

Estimation of spatial-temporal point process models using the (stochastic) Expectation Maximization algorithm and its application to California earthquakes

Alejandro Veen¹ and Frederic Paik Schoenberg¹

¹UCLA Department of Statistics, 8125 Math Sciences Building, Box 951554, Los Angeles, CA 90095-1554, USA

Corresponding Author: Alejandro Veen, veen@stat.ucla.edu

Abstract: We explore the use of the Expectation Maximization (EM) algorithm as well as its stochastic version for estimating the parameters of spatial-temporal point process models, in particular the Epidemic-Type Aftershock Sequence (ETAS) model [Ogata 1988, 1998]. We present situations in which this algorithm performs better than directly maximizing the log-likelihood function and show an application to earthquake occurrence data from Southern California using a spatial-temporal ETAS model.

1. Introduction

The Expectation Maximization (EM) algorithm was first introduced in 1977 by Dempster, Laird, and Rubin and has since proven to be an efficient algorithm for finding Maximum Likelihood estimates in settings one can describe as *incomplete data problems*. In this work, we will extend this methodology to the realm of spatial-temporal point process models.

The Epidemic-Type Aftershock Sequence (ETAS) model [Ogata 1988, 1998] can in fact be formulated as an incomplete data problem, since it can be interpreted as a sub-critical branching process with immigration, in which some earthquakes can be considered to be background events while others are treated as being triggered by a preceding event. In this framework, the *observable* quantities are the time, location, and magnitude of an earthquake and there is an *unobservable* quantity, which is the information on which earthquake triggered which aftershock or whether an earthquake is a background event. The so-called *complete data* would be the combination of the observable and unobservable quantities, if it were possible to measure them.

Note that we do not intend to make a statement about the geophysical processes of how earthquakes trigger one another. In particular, the assumption that earthquakes can be assigned to single triggering events can be disputed – at least from a geophysical point of view. However, the ETAS model allows for such an interpretation. Alternatively, the ETAS model also allows for an interpretation in which multiple preceding events trigger subsequent earthquakes as their combined effect. For a more extensive explanation of alternative interpretations of the ETAS model, see for instance the beginning of section 2.2 in Sornette and Werner, 2005.

For computational reasons, it may be favorable to think of the ETAS model as a branching process with immigration. In this case, the EM algorithm or its stochastic version can be applied with the unobservable quantity being the information on which earthquakes triggered which aftershocks or whether an earthquake is a background event.

2. The EM algorithm and its stochastic version

Let θ denote the parameter vector of a spatial-temporal ETAS model and let $\ell(\theta|D)$ be the log-likelihood function for a given a data set D . Then, the Maximum Likelihood estimate $\hat{\theta}$ can be expressed as $\hat{\theta} = \arg \max_{\theta} \ell(\theta|D)$.

The EM algorithm provides results that are asymptotically equivalent to the direct maximization of the log-likelihood function. However, it has computational advantages, does not depend very much on starting values, and is often more robust than non-linear maximization (or minimization) algorithms. The EM algorithm works in two steps, which ought to be repeated until convergence.

- Expectation step (E-step):
Using the current estimate $\hat{\theta}^{(n)}$ for the parameter vector θ , find the expectation of the complete data log-likelihood $\ell(\theta | D, U)$, where U is the set of unobservable quantities. This expectation can be written as $E_{\hat{\theta}^{(n)}}[\ell(\theta | D, U)] = \sum_{u \in U} p(u | D, \hat{\theta}^{(n)}) \ell(\theta | D, U)$.
- Maximization step (M-step):
Update the current estimate by finding the Maximum Likelihood estimate of θ using the expectation derived in the E-step, rather than the incomplete data log-likelihood. Denote the new estimate as $\hat{\theta}^{(n+1)}$.

In short, one can write the EM algorithm as $\hat{\theta}^{(n+1)} = \arg \max_{\theta} E_{\hat{\theta}^{(n)}}[\ell(\theta | D, U)]$.

Finding the Maximum Likelihood estimate of θ in the M-step can be a computationally simpler than the direct maximization of the incomplete data log-likelihood $\ell(\theta | D)$. This situation arises in cases where the partial derivatives of the complete data log-likelihood $\ell(D, U | \theta)$ are simpler than those of the incomplete data log-likelihood $\ell(\theta | D)$.

The *stochastic* EM algorithm differs in the sense that the E-step is replaced by a “Random Assignment” step. Using the current estimate $\hat{\theta}^{(n)}$, the probability that a given earthquake i is triggered by the preceding earthquake j as well as the probability that earthquake i is a background event can be computed. Using this probability vector, earthquake i is then *randomly* assigned to a triggering “parent” earthquake or classified as a background event. Note that this idea is similar to what is called “stochastic reconstruction” in Zhuang, Ogata, and Vere-Jones (2004). The M-step then maximizes the complete data log-likelihood function, using the generated random assignment of earthquakes to their triggering events in place of the unobservable quantity U .

3. Application of EM techniques to Southern California earthquake data

In a simulation study, we will present the properties of the EM algorithm for the estimation of a spatial-temporal ETAS model. This methodology is then applied to the estimation of such a model using earthquake occurrence data from Southern California.

4. Concluding Remarks

Simulation studies show that EM-type algorithms are not only computationally efficient and robust, they also seem to yield estimates that are closer to the real parameters than those derived from directly maximizing the log-likelihood function.

5. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0306526. We thank Yan Kagan and Ilya Zaliapin for helpful comments and the Southern California Earthquake Center for its generosity in sharing their data.

5. References

- Dempster, A., Laird, N., and Rubin, D., *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society, Series B, 39/1, , 1977, 1 – 38.
- Ogata, Y., *Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes*, Journal of the American Statistical Association, Vol. 83, 1988, 9 – 27.
- Ogata, Y., *Space-time point-process models for earthquake occurrences*, Annals of the Institute of Statistical Mathematics, Vol. 50, 1998, 379 – 402.
- Sornette, D. and Werner, M.J., *Apparent Clustering and Apparent Background Earthquakes Biased by Undetected Seismicity*, J. Geophys. Res., Vol. 110, No. B9, B09303, 10.1029/2005JB003621.
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), *Stochastic declustering of space-time earthquake occurrences*, Journal of American Statistical Association, Vol. 97, No. 458, 369 – 380.
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2004), *Analyzing earthquake clustering features by using stochastic reconstruction*, J. Geophys. Res., 109, B05301, doi:10.1029/2003JB002879.