MIDTERM EXAMINATION # 2

Statistics 305

Term 1, 2005-2006

Thursday, November 10, 2005

Time: 9:30am – 10:45am

Student Name (**Please print in caps**): SOLUTIONS

Student Number: ____________________

Notes:

- This midterm has 5 problems on the 6 following pages, plus 3 pages of tables. Check to ensure that you have a complete paper.
- The amount each part of each question is worth is shown in [ ] on the left-hand side of the page.
- Where appropriate, record your answers in the blanks provided on the right-hand side of the page.
- Your solutions must be justified; show the work and state the reason(s) leading to your answer for each question in the space provided immediately under the question.
- Clear and complete solutions are essential; little partial credit will be given.
- This is a closed book exam.
- A single one-sided 8.5 x 11 page of notes is allowed.
- Calculators are allowed (but not for symbolic differentiation or integration).
- No devices (including calculators) that can store text or send/receive messages are allowed.

Problem	Total Available	Score
1.	10	
2.	6	
3	12	
4.	7	
5.	15	
Total	50	

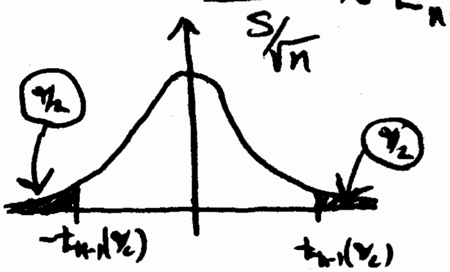
1. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from a normal distribution with mean μ and variance σ^2, where both parameters are unknown.

a) Derive the form of the exact 1 - α confidence interval for:

[3] i) the population mean μ.

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \Rightarrow 1-\alpha = P\left\{-t_{n-1}(\alpha/2) \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{n-1}(\alpha/2)\right\}$$

$$= P\left\{\bar{X}-t_{n-1}(\alpha/2)S/\sqrt{n} \le \mu \le \bar{X}+t_{n-1}(\alpha/2)S/\sqrt{n}\right\}$$



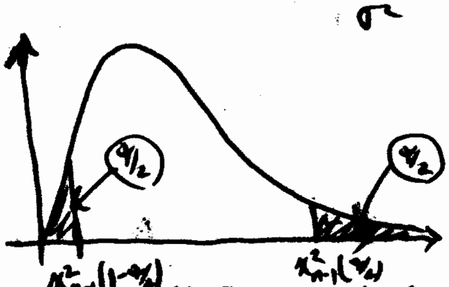
$$\Rightarrow \left(\bar{X}-t_{n-1}(\alpha/2)S/\sqrt{n},\ \bar{X}+t_{n-1}(\alpha/2)S/\sqrt{n}\right) \text{ is } 1-\alpha \text{ CI for } \mu$$

[3] ii) the population standard deviation σ.

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \Rightarrow 1-\alpha = P\left\{\chi^2_{n-1}(1-\alpha/2) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1}(\alpha/2)\right\}$$

$$= P\left\{\frac{1}{\chi^2_{n-1}(1-\alpha/2)} \ge \frac{\sigma^2}{(n-1)S^2} \ge \frac{1}{\chi^2_{n-1}(\alpha/2)}\right\}$$

$$= P\left\{\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right\}$$



$$\Rightarrow \left(\sqrt{\frac{n-1}{\chi^2_{n-1}(1-\alpha/2)}}\cdot S,\ \sqrt{\frac{n-1}{\chi^2_{n-1}(\alpha/2)}}\cdot S\right) \text{ is } 1-\alpha \text{ CI for } \sigma$$

b) Suppose a simple random sample of $n = 16$ from this distribution leads to a sample average of $\bar{x} = 10$ and a sample standard deviation of $s = 5$. Evaluate exact 90% confidence intervals for:

[2] i) the population mean μ. $(7.8, 12.2)$

$$t_{15}(0.05) = 1.753 \Rightarrow 10 \pm 1.753 \cdot 5/\sqrt{16}$$

$$\Leftrightarrow 10 \pm 1.753 \cdot 1.25$$

$$\Leftrightarrow 10 \pm 2.19125 \Rightarrow (7.80875, 12.19125)$$

[2] ii) the population standard deviation σ. $(3.9, 7.2)$

$$\chi^2_{15}(0.95) = 7.26$$

$$\chi^2_{15}(0.05) = 25.00$$

$$\Rightarrow \left(\sqrt{\frac{15}{25.00}}\cdot 5,\ \sqrt{\frac{15}{7.26}}\cdot 5\right)$$

$$\Rightarrow (0.7746 \cdot 5,\ 1.4374 \cdot 5)$$

$$\Rightarrow (3.873,\ 7.187)$$

2. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from the uniform distribution on the interval from 0 to θ; that is, from the population with density function given by:

$$f_\theta(x) = 1/\theta \qquad \text{for } 0 \le x \le \theta.$$

[2] a) Find $\hat{\theta}_{MM}$, the method of moments estimator (MME) of θ for this example. $\underline{\quad 2\overline{X} \quad}$

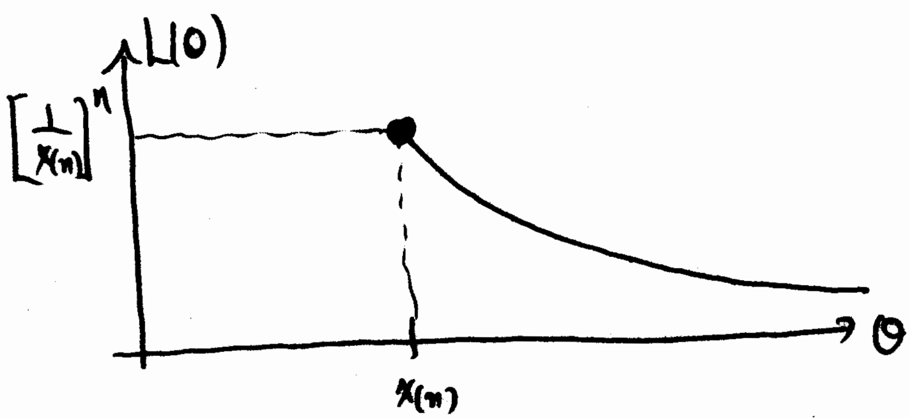
$$\mu = E(x) = \int_0^\theta x \, \frac{1}{\theta} \, d\theta = \frac{1}{\theta}\left[\frac{x^2}{2} \Big|_0^\theta \right] = \frac{\theta}{2}$$

$$\Leftrightarrow \quad \theta = 2\mu \quad \Rightarrow \quad \hat{\theta}_{MM} = 2\overline{X}$$

[3] b) What is $L(\theta)$, the likelihood function, for this example? Provide a clear sketch. ____________

$$f_\theta(x_1, x_2, \ldots, x_n) = \left(\frac{1}{\theta}\right)^n \quad \text{for } 0 \le x_i \le \theta$$

$$\Rightarrow L(\theta) = \left(\frac{1}{\theta}\right)^n \quad \text{for } \theta \ge \text{all of the } x_i \Leftrightarrow \theta \ge \underbrace{\max\{x_1, x_2, \ldots, x_n\}}_{X_{(n)}}$$



[1] c) Find $\hat{\theta}_{ML}$, the maximum likelihood estimator (MLE) of θ for this example. $\underline{\quad X_{(n)} \quad}$

from plot $\Rightarrow \hat{\theta}_{ML} = X_{(n)}$

3. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from the distribution:

$$f_\theta(x) = \theta x^{\theta-1} \qquad \text{for } 0 \le x \le 1.$$

Note that this is a density function provided that $\theta > 0$.

$$\frac{\bar{X}}{1-\bar{X}}$$

[3] a) Find $\hat{\theta}_{MM}$, the method of moments estimator (MME) of θ.

$$\mu = E(X) = \int_0^1 x \, \theta x^{\theta-1} \, dx = \theta \int_0^1 x^\theta dx = \frac{\theta}{\theta+1}\left[x^{\theta+1}\Big|_0^1\right] = \frac{\theta}{\theta+1}$$

$$\Leftarrow \mu = \frac{\theta}{\theta+1} \Leftrightarrow \theta = \frac{\mu}{1-\mu}$$

$$\Rightarrow \hat{\theta}_{MM} = \frac{\bar{X}}{1-\bar{X}}$$

$$\frac{1}{n}\frac{\theta(\theta+1)}{(\theta+2)}$$

[6] b) Find a second-order approximation to the bias of the MME $\hat{\theta}_{MM}$.

$$\hat{\theta}_{MM} = g(\bar{X}), \quad \text{where } g(x) = \frac{x}{(1-x)} = \frac{x-1+1}{(1-x)} = \frac{1}{(1-x)} - 1$$

By Delta Method:

$$E(\hat{\theta}_{MM}) \hat{=} g(E(\bar{X})) + \frac{1}{2}\text{Var}(\bar{X})\, g''(E(\bar{X}))$$

$$\Rightarrow g(x) = (1-x)^{-1} - 1$$

$$\Rightarrow g'(x) = +(1-x)^{-2}$$

$$\Rightarrow g''(x) = +2(1-x)^{-3}$$

$$= g(\mu) + \frac{1}{2}\frac{1}{n}\text{Var}(X)\, g''(\mu)$$

$$= \theta + \frac{1}{2}\frac{1}{n}\text{Var}(X)\frac{2}{\left(1-\frac{\theta}{\theta+1}\right)^3}$$

$$= \theta + \frac{1}{n}(\theta+1)^3\,\text{Var}(X)$$

$$= \theta + \frac{1}{n}(\theta+1)^3 \frac{\theta}{(\theta+1)^2(\theta+2)}$$

$$\boxed{\begin{aligned} \text{But } E(X^2) &= \int_0^1 x^2 \theta x^{\theta-1} dx = \frac{\theta}{\theta+2} \\ \Rightarrow \text{Var}(X) &= \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 \\ &= \frac{\theta}{(\theta+1)^2(\theta+2)} \end{aligned}}$$

$$\Rightarrow \text{Bias} = \frac{1}{n}\frac{\theta(\theta+1)}{(\theta+2)}$$

$$\frac{1}{n}\frac{\theta(\theta+1)^2}{(\theta+2)}$$

[3] c) Find the asymptotic variance of the MME $\hat{\theta}_{MM}$.

By Delta Method:

$$\text{Var}(\hat{\theta}_{MM}) \approx \text{Var}(\bar{X}) \cdot \left[g'(E(\bar{X}))\right]^2$$

$$= \frac{1}{n}\frac{\theta}{(\theta+1)^2(\theta+2)}\left[\frac{1}{\left(1-\frac{\theta}{\theta+1}\right)^2}\right]^2$$

$$= \frac{1}{n}\frac{\theta}{(\theta+1)^2(\theta+2)}(\theta+1)^4 = \frac{1}{n}\frac{\theta(\theta+1)^2}{(\theta+2)}$$

4. Suppose X is a normally distributed random variable with mean μ and variance σ^2, where both parameters are unknown. Evaluate the Fisher Information matrix for

[7] the pair of parameters μ and σ^2 based on the single random variable X.

$$f_{\underset{\sim}{\theta}}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \text{ where } \underset{\sim}{\theta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

$$\log f_{\underset{\sim}{\theta}}(x) = \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\Rightarrow \frac{\partial}{\partial\mu}\log f_{\underset{\sim}{\theta}}(x) = +\frac{2}{2\sigma^2}(x-\mu) = \frac{x-\mu}{\sigma^2}$$

$$\text{and } \frac{\partial}{\partial\sigma^2}\log f_{\underset{\sim}{\theta}}(x) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2$$

$$\Rightarrow \frac{\partial^2}{\partial\mu^2}\log f_{\underset{\sim}{\theta}}(x) = \frac{\partial}{\partial\mu}\left[\frac{\partial}{\partial\mu}\log f_{\underset{\sim}{\theta}}(x)\right] = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2}{\partial\mu\,\partial\sigma^2}\log f_{\underset{\sim}{\theta}}(x) = \frac{\partial}{\partial\sigma^2}\left[\frac{\partial}{\partial\mu}\log f_{\underset{\sim}{\theta}}(x)\right] = -\frac{x-\mu}{\sigma^4}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log f_{\underset{\sim}{\theta}}(x) = \frac{\partial}{\partial\sigma^2}\left[\frac{\partial}{\partial\sigma^2}\log f_{\underset{\sim}{\theta}}(x)\right] = \frac{1}{2\sigma^4} - \frac{2}{2\sigma^6}(x-\mu)^2$$

$$\Rightarrow E\left[-\ddot{\ell}(\underset{\sim}{\theta})\right] = \begin{bmatrix} E\left(+\frac{1}{\sigma^2}\right) & E\left[+\frac{(X-\mu)}{\sigma^4}\right] \\ & E\left[-\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}(X-\mu)^2\right] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix},$$

the Fisher Information matrix based on X

5. According to the Hardy-Weinberg law, if gene frequencies are in equilibrium, the genotypes *AA*, *Aa* and *aa* occur with relative frequencies $(1-\theta)^2$, $2\theta(1-\theta)$ and θ^2, respectively. Plato et al. (1964) published data on haptoglobin type in a random sample of $n = 190$ individuals with corresponding counts of 10, 68 and 112. Let X_1, X_2 and X_3 denote the random variables leading to these observed counts (Note: $X_1 + X_2 + X_3 = n$).

[6] a) Find an expression for $\hat{\theta}_{ML}$, the maximum likelihood estimator (MLE) of θ, in terms of the random variables X_1, X_2 and X_3. $\underline{\dfrac{X_2+2X_3}{2n}}$

Multinomial problem $\Rightarrow L(\theta) = \dfrac{n!}{x_1!x_2!x_3!}\left[(1-\theta)^2\right]^{x_1}\left[2\theta(1-\theta)\right]^{x_2}\left[\theta^2\right]^{x_3}$

$$\Rightarrow \ell(\theta) \equiv \log L(\theta) = \text{constant} + 2x_1 \log(1-\theta) + x_2 \log[\theta(1-\theta)] + 2x_3 \log\theta$$

$$= \text{constant} + 2x_1 \log(1-\theta) + x_2 \log\theta + x_2 \log(1-\theta) + 2x_3 \log\theta$$

$$= \text{constant} + (2x_1+x_2)\log(1-\theta) + (x_2+2x_3)\log\theta$$

$$\Rightarrow \ell'(\theta) = -\frac{(2x_1+x_2)}{(1-\theta)} + \frac{(x_2+2x_3)}{\theta}$$

$\boxed{\text{Note: } 0<\theta<1, \text{ so both } \theta \text{ and } 1-\theta \text{ are} > 0}$

$$\Rightarrow \ell'(\theta) > 0 \Leftrightarrow \frac{x_2+2x_3}{\theta} > \frac{2x_1+x_2}{(1-\theta)} \Leftrightarrow x_2+2x_3 > \theta(2x_1+x_2+x_2+2x_3)$$

$$\Leftrightarrow \theta < \frac{x_2+2x_3}{2n}$$

$\Rightarrow \ell'(\theta)=0 \Leftrightarrow \theta = \dfrac{x_2+2x_3}{2n}$ and this root does correspond to the maximum of $\ell(\theta) \Rightarrow \hat{\theta}_{ML} = \dfrac{X_2+2X_3}{2n}$

[1] b) Evaluate the MLE $\hat{\theta}_{ML}$ for the given data. $\underline{0.77}$

$$\hat{\theta}_{ML} = \frac{68+2(112)}{2(190)} = 0.768,421$$

5. (continued)

[4] c) Find an expression for the asymptotic variance of $\hat{\theta}_{ML}$ in terms of n and θ. $\frac{\theta(1-\theta)}{2n}$

Evaluate the Fisher information in the sample

$$\Rightarrow \ell''(\theta) = -\frac{(2x_1+x_2)}{(1-\theta)^2} - \frac{(x_2+2x_3)}{\theta^2}$$

$$\Rightarrow E[-\ell''(\theta)] = E\left[\frac{2X_1+X_2}{(1-\theta)^2} + \frac{X_2+2X_3}{\theta^2}\right]$$

But:

$X_1 \sim B(n, (1-\theta)^2)$

$X_2 \sim B(n, 2\theta(1-\theta)) \Rightarrow$

$X_3 \sim B(n, \theta^2)$

$$= n\left[\frac{2(1-\theta)^2+2\theta(1-\theta)}{(1-\theta)^2} + \frac{2\theta(1-\theta)+2\theta^2}{\theta^2}\right]$$

$$= 2n\left[\frac{1-\theta+\theta}{(1-\theta)} + \frac{1-\theta+\theta}{\theta}\right] = 2n\left[\frac{1}{1-\theta} + \frac{1}{\theta}\right] = \frac{2n}{\theta(1-\theta)}$$

[1] d) Evaluate the estimated standard error of the MLE $\hat{\theta}_{ML}$ for the given data. 0.022

$$\hat{A.Var}(\hat{\theta}_{ML}) = \frac{\hat{\theta}_{ML}(1-\hat{\theta}_{ML})}{2n}$$

$$= \frac{0.768421(1-0.768421)}{2(190)} = 0.000,468,29\text{---}$$

$$\Rightarrow \hat{SE}(\hat{\theta}_{ML}) = \sqrt{0.000,468\text{---}} = $$

$$= 0.021,640\text{---}$$

[3] e) Find an approximate 99% confidence interval for θ based on the MLE $\hat{\theta}_{ML}$. (0.71, 0.82)

$$\hat{\theta}_{ML} \pm z(\alpha/2) \cdot \hat{SE}(\hat{\theta}_{ML})$$

$$\Rightarrow 0.768,421 \pm 2.575 * 0.021,640$$

$$\Leftrightarrow 0.768,421 \pm 0.055,723$$

$$\Leftrightarrow (0.712,698,\ 0.824,144)$$