# UNIVERSITY OF TORONTO

## Faculty of Arts and Science

## APRIL/MAY EXAMINATIONS 2004

## STA261H1 S

Duration – 3 hours

**NO AIDS ALLOWED – NOT EVEN CALCULATORS.**

## SOLUTIONS

**1.** [10 points] Suppose $\Omega = [0, \infty)$ and $S = [0, 1]$, and $P_\theta$ has density $f_\theta(s) = (\theta + 1)s^\theta$ for all $\theta \in \Omega$, and we observe the values $x_1 = 1/2$, $x_2 = 2/3$, $x_3 = 3/4$. What is the MLE for $\theta$?

> **Solution.** *Here the likelihood function is* $L(\theta \,|\, x_1, x_2, x_3) = [(\theta+1)(1/2)^\theta]\,[(\theta+1)(2/3)^\theta]\,[(\theta+1)(3/4)^\theta] = (\theta+1)^3/4^\theta$, *for* $\theta \in \Omega$. *Hence, the log-likelihood is* $\ell(\theta \,|\, x_1, x_2, x_3) = \log[(\theta+1)^3/4^\theta] = 3\log(\theta+1) - \theta\log(4)$. *So, the Score Function is* $S(\theta \,|\, x_1, x_2, x_3) = \frac{\partial}{\partial\theta}\ell(\theta \,|\, x_1, x_2, x_3) = 3/(\theta+1) - \log(4)$. *This equals 0 iff* $\theta + 1 = 3/\log(4)$, *i.e.* $\theta = [3/\log(4)] - 1$. *Furthermore, the second derivative is* $(\frac{\partial}{\partial\theta})^2\ell(\theta \,|\, x_1, x_2, x_3) = -3/(\theta+1)^2 < 0$ *for all* $\theta \in \Omega$. *Hence, this critical point is a maximum and therefore must be the MLE. Hence, the MLE is* $\hat{\theta} = [3/\log(4)] - 1 \doteq 1.164$.

**2.** [10 points] Suppose $S = \Omega = [0, 2]$, and we observe two observations $x_1$ and $x_2$, and obtain the following likelihood function: $L(\theta \,|\, x_1, x_2) = (\theta - x_1 - x_2)^4$. Prove or disprove that $T(x_1, x_2) = x_1 + x_2$ is a minimal sufficient statistic for $\theta$.

> **Solution.** *Yes,* $T$ *is a minimal sufficient statistic for* $\theta$.
>
> *To prove that* $T$ *is sufficient: In terms of the Factorisation Theorem, we can write* $L(\theta \,|\, s) = g_\theta(T(s))\,h(s)$ *where* $g_\theta(r) = (\theta - r)^4$ *and* $h(s) \equiv 1$. *Hence, from the Factorisation Theorem,* $T$ *is sufficient.*
>
> *To prove that* $T$ *is minimal: Note that* $\frac{\partial}{\partial\theta}L(\theta \,|\, x_1, x_2) = \frac{\partial}{\partial\theta}(\theta - x_1 - x_2)^4 = 4(\theta - x_1 - x_2)^3$, *which equals 0 iff* $\theta = x_1 + x_2 = T(x_1, x_2)$. *Thus,* $T(x_1, x_2)$ *is the unique critical point of the likelihood function. It follows that we can compute* $T(x_1, x_2)$ *from the likelihood function, i.e.* $T(x_1, x_2)$ *is uniquely determined by the function* $L(\theta \,|\, x_1, x_2)$. *So,* $T$ *must be a minimal sufficient statistic.*

**3.** [10 points] Suppose the disease "Examitus" has a 40% fatality rate. A company then developes a drug which they claim <u>reduces</u> this fatality rate. The drug is given to 10 Examitus patients, of whom 2 die and 8 live. Compute (with explanation) a P-value for testing the null hypothesis that the drug had no effect, versus the alternative hypothesis that the drug reduced the fatality of Examitus. [You do not need to simplify arithmetic expressions.]

**Solution.** *The P-value is the probability of obtaining as few or fewer deaths, under the null hypothesis of 40% fatality rate. Thus, it equals* $P_{40\%}[\leq 2 \text{ die}] = P_{40\%}[2 \text{ die}] + P_{40\%}[1 \text{ die}] + P_{40\%}[0 \text{ die}] = \binom{10}{2}(0.4)^2(0.6)^8 + \binom{10}{1}(0.4)^1(0.6)^9 + \binom{10}{0}(0.4)^0(0.6)^{10} = 45(0.4)^2(0.6)^8 + 10(0.4)^1(0.6)^9 + (0.4)^0(0.6)^{10}$. *[In fact this* $\doteq 0.16729$*, but you don't need to know that.]*

**4.** Consider the statistical model where $S = \{0, 1, 2, \ldots\}$, $\Omega = (0, \infty)$, and for $\theta \in \Omega$, $P_\theta$ is the Poisson($\theta$) distribution. [Thus, $P_\theta\{s\} = e^{-\theta}\theta^s \,/\, s!$ for $s \in S$, and $P_\theta$ has mean $\theta$ and variance $\theta$.] Suppose we observe a single observation $s \in S$. Consider the following two estimators for $\theta$: $\hat{\theta}_1 = s$, and $\hat{\theta}_2 = s + 2$.

**(a)** [3 points] Compute (with explanation) $\text{Bias}_\theta(\hat{\theta}_1)$, $\text{Var}_\theta(\hat{\theta}_1)$, and $MSE_\theta(\hat{\theta}_1)$.

**Solution.** $\text{Bias}_\theta(\hat{\theta}_1) = E_\theta(s) - \theta = \theta - \theta = 0$. $\text{Var}_\theta(\hat{\theta}_1) = Var_\theta(s) = \theta$. $MSE_\theta(\hat{\theta}_1) = \text{Var}_\theta(\hat{\theta}_1) + [\text{Bias}_\theta(\hat{\theta}_1)]^2 = \theta + 0^2 = \theta$.

**(b)** [3 points] Compute (with explanation) $\text{Bias}_\theta(\hat{\theta}_2)$, $\text{Var}_\theta(\hat{\theta}_2)$, and $MSE_\theta(\hat{\theta}_2)$.

**Solution.** $\text{Bias}_\theta(\hat{\theta}_2) = E_\theta(s+2) - \theta = \theta + 2 - \theta = 2$. $\text{Var}_\theta(\hat{\theta}_2) = Var_\theta(s+2) = \text{Var}_\theta(s) = \theta$. $MSE_\theta(\hat{\theta}_2) = \text{Var}_\theta(\hat{\theta}_2) + [\text{Bias}_\theta(\hat{\theta}_2)]^2 = \theta + 2^2 = \theta + 4$.

**(c)** [3 points] Is $\hat{\theta}_1$ consistent? (Explain your reasoning.)

**Solution.** *[Note: To make sense of this question, we need to interpret* $\hat{\theta}_1$ *for* $n$ *observations as* $\overline{x}$*.] Yes,* $\hat{\theta}_1$ *is consistent, since* $\lim_{n\to\infty} MSE_\theta(\hat{\theta}_1) = \lim_{n\to\infty} MSE_\theta(\overline{x}) = \lim_{n\to\infty} \theta \,/\, n = 0$.

**(d)** [3 points] Is $\hat{\theta}_2$ consistent? (Explain your reasoning.)

**Solution.** *[Note: To make sense of this question, we need to interpret* $\hat{\theta}_2$ *for* $n$ *observations as* $\overline{x} + 2$*.] Here* $MSE_\theta(\hat{\theta}_2) \not\to 0$*, which does not tell us whether or not* $\hat{\theta}_2$ *is consistent. However, in fact* $\hat{\theta}_2$ *is not consistent. Indeed, for* $0 < \epsilon < 1$, $P[|\hat{\theta}_2 - \theta| \geq \epsilon] \geq P[\hat{\theta}_2 \geq \theta + 1] = P[\hat{\theta}_1 + 2 \geq \theta + 1] = P[\hat{\theta}_1 \geq \theta - 1] \geq P[|\hat{\theta}_1 - \theta| \leq 1] = 1 - P[|\hat{\theta}_1 - \theta| > 1]$*, which goes to* $1 - 0 = 1 \neq 0$ *as* $n \to \infty$ *(since* $\hat{\theta}_1$ *is consistent). So,* $\hat{\theta}_2$ *is not consistent.*

**(e)** [3 points] Which estimator is "better"? (Explain your reasoning.)

**Solution.** $\hat{\theta}_1$ *is better, because it has smaller MSE (which is good), and also* *it is consistent (which is good) whereas $\hat{\theta}_2$ is not consistent (which is bad).*

**5.** [10 points] Suppose $\Omega = (0, \infty)$ and $S = \mathbf{R}$, and we observe the three values 1, 3, and 4. Suppose for $\theta \in \Omega$, the distribution $P_\theta$ has density function $f_\theta(s) = (3/2)\theta^{-3}s^2$ for $-\theta \le s \le \theta$, with $f_\theta(s) = 0$ otherwise. Compute (with explanation) a method-of-moments estimate for $\theta$.

**Solution.** *Here $P_\theta$ has mean $\int_{-\theta}^{\theta}(s)(3/2)(\theta^{-3}s^2)ds = (3/2)\theta^{-3}\int_{-\theta}^{\theta}s^3 ds = (3/2)\theta^{-3}[(\theta^4/4) - ((-\theta)^2/4)] = 0$, which does not specify a choice of $\theta$. Therefore, we have to use second moments instead. $P_\theta$ has second moment $\int_{-\theta}^{\theta}(3/2)(s^2)(\theta^{-3}s^2)ds = (3/2)\theta^{-3}[\int_{-\theta}^{\theta}s^4 ds = (3/2)\theta^{-3}[(\theta^5/5) - ((-\theta)^5/5)] = 3\theta^{-3}\theta^5/5 = 3\theta^2/5$. Also $\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 = (1+9+16)/3 = 26/3$. So, we want to solve for $3\theta^2/5 = 26/3$, i.e. $\hat{\theta} = \sqrt{130/9} \doteq 3.8$.*

**6.** Let $\Omega = \{1, 2\}$, with $P_1\{2\} = P_1\{3\} = 1/2$, and $P_2\{4\} = P_2\{9\} = 1/2$. Suppose we observe two observations, $X_1$ and $X_2$. Determine (with explanation) whether or not each of the following statistics is <u>ancilliary</u>.

**(a)** [5 points] $D_1 = X_1 - X_2$.

**Solution.** *Here $P_1[D_1 = 1] = P_1[X_1 = 3,\ X_2 = 2] = (1/2)(1/2) = 1/4$. On the other hand, none of $9-9$, $9-4$, $4-9$, or $4-4$ equals 1. So, $P_2[D_1 = 1] = 0$. Hence, $P_\theta[D_1 = 1]$ depends on $\theta$. Thus, the distribution of $D_1$ depends on $\theta$, and thus is <u>not</u> ancilliary.*

**(b)** [5 points] $D_2 = X_1 \,/\, X_2$.

**Solution.** *Here $P_1[D_2 = 3/2] = P_1[X_1 = 3,\ X_2 = 2] = (1/2)(1/2) = 1/4$. On the other hand, none of $9/9$, $9/4$, $4/9$, or $4/4$ equals $3/2$. So, $P_2[D_2 = 3/2] = 0$. Hence, $P_\theta[D_2 = 3/2]$ depends on $\theta$. Thus, the distribution of $D_2$ depends on $\theta$, and thus is <u>not</u> ancilliary.*

**7.** Suppose we flip a coin 1000 times, and obtain 444 heads and 556 tails. We want to test the (null) hypothesis that the coin is fair, i.e. that heads and tails each have probability 1/2.

**(a)** [6 points] Compute the observed value of the $\chi^2$-statistic associated with the null hypothesis. [You do not need to simplify arithmetic expressions.]

**Solution.** *Here the $\chi^2$-statistic is $x^2 = \sum_{i=1}^{2}(c_i - np_i)^2/(np_i) = (444 - 1000(1/2))^2/(1000(1/2)) + (556 - 1000(1/2))^2/(1000(1/2)) = 56^2/500 + 56^2/500 = 56^2/250$. [In fact this $\doteq 12.5$, but you don't need to know that.]*

**(b)** [4 points] Explain how to use this $\chi^2$-statistic to compute the P-value for the null hypothesis. [You do not need to compute the actual value, you just need to explain precisely how you would compute it, in terms of precisely which probabilities corresponding to precisely which distributions.]

**Solution.** *Under the null hypothesis,* $X^2 \sim \chi^2(k-1) = \chi^2(2-1) = \chi^2(1)$. *So, the P-value is the probability that* $X^2$ *would be at least as large as the observed value of* $x^2$. *Hence, the P-value is equal to* $P[X^2 \geq x^2]$, *where* $X^2 \sim \chi^2(1)$, *and* $x^2$ *is the observed value from part (a).*

**8.** [10 points] (Bayesian inference.) Suppose $\Omega = \{1, 2\}$ and $S = \{6, 8\}$, with $P_1\{6\} = 1/4$, $P_1\{8\} = 3/4$, $P_2\{6\} = 2/3$, $P_2\{8\} = 1/3$. Suppose our prior distribution for $\theta$ is given by $\Pi\{1\} = 1/5$, and $\Pi\{2\} = 4/5$. Suppose we observe the single observation $s = 8$. Compute the posterior probability that $\theta = 1$, i.e. compute $\Pi(1 \mid 8)$.

**Solution.** *Here*

$$\Pi(1 \mid 8) = \frac{\Pi\{1\}\, f_1(8)}{m(8)} = \frac{\Pi\{1\}\, f_1(8)}{\Pi\{1\}\, f_1(8) + \Pi\{2\}\, f_2(8)}$$

$$= \frac{(1/5)\,(3/4)}{(1/5)\,(3/4) + (4/5)\,(1/3)} = \frac{3/20}{(9/60) + (16/60)} = \frac{3/20}{25/60} = 9/25\,.$$

**9.** [10 points] Suppose we have the following model (instead of the usual linear regression model): that $E[Y \mid X = x] = \beta\, x^3$, where $\beta$ is unknown. Suppose we observe the pairs $(x_1, y_1), \ldots, (x_n, y_n)$. Derive an appropriate least-squares estimate of $\beta$. [Hint: First write down a formula for the squared error.]

**Solution.** *Here the squared-error is* $SE = \sum_{i=1}^n (y_i - \beta(x_i)^3)^2$. *Since this is a quadratic function of* $\beta$ *with positive leading coefficient, it is minimised where* $\frac{\partial}{\partial \beta} SE = 0$, *i.e.* $-2\sum_{i=1}^n (x_i)^3 (y_i - \beta(x_i)^3) = 0$, *i.e.* $\sum_{i=1}^n (x_i)^3 y_i = \beta \sum_{i=1}^n (x_i)^6$, *i.e.* $\beta = \hat{\beta} \equiv [\sum_{i=1}^n (x_i)^3 y_i] \;/\; [\sum_{i=1}^n (x_i)^6]$. *This is the least-squares estimate of* $\beta$.

**10.** [10 points] Suppose we conduct a poll of 400 citizens, and find that 250 support candidate A, while 150 support candidate B. Provide (with explanation) a 90% confidence interval for the fraction $\theta$ of the population that supports candidate A. [You may use the facts that if $Z \sim N(0, 1)$, then $P[Z \leq -0.524] \doteq 0.3$, $P[Z \leq -0.842] \doteq 0.2$, $P[Z \leq -1.28] \doteq 0.1$, $P[Z \leq -1.64] \doteq 0.05$, and $P[Z \leq -1.96] \doteq 0.025$.]

**Solution.** *Let* $C$ *be the number in the poll that support A. Then* $C \sim \text{Binomial}(400, \theta)$. *Thus,* $C$ *has mean* $400\theta$, *and variance* $400\theta(1-\theta)$. *Thus,* $Z \equiv (C - 400\theta)/\sqrt{400\theta(1-\theta)}$ *has mean 0 and variance 1, and for large* $n$, $Z \approx N(0, 1)$. *Hence,* $P[-1.64 < Z < +1.64] = 1 - P[Z \leq -1.64] - P[Z \geq$

$+1.64] = 1 - 2\,P[Z \le -1.64] \doteq 1 - 2(0.05) = 1 - 0.1 = 0.9 = 90\%$. *Thus,* $90\% = P[-1.64 < (C - 400\theta)/\sqrt{400\theta(1-\theta)} < 1.64] = P[-1.64\sqrt{400\theta(1-\theta)} < (C - 400\theta) < 1.64\sqrt{400\theta(1-\theta)}]$, *so* $90\% = P[(C/400) - 1.64\sqrt{\theta(1-\theta)/400} < \theta < (C/400) + 1.64\sqrt{\theta(1-\theta)/400}]$. *This gives a 90% confidence interval for* $\theta$ *of* $(c/400) \pm 1.64\sqrt{\theta(1-\theta)/400} = (250/400) \pm (1.64/20)\sqrt{\theta(1-\theta)} = (5/8) \pm (0.082)\sqrt{\theta(1-\theta)}$.

*But* $\theta$ *is unknown. Thus, we can use the Plug-In Estimate, i.e. replace* $\theta$ *by* $\hat{\theta} = 250/400 = 5/8$, *to get a 90% confidence interval of* $(5/8) \pm (0.082)\sqrt{(5/8)(3/8)} = (5/8) \pm (0.082/8)\sqrt{15}$. *Alternatively, we can use the Conservative Option that* $\theta(1-\theta) \le 1/4$, *to get a 90% confidence interval of* $(5/8) \pm (0.082)(1/2) = (5/8) \pm 0.041 = 0.625 \pm 0.041 = (0.584,\, 0.666)$.

**11.** Consider the usual normal linear regression model, with the conditional distribution of $Y$ given that $X = x$ equal to $N(\beta_1 + \beta_2 x,\, \sigma^2)$, with $\beta_1$ and $\beta_2$ and $\sigma^2$ unknown. Suppose we observe the following three pairs $(x_i, y_i)$: $(2,5)$, $(4,5)$, $(6,2)$. It is a fact (which you may use) that this leads to the least-squares estimates $b_1 = 7$ and $b_2 = -3/4$.

**(a)** [6 points] Compute $s^2$, $RSS$, and $ESS$.

**Solution.** $s^2 = \frac{1}{n-2}\sum_i (y_i - b_1 - b_2 x_i)^2 = \frac{1}{3-2}[(5 - 7 - (-3/4)2)^2 + (5 - 7 - (-3/4)4)^2 + (2 - 7 - (-3/4)6)^2] = (-1/2)^2 + (1)^2 + (-1/2)^2 = 3/2$.

$RSS = (b_2)^2 \sum_i (x_i - \overline{x})^2 = (-3/4)^2[(2-4)^2 + (4-4)^2 + (6-4)^2] = (9/16)[8] = 9/2$.

$ESS = (n-2)s^2 = (3-2)(3/2) = 3/2$.

**(b)** [3 points] Compute $\sum_{i=1}^n (y_i - \overline{y})^2$ directly, and then check whether or not $RSS + ESS = \sum_{i=1}^n (y_i - \overline{y})^2$ in this case.

**Solution.** $\overline{y} = (5+5+2)/3 = 4$, *so* $\sum_i (y_i - \overline{y})^2 = (5-4)^2 + (5-4)^2 + (2-4)^2 = 1 + 1 + 4 = 6$. *And, we verify that* $RSS + ESS = (9/2) + (3/2) = 12/2 = 6 = \sum_i (y_i - \overline{y})^2$, *as it must.*

**(c)** [2 points] Compute the F statistic.

**Solution.** $F = RSS/s^2 = (9/2)/(3/2) = 3$.

**(d)** [4 points] Explain how you would use the F statistic to obtain a P-value for testing the (null) hypothesis that $\beta_2 = 0$, versus the alternative hypothesis that $\beta_2 \neq 0$.

**Solution.** *Under the null hypothesis that* $\beta_2 = 0$, *we have* $F \sim F(1, n-2) = F(1,1)$. *Thus, the P-value is given by* $P[W \ge 3]$, *where* $W \sim F(1,1)$.

**12.** [10 points] Suppose Mary and Jane are two long-jump athletes. Mary makes 4 jumps, of the following distances (in decimeters): 12, 13, 15, 12. Jane makes 3 jumps, of the following distances: 11, 12, 13. Assume that Mary's distances have distribution $N(\beta_1, \sigma^2)$, and Jane's distances have distribution $N(\beta_2, \sigma^2)$, with $\beta_1$ and $\beta_2$ and $\sigma^2$ unknown. Compute (with explanation) a 95% confidence interval for the difference of means $\beta_1 - \beta_2$. [You do not need to simplify arithmetic expressions, but your answer should be purely numerical, i.e. it should not involve any unknown or unspecified quantities. You may use the facts that if $T_2 \sim t(2)$, $T_3 \sim t(3)$, $T_4 \sim t(4)$, and $T_5 \sim t(5)$, then $P[T_2 \le -2.92] \doteq P[T_3 \le -2.35] \doteq P[T_4 \le -2.13] \doteq P[T_5 \le -2.02] \doteq 0.05$, and $P[T_2 \le -4.30] \doteq P[T_3 \le -3.18] \doteq P[T_4 \le -2.78] \doteq P[T_5 \le -2.57] \doteq 0.025$.]

**Solution.** *We compute that* $\overline{y}_1 = (12 + 13 + 15 + 12)/4 = 13$ *and* $\overline{y}_2 = (11+12+13)/3 = 12$. *Also* $n_1 = 4$, $n_2 = 3$, *and* $N = n_1+n_2 = 7$, *and* $a = 2$. *Thus,* $s^2 = \frac{1}{N-a}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2 = \frac{1}{7-2}[(12-13)^2 + (13-13)^2 + (15-13)^2 + (12-13)^2 + (11-12)^2 + (12-12)^2 + (13-12)^2] = (1/5)[1+0+4+1+1+0+1] = 8/5$.

*Then* $T \sim t(N-a) = t(7-2) = t(5)$, *where* $T = [(\overline{Y}_1 - \overline{Y}_2) - (\beta_1 - \beta_2)] \,/\, \sqrt{S^2((1/n_1) + (1/n_2))}$. *Thus,* $P[-2.57 \le T \le +2.57] = 1 - 2\,P[T \le -2.57] \doteq 1 - 2(0.025) = 0.95 = 95\%$. *Thus,* $P[(\overline{Y}_1 - \overline{Y}_2) - 2.57\sqrt{S^2((1/n_1) + (1/n_2))} < \beta_1 - \beta_2 < (\overline{Y}_1 - \overline{Y}_2) - 2.57\sqrt{S^2((1/n_1) + (1/n_2))}] \doteq 0.95$, *so a 95% confidence interval for* $\beta_1 - \beta_2$ *is* $(\overline{y}_1 - \overline{y}_2) \pm 2.57\sqrt{s^2((1/n_1) + (1/n_2))} = (13 - 12) \pm 2.57\sqrt{(8/5)((1/4) + (1/3))} = 1 \pm 2.57\sqrt{56/60}$. *[In fact this* $\doteq (-1.48,\ 3.48)$, *but you don't need to know that.]*