a) Show that the log-likelihood function in p can be written as

   $$\ell_x(p) = n(\bar{x} - 1)\log(1-p) + n \log p$$

   (b) For recorded data with $n = 10, \bar{X} = 67.2, X_{\text{med}} = 35.5$ and $s_X = 77.85$, show that $\hat{p}_{\text{MLE}} = 0.015$.

   (c) The geometric model is a regular model and it turns out that $I_{\text{OBS}}(x) = n\bar{x}^3/(\bar{x} - 1)$. For the data given in part (b) would the size 5% ML test reject $H_0 : p \leq 1\%$ against $H_1 : p > 1\%$? Justify.