STA561/CS571 — Fall 2013 **Homework 1** Due: September 9, 2013

Instructions: Please put all answers in a single PDF with your name and NetID and upload to SAKAI before class on the due date (there is a LaTeX template on the course web site for you to use). Do not hand in your R code this week, only the results. Definitely consider working in a group; please include the names of the people in your group and write up your solutions separately. If you look at any references (even wikipedia), cite them. If you happen to track the number of hours you spent on the homework, it would be great if you could put that at the top of your homework to give us an indication of how difficult it was.

Problem 1

Probability. If you throw two fair, six-sided dice, what is the probability that there will be at least one 5?

Problem 2

Bayes Rule. There is a 0.1% chance that I have a certain disease. The test for this disease is 90% accurate for positive test results (i.e., p(test positive | have disease) = 0.9) and 80% accurate for negative test results (i.e., p(test negative | don't have disease) = 0.8). What is the probability that I have the disease given that I have tested positive?

Problem 3

Uniform Distribution. Let X be an independent and identically distributed (i.i.d.) collection of random variables from a Uniform distribution with parameters a and b, $X \sim uniform(x|a, b)$, where $a = 0$ and $b = \frac{1}{2}$.

(a) What out the probability density function (pdf) of X?

(b) What is the p($X = 0.00027|a, b$) (according to the pdf)?

(c) What is the Pr($X = 0.00027|a, b$) (the probability that X=0.00027)?

Problem 4

Poisson Distribution. Let X be an i.i.d. collection of random variables form a Poisson distribution with parameter λ, $X \sim pois(\lambda)$, where $\lambda > 0$.

(a) Write out the exponential family form of X.

(b) Determine the sufficient statistic of X for the Poisson distribution ($T(x)$).

(c) Write out the log-partition function of X ($A(\eta)$).

(d) Determine the response function (*hint:* find the inverse of the link function).

(e) Determine $\mathbb{E}(X)$ and $var(X)$ (you can either derive from the Poisson distribution or use the log-partition function).

(f) Write out the maximum likelihood estimate (MLE) of λ for data $X = \{x_1, ..., x_n\}$.

(g) Now, let $\lambda \sim Ga(\alpha, \beta)$, where $\alpha, \beta > 0$. The gamma distribution is conjugate to the Poisson distribution, $Ga(x|\alpha, \beta) \propto x^{\alpha-1}/e^{\beta x}$. Write out the MAP of λ given $X = \{x_1, ..., x_n\}$.

Problem 5

MAP and MLE simulation. For this problem, please download the data from Sakai or the course website labelled HW1.txt. This file contains a column array of $10,000$ integer values, where each value represents the number of customers that entered a 24 hour laundromat in one hour time intervals, over 10,000 hours. Let $X = \{x_1, ..., x_t, ..., x_n\})$ represent this column array, where x_t is the number of customers that entered during the t^{th} hour. The hourly arrival of customers can be modeled as a collection of i.i.d. random variables drawn from a Poisson distribution, $X \sim pois(\lambda)$, where λ is the *hourly arrival rate.*

(a) Plot a histogram of X using 25 bins.

(b) Using your answer from Problem 4(f), compute the MLE of λ for the observed data X.

For parts c - e, model the *hourly arrival rate* λ as having a Gamma distribution, $\lambda \sim Ga(\alpha, \beta)$, where $\alpha, \beta > 0$. Use your answer from Problem 4(g) to:

(c) Compute the MAP of λ given X for $\alpha = 1$ and $\beta = 1$.

(d) Compute the MAP of λ given X for $\alpha = 100$ and $\beta = 1$.

(e) Compute the MAP of λ given X for $\alpha = 10$ and $\beta = 1$.

(f) Which approximation of λ in parts b - e do you think is the best? How much does the prior distribution, and parameterizations of the prior in particular, impact the MAP estimates of λ? (one or two sentences)