Evaluation report on Jiři Dvořak's Habilitation thesis

Summary of the thesis

In his thesis, Jiři Dvořak developes new non-parametric Monte Carlo tests to analyse point patterns, random fields, and independence between two random variables. The null distribution is constructed by using permutations, random shifts or stochastic reconstruction and for functional test statistics, global envelope tests are applied.

The thesis consists of a well written introduction (Chapters 1-4) and five papers. In Chapter 1, some basic definitions concerning point processes, Monte Carlo tests, and global envelopes are recalled. Then, an overview of Paper 2 and Paper 3 on permutation tests is given in Chapter 2 and of Paper 1, Paper 4, and Paper 5 on random shift tests in Chapter 3. Tests based on stochastic reconstruction are briefly discussed in Chapter 4. Finally, the five papers are given in Appendix. In all papers, the methods are illustrated by simulations and real examples and the computer codes are made available. Below, I give a short summary of the main results of the thesis. The papers included in the thesis are all jointly authored with other researchers. Therefore, when I write "Jiři" or "he" in what follows, I mean "Jiři and his collaborators".

In Paper 1, Jiři suggests a new non-parametric test of independence of two components of a stationary bivariate process, either a random field or a point process, based on the random shift approach. The torus and border corrections are very popular methods to account for the unobserved data outside the observation window. However, the torus correction breaks the marginal spatial correlation structure resulting in liberal tests and is restricted to rectangular windows. The border correction, i.e. erosion of the window, does not have the problem of breaking the autocorrelation structure but it is suitable only if the number of data points is large enough. In his thesis, Jiři suggests a new variance correction which preserves the autocorrelation structure and can be used for any shape windows. The test statistic is computed from the data in the intersection of the window and a shifted window which results in it being based on different amounts of data due to the different shifts of the window. Therefore, the values are standardized to have the same mean and variance. It is shown in the paper that the asymptotic order of the variance of the sample covariance of a random field is one over the number of sampling locations. In the point process case, the cross K-function or the mean cross nearest neigbour distance is used as a test statistic. The variance of the estimator of the cross K-function is computed by using different versions of the Campbell theorem. In a simulation study, it is shown that the variance correction reduces the liberality of the test and works better than the torus correction for random fields. In the point process case, it is demonstrated that the cross K-function is an appropriate test statistic in the Poisson and regular cases. However, if the point pattern is clustered, the mean cross nearest neigbour distance is recommended.

Three different tests to investigate the first-order separability of a homogeneous spatio-temporal point process, i.e. whether the spatial and temporal intensity functions of the process can be considered separately, are proposed in Paper 2. Two of tests are non-parametric Monte Carlo tests, one based on permutations and one on stochastic reconstruction, where the test statistics are based on the nonparametric estimates of the nonseparable and separable intensity functions. The global envelope tests are used to provide a graphical interpretation and information on where and when the possible non-separability occurs. The third test is an asymptotic χ2-test. It is demonstrated in a simulation study that the permutation test and the asymptotic χ2-test work well under the Poisson assumption but not for clustered point patterns. The test based on stochastic reconstruction, an approach to obtain replications that have the same properties as the observed data, and the similarity between the observed and reconstructed point patterns is measured by using an energy functional chosen by the user. This test works well for clustered point patterns as demonstrated with a real application.

In Paper 3, general non-parametric Monte Carlo tests of independence between two random variables that can be applied for any continuous or discrete distributions as well as for mixtures of the two are proposed. The work was motivated by the fact that the existing methods only summarize information from the whole distribution and do not give any information on which quantiles cause the possible rejection of the null hypothesis. The null distribution under independence is obtained by randomly permuting the order in one of the two samples. Two test statistics are suggested. One of them is the sample cumulative distribution function computed on a fine grid of quantiles. The other one is constructed by using a two-dimensional QQ-plot computed on a fine grid of quantiles which is regarded as a realization of a point process and the kernel estimator of the intensity function of the process is used as the test statistic. Under the assumption of independence, the QQ-plot is a homogeneous point process in the unit square. Using global envelopes, both tests provide graphical visualization of the results and information on not only whether the two random variables are independent but also for which quantiles the possible dependence occurs for continuous distributions. For categorical distributions, adapted versions of these tests reveal which combinations of the categories cause the possible rejection of the null hypothesis. Jiři provides theoretical proofs for that the permutation test rejects the null hypothesis at the correct significance level and for how the power of the tests depends on the number of permutations. According to a simulation study, the tests seem to perform competitively or better than the existing methods.

The main question in Paper 4 is how to test the hypothesis of independence between a covariate and the marks in a marked point process. Since the underlying unmarked point process is typically not independent of the covariate and the marks, the dependence structure in the triangle points-marks-covariates is investigated in three steps: independence between the points and marks, between points and the covariate, and between the marked points and the covariate. The tests are constructed using the non-parametric random shift method with variance correction introduced in Paper 1. To test independence between points and marks, the l_2 norm of the difference between the estimated conditional expectation of the mark and the mean observed mark is used. It is shown by using, among other tools, the Campbell theorem and the Cauchy-Schwarz inequality that the variance of the test statistic is of order 1/n, where n is the number of points. The tests of independence between the points or marked points and the covariate are based on the random shift approach given that the covariate values are available in the entire observation window. When independence between the points and the covariate is tested, the sample mean of the covariate values observed at the point locations is used as the test statistic. If the two are not independent, the points of the process are more likely to appear in locations with high/low covariate. In the case of marked points and covariates, the test statistic is chosen to be either Pearson's correlation coefficient or Kendall's rank correlation coefficient since there are possible preferential sampling issues if the sample covariance is used. The asymptotic order of the variance of the test statistic is shown to be 1/n by using e.g. the Campbell theorem and Fubini's theorem. All the proposed tests are compared to some existing tests in a large simulation study. To test the independence between points and covariate or marked points and covariate, the tests introduced in the thesis perform better than the existing tests. If the mark is categorical or if more than one covariate is of interest, global envelope test can be applied.

In Paper 5, Jiři proposes a non-parametric approach for testing significance of a covariate, taking into account possible effects of nuisance covariates. It is shown in a simulation study that the existing parametric tests are often quite liberal even if the parametric model for the intensity and interaction are correct and do not perform so well if at least one of the models is misspecified. He introduces non-parametric smooth residuals by using the kernel estimated intensity function and defines a correllation coefficient between a point process and a covariate using these residuals (no nuisance). Furthermore, to measure the dependence between the point process and the covariate of interest among several possibly correlated covariates, he defines a partial correllation coefficient after having removed the possible effect of covariates, some of which can be nuisance covariates. Two Monte Carlo tests of covariate significance using the random shift approach suggested in Paper 1 are proposed. One of the tests is fully non-parametric and does not require the selection of any intensity or interaction model. The other test is semi-parametric, where an intensity function needs to be selected but no assumptions of the interaction structure are needed. A large simulation study, where the null hypothesis of independence between the point process and the covariate (given the possible nuisance covariate) is tested against various alternatives, is performed. The performance of the tests is investigated when the underlying interaction structure is completely spatially random, clustered, regular, or a combination of regular and clustered. The tests proposed in the paper do not have the problem of liberality and have powers which are comparable to the powers of the existing parametric tests when the models are selected correctly. If either the intensity or the interaction model is misspecified, the tests proposed in this thesis perform better. The tests can be used even for categorical covariates by using the global envelope test.

Evaluation of the thesis

The thesis provides a series of new non-parametric methods for testing independence between two components of a bivariate process, between a covariate and a mark in a marked point process or between a point process and a covariate, testing first-order separability of a spatio-temporal point process, and testing significance of a covariate. I would particularly like to mention the new variance correction method in the random shift approach that overcomes the problems that the previous edge correction methods have. In addition, a general test for testing independence between two random variables is provided. The construction of this test, namely regarding the two-dimensional QQ-plot as a realization of a point process and analyzing it as such, was especially clever. All the tests are compared to existing tests and in general, the tests proposed in this thesis perform very well. A huge advantage of the suggested non-parametric tests is that one does not have to make (hardly) any assumptions concerning the underlying random processes.

The originality of the thesis is very high. Four of the fives papers are already published in good journals and the fifth one has been submitted. This explains why the Turnitin system gave a rather high overall match 52%.

I find Jiři's thesis a very good and well written habilitation thesis. New non-parametric methods are developed and some drawbacks of some of the existing methods are pointed out and corrected. Proofs for theoretical results are provided and all the methods are illustrated by simulation studies and real data. In addition, the R codes are made publicly available or available upon request.

To summarize, the contributions in this thesis are very important additions to the existing literature, especially on point pattern analysis.

Gothenburg, Sweden, on May 15, 2023

Sincerely,

Aila Särkkä, professor
Department of Mathematical sciences
Chalmers University of Technology and the University of Gothenburg