The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2018-2

Multi-stage Multivariate Modeling of Temporal Patterns in Prescription Counts for Computing Drugs in a Therapeutic Category

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Discussion of 'Multi-stage Multivariate Modeling of Temporal Patterns in Prescription Counts for Computing Drugs in a Therapeutic Category'

Refik Soyer
Department of Decision Sciences
The George Washington University

I would like to congratulate Serhiyenko, Ravishanker and Venkatesan (henceforth, SRV) for proposing a novel modeling strategy for multivariate count series that arises in marketing of pharmaceuticals. I also would like to thank the editors for giving me the opportunity to discuss this interesting paper.

With increasing volume of Web-based data, modeling and analysis of discrete valued time series have gained more attention; see for example, the volume by Davis, Holan, Lund and Ravishanker (2015) for recent advances. Analysis of discrete-valued series poses modeling difficulties and computational challenges. As a result, limited attention has been given to multivariate discrete valued time series in the literature. The framework of *observation driven* and *parameter driven* models of Cox (1981) is still applicable for temporal correlation but one also needs to describe dependence of components of the vector time-series. Some of the recent observation driven models for multivariate counts include multivariate integer valued autoregressive (INAR) models of Pedeli and Karlis (2011) and the multivariate Poisson series of Ravishanker, Serhiyenko, and Willig (2014) who considered a multivariate Poisson observation model with a hierarchical dynamic setup. Computation of the multivariate Poisson likelihood requires a significant computational effort in Ravishanker et al. (2014) and also the model is highly parameterized as a result of sharing common terms. An alternative approach is to use parameter driven models as considered by Aktekin, Polson and Soyer (2017). This model is a generalization of the univariate Poisson dynamic time series considered by Aktekin and Soyer (2011) and Aktekin, Soyer and Xu (2013). Concept of *common random environment* is used to incorporate dependence among multivariate components of the series. Although, the multivariate models proposed by Aktekin et al. (2017) provide analytical forms for conditional distributions which allow them to use particle filtering methods, they are limited to positively correlated series.

An important contribution of the SRV paper is that the proposed "level corre-

lated" models provide a more flexible correlation structure for components of the vector of counts. The ability to capture both positive and negative correlations of the components of the series is an attractive feature of these models. The proposed set up is nicely motivated by the particular application considered by the authors. More specifically, SRV considers $J-$variate (*different drugs*) vector of counts collected on n subjects (*doctors*) over T time periods. The level correlated model (LCM) is given by

$$Y_{jit}|\lambda_{jit} \sim Pois(\lambda_{jit})$$

for $j = 1, \ldots, J$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ with link equation

$$log(\lambda_{jit}) = \gamma_{jt} + \beta_{ji0} + \boldsymbol{z}'_{\boldsymbol{jit}}\boldsymbol{\beta}_{ji} + \alpha_{jit}$$

where the J dimensional random effect term vector $\boldsymbol{\alpha}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ induces the correlation structure for components of the count vector $\boldsymbol{Y}_{it}$. Furthermore, the state equation for each component is given by

$$\gamma_{jt} = \gamma_{j,t-1} + w_{jt}$$

where $w_{jt} \sim N(0, 1/W_j)$. This Markovian structure introduces the temporal correlations in the model.

The authors develop a Bayesian inference for the model via the integrated Laplace approximation (INLA). They point out that for large n INLA is more efficient than MCMC methods. The proposed estimation procedure seems to be quite efficient for low values of J. Although the application involves $J = 3$, it is important to comment on performance of INLA for large dimensions. Especially, evaluation of posterior modes in implementing INLA may pose challenges as J increases.

In certain applications, one may be interested in real time updating of the parameters as new observations become available. I wonder if INLA is efficient for such sequential updating for moderate dimensions. This is especially important when one includes more dynamic parameters such as the fixed effect terms in the model.

SRV point out that the observation equation can be generalized for any other univariate distribution of counts (UDC)

$$Y_{jit}|\boldsymbol{\Theta}_{jit} \sim UDC(\boldsymbol{\Theta}_{jit})$$

where $\boldsymbol{\Theta}_{jit}$ is the vector parameters associated with the specific UDC. As discussed in the paper, the cases of negative binomial and zero-inflated Poisson are straightforward, but there is also referral to Conway-Maxwell (COM) Poisson distribution where computation of the individual likelihoods may become cumbersome as J increases. I would like to invite authors to comment on how plausible this extension is and any experience they may have with this type of models. Also, I would like to draw the attention of the authors to potential use of conjugate priors, considered by Kadane et al. (2006), for developing COM Poisson LCMs.

References

[1] Aktekin, T. and Soyer, R. (2011). Call center arrival modeling: A Bayesian state-space approach. *Naval Research Logistics*, 58:28-42.

[2] Aktekin, T., Soyer, R., and Xu, F. (2013). Assessment of mortgage default risk via Bayesian state space models. *Annals of Applied Statistics*, 7:1450-1473.

[3] Aktekin, T. Polson, N. G., and Soyer, R. (2017). Sequential Bayesian analysis of multivariate Poisson count data. *Bayesian Analysis*, to appear.

[4] Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*. 8, 93-115.

[5] Davis, R. Holan, S., Lund, R. and Ravishanker, N. (2015). *Handbook of Discrete-Valued Time Series*, Chapman and Hall/CRC.

[6] Kadane, J. B., Shmueli, G., Minka, T. P., Borle, S. and Boatwright, P. (2006). Conjugate analysis of the Conway-Maxwell-Poisson distribution, *Bayesian Analysis*, 1:363-374

[7] Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) model with application. *Statistical Modelling*, 11:325-349.

[8] Ravishanker, N., Serhiyenko, V., and Willig, M. R. (2014). Hierarchical dynamic models for multivariate times series of counts. *Statistics and Its Interface*, 7:559-570.