**STAT 611 600: Theory of Inference** **Spring 2021**

# Homework 2: due Thursday, February 4, 2021, 11:59 pm CDT

*Professor: Tiandong Wang* *Name:* *UIN:*

**Instructions**:

- Whether you write out the solution by hand or in a text document, be sure that they are neat, legible and in order (even if you choose to solve them in different order). We highly recommend that you write your solutions in **LaTeX** and print them to a **PDF file**.
- Write/Type your name, UIN at the top of the first page. Otherwise, your submission will not be graded.
- Either scan or print your solutions to a **PDF file** under 15MB in size. It must be in a single file, not separate files for separate pages. Do not take a photo of each page and then paste them into a document - this will make your file too big and the results will generally not be very readable anyway.
- All students should login to their eCampus account to upload your file. You must do this by **11:59 pm U.S. Central time**, on the due date. You can make multiple submissions, but only the last submission will be graded.
- Write down all of your problem-solving process and cite any resources you have used in addition to lecture notes and the textbook.
- It is prohibited to share or distribute the content in this document.

1. Let $(X_1, \ldots, X_n)$ be iid with common pdf

$$f(x;\theta) = \theta^{-1} e^{-(x-\theta)/\theta} \mathbf{1}_{\{x \geq \theta\}},$$

where $\theta > 0$ is an unknown parameter.

   (a) Find a statistic that is minimal sufficient for $\theta$.
   (Hints: Does this distribution belong to the exponential family? If not, you may try proving the statement using Theorem 6.2.13 in C&B. )

   (b) Is the minimal sufficient statistic in (a) complete? Justify your answer.

2. (Network Inference) Consider the presence or absence of a friendship link between each pairs of users (e.g. on Facebook). Suppose we model these observations as independent Bernoulli random variable $X_{ij} \sim \text{Bernoulli}(p_{ij})$, where $X_{ij} = 1$ means the presence of a link between users $i$ and $j$ ($i < j$; $i, j \in \{1, \ldots, n\}$), and $X_{ij} = 0$ is the absence of a link. Hence, we can represent the observations as an $n \times n$ symmetric, zero-diagonal matrix $\boldsymbol{X} = (X_{ij})$.

   (a) Find a minimal sufficient statistic for the model parameter $\theta = (p_{ij})_{i<j}$. Justify sufficiency and minimality. (Hint: Find the natural parametrization for this particular model.)

   (b) It is often the case that each user tend to interact with other users that are "popular". To capture this tendency, we consider a lower-dimensional, logistic-regression-like model family, in which

$$p_{ij} = \frac{\exp(\beta_i + \beta_j)}{1 + \exp(\beta_i + \beta_j)},$$

   where $\beta_i \in \mathbb{R}$ is a popularity measure for user $i$. Find a minimal sufficient statistics for the model parameter $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_n)$. Prove sufficiency and minimality.
   **Hints:**
      i. $\frac{p_{ij}}{1-p_{ij}} = \exp(\beta_i + \beta_j)$.
      ii. You may be able to get a simple statistic by noting $x_{ij} = x_{ji}$ as the matrix $\boldsymbol{X}$ is symmetric.

   (c) Is your statistic from part (b) complete? Justify your answer.

3. **Optional exercises**: 6.15 & 6.17 in C& B.