Two-step estimation for inhomogeneous spatial point processes - a simulation study

Rasmus Waagepetersen†
Aalborg University, Aalborg, Denmark
Yongtao Guan
Yale University, New Haven, USA

Summary.This report describes a simulation study of the results in Waagepetersen and Guan (2008). We refer to this paper and Waagepetersen (2007) for background on the simulation study.

1. Simulation study

To check how the asymptotic results in Waagepetersen and Guan (2008) apply in finite-sample settings we consider simulation studies for an inhomogeneous Thomas process and an LGCP with exponential correlation function. The set-up for the simulation study is similar to the one in Waagepetersen (2007) with altitude and slope covariates and a 1000 m by 500 m simulation window. The altitude and slope parameters β_2^* and β_3^* used in the simulations are given by the parameter estimates in Waagepetersen (2007). In the case of the Thomas process we let $\psi = (\log \kappa, \log \omega)$ while $\psi = (\log \sigma^2, \log \phi)$ for the LGCP. In the simulation study we focus on the asymptotic normality of $\hat{\psi}$, the asymptotic standard errors for $\hat{\psi}$, and the coverage properties of approximate confidence intervals based on the asymptotic normality of the parameter estimates. In all cases, the model fitted to the simulated data coincides with the model that was used to generate the simulated data. That is, issues of model misspecification is not considered in the simulation study.

In the case of an inhomogeneous Thomas process we vary (κ^*, ω^*) and the expected number of points μ^* to reflect varying degrees of clustering and tree abundance. We let ω^* equal to 10 or 20 while κ^* is 1×10^{-4} or 5×10^{-4} corresponding to expected numbers 50 or 250 of mother points within the plot and recall that larger κ^* and ω^* results in less clustering. The expected number μ^* of simulated points is either 200 or 800 corresponding to "sparse" and "moderately abundant" point patterns. For each combination of κ^*, ω^*, and μ^* we generate 1000 synthetic data sets and obtain simulated parameter estimates by applying our estimation procedure with $r = 100$ and $c = 0.25$. We compute the empirical standard deviation of the simulated parameter estimates and we evaluate for each simulation the asymptotic covariance matrix by plugging in the corresponding simulated parameter estimate. Approximate 95 % confidence intervals are constructed using standard errors extracted from the the estimated asymptotic covariance matrices. We only report results obtained with $\tilde{\Sigma}_n$ since similar results are obtained with Σ_n.

†*Address for correspondence:* Rasmus Waagepetersen, Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg
E-mail: rw@math.aau.dk

Table 1. Columns 4-6: standard deviation for $\hat{\psi}_1 = \log \hat{\kappa}$ estimated from simulations, median of standard deviations obtained from estimated asymptotic covariance matrices, coverage of nominal 95 % approximate confidence intervals. Columns 7-9: as columns 4-6 but for $\hat{\psi}_2 = \log \hat{\omega}$.

κ^*	ω^*	μ^*	sd_1	$\widehat{\mathrm{sd}}_1$	cvrg_1	sd_2	$\widehat{\mathrm{sd}}_2$	cvrg_2
$1{\times}10^{-4}$	10	200	0.28	0.28	0.94	0.15	0.16	0.96
$1{\times}10^{-4}$	10	800	0.25	0.24	0.94	0.12	0.12	0.96
$1{\times}10^{-4}$	20	200	0.39	0.38	0.95	0.19	0.2	0.96
$1{\times}10^{-4}$	20	800	0.30	0.30	0.94	0.12	0.13	0.96
$5{\times}10^{-4}$	10	200	0.47	0.48	0.98	0.30	0.32	0.95
$5{\times}10^{-4}$	10	800	0.25	0.24	0.94	0.14	0.14	0.96
$5{\times}10^{-4}$	20	200	1.72	2.24	1.00	0.72	1.44	1.00
$5{\times}10^{-4}$	20	800	0.43	0.43	0.95	0.22	0.23	0.96

Table 2. Columns 4-6: standard deviation for $\hat{\psi}_1 = \log \hat{\sigma}^2$ estimated from simulations, median of standard deviations obtained from estimated asymptotic covariance matrices using Σ_n (and $\tilde{\Sigma}_n$ in parentheses), coverage of nominal 95 % approximate confidence intervals. Columns 7-9: as columns 4-6 but for $\hat{\psi}_2 = \log \hat{\phi}$.

$\sigma^{2,*}$	ϕ^*	μ^*	sd_1	$\widehat{\mathrm{sd}}_1$	cvrg_1	sd_2	$\widehat{\mathrm{sd}}_2$	cvrg_2
0.5	15	800	0.34	0.34 (0.34)	0.96	0.39	0.41 (0.42)	0.97
0.5	30	800	0.24	0.25 (0.26)	0.97	0.35	0.36 (0.37)	0.95
1.5	15	800	0.17	0.20 (0.23)	0.98	0.21	0.23 (0.25)	0.96
1.5	30	800	0.18	0.19 (0.23)	0.95	0.24	0.24 (0.28)	0.92

Except for the 7th row, the simulation results in Table 1 shows fine agreement between the empirical standard errors and the median asymptotic standard errors for the simulated parameter estimates. The coverages of the confidence intervals are also fairly close (in general within 1%) to the nominal coverages of 95%. The problems in row 7 is probably due to that the parameter values $\kappa^* = 5{\times}10^{-4}$ and $\omega^* = 20$ corresponds to the least clustered case and with only 200 simulated points on average it may often be hard to distinguish the estimated K-function from that of a Poisson process. This leads to rather extreme values of the parameter estimates and for 3% of the simulated point patterns for row 7, the minimum contrast procedure did in fact not converge.

Quantile plots based on the simulated parameter estimates are shown in Figure 1 and Figure 2. The distributions of the parameter estimates are in general fairly close to normal. Bivariate scatter plots (omitted) of $(\log \hat{\kappa}, \log \hat{\omega})$ indicate that the joint distribution is well approximated by a bivariate normal and that $\log \hat{\kappa}$ is strongly negatively correlated with $\log \hat{\omega}$. However, for reasons discussed in the above paragraph, the case $\kappa^* = 5{\times}10^{-4}$, $\omega^* = 20$, and $\mu^* = 200$ is an exception where the distributions of both $\log \hat{\kappa}$ and $\log \hat{\omega}$ are very heavy tailed.

For the LGCP we restrict attention to the case $\mu^* = 800$ and values of $\sigma^{2,*} = 0.5, 1.5$ and $\phi^* = 15, 30$. Proceeding as for the Thomas process we obtain Table 1. The asymptotic results work rather well for $\sigma^{2,*} = 0.5$. For $\sigma^{2,*} = 1.5$ the asymptotic standard errors tend to overestimate the true standard errors and this is especially so for the asymptotic standard errors obtained with $\tilde{\Sigma}_n$. The quantile plots in Figure 3 show some deviations from normality both when $\sigma^{2,*} = 0.5$ and $\sigma^{2,*} = 1.5$ but the deviations seem rather modest.

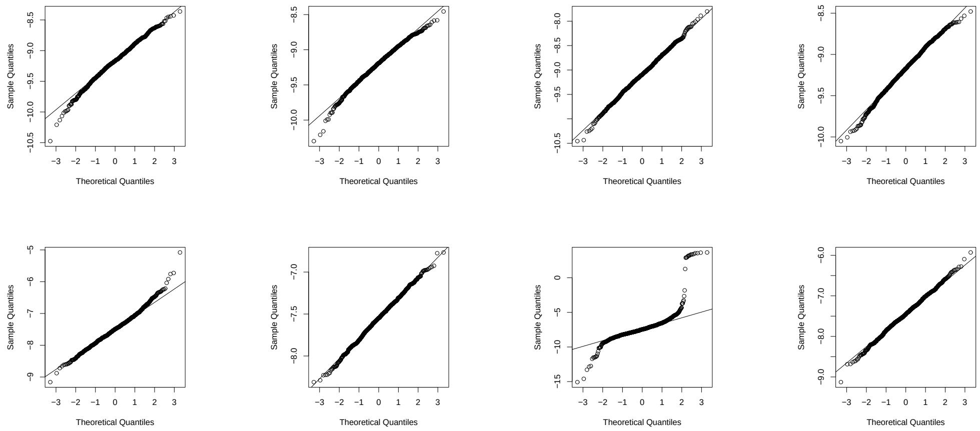

Fig. 1. Empirical quantiles of the simulated parameter estimates for $\log\kappa$ against quantiles for a standard normal distribution. Upper row, $\kappa^* = 1{\times}10^{-4}$, lower row: $\kappa^* = 5{\times}10^{-4}$. First two columns: $\omega^* = 10$ and second two columns: $\omega^* = 20$. First and third column: $\mu^* = 200$, second and fourth column: $\mu^* = 800$.

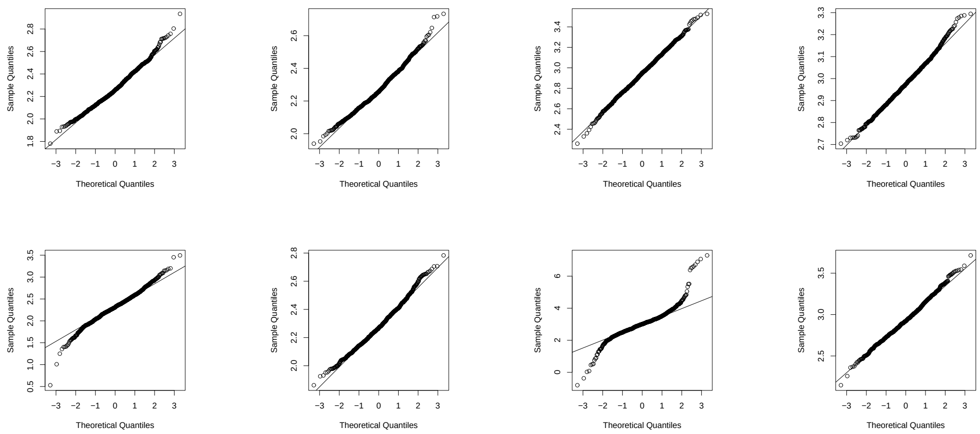

Fig. 2. Empirical quantiles for the simulated parameter estimates for $\log\omega$ against quantiles of a standard normal distribution. Upper row, $\kappa^* = 1{\times}10^{-4}$, lower row: $\kappa^* = 5{\times}10^{-4}$. First two columns: $\omega^* = 10$ and second two columns: $\omega^* = 20$. First and third column: $\mu^* = 200$, second and fourth column: $\mu^* = 800$.

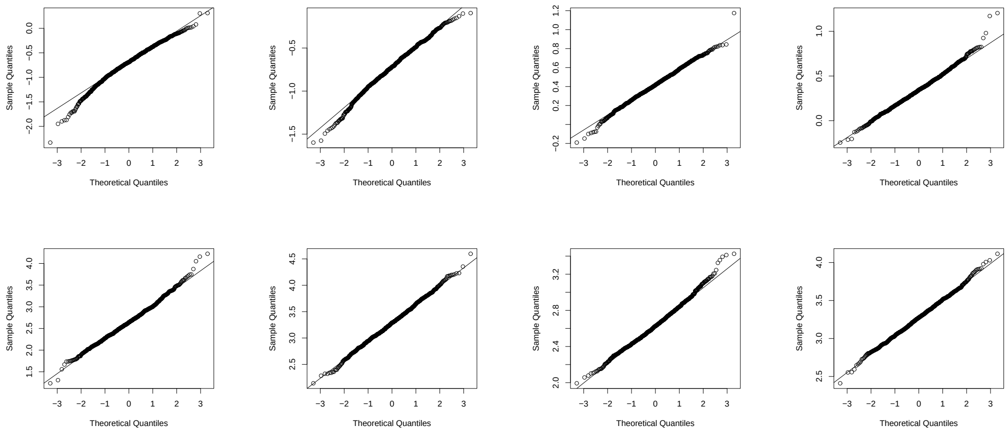

Fig. 3. Empirical quantiles for simulated parameter estimates of $\log \sigma^2$ (upper row) and $\log \phi$ (lower row) against quantiles of a standard normal distribution. From left to right: $(\sigma^{,2,*}, \phi^*) = (0.5, 15), (0.5, 30), (1.5, 15), (1.5, 30)$.

References

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.

Waagepetersen, R. & Guan, Y. (2008). Two-step estimation for inhomogeneous spatial point processes. Under revision.