NTHU MATH 2820 Final Examination Jun 17, 2008

1. (*10pts, 2pts for each*) For the following statements, please answer true or false. **If false, please explain why.**

   (a) Suppose that a 99% confidence interval for the mean μ of a normal distribution is found to be $(-2.0, 3.0)$. The probability that $(-2.0, 3.0)$ contains μ equals 0.99.

   (b) Suppose that a 99% confidence interval for the mean μ of a normal distribution is found to be $(-2.0, 3.0)$. A corresponding test of $H_0 : \mu = -3$ versus $H_A : \mu \neq -3$ would be accepted at the 0.01 significance level.

   (c) The probability that the null hypothesis is true equals the significance level α.

   (d) When the consequences of making a type I error are very serious, we should set the significant level α to a smaller value.

   (e) The p-value of a test is a statistic, i.e., a function of data only.

2. (*15pts, 3pts for each*) For each of the random observations X below, determine a joint distribution which best models X. Then, formulate the problem of interest into a hypotheses testing problem by specifying for the distribution of X the space of all possible parameters (i.e., Ω), and the space of parameters under null hypothesis (i.e., Ω_0).

   (a) It has been claimed that dying people may be able to postpone their death until after an important occasion, such as a wedding or birthday. A study was conducted for the patterns of death surrounding Passover, an important Jewish holiday, in California during the years 1966-1984. They compared the number of deaths during the week before Passover to the number of deaths during the week after Passover for 1919 people who had Jewish surnames. Of these, X occurred in the week before Passover and $1919 - X$, in the week after Passover. Use the data X to examine the claim.

   (b) A manufacturer of an integrated circuit claims that the mean time to failure of their products is over 2 years. Let $X = (X_1, \ldots, X_{20})$ be the observed lifetimes (**in days**) of 20 products. Use the data X to examine the claim.

   (c) A survey of voter sentiment was conducted in four midcity political wards to compare the fraction of voters favoring cindidate A. Random samples of 200 voters were polled in each of the four wards. Let $X = (X_1, X_2, X_3, X_4)$ be the numbers of voters favoring A in the four samples. It can be assumed that X_1, X_2, X_3, X_4 are independent. Use the data X to examine the claim that the fractions of voters favoring candidate A are the same in all four wards.

   (d) A plant manager, in deciding whether to purchase a machine of design A or design B, checks the times for completing a certain task on each machine. Eight technicians were used in the experiment, with each technician using both machine A and machine B. Let Y_i and Z_i, $i = 1, \ldots, 8$, be the times (in seconds) that the ith technician took to complete the task using machines A and machine B, respectively. Let $X = (X_1, \ldots, X_8)$, where

$$X_i = Y_i - Z_i, \quad i = 1, \ldots, 8.$$

Use the data X to examine whether there exists a difference between the mean completion times of the two machines.

(e) Nylon bars were tested for brittleness. Each of 280 bars was molded under similar conditions and was tested in five places. Assuming that each bar has uniform composition, the number of breaks on a given bar shuold be binomially distributed with five trials and an unknown probability of failure p. If the bars are all of the same uniform strength, p should be the same for all of them; if they are of different strengths, p should vary from bar to bar. The following table summaries the outcome of the experiment:

Breaks/Bar	0	1	2	3	4	5
Frequency	X_0	X_1	X_2	X_3	X_4	X_5

where $X_0 + X_1 + \cdots + X_5 = 280$. Use the data $X = (X_0, \ldots, X_5)$ to examine the claim that the bars are all of the same uniform strength.

3. Let $X_1, \ldots, X_n$ be i.i.d. from a Poisson distribution $P(\lambda)$, whose pdf is

$$\frac{e^{-\lambda}\lambda^x}{x!}, \text{ where } x = 0, 1, 2, \ldots, \text{ and } \lambda > 0.$$

(a) (*15 pts*) Use Neyman-Pearson lemma to find the likelihood ratio for testing

$$H_0 : \lambda = \lambda_0 \text{ versus } H_A : \lambda = \lambda_1,$$

where $\lambda_1 > \lambda_0$. Use the fact that the sum of independent Poisson random variables follows a Poisson distribution, i.e.,

$$\sum_{i=1}^{n} X_i \sim P(n\lambda),$$

to explain how to determine a rejection region for a *randomized* test at level α.

(b) (*5 pts*) Show that the test in (a) is uniformly most powerful (UMP) for testing

$$H_0 : \lambda = \lambda_0 \text{ versus } H_A : \lambda > \lambda_0.$$

4. Suppose that $X_1, \ldots, X_n$ are i.i.d. from Uniform distribution $U(0, \theta)$, where $\theta \in \Omega = (0, \infty)$. Let

$$X_{(n)} = \max\{X_1, \ldots, X_n\},$$

then the joint pdf of $X_1, \ldots, X_n$ can be written as:

$$f(x_1, \ldots, x_n|\theta) = \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}), \tag{1}$$

where $x_{(n)} = \max\{x_1, \ldots, x_n\}$ and I is the indicator function, i.e.,

$$I_{[0,\theta]}(t) = \begin{cases} 1, & \text{if } t \in [0, \theta], \\ 0, & \text{otherwise.} \end{cases}$$

(a) (*7pts*) For a fixed $\theta_0 \in (0, \infty)$, use (1) to show that the generalized likelihood ratio (GLR) Λ for testing

$$H_0 : \theta = \theta_0 \text{ versus } H_A : \theta \neq \theta_0$$

is

$$\Lambda(x_1, \dots, x_n) = \begin{cases} \left[\frac{x_{(n)}}{\theta_0}\right]^n, & \text{if } 0 \leq x_{(n)} \leq \theta_0, \\ 0, & \text{if } x_{(n)} > \theta_0. \end{cases} \tag{2}$$

[**Hint:** (i) Under Ω, the MLE of θ is $\hat{\theta} = X_{(n)}$. (ii) $I_{[0,x_{(n)}]}(x_{(n)}) = 1$.]

(b) (*4pts*) Use (2) to show that

$$\Lambda(X_1, \dots, X_n) < s,$$

where $0 < s < 1$, is equivalent to

$$X_{(n)} > \theta_0 \text{ or } X_{(n)} < s^{1/n}\theta_0.$$

[**Hint**: draw the graph of Λ as a function of $x_{(n)}$ to help you identify the region.]

(c) (*9pts*) Use (b) to derive the rejection region of the GLR test at the significance level $\alpha = 0.05$. **Please express the rejection region in terms of $X_{(n)}$ and find the value of s for $\alpha = 0.05$.**
[**Hint:** Under H_0, the pdf of $X_{(n)}$ is

$$f(x) = \begin{cases} \frac{nx^{n-1}}{\theta_0^n}, & \text{for } 0 \leq x \leq \theta_0, \\ 0, & \text{otherwise.} \end{cases} \quad]$$

(d) (*4pts*) Use the acceptance region of the GLR test in (c) to construct a 95% confidence interval for θ.

5. Let $X_i \sim B(n_i, p_i)$, $i = 1, \dots, m$, be independent random variables, then the joint pmf of $X_1, \dots, X_m$ is:

$$f(x_1, \dots, x_m | p_1, \dots, p_m) = \prod_{i=1}^{m} \binom{n_i}{x_i} p_i^{x_i}(1 - p_i)^{n_i - x_i}.$$

Suppose that we are interested in testing the hypotheses

$$H_0 : (p_1, p_2, \dots, p_m) \in \Omega_0 \quad \text{v.s.} \quad H_A : (p_1, p_2, \dots, p_m) \in \Omega \backslash \Omega_0,$$

where

$$\Omega_0 = \{(p_1, \dots, p_m) |\ 0 < p_1 = \cdots = p_m < 1\}$$

and

$$\Omega = \{(p_1, \dots, p_m) |\ 0 < p_1 < 1, 0 < p_2 < 1, \dots, 0 < p_m < 1\}.$$

(a) (*2pts*) What are the dimensions of Ω_0 and Ω?

(b) (*8pts*) Denote $O_{i1} = n_i\hat{p}_i$, $O_{i2} = n_i(1-\hat{p}_i)$, $E_{i1} = n_i\hat{p}$, and $E_{i2} = n_i(1-\hat{p})$, where $\hat{p}_i$'s and $\hat{p}$ are defined in **Hint** below. Show that the test statistic $-2\log\Lambda$, where Λ is the GLR, is:

$$2\sum_{i=1}^{m}\sum_{j=1}^{2} O_{ij}\log\left(\frac{O_{ij}}{E_{ij}}\right).$$

[**Hint:** (i) Under Ω_0, the MLE is

$$\hat{p} = \frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m} n_i}.$$

(ii) Under Ω, the MLE of p_i is

$$\hat{p}_i = \frac{X_i}{n_i}, \quad i = 1, \ldots, m.$$]

(c) (*2pts*) Under H_0, what is the large sample distribution of the $-2\log\Lambda$ in (b)?

(d) (*4pts*) Suppose that the sample sizes n_i's are large enough. Use the results in (b) and (c) to find the rejection region of the GLR test at significance level α.

6. Let $X_1, \ldots, X_n$ be an i.i.d. sample from Poisson $P(\theta)$. The pdf of $P(\theta)$ is:

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}.$$

Let θ have a Gamma prior distribution $\Gamma(\alpha, \lambda)$, whose pdf is:

$$g(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta},$$

where α and λ are assumed known.

(a) (*5pts*) Show that the posterior distribution of θ is Gamma $\Gamma(n\overline{X} + \alpha, n + \lambda)$, where $\overline{X} = \sum_{i=1}^{n} X_i/n$.

(b) (*2pts*) Is Gamma a conjugate prior of the Poisson (Yes or No)? Please explain why.

(c) (*5pts*) Determine the Bayes estimator of θ under squared error loss, and show that it is a weighted average of the prior mean and $\overline{X}$.
[**Hint:** Let $Y \sim \Gamma(a, b)$, then $E(Y) = \frac{a}{b}$.]

(d) (*3pts*) Use (c) to show that when the sample size n is large, the information from the sample will dominate the estimation.
[**Note.** $\overline{X}$ is the MLE under Frequentist approach.]