Spatial Elements in Poisson Regression Using Bayesian Methods: Applying the Besag-York-Mollié Model

Wendy Olsen
August 2022
Department of Social Statistics
School of Social Sciences
University of Manchester

https://github.com/WendyOlsen/SpatialRegressionBayesIndia2022

www.socialsciences.manchester.ac.uk/social-statistics/

Contents

- First, an applied regression model.
- Samples of the ‘research question’.

METHODS SECTIONS:

- 3rd, a Poisson model as a statistical distribution for the dependent variable, allowing generalised linear modelling with hierarchical elements
- 4th, a spatial element shown in BYM2 format
  - (Besag-York-Mollié v. 2)
- Conclusion: Summary plus a tutorial task

‘Labour-Force Active’ People = 1, non-active = 0. Young people join the labour market…

- The debate about productivity has assumed that it is normal for men to work…but what about youths and children and women?
- We make a model of the risk of ‘being labour-market inactive’ vs.

Active in the labour market. Inactive is MUCH more common that Unemp.

- **Making a risk model is similar to looking at any problem,**

and the factors that raise the risk of that injury or illness happening.

  - Debating the components of the model on the right-hand side.
  - $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **sex + social-group**
  - $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **sex + social-group + spatial element (unexplained u_j)**
  - $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **sex + social-group + spatially-varying Z + u_j**

- We call the regression coefficients the ‘slopes **(BOLD)**

Research Questions

- Define youth as age 15-24, or you may use 16-24 or 17-24.

- RQ1 What are the social factors (like religious group or minority-group) associated with being ACTIVE in the labour market as a youth?
  - We have decided to select India's youths up to age 24 for the moment.
  - Age can then be left out of the equation.

- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **sex + social-group + religious minority status +** $[spatial\ terms]$

Research Questions

- RQ1 First one must adjust for all the child and youth [illegible] e to cultural norms around doing paid labour.
- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **+ age + soci[illegible]**
  $[spatial\ term[illegible]$

Perhaps this is the broad RQ, but it is **too general** to publish in a journal article!

Narrow down your RQ to generate testable hypotheses

A Narrower Research Question

- RQ 2 Via what routes does gender affect the risk of a youth in India being active/inactive in the labour market?
- …after taking into account spatial variations in norms
  - You could for example interact SEX with FORMAL EDUCATION

- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **gender** + *interactions of gender with* {**social-group + religious minority status**} + $[spatial\ terms]$ + *gender*spatial terms*

Innovative Research Questions (A), Use ICC

- RQ3 What interaction effects occur, a) of gender by social group? After allowing for spatial covariation as an extracted factor; b) of gender by social group without allowing for spatial covariation?
- Model 0 is the empty hierarchical (spatial and s.g.) model, S.g.=SocGroup
- $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}$+sex+age+s.g. + sex*s.g. + $[spatial\ term] + residual$ (Eq. 6)
  - Gather the Intra-class correlation for the empty model and the above model
- $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}$+sex+age+s.g. + sex*s.g. + $residual$ (Eq. 7))

  Now gather the ICC for the revised model without spatial terms

Innovative Research Questions (B): Add additional innovative variables

- You can add more variables
  - For example whether this **youth** is the eldest sibling of 2+ co-resident siblings.
  - And interact this variable with Sex of this youth.

- Another possibility: interact sex with age, within 15-24 (Females 'down' males 'up')?
- Risk of **bias/variance** trade-off (Kuhn & Johnson, chapter 5. Avoid Correlation of X with X)
  Increased bias from more parameters?! WATCH OUT.

- Could add whether it is a single-parent family (BE CAREFUL).
- Interact that with the sex of the lone parent. 0 = not single. F1=1 if a female headed household with 1+ children. M1 = 1 if male headed household with 1+ children

Summary of the Model and its Spatial Term (A) was to test whether this term *matters.*

- $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+\left[\left(\sqrt{\rho}/_{s}\right)\varphi^{*}+(\sqrt{1-\rho})\theta^{*}\right]\sigma$

In general use existing theory to choose the X variates.

- This is **confirmatory regression.**
- Comparison of models is how we draw conclusions about specific hypotheses. Compare via AIC, BIC, or LR test

Use the likelihood-ratio test statistic. (LR test) Fox, chapter 15, section 15.1.1 in 2nd edition.

- $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+[spatial\ term]$

Notation Used Here

- We follow Fox in offering the main linear part of the generalised linear model:
- $\eta_j = \beta_{0j} + \mathbf{X\beta}$ where eta is the dependent variable measured suitably.
- B_{0j} is the intercept for risk and may not be interesting.
- **XB** is the estimate for **Xβ,** a vector multiplication. **X** holds several independent variables and **Beta** holds the slope coefficients.
- In an aggregated model at District units j, we will effectively have a randomly distributed District fixed effect – also called a random slope on district.
- In such a model, the intercept B_{0j} drops out.

The Besag-York- Mollié Model (v. 2) - More details of how we interpret it. See Morris et al., 2019. Very helpful.

- …+$\left[\left(\sqrt{\rho/_s}\right)\varphi^* + (\sqrt{1-\rho})\theta^*\right]\sigma$
  Rho, ρ, measures the degree of correlation of the data from nearby and contiguous districts, such that when rho is large, the first term is larger and the second term is smaller. Rho runs from [0…1].
  - If the spatially correlated terms are greater, then RHO is larger, and if the spatially uncorrelated parts of the geography are greater, then RHO is closer to zero. (Morris, et al., 2019: 7)
- BYM introduced Phi and Theta. Here, in BYM2, ρ appears twice, with Rho weighting the two parts.
- Phi φ is 'spatial effects'.
  - Phi measures how the adjacency matrix is summarised using pairs of locations I and j; when phi is large, there is greater nearness, or greater contiguity, of the pair of locations (districts).
- Theta θ is the heterogenous spatial effects mop-up term. "Independent Error Terms"
- Sigma
  - scales-up the spatial part of the model to reflect the spatial terms' standard error. If it is large, then the spatial part plays a **greater part.**

References (1)

- Fox, John (2008), *Applied Regression Analysis and Generalized Linear Models*, London: Sage
- Kuhn, Max, and Kjell Johnson (2013), *Applied Predictive Modelling* (chapter 5 on the variance-bias tradeoff), London: Springer.
- Morris, Mitzi, K. Wheeler Martin, D. Simpson, S J. Mooney, A. Gelman, and C. DiMaggio (2019), Bayesian Hierarchical Spatial Models: Implementing the Besag-York-Mollié model in Stan, *Spatial and Spatio-Temporal Epidemiology*, 31, 100301.

Poisson Model – Used for a "Count"-Dependent Variable.

- In saying "Poisson" regression models, we mean a log Poisson distribution of the Dependent variable is a linear sum of terms.
- Poisson distributions have the feature that the mean of the distribution equals its variance, reducing key parameters from 2 to 1. But we add other parameters.
- A reference work for hierarchical, generalized linear models (GLM) by Fox 2008: Chapter 15) argues that the log Poisson model should be checked for overdispersion. We then have a quasi-Poisson model which many packages can estimate. I will present the version that expresses 'exposure' (space or time or population) and 'risk' λ.
- Sometimes, a log poisson model needs an overdispersion adjustment.
- An overdispersion adjustment is a multiplicative factor added as $c*\lambda$.
- λ is the risk of the event.

Poisson Distribution

Poisson Histogram		
$p(y) =$ $\mu^y * \frac{e^x}{y!}$ for y =0,1,2, … (Eq. 1)	Notice behavior when sample size is large, in the limit it reaches the red shape, based on Binomial distribution, but an offset shown at left helps with fitting.	

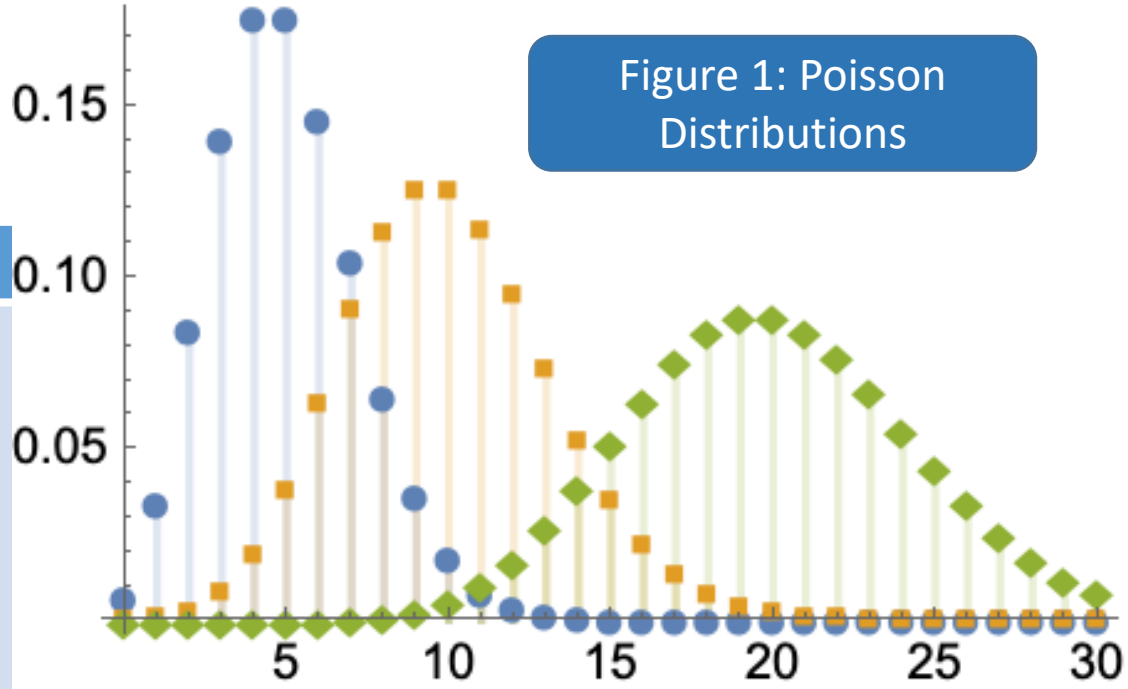

Figure 1: Poisson Distributions

Source https://mathworld.wolfram.com/PoissonDistribution.html , accessed June 2022. And see https://reference.wolfram.com/language/ref/PoissonDistribution.html

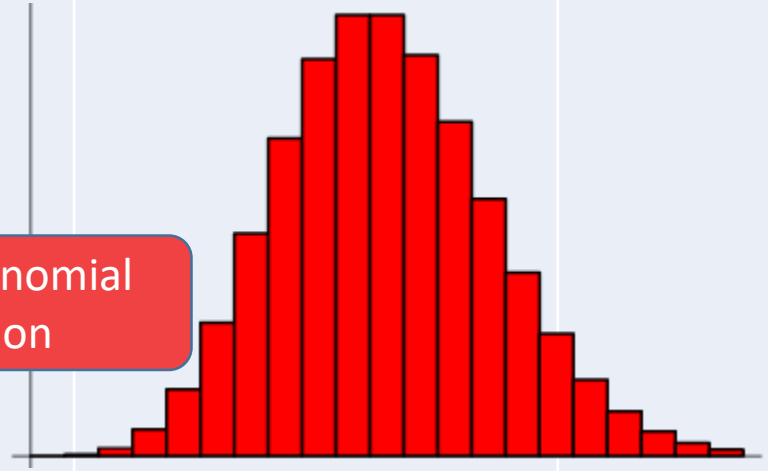

Figure 2: A Binomial Distribution

Horizontal: Count of the number of instances n of events in which Y might occur. Y is 0/1 and has mean p.
Vertical: The probability. As a probability mass function.

We fit the Poisson, for data X and Y and cases “i” within groups numbered j. Model the risk of Y.

- A nested Poisson multi-level regression model
  - Maximum likelihood estimation using R’s lme4
  - Three-level multilevel model with random intercepts for county and age and a random slope for selected county variables
- Dependent variable is count of workers in each social group in each district
  - Each district is marked j
  - You may add an offset (population of District) to get a better estimate.
  - Each individual youth in all households has subscript I initially
  - Aggregation is carried out to create groups with 0 or >0 counts of ‘active workers’
  - Weighted sums are used

There are ny successes in n trials. N(1-y) are fails, ie and the Poisson distribution is uses the factorial of Y.

$$p(y) = \mu^y * \frac{e^x}{y!} \text{ for } y = 0, 1, 2, \ldots \text{ (Eq. 1)}$$

Here when writing code, we inform Stan or R BRMS or lme4 that the canonical link function is log and the family is Poisson.

The expectation is $\mu_i = E(y_i)$ *(Eq. 2 using mu)*

The conditional variance of Y is $V(y_i|\eta_i) = \mu_i$ (Eq. 3 also mu)

Poisson(μ_i) = $\eta_i = \beta_0 + \mathbf{x\beta}$ for β from 1 to k for k coefficients

This special situation can be tested for. See Fox 2008: Chap. 15.

Offset written into Poisson Model

- Poisson($c_j\lambda_i$) = $\eta_i = \beta_0 + \underline{\mathbf{x\beta}}$ for β from 1 to k for k coefficients,
- Where c_j is exposure in group j either on average or as a total or mean. The units of the count are in logs, so using populations we log the population of the group.

Poisson($c_j\lambda_j$) = $\eta_j = \beta_{0j} + \underline{x_i}\underline{\beta_k}$ for β from 1 to k (Eq. 4) for k coefficients

- Aggregation from i to j groups for districts j for example, might give data as shown for regression:
  - This is a multilevel model with spatial groups. We often also group socially.
- When you remove the log, you get e^{c_j} as an additive offset.

How we could programme the Poisson fit for the count data using R with R2Jags or Stan (Shown: Winbugs format)

- **Here is a likelihood function, seen in the WinBUGS code format:**
- ### LIKELIHOOD ###

```
for (j in 1: N.obs){
     for (i in 1: N.X){
          X.row[i, j] <- X.Eff[i, X[j, i]]
     }
     for (i in 1: N.Z){
          Z.row[i, j] <- Z.Eff[i, Z[j, i]]
     }
     log(lambda[j]) <- Beta0 + log(Offset[j]) +
sum(X.row[, j]) + sum(Z.row[, j])
     Y[j] ~ dpois(mu[j])
} }
```

Results of a Poisson Model – Theory and Practice

- See our paper "A Bayesian Estimation of Child Labour in India" (*Child Indicators Research, A Bayesian Estimation of Child Labour in India, Kim, J. H., Olsen, W. & Wiśniowski, A., 2020, volume 13*)
- and its supplement.

- Or see more recent appendix Tables A3-A6 for WES paper:

https://github.com/WendyOlsen/normslabourindia

Notation here is from the *Child Indicators Research* paper

Having found the risk of 'child labour' rises rapidly from age 9 upwards, we modelled each year of Age.

Here a refers to the first dataset and b simply refers to a different dataset.

Types of Models	Model 1 IHDS (Poisson)	Model 2 NSS (Poisson)
Likelihood	$y.a_{ij} \sim$ Poisson($\mu.a_{ij} * n.a_{ij}$)	-
	-	$y.b_{ij} \sim$ Poisson($\mu.b_{ij} * n.b_{ij}$)
Overdispersion	-	-
	-	-
Prediction	$\hat{y}_{ij} \sim$ Poisson ($\mu.a_{ij} * N_{ij}$) $\hat{y}_{i+} = \sum_i \hat{y}_{ij}$	$\hat{y}_{ij} \sim$ Poisson ($\mu.b_{ij} * N_{ij}$) $\hat{y}_{i+} = \sum_i \hat{y}_{ij}$
Model for true child labour rate	$\log(\mu.a_{ij}) = \beta_0 + \beta_1 * x_i + \beta_2 * \log(z_{ij})$	$\log(\mu.b_{ij}) = \beta_0 + \beta_1 * x_i + \beta_2 * \log(z_{ij})$

Notes: i – age; j – state; τ is a precision (inverse variance). The *CIR* paper uses I for integer child—age-groups.

Z_{ij} reflects the population of people in that state, ie Census 2011 data on 'main workers' in that district = an offset.

Besag–York–Mollié terms added to model

- The BYM model adds terms which are built up on geographers' methods of measuring random errors that occur at a spatial level.
  - The correlated random errors are potentially located in contiguous ways.
  - The map of the places can be used to derive measures of contiguity, shared boundary, and distance from the equicentre of each district, but here in BYM, it is simply used for a Queen's matrix. This matrix is square, showing all district codes j on each edge, we do not examine it; from R mapping it is autocreated, with 1 where adjacent, and 0 where not adjacent; it is a sparse matrix.
  - We now add this term to the log Poisson GLM:

$$\eta_j = \beta_{0j} + \underline{X\beta} + \left[\left(\sqrt{\rho / s}\right)\varphi^* + \left(\sqrt{1-\rho}\right)\theta^*\right]\sigma$$

(Eq. 5)

φ^* is in units of Eta_j and it has a triangle of estimates for districts j with all other districts $k \neq j$, in the J-k matrix. We can sort that matrix.

- **Reminder**: This BYM2 model has spatial smoothing on contiguous and nearby places, as well as a random-effects component which helps with fitting the heterogenous spatial amounts of risk.
- The rho factor varies between 0 to 1, rising with greater correlated local risks in groups of areas j.
- The phi factor φ is an ICAR model to take care of correlated risks, spatially, which are contiguous. ICAR means intrinsic conditional autoregressive models. Rho appears twice in the BYM2 formula.
  - 1st term is for the spatial explanation via contiguous similar districts. 2nd term is for the unusual districts, ie heterogeneity.
- The factor s is an adjustment that can scale up/down the first term.
- Lastly, a scaling factor ***sigma*** allows the model to upscale the spatial term.

$$\eta_j = \ldots \left(\sqrt{\rho / s}\right) \varphi^* + \cdots$$

(Eq. 5)

AN ILLUSTRATION

- Illustration of the key term in Besag-York-Mollié model
  - 1st term reflects +'ve correlated contiguous similar districts.
  - **Desertification** occurs in contiguous rural districts! SO desertification would cause all those rural areas to have LOW labour-force participation. Rho would be high but so would PHI for those areas. PHI* would have clumps of 5 districts with deserts.

Conclusions

- A series of linear terms embedded in a Poisson model can attribute risk to competing factors. Some can have interaction terms.
- The fit of the model can be assessed using Bayesian Information Criterion (BIC), related measures AIC, nested-models LR test and MCMC type tests.
- The tutorial involves amending an existing model
  - Install R with packages stan, stanarm and tidyverse;
  - Step 1 run the program using the code provided
    - (see https://github.com/WendyOlsen/https/SpatialRegressionBayesIndia2022)
    - Data are also provided here for age 15-24.
    - You must only do non-commercial work with these data.
  - Step 2, replace the variable 'rural' with 'age' and run it again, interpret.
  - Step 3, test a more complex hypothesis. Good luck!

References

- Besag, J.J.Y., Mollié, A. (1991), Bayesian Image Restoration with Two Applications in Spatial Statistics, *Ann. Inst. Stat. Math.* 43, 1-59,10.1007.
- Fox, John (2008), *Applied Regression Analysis and Generalized Linear Models*, London: Sage.
- Kim, Jihye, Olsen, W.K. and A. Wisniowski (2022) Predicting Child-Labour Risks by Norms in India. *Work, Employment and Society*. doi:10.1177/09500170221091886
- Kim, Jihye, Olsen, W.K. and A. Wisniowski (2020), A Bayesian Estimation of Child Labour in India, *Child Indicators Research*, DOI https://doi.org/10.1007/s12187-020-09740-w.
- Kuhn, Max, and Kjell Johnson (2013), *Applied Predictive Modelling* (chapter 5 on the variance-bias tradeoff), London: Springer.
- Morris, Mitzi, K. Wheeler Martin, D. Simpson, S J. Mooney, A. Gelman, and C. DiMaggio (2019), Bayesian Hierarchical Spatial Models: Implementing the Besag-York-Mollié model in Stan, *Spatial and Spatio-Temporal Epidemiology*, 31, 100301.
- Olsen, Wendy, Manasi Bera, Amaresh Dubey, Jihye Kim, Arkadiusz Wisniowski, Purva Yadav (2020). Hierarchical Modelling of COVID-19 Death Risk in India in the Early Phase of the Pandemic, *European Journal of Development Research.* DOI https://link.springer.com/article/10.1057/s41287-020-00333-5

Tutorial

- You can carry out a tutorial activity based on data in our github site.
- https://github.com/WendyOlsen/SpatialRegressionBayesIndia2022

- Tutorial documents are aimed at All-India estimates.
- You can do all-age estimates or study any age group from 15 to age 100 years.
- The dates of data are 2017/8 and 2018/9.

Tutorial

- You can carry out a tutorial activity based on data in our github site.
- Your first task is find or **create an 'age' factor in R**. Age can be 0=15-19 and 1=20-24. Our hypothesis is that those who are older are more likely to work, but it is gender-specific. Men more, women less. Many people get married in this period and the gender impact on labour-force participation is reversed. (Discuss)
  - Preferred model: OMIT RURAL/URBAN, INCLUDE AGE 0/1, INCLUDE BYM2.
  - And contrast that with: OMIT RURAL/URBAN, INCLUDE AGE 0/1, OMIT BYM2.
- Your second task is **now to run the preferred models** as Poisson model, given in the code files. You can run empty models and those with the given variables (sex and age). Optionally, also include an interaction effect sex*age. You can also run lm from lme4 and the logistic model.
- You can notice the ICC calculation code file. But wait a minute –do step 3 first.
- Step 3:You now have a much better fit than we had in our Webinar**. What are your 'fit statistics**', ie measures of goodness of fit? Make a simple table (example on the next page). How do you read and assess these? What is your interpretation?
- Step 4: Now **make a table showing the ICC part** for 4 models as shown on next slide.

Tutorial Table Guidance (data https://github.com/WendyOlsen/SpatialRegressionBayesIndia2022)

- Table 1
- Define your models by number.
  - Eg 0 empty Poisson 1 Poisson with variables 2 Poisson with variables with spatial BYM2 term
  - 3 a logistic model is optional
  - Perhaps 4 empty model with spatial BYM2 if you wish.
- Headings:

  Model 0 1 2 3 4 5 6
  - Constant term if any
  - Coeff Age
  - Coeff Sex
  - Coeff Age*sex
  - Mean of the random effects j
  - Variance of the random effects j
  - Covariances must not be in this table!

BIC if available

- Table 2
- Use the same models.
- Choose only 4 of them
  - We want to see:
  - Empty Poisson model with BYM
  - Poisson model with BYM without interaction effects
  - Poisson model with BYM with interaction effects
- Table headings:

  Model 0 1 2 3

  Coefficients

  Constant

  X1

  X2

  X1*X2

  Bayesian Info Criterion BIC

  ICC κ

  Variance of the whole model (var(Y-hat)) ν

  Variance not covered in ICC ν -κ

Quiz to discuss BYM research prospectively

- Form a small group
- Introduce yourselves (Name, Place)
- What is the Count outcome in your research?
- What would 3 X variables be?

What is the available Spatial Unit?

Do you have any aggregate variables X_j ? (e g GDP per capita)

There is only time to develop one person's project! The rest are listening/commenting.