STA303H1S/STA1002HS Assignment 1
Due on 23rd July, 2020 11:59 PM in Quercus
All relevant work must be shown for credit.

Note: In any question, if you are using `R`, all `R` codes and `R` outputs must be included in your answers. You should assume that the reader is not familiar with R outputs and so explain all your findings, quoting necessary values form your outputs. Please note that academic integrity is fundamental to learning and scholarship. You may discuss questions with other students. However, the work you submit should be your own. If I feel suspicious of any assignment (e.g. if your work doesn't appear to be consistent with what we have discussed in class), I will not mark the assignment. Instead, I will ask you to present your work in my office and your grade will be assigned based on your presentation. Assignments can be hand written but the `R` codes and outputs should be printed.

1. Let $(Y_1, Y_2, ..., Y_K) \sim \text{Multinomial}(n, \pi_1, \pi_2, ..., \pi_k)$, then

   (a) Calculate the moment generating function (This will be a multivariate MGF). [3 Marks]

   (b) Using the MGF from (a),

      i. Show that $E(Y_j) = n\pi_j$ [2 Marks]

      ii. Show that $\text{Var}(Y_j) = n\pi_j(1 - \pi_j)$ [4 Marks]

      iii. Show that $\text{Cov}(Y_i, Y_j) = -n\pi_i\pi_j$ [6 Marks]

   (c) Let's assume $Y_1, Y_2, ..., Y_C \sim \text{Multinomial}(n, \pi_1, \pi_2, ..., \pi_C)$. Then show that the correlation coefficient between Y_i and Y_j is,

   $$\text{Cor}(Y_i, Y_j) = \frac{-\pi_i\pi_j}{\sqrt{\pi_i(1 - \pi_i)\pi_j(1 - \pi_j)}}$$

   [3 Marks]

   (d) Show that for $C = 2$ the $\text{Cor}(Y_1, Y_2) = -1$. Explain why? [2 Marks]

   Hint: You can use the Multivariate MGF. Recall if $\mathbf{Y} = (Y_1, Y_2, ..., Y_k)$ be a multivariate random variable. Then the MGF is defined as,

   $$M_{\mathbf{Y}}(t_1, ..., t_k) = \text{E}(\exp(t_1Y_1 + ... + t_kY_k))$$

   Now use partial derivative on t_j or t_i, t_j to achieve the results

2. If Y_1 and Y_2 are independent Poisson random variable with parameters μ_1 and μ_2 respectively. Then find the conditional distribution of Y_1 given $Y_1 + Y_2 = n$. That is calculate $P(Y_1 = k|Y_1 + Y_2 = n)$. [5 Marks]

3. Let $Y \sim \text{Bin}(n = 30, \pi = 0.9)$. Y can be interpreted as the number of successes in a sample of size $n = 30$ from a Binomial distribution with probability of success $\pi = 0.9$.

(a) Let the observed number of success after 30 trials is $y = 27$. Calculate Wald and score (Wilson) 95% confidence interval. [5 Marks]

(b) Simulate $N = 100,000$ observations of Y using `R` function `rbinom()`. Calculate the Wald and Score 95% confidence interval for each of the observations. This means you are calculating $100,000$ confidence intervals of each type. Calculate the proportion of these Wald intervals that contain 0.9 (the true value of π). Also calculate the proportion of score intervals that contain 0.9. Compare the results and comment on your findings. Which one do you feel is a more reliable CI. [10 Marks]

Note: `R` cannot generate random numbers. It only generates "pseudo" random numbers. Thus a seed needs to be provided to reproduce the results. One can fix the seed in `R` using the `set.seed()` command. The seed you are going to use is your student ID. Thus you have to start the code with `set.seed(Your student ID)`. If you don't provide the seed you will loose 3 Marks.

4. Same as the previous question Let $Y \sim \text{Bin}(n = 30, \pi)$ and $y = 27$. This time we don't know the true value of π

   (a) Find the likelihood ($\ell(\pi)$) and log-likelihood function ($L(\pi)$) [2 Marks]

   (b) Using `R`, find the maximum likelihood estimate of π and plot $\ell(\pi)$ and $L(\pi)$ over the values of π. [5 Marks]

   (c) Test $H_0 : \pi_0 = 0.5$ vs $H_a : \pi \neq 0.5$ using the likelihood ratio test. [3 Marks]

   (d) Using `R` calculate the 95% likelihood ratio confidence interval for π [5 Marks]

5. (a) Perform the following simulation (for this please set the seed to your student ID),

   - Generate 500 random values from $X_1 \sim \text{Uniform}[-10, 10]$, $X_2 \sim \text{N}(0, 4)$ and $X_3 \sim \text{Bernoulli}(0.7)$
   - Set $\boldsymbol{\beta} = (-0.8, 0.1, 0.2, 0.3)$
   - Simulate $Y_i \sim \text{Poisson}(\mu_i)$, where, $\mu_i = \exp(\sum_j x_{ij}\beta_j)$

   [10 Marks]

   (b) Estimate the βs using Iteratively Weighted Least Square (IRLS) method by writing your own function. Explain the procedure and state the W matrix as mentioned in lecture 4. Compare the results with `glm` code in `R` [15 Marks]