University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Homework 11

EXERCISE 1
Answer the following questions:

a. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter λ. Find the best critical region for testing
$H_0: \lambda = 2$
$H_a: \lambda = 5$
using the Neyman-Pearson lemma.

b. Let $Y_1, Y_2, \ldots, Y_n$ be the outcomes of n independent Bernoulli trials. Find the best critical region for testing
$H_0: p = p_0$
$H_a: p > p_0$
using the Neyman-Pearson lemma.

EXERCISE 2
Let $X_1, X_2, \ldots, X_n$ be a random sample from normal distribution with $E(X_i) = \mu, var(X_i) = \sigma^2$, and $cov(X_i, X_j) = \rho\sigma^2$, for $i \neq j$. Then $\bar{X} \sim N(\mu, \sigma\sqrt{\frac{1+(n-1)\rho}{n}})$, and a 95% confidence interval for μ will be

$$\bar{x} \pm 1.96\sigma\sqrt{\frac{1+(n-1)\rho}{n}}. \tag{1}$$

Note: If $\rho = 0$ then we get the usual confidence interval for μ when we are dealing with i.i.d. random variables: $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ (2). Now, suppose we fail to see the dependence in our random variables, and instead of using (1) we decided to use (2). What is the actual coverage of our confidence interval if $n = 25, \sigma = 3, \rho = 0.2$?
.

EXERCISE 3
Consider the simple regression model through the origin, $y_i = \beta_1 x_i + \epsilon_i$, with $\epsilon_i \sim N(0, \sigma)$, $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ independent, and σ^2 is known. Use the Neyman-Pearson lemma to find the most powerful test for testing
$H_0: \beta_1 = 0$
$H_a: \beta_1 > 0$
at the level of significance α (for a fixed value of β_1 under H_a).

EXERCISE 4
Consider the regression model $y_i = \frac{\theta}{2}x_i^2 + \epsilon_i$, for $i = 1, \ldots, n$, with $E(\epsilon_i) = 0$, $var(\epsilon_i) = \sigma^2$, $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ because ϵ_i, ϵ_j are independent, and $\epsilon_i \sim N(0, \sigma)$. Show that the maximum likelihood estimate of θ is $\hat{\theta} = \frac{2\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}$. Find the distribution of $\hat{\theta}$. Suppose σ^2 is known. Using the distribution of $\hat{\theta}$ find $1 - \alpha$ confidence interval for θ.

Feng Shiwei UID: 305256428

Ex 3. $H_0: \beta_1 = 0$

$H_a: \beta_1 = \beta > 0$ $\quad y_i \sim N(\beta_1 x_i, \sigma)$

$$\frac{L(\beta_1=0)}{L(\beta_1=\beta)} = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}\Sigma(y_i-0)^2}}{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\Sigma(y_i-\beta x_i)^2}} < k$$

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i-\beta_1 x_i)^2 - y_i^2\right] < \ln k$$

$$\Sigma x_i y_i > \frac{\beta^2 \Sigma x_i^2 - 2\sigma^2 \ln k}{2\beta}$$

Note that $\Sigma x_i y_i \sim N(\underset{0}{\underset{\parallel}{\beta_1}} \Sigma x_i^2, \sigma\sqrt{\Sigma x_i^2})$

$$\therefore \frac{\Sigma x_i y_i}{\sigma\sqrt{\Sigma x_i^2}} > k'$$

$$\therefore P(z > k') = \alpha$$

$$k' = z_{1-\alpha}.$$

Ex 4 $\quad y_i \sim N(\frac{\theta}{2} x_i^2, \sigma)$

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\Sigma(y_i - \frac{\theta}{2}x_i^2)^2}$$

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\theta}{2}x_i^2)^2$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \frac{\theta}{2}x_i^2)\cdot(-\frac{1}{2}x_i^2) = \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2(y_i - \frac{\theta}{2}x_i^2) = 0$$

$$\therefore \hat{\theta} = \frac{2\sum_{i=1}^{n} x_i^2 y_i}{\sum_{i=1}^{n} x_i^4}$$

Ex1. (a) $\frac{L(\theta_0)}{L(\theta_a)} < k$

$$\frac{2^{\sum x_i} \cdot e^{-2n} / \prod_{i=1}^{n} x_i!}{5^{\sum x_i} e^{-5n} / \prod_{i=1}^{n} x_i!} < k$$

$$\therefore \sum x_i > \frac{\ln 3k - 3n}{\ln \frac{2}{5}} \qquad (\bar{x} > k')$$

(b) $H_0: p = p_0$

$H_a: p = p_1 > p_0$

$$\frac{L(p_0)}{L(p_1)} = \frac{\binom{n}{x} p_0^{x} (1-p_0)^{n-x}}{\binom{n}{x} p_1^{x} (1-p_1)^{n-x}} < k$$

$$\therefore x > \frac{\ln k - n \ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_0(1-p_1)}{p_1(1-p_0)}} \qquad \left(\frac{x}{n} > k'\right)$$

Ex2. $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

$$= \bar{x} \pm \frac{1.96}{\sqrt{1+(n-1)\rho}} \cdot \sigma \sqrt{\frac{1+(n-1)\rho}{n}}$$

$$= \bar{x} \pm 0.813 \sigma \sqrt{\frac{1+(n-1)\rho}{n}}$$

$$Z_{0.813} \approx 72.9\%$$

$\therefore$ The actual coverage is $100\% - 2(100\% - 72.9\%) = 58.2\%$

$\because \; y_i \sim N(\frac{\theta}{2}x_i^2, \sigma)$

$$\therefore E(\hat{\theta}) = E\left(\frac{2\sum x_i^2 y_i}{\sum x_i^4}\right) = \frac{2\sum x_i^2 E(y_i)}{\sum x_i^4} = \frac{2\sum x_i^2 \cdot \frac{\theta}{2} x_i^2}{\sum x_i^4} = \theta$$

$$var(\hat{\theta}) = var\left(\frac{2\sum x_i^2 y_i}{\sum x_i^4}\right) = \frac{4\sum x_i^4 \, var(y_i)}{(\sum x_i^4)^2} = \frac{4\sigma^2}{\sum_{i=1}^{n} x_i^4}$$

$$\therefore \hat{\theta} \sim N\left(\theta, \frac{2\sigma}{\sqrt{\sum_{i=1}^{n} x_i^4}}\right)$$

$$\frac{\hat{\theta} - \theta}{\frac{2\sigma}{\sqrt{\sum_{i=1}^{n} x_i^4}}} \sim N(0,1)$$

$\therefore$ The $1-\alpha$ confidence interval for θ is $\hat{\theta} \pm z_{\alpha/2} \frac{2\sigma}{\sqrt{\sum_{i=1}^{n} x_i^4}}$