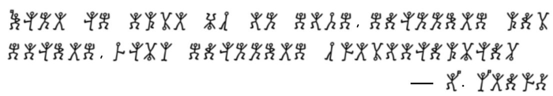
STA3020 Assignment 4.

[1] *School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)*



*You may pick 5 out of this 6 questions as your homework.

1. Composite Likelihood

•*Exercise* 1.1 (♣ **Dyadic Regression**). Consider investigating the international exports (Denoted as X_{ijt}) and imports data (Denoted as M_{ijt}) (WTO-OECD Balanced Trade in Services Dataset (BaTiS) — BPM6 data) in "A4Q1-1.rds" together with the term explanation file. Assume a gravity model, i.e.,

$$X_{ijt} = M_{jit} = \exp\Big(\beta_0 + \beta_1 \mathrm{GDP}_{it} + \beta_2 \mathrm{GDP}_{jt} + \beta_3 \text{Trade Costs}_{ij} + \beta_4 \text{Other Predictors}_{ijt}\Big)\eta_{ijt}$$

where we have η_{ijt}, $1 \le i, j \le n$ and $1 \le t \le T$ are i.i.d following Gamma(α, α^{-1}) s.t. $\mathbb{E}\eta_{ijt} = 1$, α.

The GDP data (together with the Tariffs data) can be found in

https://databank.worldbank.org/source/world-development-indicators#

The Trade Costs data (the contry-specific data geo_cepii.xls) can be found in "geo_cepii.xls".

we use three Trade Costs factors, the bilateral distance, common language, and the presence of common colonial relationship, i.e.,

$$\begin{aligned}\beta_3 \text{Trade Costs}_{ij} = \beta_{31} \cdot \Big((\mathrm{lat}_i - \mathrm{lat}_j)^2 + (\mathrm{lon}_i - \mathrm{lon}_j)^2\Big)^{1/2} \\ + \beta_{32} \cdot \mathbb{1}\Big(\mathrm{Language}_i \cap \mathrm{Language}_j \neq \emptyset\Big) \\ + \beta_{33} \cdot \mathbb{1}\Big(\mathrm{Colonizer}_i \cap \mathrm{Colonizer}_j \neq \emptyset\Big)\end{aligned}$$

Similarly, we use one other factors, the absolute difference of the tariff rate (Tariff rate, applied, weighted mean, primary products) of two contries, i.e.,

$$\beta_4 \text{Other Predictors}_{ijt} = \beta_4 \cdot \Big|\mathrm{Tariff}_{it} - \mathrm{Tariff}_{jt}\Big|$$

(i) Please estimate the parameters (α, β) using the composite likelihood approach.

(ii) Use the likelihood ratio test to test the hypothesis

$$H_0 : \tilde{\alpha} = \acute{\alpha}, \;\; \tilde{\beta} = \acute{\beta} \quad v.s. \quad H_1 : \tilde{\alpha} \neq \acute{\alpha}, \;\; \text{or} \;\; \tilde{\beta} \neq \acute{\beta}$$

where $(\tilde{\alpha}, \tilde{\beta})$ is the parameters for APEC (Asia-Pacific Economic Cooperation) counties, and $(\acute{\alpha}, \acute{\beta})$ is the parameters for EU (European Union) counties. In other words, we want to see whether there are any differences in the model parameters for these two regions.

(iii) Please visualize this network trading data for the APEC counties (visualize the counties' GDP, trading X_{ijt} or M_{ijt} with other counties (you don't have to visualize the geo distance), examples are in "https://kateto.net/network-visualization").

•*Exercise* 1.2 (Model Misspecification). For a sequence of white noise $\epsilon_1, \cdots, \epsilon_n \sim_{i.i.d} N(0,1)$, we observe a random sample $X_1, \cdots, X_n$ with each $X_i = (X_{i1}, X_{i2}, X_{i3})$ in "A4Q1-2.rds", the data were generated in the following way, $X_{i1} \sim N(\mu_1, \sigma_1^2)$, $X_{i2} \sim N(\mu_2, \sigma_2^2)$ is independent with X_{i1}, and

$$X_{i3} = X_{i1} + X_{i2} + \beta X_{i1} X_{i2} + \epsilon_i.$$

Notice that $X_{i1}|(X_{i2}, X_{i3})$, $X_{i2}|(X_{i1}, X_{i3})$ and $X_{i3}|(X_{i1}, X_{i2})$ all follow normal distribution.

(i) Please give the mean vector $\mu(\beta) = \mathbb{E}X_i$ and the covariance matrix $\Sigma(\beta) = \text{Cov}(X_i)$.

(ii) Since all marginal distribution function are normal, if we misspecified the joint model to be

$$X_i \sim N\left(\mu(\beta), \Sigma(\beta)\right),$$

please use the data "A4Q1-2.rds" to give an estimate of $\theta = (\mu_1, \mu_2, \beta, \sigma_1^2, \sigma_2^2)$.

(iii) If we only uses the marginal distribution to construct our composite likelihood, please give the MCLE of β (denote it as $\hat{\beta}$) and please use the data "A4Q1-2.rds" to give an estimate of $\theta = (\mu_1, \mu_2, \beta, \sigma_1^2, \sigma_2^2)$.

2. Quasi Likelihood

•*Exercise* 2.1 (♣ **Poisson Regression Model Using Generalized Estimating Equation**). Assume we have random sample $Y = (Y_1, \cdots, Y_n)$, each of them being a binary observation, i.e.,

$$Y_i \sim Poisson(\lambda_i), \;\; i = 1, \cdots, n,$$

and

$$\log \lambda_i = \alpha + x_i \beta, \;\; i = 1, \cdots, n,$$

where $\alpha, \beta \in \mathbb{R}$ are parameters of interests and $x_i \in \mathbb{R}$, $1 \le i \le n$ are known covariates. Please write out the generalized estimating equation (quasi-likelihood approach), the iterative equation of Newton's method, and obtain an estimate (together with the code) using the data "A4Q2-1.rds".

3. Profile Likelihood

•*Exercise* 3.1 (♣ **Bivariate Three Sample Problem**). Suppose we observed three random samples independently. For the first random sample, we observed

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(0_{2\times 1}, \begin{pmatrix} \sigma_{11}^2 & \rho\sigma_{11}\sigma_{22} \\ \rho\sigma_{11}\sigma_{22} & \sigma_{22}^2 \end{pmatrix}\right), \quad i = 1, \cdots, n,$$

where we suppose ρ is a known constant. For the second random sample, we only observed the observations from the first coordinate, i.e.,

$$X_i' \sim N\left(0, \sigma_{11}^2\right), \quad i = 1, \cdots, m.$$

For the third random sample, we only observed the observations from the second coordinate with the same sample size as the second random sample, i.e.,

$$Y_i' \sim N\left(0, \sigma_{22}^2\right), \quad i = 1, \cdots, m.$$

Please give an 95% confidence interval of $(\sigma_{11}^2, \sigma_{22}^2)$ using Wilk's theorem.

4. Generalized Profile Likelihood

•*Exercise* 4.1 (♣ **Normal Mixture**). Assume $X = \{X_1, \cdots, X_n\}$ is a random sample from a simple normal mixture distribution each with distribution function,

$$f(x_i|p, \mu_1, \mu_2, \sigma^2) = \frac{p}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x_i-\mu_1)^2}{2\sigma^2}\Big) + \frac{(1-p)}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x_i-\mu_2)^2}{2\sigma^2}\Big).$$

(i) Please find the methods of moment estimator $\hat{p}$ of p for every given (μ_1, μ_2, σ^2) .

(ii) By plug $\hat{p}$ into the likelihood to obtain a generalized profile likelihood and please obtain an estimate (together with the code) for all the parameters using this generalized profile likelihood and the data "A4Q4-1.rds".

•*Exercise* 4.2 (♣ **Gumbel Distribution**). Assume $X = \{X_1, \cdots, X_n\}$ is a random sample from the Gumbel distribution each with distribution function,

$$f(x_i|\gamma, \beta) = \frac{1}{\beta} \exp\Big[-\frac{x_i-\mu}{\beta} - e^{-\frac{x_i-\mu}{\beta}}\Big].$$

Notice that the cumulative distribution function of Weibull distribution is

$$F(x) \triangleq F(x|\mu, \beta) = \mathbb{P}\Big(X_1 \le x|\mu, \beta\Big) = \exp\Big[-e^{-\frac{x-\mu}{\beta}}\Big].$$

(i) Now, for arbitrary fixed z_1 and z_2, please construct an estimator $\hat{\beta}$ of β using

$$\hat{F}(z_1) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq z_1) \xrightarrow{p} F(z_1),$$
$$\text{and } \hat{F}(z_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq z_2) \xrightarrow{p} F(z_2).$$

(ii) By plug $\hat{\beta}$ into the likelihood to obtain a generalized profile likelihood and please obtain the maximum generalized profile likelihood estimator of μ.