Petr Volf:
Report on doctoral thesis of Mgr. Kateřina Helisová
"Models for random union of interacting discs"

The thesis concerns to important area of stereologic stochastic processes. It deals with an interesting theme – namely a model (set of models) for random process of discs and their union, with interaction of discs. Interactions are described via characteristics which are easily derived from geometric properties of union (in statistical analysis problem from geometric properties of analyzed image data) and computed with the aid of power tessellation created from set of discs.

After Introduction and part 2 bringing a set of basic definition (in rather compressed form), Chapter 3 is devoted to exact description of the model, its theoretical properties and relations between them (formulated as propositions). Further, the density of the union of discs is specified, in a simple log-linear form, so that belonging to exponential family with linear combination of geometrical characteristics mentioned above. It is actually a density with respect to reference Boolean model, a basic model of Poisson type. Part 3.4 then contains also description of procedure for MCMC simulation a representation of the model.

Chapter 4 deals with problems of statistical analysis. It is due the log-linear form of density and easily computable characteristics that the formulation and maximization of log-likelihood is not difficult - except two points. The first one concerns the edge effect, when observing window is relatively small w.r. to size of connected components of discs union. This problem is solved in the thesis both theoretically and practically, by conditional likelihood formulation. The second problem is that normalizing constants are not tractable and must be obtained (as proposed by authoress) by simulation.

And here is actually the point for my first questions and remarks concerning the proposed solution itself, but also a wider context of approach:

1. Statistical solution is based on standard MLE concept, however, the part of computation uses MCMC generation and part of input (parameters of reference process) must be selected, actually from preliminary analysis of data. Why, in that case, not to use Bayes approach consistently, with reference process as prior, its parameters being hyper-parameters etc. in a proper hierarchical structure. Are there any works using such an approach, in similar setting?

2. Page 41: How parameter θ_0 is selected. Is it changed during computations?

3. P. 45 -46: CI are actually just approximate, because likelihood is also approximate, due simulations of proportion of normalization constants.

4. Tables show strong dependence of results on selection of reference process. Could the reason be identified and, possibly, eliminated? (Again, interesting question could be whether, in Bayes solution, such a sensitivity to prior will appear, too).

5. Part 4.6. - Model validation - is very valuable, showing that really interaction model proposed by author is closer to data that simple Boolean model. On the other hand (and it is also seen from Figure 4.4/ p.50), the statement of last

sentence of this part (p.55) holds, saying approximately: '... in this sense, the models are not fitting well, possibly since the heather data are rather smooth while a discs process is naturally more rugged.'

It could also mean that another selection of reference process could be better. For instance a process of clusters, birth process, or another process more taking into account natural features (and origin) of data - even a process with growth. Actually, these variants are mentioned in conclusion of the thesis, as directions of promising model enlargements. Then, what direction will the authoress prefer in nearest future?

6. The analysis deals exclusively with P. Diggle's heather data-set. However, it is also convenient to show model properties and analysis on a set of simulated data, in order to see better a 'consistency' of model, analysis and data features. Such analysis surely has been performed, are there any references, or could such results be commented? Has the model been applied also to another real data?

7. Chapter 5 describes the algorithms and software programmed in C++, which is also a part of the thesis. Again, it is also a valuable part of whole work.

   To this, I have just a general question: Is the software able to analyze any data (of proper form), without large adaptation to them?

8. Several minor remarks to Chapter 3:

   In Part 3.3, there are 2 sets of characteristics, related to union U (p.22) and to tessellation B (p.25). It seems to me that characteristics concerning U should not contain terms as 'edges' or 'vertices', while the last characteristics N on p.25 concerns discs, not tessellation (?)

   p.23[7]: "We *often* consider $\chi = N_{cc} - N_h$." What 'often' means. Always, or is it specified, when yes and when no?

   p.29_{13}: In $\lambda(x, v)$ terms 2 and 3 are missing (?)

The whole thesis is written with evident knowledge of problems and on high mathematical level. My remarks above are inspired mainly by my curiosity and interest in problems, number of minor imprecisions is rather small and do not influence the level of thesis. This thesis, as well as other works of the authoress, demonstrate clearly her ability to scientific research and to further achievements in this area. Therefore, I recommend the thesis for defence and as a basis for conferment Mgr. Kateřina Helisová a degree PhD.

Praha, 31.07.2008