Rejoinder: A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications

Alex Reinhart

I would first like to thank the discussants for their interesting and insightful comments. I was first motivated to write a review when I noticed the diversity of applications of self-exciting processes and the interesting innovations being made in widely separate fields, so it has been rewarding to read such a broad range of perspectives from the discussants.

I broadly agree with many of the points brought up by the discussants, so here I would like to focus on a few topics that point to areas of future research.

1. CAUSALITY AND MUTUALLY EXCITING PROCESSES

Professor Ogata raises an important topic otherwise omitted from my review: "the modeling of point processes for causality analysis from a stochastic process including another point process." He discusses models whose intensity function depends not just on the past history of the process but the past history of *another* process, for example, models of earthquake events incorporating stress change or fault-weakening events, and the use of AIC and other techniques to perform model selection and test hypotheses about earthquake precursors.

This resembles the "leading indicators" incorporated by Mohler (2014) to allow events such as disorderly conduct and public drunkenness to be used as predictors for more serious crimes, though Mohler did not perform extensive inference on these leading indicators to test specific hypotheses. I agree with Ogata that modeling of such leading indicators is essential for answering important questions in many areas of application, and would like to note some recent theoretical developments which may help with the task.

If we have multiple separate processes, such as earthquake events and precursor events, or records of several different types of crimes, we can consider them as a single multivariate point process. As noted in Section 2.3 of the review, a multivariate point process is simply a marked point process in which the mark space is a finite set $\{1, \ldots, m\}$. In a mutually exciting multivariate process, instead of a single triggering function g, there is a matrix of functions g_{ij} specifying the effect of event type j on the intensity of process i.

There has been a recent surge in methods for detecting causality in multivariate point processes, though largely focused on purely temporal processes. Eichler, Dahlhaus and Dueck (2017) recently demonstrated that the triggering functions g_{ij} have a direct connection to Granger causality: events of type i do not Granger-cause events of type j if and only if $g_{ji}(t) = 0$ for all $t \in \mathbb{R}$. This can be used to define a Granger causality graph G whose directed edges satisfy the property

$$ i \to j \notin G \quad \Longleftrightarrow \quad g_{ji}(t) = 0 \quad \text{for all } t \in \mathbb{R}, $$

indicating the Granger causality relationships between the mutually exciting processes.

There are now several competing methods for estimating the causality graph G and testing hypotheses about its edges [e.g., Chen, Witten and Shojaie (2017), Xu, Farajtabar and Zha (2016), Achab et al. (2017)], again focusing on purely temporal processes. It would be quite interesting to see these methods extended to spatio-temporal mutually exciting processes and used in applications, where G may answer substantive scientific questions and the use of leading indicator processes could improve predictions.

2. STOCHASTIC RECONSTRUCTION

Professor Zhuang points out that the goodness-of-fit methods given in the review evaluate the fit of the entire model, either on all the data or over specific space-time regions (like the residual methods of Section 3.6), leaving an important gap: it may be necessary to assess

Alex Reinhart is PhD candidate, Department of Statistics & Data Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, Pennsylvania 15213, USA (e-mail: areinhar@stat.cmu.edu).

specific components of the model formulation, such as the form of the triggering function g or of the background model μ. Stochastic reconstruction [Zhuang, Ogata and Vere-Jones (2004)] promises to allow this kind of careful model evaluation.

The procedure is intuitively straightforward, exploiting the cluster process representation and the triggering probabilities in equations (9) and (10) of the review. Suppose we have fit a particular parametric model to data and wish to validate one component, like the distribution of time between a parent event and the offspring event it triggers. Following Zhuang, Ogata and Vere-Jones (2004), Section 6.2, we can reconstruct a probability density function for this distribution from the data:

$$\hat{g}(t) = \frac{\sum_{i,j} \Pr(u_i = j)\mathbb{I}(|(t_i - t_j) - t| < \Delta_t/2)}{\Delta_t \sum_{i,j} \Pr(u_i = j)},$$

where Δ_t is a small positive number. $\hat{g}$ is hence an estimate of the distribution of times between an event i and prior events j, weighted by the probability that event j triggered event i. Similar estimators can be made for the distribution of distances between parent events and offspring, or for the expected number of events triggered by any event. The reconstructed distributions can be compared to the distributions implied by the particular triggering function g assumed by the model.

An example will illustrate the procedure and also illustrate my misgivings. Consider a spatio-temporal model of the form

$$\lambda(s, t) = \nu + \sum_{i:t_i < t} g(s - s_i, t - t_i),$$

where the true triggering function g is such that offspring events are normally distributed in space and $\chi^2(6)$-distributed in time, relative to their parent event. Suppose, however, that we fit the model with an incorrect g^* which implies offspring events are normally distributed in space but *exponentially* distributed in time. The distributions chosen are shown in Figure 1.

We then simulate from the true spatio-temporal model with $\chi^2(6)$-distributed offspring times, and fit the misspecified exponential model g^* to the data. After stochastic reconstruction, we obtain the distribution $\hat{g}$ of offspring times shown in the left panel of Figure 2. The distribution matches quite closely to the exponential distribution assumed by the model, with only a few small deviations. However, it is quite different from the true $\chi^2(6)$ offspring distribution used to simulate the data.
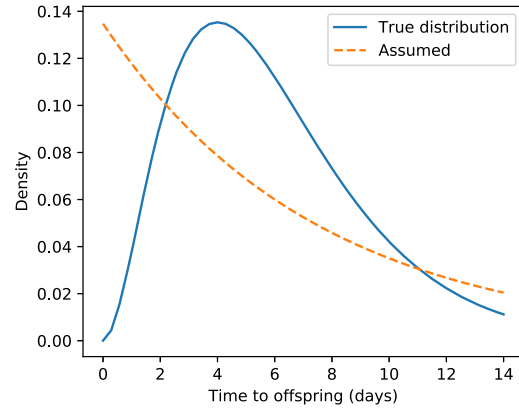

FIG. 1. *The true $\chi^2(6)$ distribution of times to offspring events, in blue, compared to the incorrect exponential distribution fit to the data, in dashed orange, for the simulation shown in the left panel of Figure 2.*

The right panel of Figure 2 shows the result of a reversed simulation: the true offspring distribution is exponential, but we choose a model g^* which assumes it is $\chi^2(6)$. (The degrees of freedom here are not fit to the data but assumed outright.) After stochastic reconstruction, the estimated $\hat{g}$ matches a $\chi^2(6)$ much better than it matches the true exponential distribution. There are deviations suggesting a problem, but from the deviations alone we could not intuit that the true distribution is exponential.

The problem is that $\hat{g}$ is constructed using $\Pr(u_i = j)$, which is calculated with the assumed g^* using equations (9) and (10). The reconstructed time distribution hence implicitly assumes the truth of g^*. These simulations suggest that it can be difficult to use stochastic reconstruction to validate specific features of a self-exciting model. Zhuang (2006) suggests an iterative procedure, repeatedly using $\hat{g}$ to recalculate $\Pr(u_i = j)$ and obtain a new $\hat{g}$, and I would be interested to see similar simulation studies to determine its ability to detect different types of model misspecification, along with further study of methods to detect model misspecification.

3. ESTIMATION

I wholly agree with Professor Schoenberg about the challenges of maximum likelihood estimation of self-exciting models: I perhaps naively implied that one can simply write down the likelihood, work out the EM update steps and forget about estimation thereafter, but it is not quite so easy. Beyond the integral term pointed out by Schoenberg, one must also worry about multimodal likelihoods, spatial and temporal boundary ef-
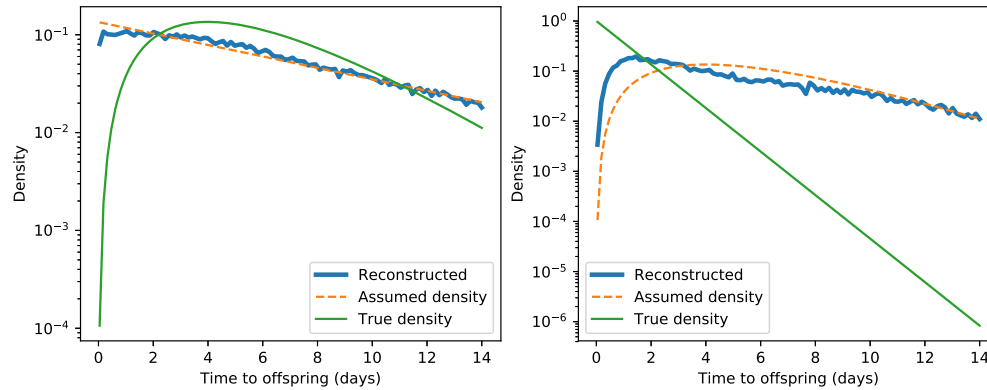
FIG. 2. *Distributions (log scale) of time to offspring events, after stochastic reconstruction. At left, the reconstructed distribution $\hat{g}$ fits quite well to the exponential distribution assumed by the model, even though the true distribution is $\chi^2(6)$. At right, the reconstructed distribution $\hat{g}$ does not quite fit the assumed $\chi^2(6)$, but its deviations do not hint at the true underlying exponential distribution.*

fects and the computational cost of repeatedly evaluating the likelihood. [Ogata notes faster procedures to evaluate the likelihood of temporal processes in Ogata, Matsu'ura and Katsura (1993), including an accurate approximation to the integral term, and there are other approximations for spatio-temporal processes, like Oates (2015).] Implementing maximum likelihood estimation for a new model and exploring its behavior may take months of work.

New estimation procedures—and flexible general-purpose software packages implementing them, like those discussed by Dr. Meyer—will be essential to make self-exciting models practical for a wider range of problems. Further understanding of boundary effects, which may be arbitrarily confounded with relevant covariates or other feature of the process, will also be essential.

Dr. Meyer notes the problems of underreporting and aggregation in public health data, which present other obstacles to estimation and bias parameter estimates. These problems also appear in crime forecasting, where crimes are reported at unknown rates which may vary across space and time, and where publicly available data may be aggregated to various levels of resolution and contain events with uncertain times or locations (e.g., when a burglary is reported, the actual time the burglary occurred may not be known). The reporting rate may be of interest on its own, since it varies with public trust in police and for many other complex reasons.

Perhaps underreporting and boundary effects can be considered manifestations of the same problem. Consider a self-exciting process $\lambda(s, t)$ and a function $r(s, t)$ representing the probability that an event occurring at (s, t) would be reported. Boundary effects can

be modeled by setting $r(s, t) = 0$ for s or t outside the observation region, implying that events occurring outside the region are never observed, and underreporting corresponds to $r(s, t) < 1$ for s and t *inside* the observation region. For a given r, how would the estimated model $\hat{\lambda}(s, t)$ from the reported data differ from the true model $\lambda(s, t)$? Can anything be learned about r from the observed data? Could r sometimes depend on the history of the process? Can its effects on $\hat{\lambda}(s, t)$ be bounded, perhaps by making assumptions about its structure?

The existing literature has explored boundary effects, but the effects of other forms of underreporting have not been widely studied, and a great deal of further work will be necessary to answer these questions.

4. CONCLUSION

Self-exciting spatio-temporal point processes are an active area of research across many fields of application, and as the discussants—major contributors to the field—have shown, there are still many interesting open questions. I can only echo the discussants in hoping that this review and discussion will trigger further theoretical development; as methods advance and practical software becomes more widely available, I hope to see new applications in many areas of science.

ACKNOWLEDGMENTS

REFERENCES

ACHAB, M., BACRY, E., GAÏFFAS, S., MASTROMATTEO, I. and MUZY, J.-F. (2017). Uncovering causality from multivariate Hawkes integrated cumulants. In *Proceedings of the 34th International Conference on Machine Learning* **70**.

CHEN, S., WITTEN, D. and SHOJAIE, A. (2017). Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process. *Electron. J. Stat.* **11** 1207–1234. MR3634334

EICHLER, M., DAHLHAUS, R. and DUECK, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *J. Time Series Anal.* **38** 225–242. MR3611742

MOHLER, G. O. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497. DOI:10.1016/j.ijforecast.2014.01.004.

OATES, C. J. (2015). Accelerated non-parametrics for cascades of Poisson processes. *Stat* **4** 183–195. MR3405400

OGATA, Y., MATSU'URA, R. S. and KATSURA, K. (1993). Fast likelihood computation of epidemic type aftershock-sequence model. *Geophys. Res. Lett.* **20** 2143–2146. DOI:10.1029/93gl02142.

XU, H., FARAJTABAR, M. and ZHA, H. (2016). Learning granger causality for Hawkes processes. In *Proceedings of the 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 1717–1726. PMLR, New York.

ZHUANG, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 635–653. MR2301012

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.* **109** B05301. DOI:10.1029/2003JB002879.