Moment Estimation Methods for Stationary Spatial Cox Processes – A Simulation Study

Jiří Dvořák and Michaela Prokešová

June 4, 2012

Abstract: This report presents complete results of the simulation study described in the paper Dvořák and Prokešová [3].

Keywords: Moment estimation methods, Spatial Cox point process, Parametric inference

1 Simulation study

The paper Dvořák and Prokešová [3] presents a comparison of empirical performance of the moment estimation methods for model fitting of spatial stationary Cox point processes. These methods were introduced in the literature in the past few years as faster, simulation-free alternatives to computationally prohibitive maximum likelihood estimation.

Direct comparison of the three estimators – minimum contrast estimation and composite and Palm likelihood aproaches – is not available in the literature. The formulas for the asymptotic variance of these estimators are complicated and depend on integrals with respect to the third and fourth-order moment measures of the point processes in question. Thus, straightforward comparison of the efficiency of these estimators is not possible.

To compare the empirical performance of the estimators we chose two different types of cluster processes – the Thomas process and the log-Gaussian Cox process, see Examples 3.1 and 3.2 in [3].

To assess the performance of the estimators on middle-sized to large point patterns exhibiting different degree of clustering we chose eight combinations of parameter values for the Thomas process ($\kappa = 25$ or 50, $\mu = 4$ or 6 and $\sigma = 0.02$ or 0.04, representing relatively strong and weak clustering, respectively) and for the log-Gaussian Cox process ($\beta = 10$ or 20 representing relatively strong and weak dependence, respectively, $\sigma^2 = 1$ and μ calculated so that the intensity of the process is 100, 150, 200 and 300 as in the case of the Thomas process).

For each process and each combination of parameters we generated 500 independent realizations and re-estimated the parameters using the three moment estimation methods (i.e. using formulas (2), (3) and (4) in [3]).

2 Parameter estimation – computational details

Realizations of the Thomas process and the log-Gaussian Cox process with appropriate parameters in a unit square window W were simulated using the package `spatstat` for `R` (for reference see [1]).

When simulating realizations of the log-Gaussian Cox process we need to use a finite representation of the random field Z in the observation window W. We chose a regular grid of 25×25 points and instead of the Gaussian random field Z we simulated a Gaussian random vector representing values of Z in the given lattice points. Then we approximated values of the random field Z in W by the value in the nearest lattice point.

We used the minimum contrast method (MCE) with both the K-function (MCEK) and the pair correlation function g (MCEg). The tuning parameters in the case of the K-function were

chosen according to the recommendation in [2, Section 6.1.1]. The same values of parameters were used for the g-function, i.e. $c = 0.25$, $R = 0.25$ and r equal to the minimum interpoint distance observed in the given realization. The estimation was conducted by the functions provided by `spatstat`.

The composite likelihood (CLE) and Palm likelihood (PLE) estimation methods have only one tuning parameter, R, determining the maximum distance of pairs of points that will be taken into account. The distance R has to be chosen carefully so that the variance of the estimators is reduced but not much information about the interaction in the point pattern is lost. Since the observation window W is a unit square we chose the values of R to be 0.1, 0.2 and 0.3, respectively.

To take into account the interpoint interactions when choosing the value of R we also consider $R = r_{data}$, where r_{data} corresponds to a range of correlation. It is determined as the smallest r for which $\hat{g}(r) < 1$ holds, where $\hat{g}$ is a nonparametric estimate of the pair correlation function g.

For CLE and PLE we applied the inner region correction for the edge effects, see (3) and (4) in [3]. It is straightforward and can be used for data observed in irregular windows. Other edge corrections could be used as well, for example the torus correction could perform better in the case of a rectangular window but it is difficult to use in practice for irregularly shaped windows.

The estimation for CLE and PLE was conducted in software `Mathematica 7`. Maximization of the appropriate log-likelihood functions was performed using a combination of two methods.

First, derivatives of the log-likelihood function with respect to the unknown parameters were calculated and a vector of parameter values was found for which all the derivatives were equal to zero using a default Newton's method.

Then, if the numerical method diverged (i.e. the estimate lies out of a generously long interval containing the true value of the parameter), the parameters are estimated again by direct maximization of the log-likelihood function using simulated annealing.

This procedure is motivated by the fact that the estimation using derivatives is fast but somewhat numerically unstable, while direct maximization is more computationally demanding but in general less likely to diverge. The combination of these two methods was a good compromise between the computational time and numerical stability.

For the minimum contrast methods and the composite likelihood method the value of the non-identifiable parameter μ was estimated from the observed intensity of the point process as described in Examples 3.1 and 3.2 in [3].

3 Results

Main results are summarized in Tables 1 to 4 for the Thomas process and the log-Gaussian Cox process, respectively. They show results for two variants of MCE (based on K-function and pair correlation function) and for CLE and PLE with different choices of the tuning parameter R.

The tables show relative mean biases of the estimators and relative mean squared errors (MSEs) (by relative we mean divided by the true value of the estimated parameter or by its square for the MSE). All the statistics were obtained from the middle 95 % of the estimates from 500 replications.

Neglecting the 5% of the most extreme estimates was motivated by the fact that in certain situations the estimation methods can be numerically unstable, see Section 5.1 in [3]. If such a situation was encountered in practice the estimates would easily be identified as unrealistic and one would alter the parameters of the underlying optimization methods or opt for an alternative estimation method. However, due to the extent of the computations involved, this was not possible in this simulation study. Note that if all the estimates were used to calculate the statistics these would be severly distorted by the numerical instability and hence completely uninformative for a prospective researchers willing to find a method of choice for their particular dataset.

3.1 Thomas process

3.1.1 Parameter σ

Among the three parameters of the Thomas process model σ is the one estimated with most accuracy by any of the compared methods. Particularly MCEg and CLE with $R = 0.1$ give very precise estimates. From the two MCE methods MCEK is always worse than MCEg. PLE with $R = 0.1$ (as well as CLE with $R = 0.2$) is comparable with MCEK for strong clustering. For weak clustering PLE is inferior to the other methods.

The reason for this is the strong positive bias (around 25% for $R = 0.1$) of PLE in case of weak clustering. PLE generally overestimates the parameter σ while MCE always underestimates it. CLE with $R = 0.1$ is virtually unbiased for strong clustering and positively biased for weak clustering. For strong clustering CLE with $R = 0.1$ generally showed the smallest variance followed by MCEg, PLE and MCEK. For weak clustering variance of CLE and PLE with $R = 0.1$ becomes larger and MCEg becomes the most stable estimator.

An important observation is the rapid deterioration of the quality of the CLE and PLE estimates with increasing value of the tuning parameter R. Increase in the variability of the estimators with growing R is to be expected, since with the employed inner region edge correction and with growing R we lose growing percentage of the data. However, for CLE and $R = 0.3$ the estimator of σ is not just bad, it is even numerically unstable, and in case of weak clustering and higher intensity of the point process the same happens also for $R = 0.2$.

An explanation for this could be found in the exact form of the score equations for CLE - the estimating functions as described in Example 3.1 in Section 4 of [3] have a steep step around the correct value of the parameter σ and then become virtually constant (and nonzero). Thus for a small number of observed pairs of points from X and a clustered point pattern (with few distant pairs observed) the estimating equation can lead to a severe overestimation of the parameter σ. Therefore it is extremely important when estimating the interaction parameters for processes with weaker clustering by CLE or PLE to choose the tuning constant R not too large. Definitely not larger than the range of correlation and reasonably small with respect to the size of the observation window so that not too much data is lost by the edge correction.

3.1.2 Parameter κ

For the other interaction parameter κ estimated directly by all three methods the best estimates are provided by PLE (with $R = 0.1$, for weak clustering also with $R = 0.2$) in the majority of cases. The exception are processes with low intensity and strong clustering, i.e. with a few number of tight well defined clusters in the observed point pattern – here MCEg gives better results. Like with the other parameters the MCEK method is consistently worse than MCEg and the worst results are generally provided by CLE (with $R = 0.1$).

Note that the good results of PLE ($R = 0.1$) are mainly caused by the comparatively small variability of the estimates because all the PLE estimates with $R = 0.1$ have considerable negative bias. The bias becomes lower for $R = 0.2$ and thus the best estimates for processes with low intensity and weak interaction (i.e. a few loose clusters in the observation window) are obtained by PLE with $R = 0.2$. This behaviour is again implied by the exact form of the estimating functions used. Note that for CLE the situation is somewhat similar like for the parameter σ – CLE with $R = 0.1$ provides nearly unbiased estimates, whereas with $R = 0.2$ we have considerable positive bias and with $R = 0.3$ the estimates become useless. Nevertheless the variability of CLE is generally higher than variability of the other estimators.

When comparing the MCEg and MCEK methods the main factor is also variability of the estimators - the MCEg method is consistently less variable than MCEK (the same situation like with the parameter σ).

3.1.3 Parameter μ

The last parameter μ is by all the methods determined by means of the observed number of points of X and the values of the already estimated parameters. It is true not just for the MCE and CLE where $\hat{\mu} = X(W)/(|W|\hat{\kappa})$ but also for the PLE method. Namely (5) in [3] is derived just from the comparison of the expected and observed number N of pairs of points from X.

Thus it is natural that the quality of the estimates of μ for MCE and CLE follows the pattern from estimation of κ with MCEg being better than MCEK and CLE providing even worse estimates. The PLE estimator uses mean number of pairs of points from X which could provide more exact estimates than just $X(W)$. On the other hand, the formula (5) in [3] includes both of the parameters κ and σ and thus can be influenced more by the bad quality of those estimates. The simulation study shows that in reality for weak clustering PLE (with $R = 0.1$) provides the best estimates of μ – again mainly due to low variability of the estimates since they are negatively biased. For strong clustering the estimates provided by MCE are better.

3.1.4 Procedure using r_{data}

The estimated values of r_{data} are for strong clustering concentrated around 0.1 as expected, with a fair proportion being smaller than 0.1 (see Figure 1 for the histograms of r_{data}). For CLE the estimates of σ using $R = r_{data}$ are comparable with those obtained with $R = 0.1$; they have only slightly larger bias and variance. The bias is nevertheless still very small (smaller than for MCE) and the variance is only a bit larger than for MCEg. Overall, the procedure with r_{data} for CLE estimation of σ provides very good estimates – better than for MCEK. Estimates of the other interaction parameter κ by CLE with r_{data} are worse but comparable with estimates produced by CLE with $R = 0.1$.

For PLE the estimates of the interaction parameters with $R = r_{data}$ are also worse than for fixed value $R = 0.1$. By closer inspection of the simulated data we see that particularly for the parameter κ the cases with $r_{data} < 0.1$ produce significantly biased estimates.

For weak clustering the estimated value of r_{data} is scattered between 0.1 and 0.2 with a few cases taking the value of 0.25 which was the upper bound for the estimated range of interaction (since the estimate of g function is not very stable for values larger than 0.25). For estimation of σ CLE with $R = r_{data}$ has significantly larger bias and variance than with $R = 0.1$ and PLE produces impractical estimates. For estimation of κ CLE with $R = r_{data}$ produces sometimes better estimates than with $R = 0.1$ but they are still worse that those produced by any other method. PLE with $R = r_{data}$ produces generally worse estimates of κ than with $R = 0.1$ but these are typically still better than those produced by the other methods.

In conclusion we can say that the method using $R = r_{data}$ is not superior to the fixed value of the tuning parameter $R = 0.1$ for CLE and PLE. Partially it can be explained by the larger variability caused by the changing value of R but very often also the bias of the estimators is larger than with the fixed value of $R = 0.1$. The question of the ideal choice of R for a given point process model and a given observation window is not a simple one. Even the amount of information expressed by the mean number of observed pairs of points with distance $\leq R$ for a clustered point process X need not be a monotone function of R (see [4, Section 3]). More sophisticated methods of the adaptive choice of R must be used to produce a considerable improvement in the quality of the estimates than just a simple choice $R =$ range of interaction.

3.1.5 Overall performance

In conclusion we can say that quality of all the compared estimators improves with higher intensity of the process X and stronger interactions in the point process (i.e. tighter clusters). Concerning minimum contrast estimation the version using pair correlation function always yields better estimates than the version using the K-function. When using the CLE and PLE methods for estimation of the interaction parameters it is important to use reasonably small values of the tuning parameter R which provide reasonably good estimates. High values of R lead to numerical

instability of the methods and impractical estimates – this is especially important for CLE, to a lesser extent also for PLE.

To address the overall performance, the minimum contrast estimation using the pair correlation function provides the best (in the sense of the relative MSE) estimates of the interaction parameters although for point patterns with strong clustering when estimating σ CLE (with $R = 0.1$) yields fully comparable and sometimes better quality estimates.

For point patterns with weak clustering PLE (with $R = 0.2$ and $R = 0.1$) yields the best estimates of κ. However, this is due to the minimal variability of the PLE estimator which has a serious negative bias in this case. The second best (according to MSE) estimator is MCEg which has half-size bias (i.e. acceptable 10%).

Parameter μ is calculated from the intensity (of points or pairs of points) of X and as such it depends on the quality of the other parameter estimates. The best values of MSE were achieved for strong clustering by MCEg and for weak clustering by (again significantly biased) PLE.

3.2 Log-Gaussian Cox process

3.2.1 Parameter β

For the log-Gaussian Cox process from Example 3.2 β is the scale parameter of the covariance function of the Gaussian driving field and the hardest one to estimate. Unlike in the case of Thomas process for minimum contrast estimation MCEK showed always better performance than MCEg. Nevertheless for stronger interaction ($\beta = 10$) the best results are obtained by CLE with $R = 0.1$ which has both minimal bias and minimal variability. Also here (as in the case of Thomas process) it is important to have reasonably small value of the tuning parameter R, i.e. $R = 0.1$. For larger values of R the CLE (and also PLE) deteriorates fast and the estimates become useless.

Quality of the estimates of course improves with growing intensity λ; the growing amount of information seems to be best used by CLE since for high intensity ($\lambda = 300$) CLE (with $R = 0.1$) outperforms MCEK even in the case of weak dependence. CLE (with $R = 0.1$) generally slightly underestimates β but the absolute value of bias is much smaller than for any other compared estimator and this holds uniformly for any log-Gaussian Cox process considered in the study. When the variability of this estimator becomes small enough (thanks to the sufficient amount of data in the large λ case) it becomes superior to MCE estimation even in the case of weak dependence.

To understand this results better it is good to note that the case $\beta = 10$ which we call stronger interaction/dependence means that values of $g(u) - 1$ (where g is the pair correlation function) are significantly positive for a larger range of values u than for the case of $\beta = 20$. Thus the observed point pattern of X for $\beta = 10$ is much more variable and may contain a few clusters (or sometimes none if λ is low) with large scale and large number of points (like the one in the upper right corner of the right panel in Figure 1 in [3]) whereas for $\beta = 20$ the observed point pattern is more homogeneous with large number of smaller (both in terms of scale and number of points) clusters. The large clusters from X with $\beta = 10$ make the estimation of the functional characteristics K and g less stable than in the case of $\beta = 20$ or the Thomas process, which influences negatively the quality of the MCE estimation. Obviously the CLE method is less influenced by occurrence of these large clusters.

The same fact may explain higher efficiency of MCEK when compared with MCEg. Even in the case of weaker dependence ($\beta = 20$) the clusters observed in X are highly variable in terms of the number of points and there are always some with a fairly high number of points. In such a situation the estimate of the K-function as a cumulative function is more stable than the estimate of the g function (which corresponds to the density of K-function).

The performance of PLE is inferior to the other methods and for weak dependence ($\beta = 20$) the estimator is unusable. For stronger dependence in the point pattern the performance for PLE with $R = 0.1$ is comparable with the MCE methods, for larger values of R the estimator becomes unusable again.

3.2.2 Parameter σ^2

The other interaction parameter σ^2 is best estimated by the MCE methods. The estimates are only slightly biased and the variance is lower than for the other estimators. Note that the sign of the bias depends on the value of β – for $\beta = 10$ (i.e. point patterns with a few large and heavy clusters) σ^2 is underestimated while for $\beta = 20$ it is overestimated by both MCEK and MCEg. The performance of the two MCE methods is very similar.

Worse but still reasonable results are obtained for PLE with $R = 0.1$ and strong dependence in the point pattern. In the case of weak dependence the PLE method needs higher intensity λ of the point process to provide usable estimates. The worst results are provided by CLE. Both CLE and PLE have considerable negative bias (larger for the case of strong dependence) but PLE shows consistently lower variability than CLE which makes it superior.

As in the case of estimation of β both PLE and CLE perform reasonably well only with the tuning parameter $R = 0.1$, for larger values both the bias and variability increase to impractical values.

3.2.3 Parameter μ

The intensity parameter μ is the easiest one to estimate and it is estimated very well by any of the compared methods. Remember that MCE and CLE estimate μ by the same formula (see Example 3.2 in Section 4 of [3]) which is influenced by the value of the estimated parameter σ^2. Thus it follows from the properties of the estimates of σ^2 that the MCE estimates of μ has to be superior to the CLE estimates and that MCEK and MCEg perform equally well. Note moreover that the CLE estimates with $R = 0.2, 0.3$ are still very good even though the quality of the estimates of σ^2 was not good at all. Since $\hat{\mu}$ is a linear function of $\hat{\sigma}^2$ the influence of the quality of the estimates of interaction parameters on the estimate of the intensity parameter μ is much smaller than for the Thomas process case.

Another interesting observation is that the efficiency of the PLE estimate which is determined by formula (6) in [3] is worse than for the other methods and improves with larger value of the tuning parameter R. The formula (6) in [3] uses the number of observed pairs of points with distance smaller than R and provides negatively biased estimates – the smaller the value of R the larger the bias of the estimates. Although the PLE estimates of μ are still fairly exact the conclusion is that the intensity parameter μ is better estimated by the simpler formula using just the observed number of points of the point process X.

3.2.4 Procedure using r_{data}

The procedure using $R = r_{data}$ was even less successfull for the log-Gaussian Cox process than for the Thomas process. This could be explained partially by a more uniform distribution of the estimated values of r_{data} over the whole range between 0 and 0.25 (see Figure 2 for the histograms of r_{data}). Moreover, when $\beta = 10$ we get quite often the estimate $r_{data} = 0.25$ and for such R both CLE and PLE become impractical.

For PLE the use of data dependent R produces impractical estimates of β and estimates of σ^2 with larger variance and for strong interaction also with larger bias than the procedure with fixed $R = 0.1$.

For CLE the estimate of the interaction parameter β using $R = r_{data}$ has in case of strong dependence generally smaller bias than with fixed $R = 0.1$ but the variance is larger and in total the estimate is worse than with fixed $R = 0.1$. Still it is superior to MCE. In case of weak dependence the small bias is the same like with fixed $R = 0.1$ and in total larger variance makes the estimate inferior to MCE. The estimate of σ^2 by CLE with $R = r_{data}$ is worse than the other estimates and for small values of the intensity it is even impractical.

3.2.5 Overall performance

The quality of all the compared estimators improves with higher intensity of the point process X and in the majority of cases also with weaker dependence (i.e. larger value of the parameter β). However, there is one important exception – the CLE with $R = 0.1$ provides more precise estimates of β for processes with stronger dependence. And in this case CLE is much better than the other estimation methods. A plausible explanation is that the score equations of CLE are less influenced by the high variability of the observed point pattern (caused by the variability of a few large and heavy clusters) than are the estimates of the functions K and g used for MCE. For point processes with $\beta = 20$ and small enough intensity ($\lambda < 300$) the observed clusters in the point pattern are more homogeneous and the MCEK becomes the best (in the sense of the relative MSE) estimator of β followed by MCEg.

The parameter σ^2 is best estimated by the minimum contrast methods, which show very similar performance, followed by PLE with $R = 0.1$. Like in the other cases small value of the tuning parameter R is crucial for estimation of the interaction parameters by both PLE and CLE; for larger $R = 0.2, 0.3$ both the bias and variability of the PLE and CLE become too large.

Estimators of the intensity parameter μ are very precise for any of the compared methods. The important conclusion from the simulation study is that for log-Gaussian Cox processes a simple estimate by means of the observed (first order) intensity of the point process is superior to the more complicated estimate provided by PLE which uses the observed intensity of pairs of points from X.

References

[1] A.J. Baddeley and R. Turner: Spatstat: an R package for analyzing spatial point patterns. J. Stat. Softw. *12* (2005), 1–42.

[2] P.J. Diggle: Statistical Analysis of Spatial Point Patterns. 2nd edition. Oxford University Press, New York 2003.

[3] J. Dvořák and M. Prokešová: Moment estimation methods for stationary spatial Cox processes – a comparison. Accepted to Kybernetika (2012).

[4] M. Prokešová and E.B. Vedel Jensen: Asymptotic Palm likelihood theory for stationary spatial point processes. Submitted (2010).

Acknowledgement

The work was supported by Czech Science Foundation, project no. P201/10/0472, and the grant SVV 265 315.

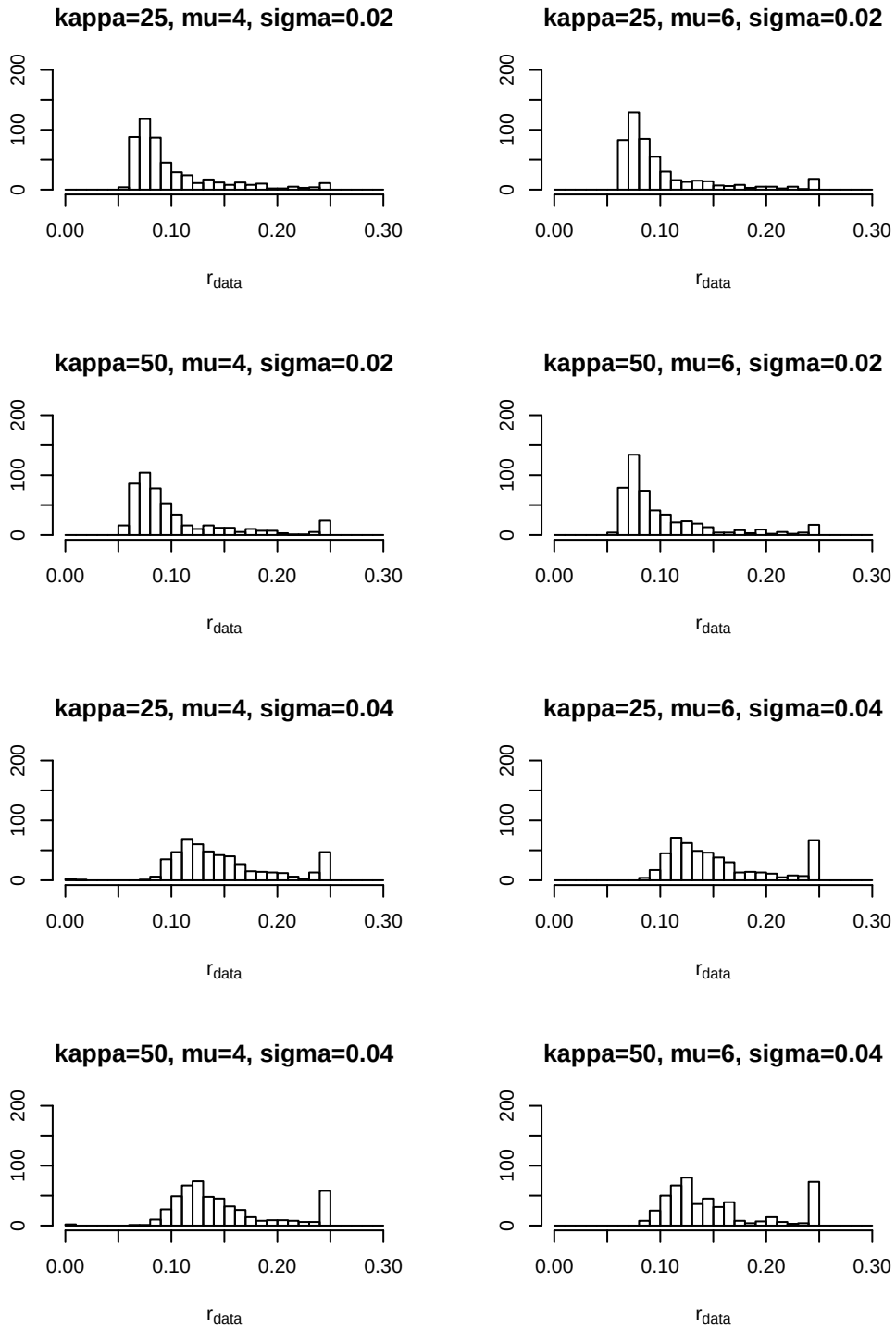

Figure 1: Histograms of estimated values of r_{data} – Thomas process.

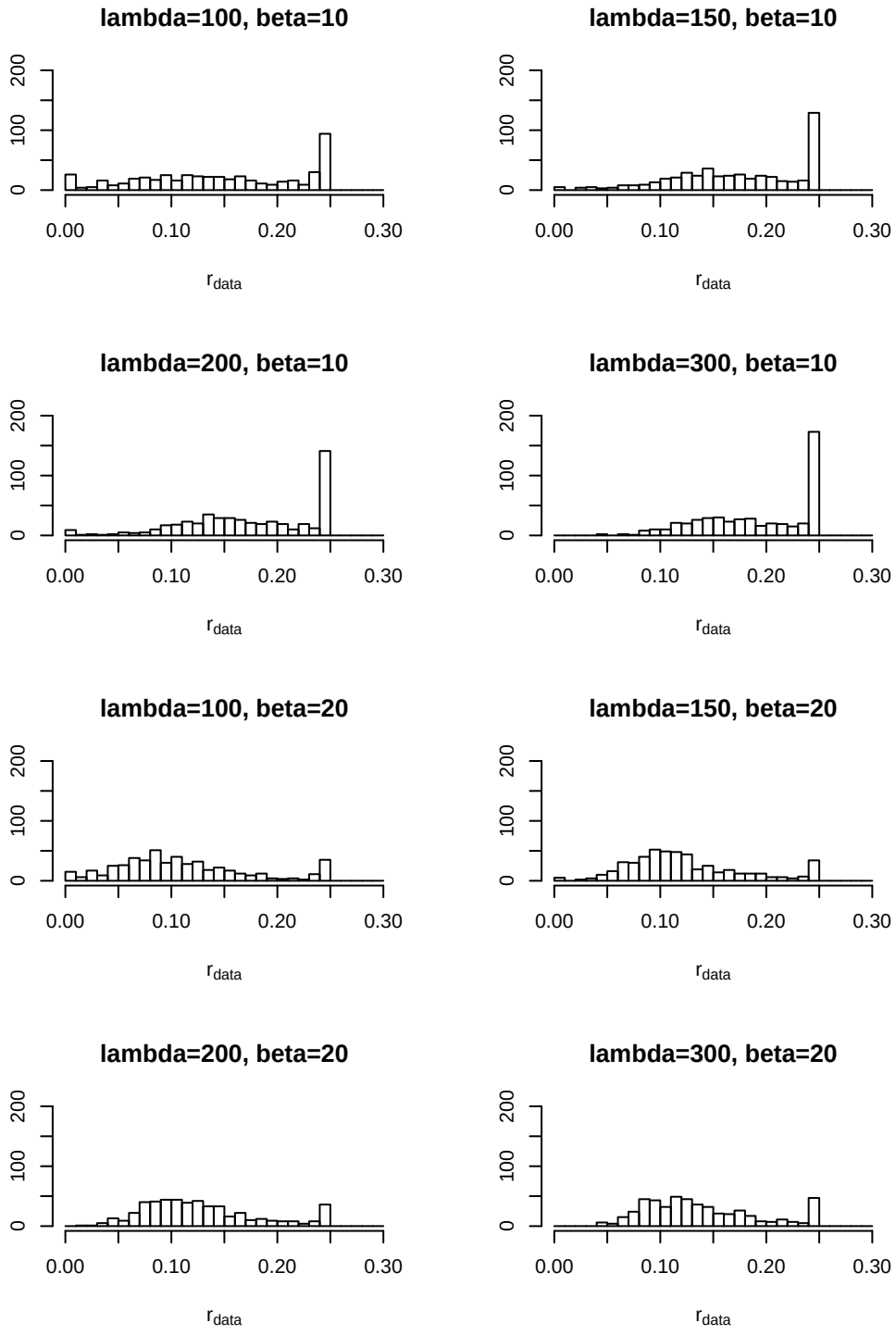

Figure 2: Histograms of estimated values of r_{data} – log-Gaussian Cox process, $\sigma^2 = 1$ in all cases.

				MCE		CLE				PLE			
	κ	μ	σ	K	pcf	.1	.2	.3	r_{data}	.1	.2	.3	r_{data}
$\hat{\kappa}$	25	4	.02	.103	.070	-.013	.271	.917	-.097	-.143	.083	.161	-.260
			.04	.148	.159	.098	.196	.987	-.087	-.331	-.178	.042	-.107
		6	.02	.093	.041	-.001	.309	.807	-.118	-.115	.145	.189	-.215
			.04	.149	.141	.039	.209	.835	-.010	-.355	-.138	.074	-.122
	50	4	.02	.076	.048	.090	.343	1.05	.028	-.129	.036	.110	-.148
			.04	.107	.129	.148	.260	.732	-.039	-.283	-.156	.0222	-.109
		6	.02	.060	.022	.071	.318	.818	.004	-.122	.115	.139	-.157
			.04	.109	.102	.064	.405	1.02	.075	-.293	-.133	.024	-.087
$\hat{\mu}$	25	4	.02	-.074	-.067	.101	-.145	-.303	.386	-.416	-.431	-.375	-.365
			.04	-.094	-.117	.098	.022	-.187	.896	-.252	-.272	-.289	-.272
		6	.02	-.071	-.047	.067	-.170	-.317	.270	-.432	-.456	-.393	-.395
			.04	-.095	-.106	.040	-.095	-.241	.138	-.256	-.292	-.276	-.254
	50	4	.02	-.061	-.056	-.073	-.217	-.339	.025	-.394	-.399	-.319	-.370
			.04	-.065	-.096	-.075	-.144	-.233	.141	-.257	-.251	-.232	-.268
		6	.02	-.054	-.038	-.057	-.199	-.338	-.002	-.405	-.434	-.344	-.378
			.04	-.068	-.076	-.035	-.230	-.260	-.011	-.257	-.274	-.254	-.266
$\hat{\sigma}$	25	4	.02	-.044	-.029	.009	-.034	-.019	.028	.068	.023	.086	.118
			.04	-.064	-.067	.077	.208	34.7	.188	.287	.494	1.27	1.22
		6	.02	-.048	-.029	.000	-.043	-.028	.027	.044	-.004	.053	.077
			.04	-.058	-.056	.065	.130	29.2	.136	.202	.343	1.44	1.20
	50	4	.02	-.037	-.035	-.006	-.043	-.105	.012	.102	.046	.112	.114
			.04	-.031	-.044	.108	34.6	199	7.90	.279	1.08	3.26	2.61
		6	.02	-.028	-.027	.001	-.045	-.086	.021	.096	.014	.052	.106
			.04	-.044	-.047	.093	25.2	160	.238	.233	.849	3.55	2.41

Table 1: Summary of simulation results (bias) – Thomas process.

				MCE		CLE				PLE			
	κ	μ	σ	K	pcf	.1	.2	.3	r_{data}	.1	.2	.3	r_{data}
$\hat{\kappa}$	25	4	.02	.125	.073	.207	.364	3.08	.244	.108	.186	.225	.233
			.04	.270	.217	.429	.576	5.67	.321	.170	.085	.192	.186
		6	.02	.111	.066	.192	.435	2.00	.193	.079	.220	.246	.214
			.04	.256	.198	.244	.396	3.11	.258	.158	.062	.226	.143
	50	4	.02	.105	.065	.148	.420	2.99	.202	.052	.066	.111	.114
			.04	.220	.172	.288	.392	3.22	.209	.090	.043	.139	.109
		6	.02	.091	.052	.097	.304	1.65	.111	.039	.100	.099	.109
			.04	.189	.148	.184	.562	4.48	.268	.093	.039	.145	.150
$\hat{\mu}$	25	4	.02	.075	.044	.266	.142	.250	1.38	.192	.225	.185	.166
			.04	.144	.116	.582	.435	.410	42.0	.076	.112	.159	.133
		6	.02	.066	.040	.198	.138	.245	.540	.202	.238	.200	.178
			.04	.133	.107	.280	.196	.313	.434	.074	.118	.166	.123
	50	4	.02	.067	.039	.082	.121	.236	.193	.167	.180	.132	.156
			.04	.144	.106	.178	.158	.195	.385	.070	.075	.106	.114
		6	.02	.055	.031	.062	.101	.200	.108	.173	.211	.140	.160
			.04	.130	.103	.133	.151	.219	.179	.071	.088	.133	.125
$\hat{\sigma}$	25	4	.02	.035	.014	.016	.032	.094	.023	.031	.045	.135	.051
			.04	.050	.034	.061	.317	$1{\cdot}10^4$	.168	.211	.764	6.85	5.47
		6	.02	.027	.011	.010	.026	.072	.015	.018	.032	.087	.026
			.04	.037	.024	.030	.139	$9{\cdot}10^3$	.100	.062	.261	8.74	5.93
	50	4	.02	.028	.012	.010	.022	.049	.015	.034	.031	.113	.034
			.04	.045	.029	.060	$2{\cdot}10^4$	$1{\cdot}10^5$	$5{\cdot}10^3$	.128	3.22	24.1	33.9
		6	.02	.022	.009	.006	.018	.036	.011	.026	.026	.055	.027
			.04	.035	.023	.036	$2{\cdot}10^4$	$1{\cdot}10^5$	.332	.077	2.45	32.6	16.8

Table 2: Summary of simulation results (MSE) – Thomas process.

			MCE		CLE				PLE			
	λ	β	K	pcf	.1	.2	.3	r_{data}	.1	.2	.3	r_{data}
$\hat{\beta}$	100	10	.281	.498	-.070	.190	.645	.087	-.221	1.76	2.44	.747
		20	.146	.216	.090	.380	.401	.095	.527	1.80	4.41	.359
	150	10	.252	.348	-.105	.169	.476	-.006	-.188	2.10	3.45	.999
		20	.118	.181	-.009	.413	.361	-.028	.485	1.75	5.63	.527
	200	10	.265	.388	-.146	.245	.618	.020	-.255	2.13	3.17	1.02
		20	.140	.195	-.023	.480	.336	.023	.609	1.95	3.31	.682
	300	10	.259	.360	-.143	-.036	.258	-.062	-.224	1.51	2.25	.847
		20	.140	.180	-.051	.437	.272	.092	.417	.798	.629	.484
$\hat{\mu}$	100	10	-.001	-.006	.026	.047	.064	-.018	-.146	-.052	-.016	-.111
		20	-.005	-.006	.017	.042	.068	-.067	-.094	-.079	-.005	-.147
	150	10	-.001	-.002	.021	.044	.067	.039	-.126	-.043	-.003	-.068
		20	-.003	-.005	.017	.035	.061	-.018	-.088	-.074	-.008	-.102
	200	10	.003	.001	.021	.043	.066	.037	-.129	-.043	-.005	-.060
		20	-.003	-.006	.018	.034	.054	.006	-.081	-.068	-.014	-.099
	300	10	.000	-.003	.019	.041	.061	.037	-.115	-.043	-.013	-.050
		20	-.006	-.008	.012	.027	.059	.014	-.086	-.062	-.011	-.082
$\hat{\sigma}^2$	100	10	-.065	-.016	-.311	-.559	-.688	.032	-.218	-.467	-.570	-.357
		20	.050	.059	-.140	-.357	-.603	.536	-.295	-.269	-.525	-.122
	150	10	-.069	-.049	-.267	-.513	-.703	-.403	-.201	-.457	-.573	-.381
		20	.038	.047	-.159	-.331	-.605	.162	-.218	-.223	-.493	-.173
	200	10	-.077	-.050	-.252	-.524	-.750	-.472	-.157	-.434	-.549	-.395
		20	.009	.035	-.185	-.390	-.609	-.092	-.216	-.260	-.443	-.158
	300	10	-.046	-.023	-.241	-.535	-.722	-.509	-.160	-.384	-.484	-.375
		20	.043	.072	-.131	-.336	-.668	-.196	-.131	-.212	-.422	-.129

Table 3: Summary of simulation results (bias) – log-Gaussian Cox process, $\sigma^2 = 1$ in all cases.

			MCE		CLE				PLE			
	λ	β	K	pcf	.1	.2	.3	r_{data}	.1	.2	.3	r_{data}
$\hat{\beta}$	100	10	.538	.825	.112	.756	2.16	.649	.513	7.59	11.0	3.35
		20	.309	.357	.428	.983	.798	.576	2.14	8.67	51.5	2.21
	150	10	.323	.380	.087	.793	1.53	.283	.435	10.2	25.4	4.79
		20	.165	.188	.245	1.08	.702	.296	1.54	9.07	117	2.12
	200	10	.267	.366	.106	1.46	2.19	.272	.340	10.8	20.8	5.36
		20	.155	.163	.188	.950	.547	.278	1.78	9.49	34.6	2.60
	300	10	.235	.299	.085	.206	.566	.089	.268	6.12	14.0	3.57
		20	.124	.132	.091	.631	.394	.220	1.13	2.10	2.00	1.24
$\hat{\mu}$	100	10	.005	.005	.007	.008	.012	.054	.038	.012	.011	.041
		20	.004	.004	.008	.010	.012	.066	.016	.016	.009	.055
	150	10	.003	.003	.004	.007	.010	.009	.028	.008	.007	.017
		20	.002	.002	.004	.006	.009	.026	.013	.011	.006	.019
	200	10	.003	.003	.004	.007	.009	.008	.028	.007	.006	.015
		20	.001	.001	.003	.005	.007	.010	.011	.009	.004	.016
	300	10	.002	.002	.003	.006	.007	.006	.021	.006	.005	.009
		20	.001	.001	.002	.003	.006	.003	.011	.006	.003	.011
$\hat{\sigma}^2$	100	10	.146	.151	.380	.548	.774	2.90	.184	.436	.623	.430
		20	.196	.198	.476	.561	.756	4.16	.294	.519	.714	.529
	150	10	.089	.086	.275	.512	.772	.495	.171	.417	.613	.361
		20	.108	.101	.314	.453	.709	1.89	.202	.372	.616	.299
	200	10	.082	.077	.241	.507	.758	.505	.132	.365	.553	.341
		20	.087	.088	.253	.431	.659	.783	.196	.332	.472	.249
	300	10	.060	.058	.184	.485	.736	.480	.132	.267	.450	.285
		20	.064	.063	.136	.354	.688	.233	.121	.226	.402	.138

Table 4: Summary of simulation results (MSE) – log-Gaussian Cox process, $\sigma^2 = 1$ in all cases.