# Assignment 1

STAT314/STAT461

Set: Tue, July-27. Due: Fri Aug-06

**Please type everything in either Word or LaTex, and submit it as a PDF file via Learn. No handwritten submissions! (And no scans of handwritten submissions, please).**

**Show your workings: equations for theoretical problems; code and (relevant) output for the computational problems. It is not sufficient to report an answer. Don't forget to include intermediate steps and explain your way of thinking. You may get points for thinking in the right direction even if you don't get the answer exactly right.**

**If you need an extention, please ask for it in advance. Late submissions will not be accepted.**

## Problem 1. Inverse Probability.

A total of 38 ancient manuscripts have so far been found in a certain area of England. Of those, 20 have been solidly attributed to the Historian A, 17 to the Historian B, and only 1 to the Historian C. All the chronicles pertain to the legendary King Arthur. Historian A tends to mention him on average 5 times per page, Historian B - 3 times per page, and Historian C is a fan and mentions King Arthur about 10 times per page.

Assume Poisson distribution for the number of mentions per page, so that you can use Poisson p.d.f. to evaluate the probability of a specific number of references to King Arthur on a page.

(a) If a single page is found from the same time period, and there is no mention of King Arthur, what is the probability that it was written by the historian C? (**1pt**)

(b) What assumptions have you made in the process of your analysis? List a couple; explain why you think they apply; and give counterexamples when they would not. (**1pt**)

## Problem 2: Maximum Likelihood and Bayesian inference.

Exponential distribution is often used to model time-to-event, such as how long is the wait until the next bus.

Let $x_1, ..., x_n$ denote observed waiting times between the arrival of the Orbiter Bus at the Tower Junction bus stop. Assume, they have an exponential distribution with some positive parameter $\lambda$:

$$x_1, ..., x_n | \lambda \sim \exp(\lambda)$$

so that the probability density function is

$$f(x_i) = \lambda \exp^{-\lambda x_i} \qquad \text{for } x_i \geq 0.$$

(a) Prove, that the function above is a valid p.d.f. (**1pt**)

(b) Derive the Maximum Likelihood Estimator (MLE) for $\lambda$. (*Remember the second-order condition!*) **1pt**

(c) Assume a gamma prior distribution for $\lambda$ with a hyperparameters $\alpha_0$ and $\beta_0$:

$$\lambda \sim \text{Gamma}(\alpha_0, \beta_0)$$

Use Bayes' Formula to derive the posterior distribution for $\lambda$. (*Use proportionality to make things easier*). **2pt**

(d) Show that as $n \to \infty$, the posterior mean $E(\lambda|x_1, ..., x_n, \alpha_0, \beta_0) \to \text{MLE}(\lambda)$ **1pt**

## Problem 3: Prior distribution.

(Con'd from Problem 2.)

The time-table for the Orbiter says that the buses come every 15 minutes. What parameters $\alpha_0$ and $\beta_0$ of the prior distribution for $\lambda$ reflects the prior knowledge, that the average waiting time should be about $15 \pm 1$ minutes? (**1pt**)

## Problem 4: Prior Predictive Distribution.

One way to check your model is to simulate some data based on it and see whether it fits with the data actually observed.

Given the parameters estimated above, produce $10^4$ realisations of $\lambda$ - average waiting times, and then produce $10^4$ realisations of $x$ - individual waiting times. The resulting sample of $x$ reflects the *prior predictive distribution* $f(\tilde{x}|\alpha_0, \beta_0)$.

You have currently been waiting for the bus for at least 20 min. Does your model support his particular observation? **2pt**.

(Hint: what proportion of your randomly simulated $x$ are at least 20 min?)