STA 312f12 Assignment One[1]

Do this review assignment in preparation for the quiz on Friday, Sept. 21st. The problems are practice for the quiz, and are not to be handed in.

The first part of this assignment is based on material that you probably know already. However, the notation used in Statistics can be an obstacle for some students when they are dming maximum likelihood problems, so we will review the following basic rules.

- The distributive law: $a(b+c) = ab + ac$. You may see this in a form like

$$\theta \sum_{i=1}^n x_i = \sum_{i=1}^n \theta x_i$$

- Power of a product is the product of powers: $(ab)^c = a^c\, b^c$. You may see this in a form like

$$\left(\prod_{i=1}^n x_i \right)^\alpha = \prod_{i=1}^n x_i^\alpha$$

- Multiplication is addition of exponents: $a^b a^c = a^{b+c}$. You may see this in a form like

$$\prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n \exp(-\theta \sum_{i=1}^n x_i)$$

- Powering is multiplication of exponents: $(a^b)^c = a^{bc}$. You may see this in a form like

$$(e^{\mu t + \frac{1}{2}\sigma^2 t^2})^n = e^{n\mu t + \frac{1}{2} n \sigma^2 t^2}$$

- The log (that's the natural log, ln on many calculators) of a product is sum of logs: $\log(ab) = \log(a) + \log(b)$. You may see this in a form like

$$\log \prod_{i=1}^n x_i = \sum_{i=1}^n \log x_i$$

- The log of a power is the exponent times the log: $\log(a^b) = b\, \log(a)$. You may see this in a form like

$$\log(\theta^n) = n \log \theta$$

- The log is the inverse of the exponential function: $\log(e^a) = a$. You may see this in a form like

$$\log \left(\theta^n \exp(-\theta \sum_{i=1}^n x_i) \right) = n \log \theta - \theta \sum_{i=1}^n x_i$$

[1]This assignment was prepared by Jerry Brunner, Department of Statistics, University of Toronto. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/~brunner/oldclass/312f12

1. Choose the correct answer.

   (a) $\prod_{i=1}^{n} e^{x_i} =$

      i. $\exp(\prod_{i=1}^{n} x_i)$
      ii. e^{nx_i}
      iii. $\exp(\sum_{i=1}^{n} x_i)$

   (b) $\prod_{i=1}^{n} \lambda e^{-\lambda x_i} =$

      i. $\lambda e^{-\lambda^n x_i}$
      ii. $\lambda^n e^{-\lambda n x_i}$
      iii. $\lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)$
      iv. $\lambda^n \exp(-n\lambda \sum_{i=1}^{n} x_i)$
      v. $\lambda^n \exp(-\lambda^n \sum_{i=1}^{n} x_i)$

   (c) $\prod_{i=1}^{n} a_i^b =$

      i. na^b
      ii. a^{nb}
      iii. $(\prod_{i=1}^{n} a_i)^b$

   (d) $\prod_{i=1}^{n} a^{b_i} =$

      i. na^{b_i}
      ii. a^{nb_i}
      iii. $\sum_{i=1}^{n} a^{b_i}$
      iv. $a^{\prod_{i=1}^{n} b_i}$
      v. $a^{\sum_{i=1}^{n} b_i}$

   (e) $\left(e^{\lambda(e^t-1)}\right)^n =$

      i. $ne^{\lambda(e^t-1)}$
      ii. $e^{n\lambda(e^t-1)}$
      iii. $e^{\lambda(e^{nt}-1)}$
      iv. $e^{n\lambda(e^t-n)}$

   (f) $\left(\prod_{i=1}^{n} e^{-\lambda x_i}\right)^2 =$

      i. $e^{-2n\lambda x_i}$
      ii. $e^{-2\lambda \sum_{i=1}^{n} x_i}$
      iii. $2e^{-\lambda \sum_{i=1}^{n} x_i}$

2. True, or False?

   (a) $\sum_{i=1}^n \frac{1}{x_i} = \frac{1}{\sum_{i=1}^n x_i}$
   (b) $\prod_{i=1}^n \frac{1}{x_i} = \frac{1}{\prod_{i=1}^n x_i}$
   (c) $\frac{a}{b+c} = \frac{a}{b} + \frac{a}{c}$
   (d) $\log(a+b) = \log(a) + \log(b)$
   (e) $e^{a+b} = e^a + e^b$
   (f) $e^{a+b} = e^a e^b$
   (g) $e^{ab} = e^a e^b$
   (h) $\prod_{i=1}^n (x_i + y_i) = \prod_{i=1}^n x_i + \prod_{i=1}^n y_i$
   (i) $\log(\prod_{i=1}^n a_i^b) = b \sum_{i=1}^n \log(a_i)$
   (j) $\sum_{i=1}^n \prod_{j=1}^n a_j = n \prod_{j=1}^n a_j$
   (k) $\sum_{i=1}^n \prod_{j=1}^n a_i = \sum_{i=1}^n a_i^n$
   (l) $\sum_{i=1}^n \prod_{j=1}^n a_{i,j} = \prod_{j=1}^n \sum_{i=1}^n a_{i,j}$

3. Simplify as much as possible.

   (a) $\log \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}$
   (b) $\log \prod_{i=1}^n \binom{m}{x_i} \theta^x (1-\theta)^{m-x_i}$
   (c) $\log \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$
   (d) $\log \prod_{i=1}^n \theta(1-\theta)^{x_i-1}$
   (e) $\log \prod_{i=1}^n \frac{1}{\theta} e^{-x_i/\theta}$
   (f) $\log \prod_{i=1}^n \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-x_i/\beta} x_i^{\alpha-1}$
   (g) $\log \prod_{i=1}^n \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-x_i/2} x_i^{\nu/2-1}$
   (h) $\log \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$
   (i) $\prod_{i=1}^n \frac{1}{\beta-\alpha} I(\alpha \leq x_i \leq \beta)$ (Express in terms of the minimum and maximum y_1 and y_n.)

4. For each of the following distributions, derive a general expression for the Maximum Likelihood Estimator (MLE). You don't have to do the second derivative test. Then use the data to calculate a numerical estimate.

   (a) $P(y|\pi) = \pi(1-\pi)^y$ for $y = 0, 1, \ldots,$ where $0 < \pi < 1$. Data: `4, 0, 1, 0, 1, 3, 2, 16, 3, 0, 4, 3, 6, 16, 0, 0, 1, 1, 6, 10`. Answer: 0.2061856
   (b) $P(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$ for $y = 0, 1, \ldots,$ where $\lambda > 0$. Data: `7, 7, 6, 4, 2, 5, 2, 3, 7, 2`. Answer: 4.5
   (c) $Pr\{Y = 1\} = \pi$ and $Pr\{Y = 0\} = 1 - \pi$, where $0 < \pi < 1$. Data: `1 1 1 1 0 1 0 1 0 0 1 0 1 0`. Answer: 0.5714286

5. Read Chapter 1 (Introduction) in the textbook, and do Problems 1.1, 1.2, 1.3, 1.7, 1.8, 1.9, 1.10, 1.12, 1.16 and 1.18. In Problem 1.16, the experiment is viewed as generating a *single* observation from a binomial distribution.

6. Nothing is perfect, and that definitely applies to medical tests. Suppose a blood test is used to detect a certain kind of thyroid disease. The *prevalence* of the disease is the probability that a randomly chosen person actually has the disease. Even with a perfect test, this could never be known exactly without testing the whole population. The *sensitivity* of the test is a conditional probability. It is the probability that a person who actually has the disease will test positive. The *specificity* of the test is another conditional probability. It is the probability that a person who does not have the disease will test negative.

   Suppose that the sensitivity of the test is 95%, the specificity is 90%, and the underlying rate of the disease is one percent. So it's a pretty good test for a rare condition.

   (a) What proportion of people in the population will test positive for the disease? The answer is a single number. Show your work. My answer is 0.1085, or 10.85%. The moral of the story is that if you confuse prevalence with the probability of testing positive (a very natural mistake), this good test can make you think the prevalence of the disease is nearly eleven times as great as it actually is.

   (b) What proportion of patents who test positive actually have the disease? My answer is $\frac{95}{1085} = 0.088$. So, a bit over 91% of the medical treatments (surgery, etc.) based on this good diagnostic test are unnecessary and maybe harmful.

7. This problem establishes a result that will be used later in our course. It requires you to remember that the sum of two independent Poisson random variables also has a Poisson distribution. So, let the number of girls born on one day in Toronto have a Poisson distribution with parameter λ_1, and let let the number of boys have a Poisson distribution with parameter λ_2. These two random variables are independent, and the whole thing is based upon the reasonable assumption that boys and girls are being born according to independent Poisson processes – see class notes.

   Anyway, *given* that n babies were born on a particular day, what is the probability distribution of the number of girls born on that day? Show all your work.

   Here are two hints. First, the word "given" is a clue that you are being asked for a conditional probability. The second hint is that you should start by specifying the possible values of the number of girls born, given that a total of n babies were born. This is the *support* of the distribution for which you are being asked.