Statistical Science
2018, Vol. 33, No. 3, 325–326
https://doi.org/10.1214/18-STS652
Main article: https://doi.org/10.1214/17-STS629

Comment on "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications" by Alex Reinhart

Frederic Paik Schoenberg

This is an excellent and extremely well-written summary of recent research on self-exciting spatial-temporal point processes. It contributes very nicely to the literature and I will use it personally to teach my graduate students about the topic. The author should be congratulated for his excellent writing.

I would like to comment briefly on estimation. Again, Reinhart provides a superb review, and seeing the current state of knowledge regarding maximum likelihood estimation (MLE) and its variants, one may walk away from this article with the misleading impression that estimation for spatial-temporal point processes is a solved problem that can readily be attacked not only by MLE but also by various other techniques such as E-M or stochastic reconstruction. However, in practice there are real problems with the implementation of many of the methods here.

The first and in my opinion main shortcoming of MLE is the integral term in Reinhart's equation (8). For some very simple models, this integral can be computed numerically as a function of the parameters being estimated, but this is rare. In practice, one must approximate this integral numerically. The problem is that, in MLE, one is searching over a vast parameter space, and the numerical approximation to the integral must be a close approximation for *all* of the parameter space, or else the optimization function may choose some parameter vector where the approximation is poor. Anyone who has dealt with MLE knows the sort of Murphy's Law to which I am referring. If anything can possibly go wrong with the approximation to the likelihood function, MLE seems to have a way of gravitating to it. Harte (2013) comments nicely on the importance of the issue of integral approximation in MLE in practice. Another issue with the integral is programming. In practice, it is not easy to program a function to compute an accurate approximation to the integral term in Reinhart's equation (8) as a function of the parameters being estimated. One reason this can be particularly difficult is that, in many useful cases, the triggering function being estimated is highly volatile, especially for realistic values of the parameters being estimated, and integrating a highly variable function accurately is difficult. Again, with MLE Murphy's Law seems to apply, and even a very small error in programming or approximating the integral term, including an error that is only relevant for certain values of the parameters, will tend to be exploited by the optimization routine in MLE.

Reinhart is correct that the use of the E-M algorithm in conjunction with MLE can help, but it is quite unclear why it helps. The theory surrounding the desirable asymptotic properties of the MLE are well known, and since the E-M modification is an approximation to MLE, the E-M method of Veen and Schoenberg (2008) should have similar properties, but it is entirely unclear why the E-M method should outperform ordinary MLE. In private communication, Bin Yu has expressed the belief that any practical advantages to the method of Veen and Schoenberg (2008) may be attributable merely to the stopping routine. That is, it may be that the default stopping routine for the E-M method may simply be better than that for the ordinary MLE. Even if this is not the case, it should be pointed out that the E-M MLE still requires computation or approximation of the integral term and, therefore, is still susceptible to the problems pointed out above. The same is true for parametric and semiparametric methods I have seen, such as those described by Reinhart in Section 3.2. Nonparametric estimation methods are fantastic, but in many cases estimating parametric models is also desirable. As Reinhart mentions, in Schoenberg (2013) I have tried to avoid the computation of the integral term by noting that for some Hawkes' processes it can be well approximated, and its approximation very simply computed, by integrating over all of $\mathbf{R}^d$ rather than

Frederic Paik Schoenberg is Professor of Statistics, University of California, Los Angeles, 8152 MS building, Los Angeles, California 90095, USA. (e-mail: frederic@stat.ucla.edu).