Variational Bayesian Inference for Spatial Negative Binomial Count Data Models with Unobserved Heterogeneity

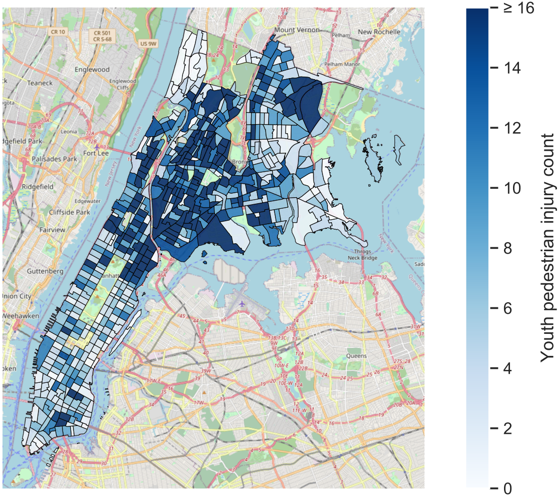
PRATEEK BANSAL[1,*], **RICO KRUEGER**[2,*], MICHEL BIERLAIRE[2], DANIEL J. GRAHAM[1]

[1]Imperial College London, [2]Ecole Polytechnique Fédérale de Lausanne
[*] *Equal contribution*

20th Swiss Transport Research Conference

Motivation i

- **Spatial count data models** are used to explain and predict frequencies of aggregate phenomena such as traffic accidents in geographically distinct entities (e.g. census tracts).
- Spatial count data may exhibit ...
    - **spatial dependence** due to systematic correlation in unobserved factors.
    - **spatial heterogeneity** due to spatially varying effects of covariates on dependent variable.



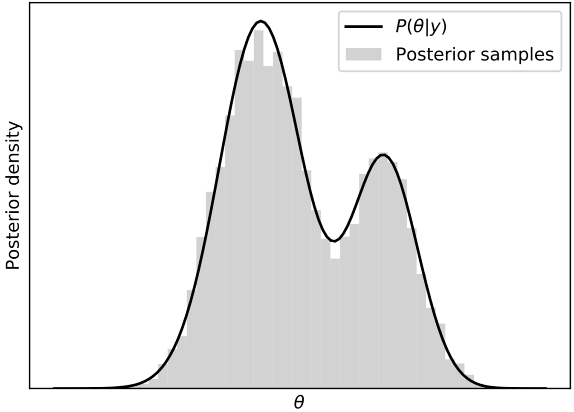
Motivation ii

Challenges

- Accounting for spatial effects complicates model estimation.
- Datasets are growing in size and are becoming available on a streaming basis.

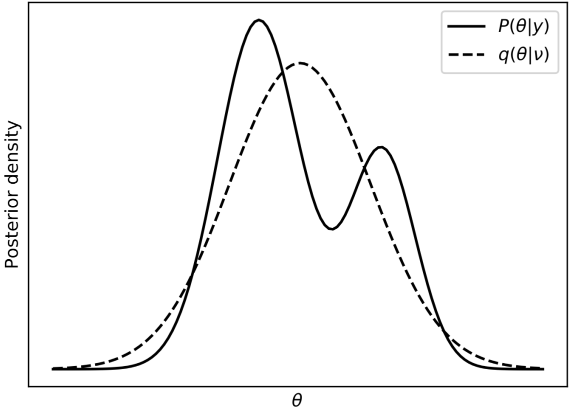
Markov chain Monte Carlo (MCMC)

- Approximate posterior $P(\theta|\boldsymbol{y})$ numerically through samples from a Markov chain.
- Use Metropolis-Hastings algorithm and Gibbs sampling to construct Markov chain.
- Issues:
  - computation times,
  - storage of posterior draws,
  - convergence assessment,
  - serial correlation.



Variational Bayes (VB)

- Recast Bayesian estimation as an optimisation problem.
- Approximate posterior $P(\theta|\boldsymbol{y})$ analytically through a parametric variational distribution $q(\theta|\boldsymbol{\nu})$.
- Advantages:
    - Reduced storage requirements
    - Straightforward convergence assessment
    - Serial correlation no longer a concern
    - Stochastic optimisation



Research objective

Develop VB method for fast estimation of negative binomial model with spatial dependence and heterogeneity.

Model formulation

Negative binomial likelihood

$$y_i \sim \mathrm{NB}(r, p_i), \quad p_i = \frac{\exp(\psi_i)}{1 + \exp(\psi_i)}, \quad \psi_i = \boldsymbol{X}_i^\top \beta_i + \phi_i, \quad i = 1, \dots, N$$

Hierarchical prior for link function parameters to accommodate spatial heterogeneity

$$\beta_i \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, \dots, N$$

Matrix exponential spatial specification (MESS; LeSage and Pace, 2007) of spatial dependence

$$\boldsymbol{S}\phi = \exp(\tau \boldsymbol{W})\phi = \epsilon, \quad \epsilon \sim \mathrm{Normal}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$$

Pólya-Gamma data augmentation (Polson et al., 2013; Zhou et al., 2012)

Issue

- Negative binomial distribution does not have a conjugate prior.
- Thus, the conditional distributions of the link function parameters do not constitute known distributions.

Remedy

- Introduce Pólya-Gamma-distributed auxiliary variables $\omega_i \sim \mathrm{PG}(y_i + r, 0), i = 1, \ldots, N$.
- Conditional on ω and r, the likelihood of observed counts is translated into a heteroskedastic Gaussian likelihood.

Variational Bayes (e.g. Blei et al., 2017)

- VB seeks to minimise the KL divergence (probability distance) between an approximating variational distribution $q(\boldsymbol{\theta})$ and the posterior of interest $P(\boldsymbol{\theta}|\mathbf{y})$:

$$q^*(\boldsymbol{\theta}) = \arg\min_q \left\{ \mathrm{KL}\left(q(\boldsymbol{\theta}) || P(\boldsymbol{\theta}|\mathbf{y})\right) \right\}.$$

- q must be selected by the analyst. Its expressiveness determines the quality of the variational approximation and the complexity of the estimation problem.

Mean-field variational Bayes (MFVB)

- Impose posterior independence between parameter blocks $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_J$:

$$q(\mathbf{\Theta}) = \prod_{j=1}^{J} q(\mathbf{\Theta}_j)$$

- For conditionally-conjugate models, MFVB leads to a simple coordinate ascent algorithm.
- τ and σ^2 are recovered poorly under MFVB assumption.

Integrated nonfactorised variational Bayes (INFVB; Han et al., 2013; Wu, 2018)

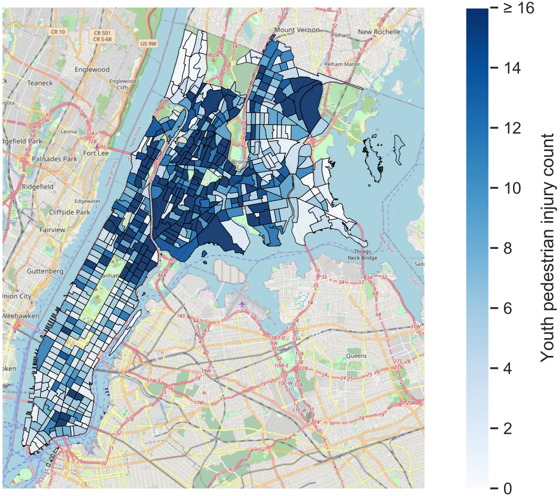
- Decompose parameters $\mathbf{\Theta}$ into two disjoint subsets $\{\mathbf{\Theta}_c, \mathbf{\Theta}_d\}$ to specify more flexible variational distribution:

$$q_{\text{INFVB}}(\mathbf{\Theta}) = q(\mathbf{\Theta}_c|\mathbf{\Theta}_d)q(\mathbf{\Theta}_d)$$

- Direct minimisation of KL divergence between $q_{\text{INFVB}}(\mathbf{\Theta})$ and posterior is challenging.
- Conditional on $\mathbf{\Theta}_d$, INFVB involves a simple coordinate ascent algorithm.
- Define a grid with points $\mathbf{\Theta}_d^{(g)}$ and run MFVB separately for each grid point. Then, compute weight of each grid point using the Boltzmann distribution.
- Select $\mathbf{\Theta}_d = \{\tau, \sigma^{-2}\}$.

Data

- Benchmark INFVB against MCMC.
- Youth pedestrian injury counts in 603 census tracts of the Bronx and Manhattan from 2005–14 (Morris et al., 2019).



Sample description (N = 603)

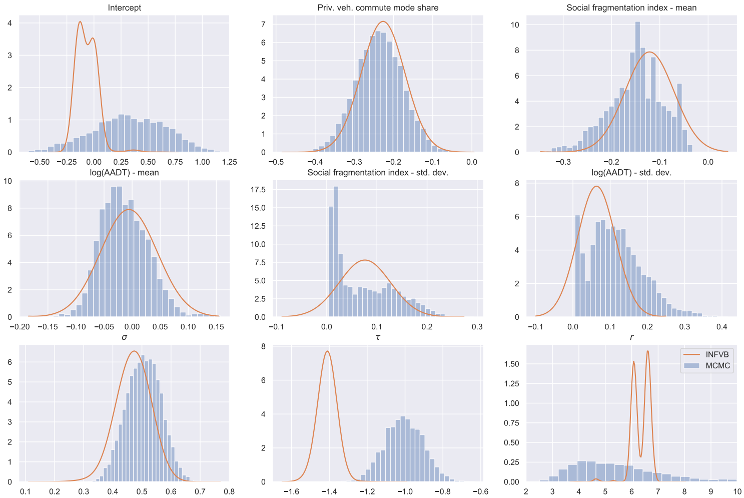
	Mean	Std.	Min.	Max.
Youth pedestrian injury count, 2005-14	9.69	8.35	0.00	44.00
Social fragmentation index	2.02	2.73	-4.50	18.67
Avg. annual daily traffic (AADT) in 1000, 2015	44.48	46.79	2.09	276.48
Private vehicle commute mode share, 2010-14	0.19	0.15	0.00	0.76

Results: Estimation summary

	Estimation time [s]	LPPD	RMSE
MCMC	1733.77	-1713.24	**3.79**
INFVB	**20.24**	**-1704.40**	3.88

Note: LPPD = log-posterior predictive density. RMSE = root mean square error.

Results: Marginal posterior distributions



Conclusion

Key points

- *Integrated nonfactorised variational Bayes (INFVB)* method for fast Bayesian estimation of spatial count data models.
- In case study, INFVB is more than 80 times faster than MCMC, while offering similar estimation accuracy.

Next steps

- Further testing
- Extension to models with spatio-temporal dependencies
- Online estimation

Thank you

References I

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877.

Han, S., Liao, X., and Carin, L. (2013). Integrated non-factorized variational inference. In *Advances in Neural Information Processing Systems*, pages 2481–2489.

LeSage, J. P. and Pace, R. K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, 140(1):190–214.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spatial and spatio-temporal epidemiology*, 31:100301.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.

Wu, G. (2018). Fast and scalable variational bayes estimation of spatial econometric models for gaussian data. *Spatial statistics*, 24:32–53.

Zhou, M., Li, L., Dunson, D., and Carin, L. (2012). Lognormal and gamma mixed negative binomial regression. In *Proceedings of the International Conference on Machine Learning. International Conference on Machine Learning*, volume 2012, page 1343. NIH Public Access.