Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Problem class sheet 4 (DRAFT, TO BE REFINED AFTER THE PROBLEM CLASS)

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Exercise 1. ($\star\star$) Consider the model

$$Z = BZ + (I - B)\, X\beta + E$$

where $E \sim \mathrm{N}\left(0, \sigma^2 I\right)$, X is a $n \times p$ design matrix X, $\beta \in \mathbb{R}^p$, B is an $n \times n$ matrix with $[B]_{i,i} = 0$, and $(I - B)$ is non-singular.

(1) Show that

$$\mathrm{E}(Z) = X\beta$$
$$\mathrm{Var}(Z) = \sigma^2 (I - B)^{-1} \left(I - B^\top\right)^{-1}$$

(2) Show that the above model is SAR for $Z - \mathrm{E}(Z)$

(3) Compute the Maximum Likelihood Estimators (MLE) $\hat{\beta}$, and $\hat{\sigma}^2$ of β, and σ^2. Assume that $((I - B)\, X)^\top (I - B)\, X$ is non-singular.

(4) Find the sampling distribution of $\hat{\beta}$ given X.

Solution.

(1) It is

$$\begin{aligned}
\mathrm{E}(Z) &= \mathrm{E}(BZ + (I - B)\, X\beta + E) \iff \\
\mathrm{E}(Z) &= \mathrm{E}(BZ) + (I - B)\, X\beta + \mathrm{E}(E) \iff \\
(I - B)\,\mathrm{E}(Z) &= (I - B)\, X\beta + \overset{=0}{\cancel{\mathrm{E}(E)}} \iff \\
\mathrm{E}(Z) &= X\beta
\end{aligned}$$

also

$$\begin{aligned}
\mathrm{Var}((I - B)\, Z) =& \mathrm{Var}((I - B)\, X\beta + E) \\
\mathrm{Var}((I - B)\, Z) =& \mathrm{Var}(E) \\
(I - B)\,\mathrm{Var}(Z)\,(I - B)^\top =& \sigma^2 I \\
\mathrm{Var}(Z) =& (I - B)^{-1} \sigma^2 I \,(I - B)^{-\top}
\end{aligned}$$

(2) It is

$$Z - \mathrm{E}(Z) = B\,(Z - X\beta) + E$$

where $E \sim \mathrm{N}(0, \sigma^2 I)$, hence $Z - \mathrm{E}(Z)$ is a SAR model given the assumptions taken.

(3) The likelihood of Z given the parameters β, and σ^2 is

$$\begin{aligned} L\left(Z;\beta,\sigma^2\right) &= \mathrm{N}\left(Z|\mathrm{E}\left(Z\right), \mathrm{Var}\left(Z\right)\right) \\ &= \mathrm{N}\left(Z|X\beta, \left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right) \end{aligned}$$

Hence

$$\begin{aligned} -2\log\left(L\left(Z;\beta,\sigma^2\right)\right) =& -2\log\left(\mathrm{N}\left(Z|X\beta, \left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right)\right) \\ =& \log\left(\det\left(\left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right)\right) \\ &+ \left(Z-X\beta\right)^\top\left(\left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right)^{-1}\left(Z-X\beta\right) \\ =& \log\left(\det\left(\left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right)\right) \\ &+ \frac{1}{\sigma^2}\left(Z-X\beta\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right) \end{aligned}$$

The likelihood equations are

$$\begin{aligned} 0 &= \nabla_{(\beta,\sigma^2)}\left(-2\log\left(L\left(Z;\beta,\sigma^2\right)\right)\right)\Big|_{(\beta,\sigma^2)=\left(\hat{\beta},\hat{\sigma}^2\right)} \\ &= \begin{bmatrix} \frac{\partial}{\partial\beta}\left(-2\log\left(L\left(Z;\beta,\sigma^2\right)\right)\right) \\ \frac{\partial}{\partial\sigma^2}\left(-2\log\left(L\left(Z;\beta,\sigma^2\right)\right)\right) \end{bmatrix}_{(\beta,\sigma^2)=\left(\hat{\beta},\hat{\sigma}^2\right)} \\ &= \begin{bmatrix} X^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right) \\ -\frac{n}{\sigma}+\frac{1}{\sigma^3}\left(Z-X\beta\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right) \end{bmatrix}_{(\beta,\sigma^2)=\left(\hat{\beta},\hat{\sigma}^2\right)} \end{aligned}$$

So the likelihood equations are

$$\begin{aligned} 0 =& X^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right) \\ 0 =& -\frac{n}{\hat{\sigma}}+\frac{1}{\hat{\sigma}^3}\left(Z-X\hat{\beta}\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right) \end{aligned}$$

Solving the first equation wrt $\hat{\beta}$ I get

$$\begin{aligned} 0 =& X^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right) \iff \\ 0 =& X^\top\left(I-B\right)^\top\left(I-B\right)Z - X^\top\left(I-B\right)^\top\left(I-B\right)X\hat{\beta} \iff \\ X^\top\left(I-B\right)^\top\left(I-B\right)X\hat{\beta} =& X^\top\left(I-B\right)^\top\left(I-B\right)Z \iff \\ \hat{\beta} =& \left(X^\top\left(I-B\right)^\top\left(I-B\right)X\right)^{-1}X^\top\left(I-B\right)^\top\left(I-B\right)Z \end{aligned}$$

provided that $X^\top\left(I-B\right)^\top\left(I-B\right)X$ is non-singular (this is given, anyway).

Solving the second equation wrt $\hat{\sigma}^2$ I get

$$\begin{aligned}
0 =& -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\left(Z - X\hat{\beta}\right)^\top (I-B)^\top (I-B)\left(Z - X\hat{\beta}\right) \iff \\
0 =& -n + \frac{1}{\hat{\sigma}^2}\left(Z - X\hat{\beta}\right)^\top (I-B)^\top (I-B)\left(Z - X\hat{\beta}\right) \iff \\
\hat{\sigma}^2 =& \frac{1}{n}\left(Z - X\hat{\beta}\right)^\top (I-B)^\top (I-B)\left(Z - X\hat{\beta}\right)
\end{aligned}$$

(4) It is Normal as a linear combination of Normally distributed random variables. Its moments (mean and variance) are

$$\begin{aligned}
\mathrm{E}\left(\hat{\beta}|X\right) =& \mathrm{E}\left(\left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B) Z|X\right) \\
=& \left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B) \mathrm{E}(Z|X) \\
=& \left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B) X\beta \\
=& \beta
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}|X\right) =& \mathrm{Var}\left(\left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B) Z|X\right) \\
=& \left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B) \mathrm{Var}(Z|X) \\
& \left(\left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B)\right)^\top \\
=& \left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top \cancel{(I-B)^\top (I-B)} \\
& \sigma^2 \cancel{(I-B)^{-1} (I-B^\top)^{-1}} \\
& \left(\left(X^\top (I-B)^\top (I-B) X\right)^{-1} X^\top (I-B)^\top (I-B)\right)^\top \\
=& \sigma^2 \cancel{\left(X^\top (I-B)^\top (I-B) X\right)^{-1}} \cancel{X^\top (I-B)^\top (I-B) X} \left(X^\top (I-B)^\top (I-B) X\right)^{-1} \\
=& \sigma^2 \left(X^\top (I-B)^\top (I-B) X\right)^{-1}
\end{aligned}$$

- Notice that, in Frequentist Statistical framework, once we have computed the sampling distributions (those above), we can produce inference tools in the similar manner to Normal Linear regression.

Exercise 2. ($\star$) Suppose that $\mathcal{S}$ is a finite set that contains at least two elements and is equipped with a symmetric relation $\sim$. Consider the Poisson auto-regression model defined as

$$\begin{cases} y_i|y_{\mathcal{S}\setminus\{i\}} \sim & \text{Poisson}\left(\mu_i\right) \\ \log\left(\mu_i\right) = & \theta\sum_{i\sim j, j\neq i} y_j \end{cases}$$

for $y \in \mathbb{N}^{\mathcal{S}}$.

Hint: You can use that if $X \sim \text{Poisson}(\mu)$ then X has PMF

$$\Pr_X(x|\mu) = \frac{1}{x!}\exp(-\mu)\,\mu^x 1\left(x \in \{0,1,2,...\}\right)$$

(1) Show that the above model is well-defined if and only if $\theta \leq 0$.

(2) Find the canonical potential with respect to $\zeta = 0$.

Solution. It is

$$\Pr_i\left(y_i|y_{\mathcal{S}\setminus\{i\}}\right) = \frac{1}{y_i!}\exp\left(-\mu_i\right)\mu_i^{y_i} 1\left(y_i \in \mathbb{N}\right)$$

(1) It is

$$\Pr_i\left(y_i = 0|y_{\mathcal{S}\setminus\{i\}}\right) = \exp\left(-\mu_i\right) =$$

and

$$\Pr_i\left(y_i = \ell|y_{\mathcal{S}\setminus\{i\}}\right) = \frac{1}{\ell!}\exp\left(-\mu_i\right)\mu_i^{\ell}$$

for $\ell \in \mathbb{N}$ Then by the factorization theorem wrt reference 0 it is

$$\begin{aligned}
\frac{\Pr_Y(y)}{\Pr_Y(0)} &= \prod_{i\in\mathcal{S}} \frac{\Pr_i\left(y_i|y_1,...,y_{i-1},0,...,0\right)}{\Pr_i\left(0|y_1,...,y_{i-1},0,...,0\right)} \\
&= \prod_{i\in\mathcal{S}} \frac{\frac{1}{y_i!}\cancel{\exp(-\mu_i)}\mu_i^{y_i}}{\cancel{\exp(-\mu_i)}} = \prod_{i\in\mathcal{S}} \frac{1}{y_i!}\mu_i^{y_i} \\
&= \prod_{i\in\mathcal{S}} \frac{1}{y_i!}\exp\left(\theta\sum_{i\sim j, j\neq i} y_j\right)^{y_i} \\
&= \exp\left(\theta\sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}}\log\left(y_i!\right)\right)
\end{aligned}$$

That is

$$\Pr_Y(y) = \exp\left(\theta\sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}}\log\left(y_i!\right)\right)\Pr_Y(0)$$

Now, if $\theta \leq 0$ then $\theta \sum_{i\sim j, j\neq i} y_i y_j \leq 0$ hence the constant is

$$\begin{aligned}
\sum_{y\in\mathbb{N}^{\mathcal{S}}} \frac{\Pr_Y(y)}{\Pr_Y(0)} &= \sum_{y\in\mathbb{N}^{\mathcal{S}}} \exp\left(\theta \sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}} \log(y_i!)\right) \\
&\leq \sum_{y\in\mathbb{N}^{\mathcal{S}}} \exp\left(-\sum_{i\in\mathcal{S}} \log(y_i!)\right) \\
&= \sum_{y\in\mathbb{N}^{\mathcal{S}}} \prod_{i\in\mathcal{S}} \frac{1}{y_i!} \\
&= \left(\sum_{\ell=0}^{\infty} \frac{1}{\ell!}\right)^{\mathrm{Card}(\mathcal{S})} < \infty
\end{aligned}$$

If $\theta > 0$ without loss of generality consider the first two sites and suppose that $1 \sim 2$, then

$$\frac{\Pr_Y\left((y_1, y_2, 0, ..., 0)^\top\right)}{\Pr_Y(0)} = \frac{\exp(\theta y_1 y_2)}{y_1! y_2!}$$

should be summable. However, the series

$$\sum_{y_1=1}^{\infty} \sum_{y_2=1}^{\infty} \frac{\exp(\theta y_1 y_2)}{y_1! y_2!} = \infty$$

diverges as the general term does not go to zero.

(2) By definition, $V_\emptyset = 0$.

Then I will use Theorem **??**, and $\zeta = 0$.

For $\mathcal{A} = \{i\}$, it is

$$V_{\{i\}}(y) = \log\left(\Pr_i(y_i | 0, ..., 0)\right) - \log\left(\Pr_i(0|0, ..., 0)\right) = -\log(y_i!)$$

For $\mathcal{A} = \{i, j\}$, it is

$$\begin{aligned}
V_{\{i,j\}}(y) =& \log\left(\Pr_i(y_i | y_j, 0, ..., 0)\right) \\
& - \log\left(\Pr_i(y_i|0, ..., 0)\right) - \log\left(\Pr_i(y_j|0, ..., 0)\right) \\
& + \log\left(\Pr_i(0|0, ..., 0)\right) \\
=& - y_i y_j 1(i \sim j)
\end{aligned}$$

So

$$V_{\{i,j\}}(y) = -y_i y_j 1(i \sim j)$$

Since the joint distribution is proportional such as

$$\Pr_{Y}(y) = \exp\left(\theta \sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}} \log(y_i!)\right) \Pr_{Y}(0)$$
$$\propto \exp\left(\theta \sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}} \log(y_i!)\right)$$

all the other potentials are zero.

- Perhaps not the most elegant derivation. In the next lecture, we will use a more general tool to compute such stuff which is based on the "exponential distribution family".

($\star$) Show that the local characteristics

$$\Pr_{1}(x|y) = \Pr_{2}(x|y) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}(y-x)^2\right)$$

do not define a proper joint distribution on $\mathbb{R}^{\{1,2\}}$

Solution. The local characteristics are well-defined normal densities. Using Besag's factorization theorem, with reference $(0,0)^\top$, the joint density would be proportional to

$$\frac{\Pr_1(x|0)}{\Pr_1(0|0)} \frac{\Pr_2(y|x)}{\Pr_2(0|x)} = \exp\left(-\frac{1}{2}(y-x)^2\right) = G(x,y)$$

for $(x,y) \in \mathbb{R}^2$. But $G(x,y)$ is not integratable,

$$\int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}(y-x)^2\right) = \infty$$
