Draft proof MNL algorithm is a consistent network estimator

Marcela I. Cespedes

1 Model specification

Following PLOS ONE reviewer comments, here is a draft proof to show the MNL is a consistent estimator. The MNL model is as follows

$$p(\mathbf{b_i}|\sigma^2, \gamma, W) \sim MVN(\mathbf{0}, \sigma_s^2 Q)$$
$$Q^{-1} = \gamma(\hat{W} - W) + (1-\gamma)\mathbb{I} \tag{1}$$

where a set of spatial observations $\mathbf{b_i}$, for individual i is conditional on the spatial scale (σ_s^2), spatial strength $(0 \leq \gamma < 1)$ and connectivity matrix W. The set of observations $\mathbf{b_i}$ follow a multivariate normal distribution with mean zero and covariance matrix $\sigma_s^2 Q$. The identity matrix is denoted by $\mathbb{I}$ and all matrices W, $\hat{W}$ and $\mathbb{I}$ have $K \times K$ dimensions.

Symmetric matrix W has zero diagonals and binary off-diagonal values $w_{kj} = w_{jk} = 1$ to denote regions k and j are connected and 0 otherwise. Matrix $\hat{W}$ is a diagonal matrix whose elements are the row sums of W expressed as $\sum_{i \neq j}^{K} w_{ji}$ for the j^{th} row.

For example, a four region connectivity matrix W could be

$$W = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \text{ then } \hat{W} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

This covariance parametrisation was first proposed by Leroux et al. (2000). While the work by Leroux and collegues discussed several desirable properties of $\sigma_s^2 Q^{-1}$, such as separate parameters for overdispersion and spatial dependence, the authors have yet to discuss the statistical properties of this covariance estimator as the sample size increases[1].

In the sections that follow we will provide a proof-by-cases that the parametrisation of the covariance matrix, $\sigma_s^2 Q$, is a positive definite matrix for all values of γ. If our covariance estimator is positive definite, then it satisfies a softer condition that covariance matrices must be positive semi-definite in order to be a valid covariance matrix (Fujikoshi et al., 2011).

There is a large body of literature which states that the maximum likelihood estimation of a parameter is a consistent estimator, and proofs of this fundamental result will be discussed at the end of the proof presented here.

[1] I have yet to find any further work by Brian Leroux and colleagues as to how he came to derive this particular parametrisation of the precision matrix. To my knowledge the authors have not demonstrated the properties of this covariance expression being positive definite.

1.1 Case 1: W has no neighbours

Theorem 8.4.6 in Anton and Busby (2003) (page 491) state that positive diagonal matrices are positive definite. We will rely on this theorem to show that when matrix W is a zero matrix, signifying no connections or spatial associations among regions, then $\sigma_s^2 Q$ is a positive definite covariance matrix.

When W is a zero matrix, model (1) becomes

$$\begin{aligned} p(\mathbf{b_i}|\sigma^2, \gamma, W) &\sim MVN(\mathbf{0}, \sigma_s^2 Q) \\ Q^{-1} &= (1-\gamma)\mathbb{I} \end{aligned} \tag{2}$$

where Q^{-1} is a diagonal matrix with values $1-\gamma$. The inverse of a diagonal matrix is equal to the inverse of each diagonal element, hence

$$\sigma_s^2 Q = \sigma_s^2 \begin{bmatrix} 1/(1-\gamma) & 0 & 0 & ... & 0 \\ 0 & 1/(1-\gamma) & 0 & ... & 0 \\ \vdots & & \ddots & & \\ 0 & & & & 1/(1-\gamma) \end{bmatrix}$$

The diagonal elements of the covariance matrix are $\sigma_s^2/(1-\gamma)$. As this is greater than zero for all values of γ and σ_s^2 then by Theorem 8.4.6 in Anton and Busby (2003), matrix $\sigma_s^2 Q$ is a positive definite matrix.

1.1.1 Sub-proof: Positive diagonal matrices are positive definite

Let $\sigma_s^2/(1-\gamma) = \delta$ be the diagonals of matrix A and $\delta > 0$ for all $0 \leq \gamma < 1$. Let arbitrary real and non-zero vector of length K be $\mathbf{z} = [z_1, z_2, ..., z_K]^T$, then

$$\begin{aligned} \mathbf{z}^T A \mathbf{z} &= \delta \boldsymbol{z}^T \boldsymbol{z} \\ &= \delta \left(\sum_{i=1}^{K} z_i^2 \right) \end{aligned}$$

and this is greater than zero, hence $\mathbf{z}^T A \mathbf{z} > 0$ and A is a positive definite matrix.

1.2 Case 2: W has one neighbour

We first introduce a theorem on diagonally dominant martices, as described in Cheney and Kincaid (2009) (page 654) and Briggs (1999).

Theorem 1. *Matrix A is said to be strictly diagonally dominant if*

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \qquad \forall i.$$

It follows that a strictly diagonally dominant symmetric real matrix with positive diagonal entries are positive definite (Anton and Busby, 2003; Briggs, 1999). An alternative proof of this result can also be derived using Gerschgorin's theorem (Saad (2003), page 118-119). A well known property of positive definite matrices is that their inverse is also positive matrix (Anton and Busby, 2003) hence Q is also a positive definite matrix and $\mathbf{z}^T Q \mathbf{z} > 0$ holds for all real $\mathbf{z}$.

Without loss of generality, let W have a single neighbour and let that neighbour be $w_{34} = w_{43} = 1$, that is

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & \dots & & & & 0 \end{bmatrix}, \tag{3}$$

then matrix $\hat{W}$ is a zero matrix, whose only non-zero diagonal elements are $\hat{w}_{33} = \hat{w}_{44} = 1$.

Let Q_b^{-1} denote the b^{th} vector of Q^{-1} and correspond to a row where W has no neighbours, that is a row of W with zero elements. Then

$$Q_b^{-1} = [0, 0, ..., 0, 1 - \gamma, 0, \dots, 0]$$

and $1 - \gamma$ is the b^{th} element or the diagonals of Q^{-1}. Note: that all the other elements are zero, and $1 - \gamma$ is the only non-zero element. Hence for rows $b = 1, 2, 5, 6, ..., K$, the absolute values of the diagonal elements are greater than the sum of the off-diagonal elements.

The third and fourth rows of Q^{-1} are as follows

$$Q_3^{-1} = [0, 0, \gamma \sum_{i \neq 3}^{K} w_{3i} + 1 - \gamma, -\gamma, 0, \dots, 0]$$
$$Q_4^{-1} = [0, 0, -\gamma, \gamma \sum_{i \neq 4}^{K} w_{4i} + 1 - \gamma, 0, \dots, 0]$$

where the diagonal elements for rows 3 and 4 are $\gamma \sum_{i \neq 3}^{K} w_{3i} + 1 - \gamma$ and $\gamma \sum_{i \neq 4}^{K} w_{4i} + 1 - \gamma$ respectively.

As shown by matrix (3), since there is only one neighbour then $\sum_{i \neq 3}^{K} w_{3i} = \sum_{i \neq 4}^{K} w_{4i} = 1$ so the above simplifies to

$$Q_3^{-1} = [0, 0, 1, -\gamma, 0, \dots, 0]$$
$$Q_4^{-1} = [0, 0, -\gamma, 1, 0, \dots, 0]$$

Since $0 \leq \gamma < 1$, then the sum of the absolute values of the off-diagonal elements is γ for rows 3 and 4 and this value is less than one.

The results above show that in the case W has a single neighbour, then Q^{-1} is a positive definite matrix. Furthermore, this implies that Q is also positive definite and we can easily show that if we scale Q by a positive value, like σ_s^2, then we preserve the relationship $\mathbf{z}^T(\sigma_s^2 Q)\mathbf{z} > 0$ as $\mathbf{z}^T(\sigma_s^2 Q)\mathbf{z} = \sigma_s^2(\mathbf{z}^T Q \mathbf{z}) > 0$.

1.3 Case 3: W has $1 < n < K-1$ neighbours

Assume connectivity matrix W has $1 < n < K-1$ nieghbours, then the precision matrix is

$$Q^{-1} = \begin{bmatrix} \gamma\sum_{i\neq 1}^{K} w_{1i} + 1 - \gamma & -\gamma w_{12} & -\gamma w_{13} & -\gamma w_{14} & \dots & -\gamma w_{1K} \\ -\gamma w_{21} & \gamma\sum_{i\neq 2}^{K} w_{2i} + 1 - \gamma & -\gamma w_{23} & -\gamma w_{24} & \dots & -\gamma w_{2K} \\ -\gamma w_{31} & -\gamma w_{32} & \gamma\sum_{i\neq 3}^{K} w_{3i} + 1 - \gamma & -\gamma w_{34} & \dots & -\gamma w_{3K} \\ \vdots & & & & & \vdots \\ -\gamma w_{K1} & & \dots & & & \gamma\sum_{i\neq K}^{K} w_{Ki} + 1 - \gamma \end{bmatrix}, \tag{4}$$

where the off-diagonal elements w_{mj} are equal to one if regions m and j are neighbours and zero otherwise. We begin by considering generic row m and consider the diagonal term and the sum of the off-diagonals at the extreme values of the range $0 \leq \gamma < 1$.

Consider generic row m, then its corresponding row vector is

$$Q_m^{-1} = \left[-\gamma w_{m1}, -\gamma w_{m2}, \dots, \gamma \sum_{i\neq m}^{K} w_{mi} + 1 - \gamma, \dots, -\gamma w_{mK} \right],$$

where $Q_{mm}^{-1} = \gamma \sum_{i\neq m}^{K} w_{mi} + 1 - \gamma$ for n connected regions, then the sum $\sum_{i\neq m}^{K} w_{mi} = n$.

Question: Is the n^{th} row strictly diagonally dominant? Does the inequality

$$|n\gamma + 1 - \gamma| > |-n\gamma| \tag{5}$$

hold for $0 \leq \gamma < 1$?

When $\gamma = 0$ the above simplifies to $1 > 0$, so the inequality (5) holds.

Let's observe this expression as $\gamma \to 1$. Since the LHS of (5) is always positive because $\gamma \neq 1$, then (5) simplifies to $n\gamma + 1 - \gamma > n\gamma$. By combining the like-terms, the inequality becomes $1 - \gamma > 0$ and this holds for the range of γ. We conclude that when W has $1 < n < K-1$ neighbours, Q^{-1} is positive definite. This implies that $\sigma_s^2 Q$ is also positive definite.

1.4 Case 4: W has $K-1$ neighbours

This final case applies to matrix Q^{-1} where each region has $K-1$ neighbours. Here matrix W looks like

$$W = \begin{bmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & 0 & 1 & \dots & 1 \\ & & & \ddots & & 1 \\ 1 & 1 & 1 & 1 & \dots & 0 \end{bmatrix},$$

and every diagonal of element of $\hat{W}$ will be $K-1$ and off-diagonals are zero. The expression for Q^{-1} becomes

$$\gamma(\hat{W}-W)+(1-\gamma)\mathbb{I} = \begin{bmatrix} \gamma\sum_{i\neq 1}^{K} w_{1i}+1-\gamma & -\gamma & -\gamma & -\gamma & \dots & -\gamma \\ -\gamma & \gamma\sum_{i\neq 2}^{K} w_{2i}+1-\gamma & -\gamma & -\gamma & \dots & -\gamma \\ -\gamma & -\gamma & \gamma\sum_{i\neq 3}^{K} w_{3i}+1-\gamma & -\gamma & \dots & -\gamma \\ \vdots & & & & & \vdots \\ -\gamma & -\gamma & & & \dots & \gamma\sum_{i\neq K}^{K} w_{Ki}+1-\gamma \end{bmatrix}$$

As each row has the same number of connections, then $\sum_{i\neq j}^{K} w_{ji} = K-1$ for all j. The above simplifies to

$$\gamma(\hat{W}-W)+(1-\gamma)\mathbb{I} = \begin{bmatrix} \gamma(K-2)+1 & -\gamma & -\gamma & -\gamma & \dots & -\gamma \\ -\gamma & \gamma(K-2)+1 & -\gamma & -\gamma & \dots & -\gamma \\ -\gamma & -\gamma & \gamma(K-2)+1 & -\gamma & \dots & -\gamma \\ \vdots & & & & & \vdots \\ -\gamma & -\gamma & & & \dots & \gamma(K-2)+1 \end{bmatrix}$$

For this matrix to be strictly diagonally dominant, and hence a positive definite matrix, then for each row we need to satisfy

$$|\gamma(K-2)+1| > |-\gamma(K-1)|.$$

For $K > 2$ the LHS is positive[2], hence we can simplify the above to $\gamma(K-2)+1 > \gamma(K-1)$. As $0 \leq \gamma < 1$, then

$$\begin{aligned} \gamma K - 2\gamma + 1 &> \gamma K - \gamma \\ -2\gamma + 1 &> -\gamma \\ -\gamma + 1 &> 0 \end{aligned} \tag{6}$$

and this is true for all values of γ, as γ is bounded to be positive and less than one. Since we have shown that the covariance matrix $\sigma_s^2 Q$ is a positive definite covariance matrix for all possible number of neighbours, it remains to show that the estimation of positive definite covariance matrices by the maximum likelihood estimator (MLE) is consistent.

1.5 Literature which describe consistent properties of MLEs

There is ample literature which describes desirable properties of multivariate normal density, such as the density being at least twice differentiable (continuity), well defined in the real domain and has a single global maximum (compactness) (Wackerly et al., 2007; Casella and Berger, 2002). These and other features are attractive for asymptotic properties (Casella and Berger, 2002) and affect the ability of the MLE estimators ability to converge.

There is a large body of literature which show MLEs are consistent, such as the work by Casella and Berger (2002), Anderson (1973) and in the textbook by Anderson (1984) as well as by Rice (2006) among others. Recent literature such as Zwiernik et al. (2017) suggest that the MLE for a positive definite covariance matrix is a consistent estimator. Alternatively, I have found a lot of literature, like the work by Pourahmadi (1999) and Pourahmadi (2000) which focus on LASSO and other sparse methods for the estimation of covariance matrices, and how the aim of these estimators is to formulate a semi-positive definite parametrisation of the covariance matrix. However, statements as to these high dimensional estimators being consistent is a little harder to ascertain as the problem is posed as either the sparsity parameter $\lambda \to 0$, $\sqrt{n}\lambda \to \infty$ or simply as $n \to \infty$. In the latter case, if $\lambda = 0$, then the problem can be formulated back as a MLE process.

[2] As it is straightforward to show that for $K = 2$ the corresponding precision matrix Q^{-1} is positive definite, this result has been omitted.

References

TW Anderson. Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, pages 135–141, 1973.

TW Anderson. An Introduction to Multivariate Analysis, 1984.

Howard Anton and Robert C Busby. *Contemporary Linear Algebra*. Wiley, 2003. ISBN 9780471163626.

Keith Briggs. Diagonally Dominant Matrix, 1999. URL `http://mathworld.wolfram.com/DiagonallyDominantMatrix`

George Casella and Roger L Berger. *Statistical Inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Ward Cheney and David Kincaid. Linear algebra: Theory and applications. 2009.

Yasunori Fujikoshi, Vladimir V Ulyanov, and Ryoichi Shimizu. *Multivariate statistics: High-dimensional and large-sample approximations*, volume 760. John Wiley & Sons, 2011. ISBN 9780470411698.

Brian G. Leroux, Xingye Lei, and Norman Breslow. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In Elizabeth M. Holloran and Donald Berry, editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer, 2000.

Mohsen Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.

Mohsen Pourahmadi. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87(2):425–435, 2000.

John Rice. *Mathematical statistics and data analysis*. Nelson Education, 2006. ISBN 9788131519547.

Yousef Saad. *Iterative methods for sparse linear systems*, volume 82. SIAM, 2 edition, 2003. ISBN 9780898715347.

Dennis Wackerly, William Mendenhall, and Richard Scheaffer. *Mathematical Statistics with Applications*. Nelson Education, 2007. ISBN 9780495110811.

Piotr Zwiernik, Caroline Uhler, and Donald Richards. Maximum likelihood estimation for linear Gaussian covariance models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1269–1292, 2017.