Stat 579: Homework 9 – December 3, 2007

1. A psychology student walked in one morning to a drop-in statistical consulting session with the following realizations from the Poisson distribution with parameter λ:

   `1, 2, 5, 3, 2, 6, 1, 5, 7, 8, 9, 4, 7, 2, 8, 6, 8.`

   The student is interested in making inference on λ and to test whether it is at least 7. She wants to use as her test statistic the maximum likelihood estimate of λ. In passing, she makes the following comment:

   Oh, I got rid of all the zeros. Why carry unnecessary baggage? Zeros never add up to anything.

   Thus, it transpires that the distribution she samples from is actually a Poisson that is truncated to be positive. Hence, the density for this distribution is given by:

   $$f(x;\lambda) = \frac{\exp(-\lambda)\lambda^x}{x!(1-\exp(-\lambda))}; \qquad x = 1, 2, \ldots.$$

   To calculate the MLE for an observed sample $x_1, x_2, \ldots, x_n$, we have the following loglikelihood:

   $$\ell(\lambda; x_1, x_2, \ldots, x_n) = -n\lambda - n\log(1-\exp(-\lambda)) + \sum_{i=1}^{n} x_i \log\lambda$$

   (a) Write a function with *named* input of arguments λ and the observation vector $\boldsymbol{x}$ which returns the value of the above log-likelihood function. Make sure that your function works with the above dataset and some test values of λ (your choice). [*15 points*]

   (b) Using the R function `optimize()`, obtain a MLE for the given dataset, by maximizing the above function. [*Note that you may choose to use Newton-Raphson iterations, but it is somewhat slightly trickier*]. As an initializing value, you may consider using the mean of the dataset. [*10 points*]

   (c) Formalize the process in (1b) above by writing a function `mymlefunc` which takes in *only the dataset vector* and returns the MLE. Note that this function must initialize the maximizer using the mean of the given dataset and then use `optimize()` from this initialized value to return the MLE (and nothing else). [*20 points*]

   (d) Our objective is to test whether $\lambda = 7$ or less than 7. To do so, we develop a sampling distribution for the MLE of a sample from the positively constrained Poisson distribution with $\lambda = 7$. We do so via the following steps:

      i. Write down a function called `rpospois` which takes in a parameter value λ and sample size n, and returns a sample of size n from the positively constrained Poisson distribution with parameter λ. (Note that a single realization from a positively constrained Poisson distribution with parameter λ

can be obtained by reporting the first positive realization from the standard Poissom(λ) distribution.) [*15 points*]

ii. Create a matrix of size `1,000 x 17` which stores 1,000 datasets of 17 observations each from calls to `rpospois` with $\lambda = 7$. [*10 points*]

iii. Use `apply` and `mymlefunc` to get MLEs for each of the 1,000 datasets. [*10 points*]

iv. Plot a histogram of the MLEs. On the same plot, draw a vertical line corresponding to the observed test statistic in (1b). [*10 points*]

v. Calculate the p-value for the observed test statistic in (1b) by reporting the proportion of values in (1(d)iii) that are below the observed test statistic. State your conclusion on whether λ for the dataset is significantly smaller than 7. [*10 points*]