The Hong Kong University of Science & Technology

MATH243 – Statistical Inference

Final Examination – Fall 08/09

Answer <u>ALL</u> Questions | **Date: 12 December 2008 (Fri)**

Time allowed: 3 hours

---

1. Assume that we have an i.i.d. random sample $X_1, X_2, ..., X_n$ from a Poisson distribution with mean parameter μ.

   (a) **(1 mark)** Find the maximum likelihood estimator of μ^2.

   (b) **(2 marks)** Is the maximum likelihood estimator of μ^2 unbiased?

   (c) **(3 marks)** Compute the mean square error of the maximum likelihood estimator in (a). Hint: Find moments from the moment generating function of a Poisson random variable.

   (d) **(2 marks)** Find CRLB of all unbiased estimators for μ^2.

   (e) **(2 marks)** Find the UMVUE of μ^2. Is its variance equal to CRLB in (d)? Explain. No need to find the variance of the UMVUE of μ^2.

2. Suppose X_1, X_2, …., X_n a random sample with Bernoulli (θ), i.e., they are independently and identically distributed and

$$X_i = \begin{cases} 1 & \text{with probability} \quad \theta \\ 0 & \text{with probability} \quad 1-\theta \end{cases}$$

   (a) **(2 marks)** Show that $X_1 - X_2$ is <u>not</u> a complete statistic.

   (b) **(2 marks)** Find the complete and sufficient statistic for θ. What is its distribution?

   (c) **(1 mark)** Show that X_1 is an unbiased estimator of θ.

   (d) **(2 marks)** Rely on the Rao-Blackwell theorem to find a better unbiased estimator of θ than the one considered in (c).

   (e) **(3 marks)** Find the UMVUE for θ^m, where *m* is a positive integer less than or equal to *n*.

3. Let $X_1, ..., X_n$ be an i.i.d. random sample from a location distribution family

$$f(x;\theta) = \exp(-(x-\theta))I(x \geq \theta)$$

   with $\theta \in R$.

   (a) **(1 mark)** Find the minimum sufficient statistic of the unknown parameter θ.

   (b) **(2 marks)** Find the distribution of the minimum sufficient statistic.

   (c) **(2 marks)** Show that the minimum sufficient statistic is complete.

   (d) **(2 marks)** Compute mean and variance of the minimum sufficient statistic.

   (e) **(1 mark)** Find the UMVUE of θ.

   (f) **(2 marks)** Find the UMVUE of θ^2.

4. (a) Let X_1, …, X_n be i.i.d. random variables, each with the Poison distribution of parameter μ.

(i) **(2 marks)** Show that, by the Neyman-Pearson Theorem, the best test of $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 (> \mu_0)$ is $C_1 = \left\{ \underset{\sim}{x} : \sum_{i=1}^{n} x_i \geq k \right\}$.

(ii) **(3 marks)** By using the central limit theorem to approximate the distribution of $\sum_{i=1}^{n} x_i$, find the smallest value of n required to obtain power at least 0.9 against the alternative $\mu_1 = 1.21$ when $\mu_0 = 1$ and $\alpha = 0.05$.

(b) Let $X_1, ..., X_n$ be i.i.d. r.v.'s from $N(\mu, \theta)$ with μ known.

(i) **(3 marks)** Construct the UMP test for $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ at level of significance α.

(ii) **(2 marks)** Calculate the power at $\theta = 12$ when $\theta_0 = 4$, $\alpha = 0.05$ and $n = 25$.

5. Let $\underset{\sim}{X} = (X_1, ..., X_n)$ where X_i denoted as the number of occurrences has multinomial distribution with parameters n, θ_1, …, θ_m. Test $H_0 : \theta_1 = \theta_{01}, \theta_2 = \theta_{02}$, …, $\theta_m = \theta_{0m}$ against $H_1 : (\theta_1, \theta_2, ..., \theta_m)$ takes any other value in $[0,1]^m$ and $\sum_{i=1}^{m} \theta_i = 1$.

(a) **(5 marks)** Find the likelihood ratio and then derive the approximate large sample likelihood ratio test.

(b) The following 3839 seedlings are progeny of self-fertilized heterozygotes. Each seedling can be classified as either "Starchy" or "Sugary" and either "Green" or "White":

Numbers of Seedlings	Green	White	Total
Starchy	1997	906	2903
Sugary	904	32	936
Total	2901	938	3839

(i) **(3 marks)** Use the result in (a), test the ratios of number of seedlings ("starchy" and "green"; "starchy" and "white"; "sugary" and "green"; "sugary" and "white") are 9:3:3:1 at $\alpha = 0.05$.

(ii) **(2 marks)** Re-do the analysis by Pearson goodness of test. Do you get the same conclusion?

6. Assume that we have two independent samples, $Y_{11}, \ldots, Y_{1n}$ and $Y_{21}, \ldots, Y_{2n}$, such that Y_{1i} has $N(\gamma_{10}+\gamma_{11}(x_{1i}-\bar{x}_1),\sigma^2)$ distribution and Y_{2i} has $N(\gamma_{20}+\gamma_{21}(x_{2i}-\bar{x}_2),\sigma^2)$ distribution, respectively, where x_{1i} and x_{2i} are fixed values. The data analyst has to decide whether the coefficients of x_{1i} and x_{2i} are equal, i.e., $H_0:\gamma_{11}=\gamma_{21}=\gamma$.

(a) **(3 marks)** Find the maximum likelihood estimators for unknown parameters under the alternative hypothesis, i.e., γ_{10}, γ_{11}, γ_{20}, γ_{21} and σ^2. Write down maximum likelihood estimators of γ_{10}, γ_{11}, γ_{20} and γ_{21} in terms of $\bar{y}_i$, $\bar{x}_i$, $S_{y_ix_i}$ and $S_{x_ix_i}$; maximum likelihood estimator of σ^2 in terms of $S_{y_iy_i}$, $S_{y_ix_i}$, $S_{x_ix_i}$ and/or maximum likelihood estimators of γ_{11} and γ_{21}, where $\bar{y}_i=\sum_{j=1}^{n} y_{ij}/n$, $\bar{x}_i=\sum_{j=1}^{n} x_{ij}/n$, $S_{y_iy_i}=\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$, $S_{y_ix_i}=\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(y_{ij}-\bar{y}_i)$ and $S_{x_ix_i}=\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$ for i = 1, 2.

(b) **(3 marks)** Find the maximum likelihood estimators for unknown parameters under the null hypothesis, i.e., γ_{10}, γ_{20}, γ and σ^2. Write down maximum likelihood estimators of γ_{10}, γ_{20}, γ in terms of $\bar{y}_i$, $\bar{x}_i$, $S_{y_ix_i}$ and $S_{x_ix_i}$; maximum likelihood estimator of σ^2 in terms of $S_{y_iy_i}$, $S_{y_ix_i}$, $S_{x_ix_i}$ and/or maximum likelihood estimator of γ, where $\bar{y}_i=\sum_{j=1}^{n} y_{ij}/n$, $\bar{x}_i=\sum_{j=1}^{n} x_{ij}/n$, $S_{y_iy_i}=\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2$, $S_{y_ix_i}=\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(y_{ij}-\bar{y}_i)$ and $S_{x_ix_i}=\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$ for i = 1, 2.

(c) **(1 mark)** Construct the likelihood ratio for testing $H_0:\gamma_{11}=\gamma_{21}=\gamma$ against $H_1:\gamma_{11}\neq\gamma_{21}$.

(d) **(3 marks)** Derive the approximate large sample likelihood ratio test. Perform the hypothesis testing for the data below and make your conclusion.

(e) **(Bonus: 4 marks)** Construct the exact likelihood ratio test with level of significance α. Do you get the same conclusion on the data as part (d)?

Data: n = 10

$\bar{x}_1$ = 9.3,	$\bar{y}_1$ = 5.3,	$S_{x_1x_1}$ = 204.1,	$S_{x_1y_1}$ = 152.1,	$S_{y_1y_1}$ = 194.1
$\bar{x}_2$ = 12.9,	$\bar{y}_2$ = 12.3,	$S_{x_2x_2}$ = 140.9,	$S_{x_2y_2}$ = 168.3,	$S_{y_2y_2}$ = 460.1

****** **END** ******