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

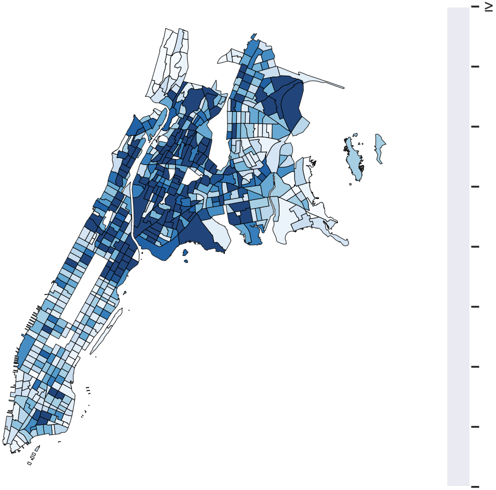
PRATEEK BANSAL[1,*], **RICO KRUEGER**[2,*], MICHEL BIERLAIRE[2], DANIEL J. GRAHAM[1]

[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 e€ects complicates model estimation.
- Datasets are growing in size and are becoming available on a streaming basis.

Markov chain Monte Carlo (MCMC)

- Approximate posterior P(jy) 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(jy) analytically through a parametric variational distribution q(j).
- 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

$$y_i \quad NB(r; p_i); \quad p_i = \frac{\exp(\ _i)}{1 + \exp(\ _i)}; \qquad _i = X_i^{>} \ _i + \ _i; \quad i = 1; \ldots ; N$$

Hierarchical prior for link function parameters to accommodate spatial heterogeneity

$$_i \quad \mathrm{Normal}(\ ;\ ); \quad i = 1; \ldots ; N$$

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

$$S \ = \exp(\ W) \ = \ ; \qquad \mathrm{Normal}(0;\ ^2 I_N)$$

Polya-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 $q(\ )$ and the posterior of interest $P(\ |y)$:

$$q\ (\ ) = \arg\min_{q} \{KL(q(\ )||P(\ |y))\}:$$

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

Mean- eld variational Bayes (MFVB)

Impose posterior independence between parameter blocks $\ _1; \ldots; \ _J$:

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

For conditionally-conjugate models, MFVB leads to a simple coordinate ascent algorithm.

and $\ ^2$ are recovered poorly under MFVB assumption.

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

Decompose parameters into two disjoint subsets f $_c$; $_d$g to specify more exible variational distribution:

$$q_{INFVB}(\quad) = q(\quad_c j \quad_d) q(\quad_d)$$

Direct minimisation of KL divergence between $q_{INFVB}(\quad)$ and posterior is challenging.

Conditional on $_d$, INFVB involves a simple coordinate ascent algorithm.

De ne a grid with points $^{(g)}_d$ and run MFVB separately for each grid point. Then, compute weight of each grid point using the Boltzmann distribution.

Select $_d$ = f ; ^{2}g.

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 tra•c (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 o€ering 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.