Report on the doctoral thesis *Statistical inference for spatial and space-time Cox point processes* by Jiří Dvořák.

Jan M. Swart

August 4, 2014

Statistics for spatial point processes is a lively subject field. Its main aim is to describe observed random point patterns, i.e., to develop models for the random distribution of points in space and then fit these to real data, allowing for quantitative statements about such data.

The present thesis, which is based of work of the author and his supervisor, starts with an overview of some known models for clustered point patterns. There are two competing approaches to parameter estimation based on such models. The first approach simply tries to find the maximum likelihood estimator given the full observed point pattern. Since this is computationally very demanding, especially if there are many observed points, one typically resorts to Markov Chain Monte Carlo methods. The second, more pragmatic approach uses moment characteristics, typically second moment characteristics like the covariance function, to define estimators that one hopes will perform well. A possible advantage of this approach is that by basing oneself on physically relevant quantities like the covarance function, one may hope that the estimators give reasonable information even when the choice of the model is not entirely correct, as may be expected in many practical situations. In contrast, the direct maximum likelihood method is like a black box method where one does not have much insight into what aspects of the data are used.

The heart of the thesis is formed by Chapter 3, where the author tries to adapt methods for clustered spatial point processes to space-time point processes, where one coordinate (time) plays a special role. Chapter 4 contains suggestions for further work. As usual, statistics lives from the combination of three ingredients:

1. Calm, systematic reasoning.

2. Experience, gut feeling, and the “right” choice of tuning constants.

3. Inspiration and imagination.

In this particular instance, the first two ingredients are present throughout the thesis, but the third one (inspiration) seems to have come quite late. Indeed, it seems to me that in retrospect, the right method for parameter estimation for space-time point processes, that should be the main novel contribution of the present thesis, is the method suggested (but not worked out) in Section 4.1.1.

Instead, the main part of the thesis focusses on a method based on projections and a "refined method" that holds a middle position between the method using projections and the (in my point of view best) method suggested in Section 4.1.1.

To explain this, I need to go a bit into detail. The main models of interest are built up in the following way. First, one creates a Poissonian collection of "mother points" in space and time, where each mother point has a weight. The collection of mother points may be dense, but the total weight of all mother points in a bounded set should be a.s. finite. Next, one constructs a random function Λ as an infinite sum of kernels k, centered at the position of the mother points and multiplied by their weights. Then, one constructs a Poisson point set with random intensity Λ. In the final step, this Poisson point set is thinned by a slowly varying function of space and time, to allow for some local variation in the overall density, as often observed in practise. The difference between a purely spatial model and a space-time model is in the choice of the kernel k. In the spatial case, one usually chooses k rotationally symmetric. In the space-time case, it is more natural to choose for k the product of a rotationally symmetric spatial kernel and a completely different temporal kernel.

In the spatial case, when k is rotationally symmetric, parameter estimation is based on the observed correlation function or an equivalent quantity (such as $\hat{K}(r)$, the average number of points within a distance r of a given point), which because of rotation symmetry is a function of one variable. In the space-time case, the correlation function depends on two variables, space and time, which play different roles. In this case, one can base the estimation on the average $\hat{K}(r,t)$, over all points in the observed point set, of the number of other points within a certain spatial distance r and temporal distance t of the given point. In fact, it is suggested in Section 4.1.1 that to estimate the spatial parameters, it is sufficient to use the function $\hat{K}(\,\cdot\,,t_0)$ for one, suitably chosen value of t_0, and to estimate the temporal parameters, it is sufficient to use the function $\hat{K}(r_0,\,\cdot\,)$ for one, suitably chosen value of r_0.

Surprisingly, this is not the approach chosen by the author in the bulk of the thesis. Instead, he bases the estimation of the spatial parameters on the projection of the observed point set to the spatial domain (i.e., forgetting the time coordinate of each point), and he bases the estimation of the temporal parameters on the projection of the observed point set to the temporal domain. As long as the observation window is not too big, this works quite well. Indeed, when the spatial size of the observation window is close to the optimal choice of the constant r_0 mentioned above, and the temporal size of the observation window is close to t_0, this is very similar to the method suggested above, as confirmed by simulation studies carried out by the author.

When the space-time observation window is large, however, estimation based on the projection processes becomes increasingly bad. In fact, consistency doesn't even hold. Indeed, for concreteness, let space be two-dimensional, let W_n $(n \geq 1)$ be a sequence of spatial observation windows, increasing to $\mathbb{R}^2$, and let T_n $(n \geq 1)$ be a sequence of observation time intervals increasing to $\mathbb{R}$. Then, as $n \to \infty$, the projection processes on both the spatial and temporal domain become increasingly dense, because of the overlapping of clusters with almost the same spatial, but different temporal coordinates, and vice versa. As the author observes correctly on page 54, the result of this is that the estimates for the spatial and temporal correlation functions become increasingly bad.

He solves this by throwing away a large part of his information. Instead of using the whole observation window $W_n \times T_n$, he bases his estimates of the spatial parameters on the window $W_n \times T_1$ and the estimates of the temporal parameters on the window $W_1 \times T_n$. In this way, consistency and much more, such as asymptotic normality of the estimators, can be proved, albeit with a much smaller rate of convergence than one would have when making optimal use of the whole observation window $W_n \times T_n$, as in the method suggested in Section 4.1.1. I think the greatest shortcoming of the thesis is that this point is insufficiently explained. In fact, the beginning of Section 3.1.3 seems to suggest that information from the whole window $W_n \times T_n$ will be used while one finds only out what is really happening one page later, where it is stated rather vaguely that "Note that, for all $n \geq 1$, we use the same temporal projection process X_t (projected from the fixed spatial region W)…". In fact, there is nothing to "note" here. The text before suggested something different from what the author actually does.

Apart from this, there is also much worthy of praise in the thesis. The three simulation studies are set up and discussed very carefully. Given the complicated nature of the processes and the estimation procedure, there is a great deal of work hidden in each table of results. The proofs of the asymptotic properties of the estimators (when using different observation windows to estimate the spatial and temporal parameters, as discussed above) seem to be carefully written. As the author points out, he follows rather closely a similar proof in a purely spatial setting from a paper by Waagepetersen and Guan, but he seems to add quite a bit of detail. I find it a bit of a pity that the theorems remain rather abstract. To apply Theorem 3.11, one has to check no less than 15 conditions, some of which are not at all trivial. Even though the author gives some comments on how these conditions can be checked and how hard this is, I very much missed a completely concrete example (including a concrete choice for the inhomogeneity function and its parametrization) where all assumptions are satisfied.

One thing I also missed in the thesis was an example (if only one!) of real space-time data. As this thesis is clearly written from a practical perspective and often (for example, on page 45) practical applications are mentioned as a motivation, I would encourage the author to actively seek scientific data that someone would really like to analyze.

The thesis is generally well-written. The style is clear, and the English is good (certainly beter than my Czech) and generally easy to understand. The interpunction could sometimes be better and the author makes frequent mistakes in the use of the articles. But this does not matter very much for readability.

Concluding, my general impression of the thesis is positive. The author has clearly demonstrated his capability to independently perform original research.

Below, I conclude my report with a list of more detailed comments, aimed at improving the presentation skills of the author.

Small corrections

- Page 5: "Hence, devoted space-time methods should be used...". This sounds odd. "Methods devoted to a space-time setting" would sound fine. (Also in Czech "venovane prostorove-casove metody" sounds strange to me.) On page 20, "dedicated space-time approach" is a similarly strange construction.
- Page 10: When discussing clustering and repulsion, it could be remarked that the rest of the thesis is devoted only to models with clustering and that models with repulsion will not be discussed further.
- Page 11: "the original Ripley's K-function" should be "Ripley's original K-function".
- Page 11: "The weights constitute of reciprocals" The preposition "of" is wrong here and even without it, it is hard to understand what you mean. Don't you simply want to say that the weights *are* the reciprocals?
- Page 11: "On the other hand, a non-parametric" The expression "on the other hand" means: by contrast. I see no contrast here.
- Page 14: "clustering or regularity" Don't you mean: "clustering or repulsion"?
- Page 14: In the last displayed formula, the right bracket should be before the $u_1, \ldots, u_k$.
- Page 15 "models of Cox point processes" Since a Cox point processes is a model, this sounds like you are discussing a model of a model.
- Page 17 "If we condition by the positions of the points" It is to condition *on*. Also, you mean mother points, not points.
- Page 20 "points both with spatial and temporal coordinate" should be "points with both a spatial and temporal coordinate".
- Page 22 "(with both $u, v \in W$ or both $t, s \in T$)" It seems to me you want to say "and" instead of "or".
- Page 25 "that can be easily revealed" should be: "that can easily be revealed". This mistake occurs more often. (Page 62: "can be easily generalized" should be "can easily be generalized".)
- Page 25 fottowing should be following.
- Page 27 "each Y_x is again a point process ... with intensity function $\lambda_o(\cdot; \theta)$" This is only approximately true, if one ignores boundary effects. This is clear from your later discussion, but should be mentioned at this point.
- Page 31 "For simulation of the Thomas process realizations" This construction is very hard to read and understand. "For the simulation of the realizations of Thomas processes" is much better.

- Page 32 "(i.e. the estimate lied out of a generously...)" The past form of to lie [ležet] is lay, as opposed to to lie [lhát], where it is lied. So correct is: "the estimate lay out of" although "the estimate was lying outside" is perhaps preferable.

- Page 47 eventhough should be even though

- Page 48 "Henceforth" means: "from now on" unlike "hence" which means "therefore". I have the feeling you wanted to say "hence" here, or even better: "by (3.6), (3.7) and (3.10)". On the other hand, on page 5: "Hence we restrict our attention...", it seems you wanted to say henceforth.

- Page 56 In this epsilon-delta definition, delta is actually a large number, which is a bit confusing.

- Page 57 "denote $\mathcal{F}^X(A)$ the σ-algebra" You should say: "denote *by* $\mathcal{F}^X(A)$ the σ-algebra" or "let $\mathcal{F}^X(A)$ denote the σ-algebra". This occurs in more places.

- Page 62 "does not appear neither in..." This is a double negation. Correct is: "appears neither in...".

- Page 63 When discussing the triangular kernel as an alternative choice, it is helpful for the reader to mention whether this actually satisfies (A6).

- Page 63 "has only a single parameter which considerably affects" I had a lot of trouble understanding this until I read on. It was not even clear if the parameter affected something, or the fact that there is only one had some effect. You wanted to say it makes no sense to choose a two-parameter family of kernels when one of these parameters cannot be estimated unless the data are almost perfect (see Figure 3.2).

- Page 66 (below): "If such a situation was encountered". Better is "If such a situation were to be encountered"

- Page 67 "When t_* is too large, many disjoint clusters ... appear to be one large cluster". I think that when t_* is too large, these clusters overlap, i.e., they are *not* disjoint.

- Page 69, Section 3.2 Your claim that cluster overlapping is not a serious issue depends on the size of the temporal interval. Observing your process for too long (which should mean collecting good data!) will still result in cluster overlapping and loss of information. Since you do not even give one example of real data, it is hard for me to believe nobody will observe his or her process for too long.

- Page 75 "Different normalization ... makes the formulas ... somewhat complicated, but it is needed ...". Here one should rather say "*this* is needed". "It" points to a concrete object, while "this" points to statements, ideas, etc.

- Page 75 In the same sentence as mentioned above, it should be "*The* different normalization" and "to keep *the* variances of the respective terms".

One could say: “to keep variances controlled”, but once one talks about the variances *of the respective terms* these are not just any variances, hence the determined article. This sort of mistakes with articles occur frequently in the thesis.

- Page 90 “the estimates precision” This is a rather heavy construction. “the precision of the estimates” is much more pleasant to read and understandable.
- Page 91 precedures should be procedures