STA 261S, Winter 2004, Test #2

(March 24, 2004, 3:10 p.m. Duration: 100 minutes.)

SOLUTIONS

1. Suppose $\Omega = (0, \infty)$ and $S = [0, \infty)$, and we observe the three values 1, 3, and 4. Suppose that for $\theta \in \Omega$, the distribution P_θ has density function $f_\theta(s) = 2\theta^{-2}s$ for $0 \le s \le \theta$, with $f_\theta(s) = 0$ otherwise. Suppose we have a prior distribution Π for θ, with density function $\pi(\theta) = \theta/18$ for $0 \le \theta \le 6$, with $\pi(\theta) = 0$ otherwise.

(a) [7 points] Compute the posterior density function $\pi(\theta \,|\, 1, 3, 4)$, up to a normalisation constant (i.e. you may ignore factors which do not depend on θ).

> **Solution.** *Here the likelihood function equals* $L(\theta \,|\, x_1, x_2, x_3) = [2\theta^{-2}(1)][2\theta^{-2}(3)][2\theta^{-2}(4)] = 96\,\theta^{-6}$ *for* $\theta \ge 4$, *otherwise* $L(\theta \,|\, x_1, x_2, x_3) = 0$ *(since if* $\theta < 4$, *then we wouldn't have* $s \le \theta$ *for* $s = 4$*). Now, the posterior density equals* $\pi(\theta)\, L(\theta \,|\, x_1, x_2, x_3) \,/\, m(x_1, x_2, x_3)$. *This equals 0 for* $\theta < 4$, *and also for* $\theta > 6$ *(since if* $\theta > 6$ *then we wouldn't have* $0 \le \theta \le 6$*). Otherwise, for* $4 \le \theta \le 6$, *it equals* $(\theta/18)\,(96\,\theta^{-6}) \,/\, m(x_1, x_2, x_3) = K\,\theta^{-5}$, *where* $K = 96 \,/\, 18\, m(x_1, x_2, x_3)$ *does not depend on* θ.

(b) [3 points] Compute the normalisation constant.

> **Solution.** *We need to have* $\int \pi(\theta)\, d\theta = 1$, *i.e.* $1 = \int_4^6 K\,\theta^{-5}\, d\theta = K\,[6^{-4} - 4^{-4}]/(-4)$. *Thus,* $K = 4/[4^{-4} - 6^{-4}]$. *[Therefore* $K = 4/[(1/4^4) - (1/6^4)] = 4/[(6^4 - 4^4)/24^4] = 4/[(1296 - 256)/331776] = 1327104/1040 = 82944/65$, *but you don't need to simplify it that much.] Hence,* $\pi(\theta) = \left(4/[4^{-4} - 6^{-4}]\right)\theta^{-5}$ *for* $4 \le \theta \le 6$, *otherwise 0.*

2. Suppose your friend goes bowling, and your (null) hypothesis is that she gets a "Strike" 1/10 of the time, a "Spare" 1/5 of the time, and "Neither" 7/10 of the time. You observe that she gets 7 Strikes, 11 Spares, and 62 Neithers.

(a) [6 points] Compute the χ^2-statistic associated with your hypothesis. [You do not need to simplify arithmetic expressions.]

> **Solution.** *Here the* χ^2*-statistic is* $x^2 = \sum_i [c_i - np_i]^2/[np_i] = [7 - (80)(1/10)]^2/[(80)(1/10)] + [11 - (80)(1/5)]^2/[(80)(1/5)] + [62 - (80)(7/10)]^2/[(80)(7/10)]$. *[In fact this* $\doteq 2.33$, *but you don't need to know that.]*

(b) [4 points] Explain how to use this χ^2-statistic to compute the P-value for your hypothesis. [You do not need to compute the actual numerical 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 your hypothesis, we expect* $X^2 \sim \chi^2(k-1) = \chi^2(2)$, *with larger values indicating less agreement with your hypothesis. Thus, the P-value is equal to* $P[X^2 \geq x^2]$, *where* $X^2 \sim \chi^2(2)$, *and* x^2 *as in part (a).*

3. Suppose you can play darts with either your Right hand or your Left hand. Your (null) hypothesis is that you play equally well with either hand, i.e. that your Wins and Losses are <u>independent</u> of which hand you use. One evening, you choose your hand <u>randomly</u> each time. With your Right hand you get 8 Wins and 3 Losses, and with your Left hand you get 6 Wins and 4 Losses.

(a) [6 points] Compute the χ^2-statistic associated with your hypothesis. [You do not need to simplify arithmetic expressions.]

Solution. *Here* $c_{1\cdot} = 8+3 = 11$, $c_{2\cdot} = 6+4 = 10$, $c_{\cdot 1} = 8+6 = 14$, *and* $c_{\cdot 2} = 3+4 = 7$, *and* $n = 8+3+6+4 = 21$. *Thus, the* χ^2*-statistic is* $x^2 = \sum_{i,j}(c_{ij} - c_{i\cdot}c_{\cdot j}/n)^2 \,/\, (c_{i\cdot}c_{\cdot j}/n) = (8-(11)(14)/21)^2/((11)(14)/21) + (3-(11)(7)/21)^2/((11)(7)/21) + (6-(10)(14)/21)^2/((10)(14)/21) + (4-(10)(7)/21)^2/((10)(7)/21)$. *[In fact, this* $= 21/55 \doteq 0.38$, *but you don't need to know that.]*

(b) [4 points] Explain how to use this χ^2-statistic to compute the P-value for your 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,* $X^2 \sim \chi^2((a-1)(b-1)) = \chi^2((2-1)(2-1)) = \chi^2(1)$. *So, the P-value is the probability that* X^2 *would be at least as large as the observed* x^2, *i.e. is equal to* $P[X^2 \geq x^2]$, *where* $X^2 \sim \chi^2(1)$, *and* x^2 *is as in part (a).*

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

Solution. *Here* P_θ *has mean* $\int_0^\theta (s)(2\,\theta^{-2}s)ds = 2\,\theta^{-2}(\theta^3/3) = 2\,\theta/3$. *Also* $\overline{x} = (1+3+5)/3 = 3$. *So, we want to solve for* $2\,\hat{\theta}/3 = 3$, *i.e.* $\hat{\theta} = 9/2$.

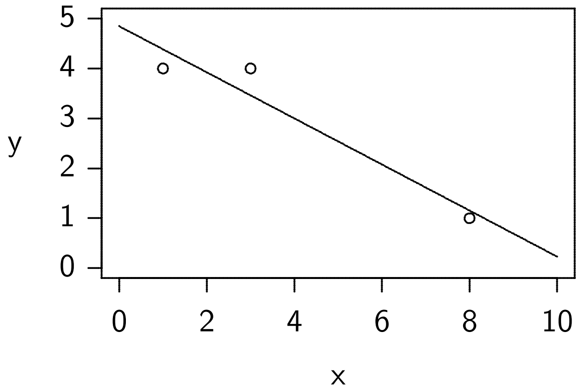
5. [5 points] Suppose we know that $E[Y \mid X = x] = 3 + 4x$, and also that $E[X] = 5$. Compute (with explanation) the value of $E[Y]$.

Solution. *Using the double-expectation formula,* $E[Y] = E\big[E[Y \mid X]\big] = E[3+4X] = 3 + 4\,E[X] = 3 + 4(5) = 23$.

6. Suppose we have a linear regression model $E[Y \mid X = x] = \beta_1 + \beta_2 x$, and we observe the following three pairs (x_i, y_i): $(1, 4)$, $(3, 4)$, $(8, 1)$.

(a) [2 points] Draw a rough graph of these three observations together with a rough, approximate line of best fit. [Your graph does not have to be perfectly accurate, but it should be accurate enough to illustrate what it is that you are graphing.]

Solution. *Your graph should look* <u>*approximately*</u> *like this:*



(b) [1 point] Compute $\overline{x}$ and $\overline{y}$.

Solution. $\overline{x} = (1 + 3 + 8)/3 = 12/3 = 4$. $\quad \overline{y} = (4 + 4 + 1)/3 = 9/3 = 3$.

(c) [5 points] Compute b_1 and b_2 (the least-squares estimates for β_1 and β_2).

Solution. $b_2 = [\sum_i (x_i - \overline{x})(y_i - \overline{y})] \,/\, [\sum_i (x_i - \overline{x})^2] = [(1-4)(4-3) + (3-4)(4-3) + (8-4)(1-3)] \,/\, [(1-4)^2 + (3-4)^2 + (8-4)^2] = [-3 - 1 - 8] \,/\, [9 + 1 + 16] = -12/26 = -6/13$.

$b_1 = \overline{y} - b_2 \overline{x} = 3 - (-6/13)4 = 3 + 24/13 = 63/13$.

(d) [2 points] In terms of the above estimates, provide an estimate of the value of $E[Y \mid X = 5]$.

Solution. *We estimate* $E[Y \mid X = 5]$ *by* $b_1 + b_2(5) = (63/13) - (30/13) = 33/13$.

7. Suppose we have a linear regression model with $E[Y \mid X = x] = \beta_1 + \beta_2 x$ and $\mathrm{Var}[Y \mid X = x] = \sigma^2$, with β_1 and β_2 and σ^2 unknown. Suppose we observe the following three pairs (x_i, y_i): $(3, 7)$, $(4, 7)$, $(5, 10)$. It is a fact (which you may use) that this leads to the least-squares estimates $b_1 = 2$ and $b_2 = 3/2$.

(a) [3 points] In terms of these valus, what is the value of the estimate s^2 for σ^2?

Solution. $s^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2 = \frac{1}{1}[(7 - 2 - (3/2)(3))^2 + (7 - 2 -$

$(3/2)(4))^2 + (10 - 2 - (3/2)(5))^2] = (0.5)^2 + (-1)^2 + (0.5)^2 = (1/4) + 1 + (1/4) = 3/2$.

(b) [3 points] In terms of these valus, what are the values of RSS and ESS?

Solution. *Here* $\overline{x} = (3+4+5)/3 = 12/3 = 4$. *Thus,* $RSS = (b_2)^2 \sum_{i=1}^{n}(x_i - \overline{x})^2 = (3/2)^2[(3-4)^2 + (4-4)^2 + (5-4)^2] = (9/4)[1+0+1] = 18/4 = 9/2$.

$ESS = (n-2)s^2 = (1)s^2 = (1)(3/2) = 3/2$.

(c) [3 points] In terms of these valus, what is the value of the F statistic?

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

(d) [1 point] In general, if in fact we had $\beta_2 \approx 0$, then which would we expect the F statistic to be: (i) negative; (ii) close to 0; (iii) close to 1; or (iv) very large?

Solution. *If* $\beta_2 \approx 0$, *then usually* $F \approx 1$, *so (iii) is correct.*