

Geostatistical inference under preferential sampling: questions

John Paige

Department of Statistics, University of Washington Seattle, WA, 98195, USA

1. The next couple of questions will be about K functions. I am trying to recreate Figure 6 from the paper Diggle et al. [2010]. In order to do so, I'm trying to use the `Kest` and `Kinhom` functions from the `spatstat` package in R. However, neither seem to get the correct result. The `Kest` function estimates the K function for homogeneous Poisson processes, so it clearly shouldn't work, but somehow it actually looks better than the result I get when using `Kinhom`. Interestingly, the documentation for `Kinhom` says it uses a theoretical estimator for the K function of an inhomogeneous Poisson process of $K(r) = \pi r^2$, but this is the same value that is claimed to be used in `Kest` function. In addition, the documentation of `Kinhom` says that if $K(r) > \pi r^2$, it is indicative of clustering, so it seems to make little sense that all inhomogeneous Poisson processes should have theoretical K functions of πr^2 .

Further confusing me is the fact that Diggle himself seems to give multiple definitions of the K function of a log-Gaussian Cox process. In Diggle et al. [2010], he says that for log-Gaussian Cox processes the theoretical K function is given by:

$$K(s) = \pi s^2 + 2\pi \int_0^s \gamma(u)u \, du$$

where $\gamma(u)$ is the covariance function of the random intensity Λ . Yet in Diggle et al. [2013] in Equation 2, he states that if $\Lambda(x) = \exp\{\mathcal{S}(x)\}$ with \mathcal{S} being a Gaussian process with mean μ and variance σ^2 (so Λ is a log-Gaussian Cox process), then the associated K function is instead:

$$K(u) = \pi u^2 + 2\pi \lambda^{-2} \int_0^u C(v)v \, dv$$

where $C(v)$ is the covariance function of random intensity Λ and

$$\lambda = E(\Lambda(x)) = \exp\{\mu + \sigma^2/2\}.$$

MLE	Value
$\hat{\kappa}$	0.5
$\hat{\sigma}^2$	0.138
$\hat{\phi}$	0.313
$\hat{\mu}_{97}$	1.515
$\hat{\beta}$	-2.198

Table 1: The MLEs given in Diggle et al. [2010]

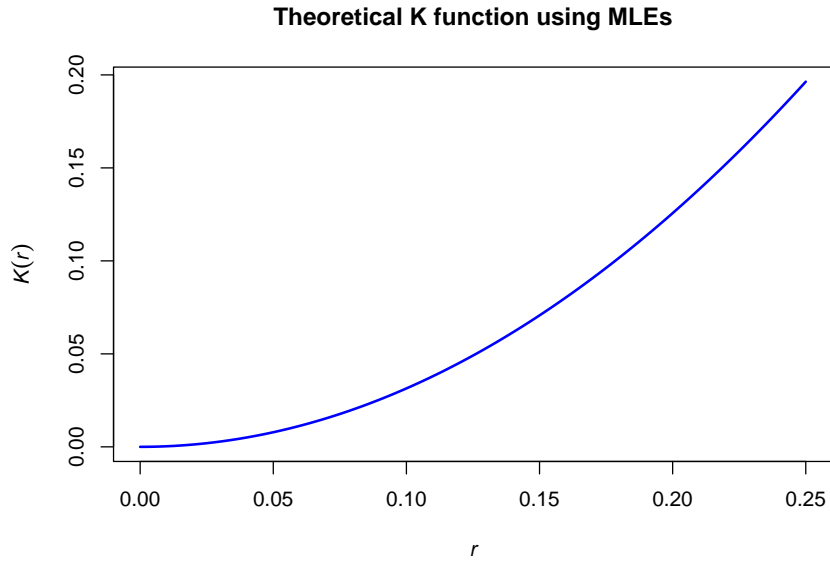


Figure 1: Theoretical K function based on MLEs and formula given in Diggle et al. [2010].

I have therefore seen three seemingly different definitions of what a K function of a log-Gaussian Cox process should be, and none of them seem to be what I find in Figure 6 of the paper Diggle et al. [2010]. Based on the estimated values given in the paper and shown in Table 1, I plotted the theoretical K function given by the definition in Diggle et al. [2010], and the results are shown in Figure 1. The K function in Figure 1 appears to be higher than the K function in Figure 6 in [Diggle et al., 2010], and I am not sure what could be going wrong.

My question is this: which definition of the K function is correct, and why might none of them seem to agree with Figure 6 in the paper?

2. A related question is this: consider a log-Gaussian Cox process with random intensity satisfying

$$\Lambda(x) = \exp \{ \alpha + \beta \mathcal{S}(x) \}$$

where \mathcal{S} is a Gaussian process. In order to generate realizations of this process with exactly n points, is it possible to raise α arbitrarily high, generate a large number of points, and then using random thinning (*i.e.* remove points with equal probability) until the number of points is just n ? I noticed that Møller claims at the end of page 3 that if all points are independently retained with equal probability, π , then the resulting intensity is $\lambda \cdot \pi$, where λ is the original intensity before thinning. If my interpretation of this were true, it suggests that the ratio of intensities at different locations would be the same, preserving β , but altering α so as to have the correct number of points.

The reason I am trying to do this is to simulate log-Gaussian Cox processes with the same number of points as in the data set from the paper to ensure it has the same statistical properties. I am trying to use this method to generate the 99 simulations generated to produce Figure 6 in the paper.

3. My last question related to K functions is this: on page 205 from Diggle et al. [2010], the authors describe the test statistic:

$$T = \int_0^{0.25} \frac{(\hat{K}(s) - K(s))^2}{\nu(s)} ds$$

They then cite that for the '97 data, the test statistic results in a p -value of 0.03. How did they get this p -value? I cannot find the name of the test they use. I thought it might have to do with a sum of χ_1^2 in a numerical integral approximation, but it seems like the better the integral approximation, the closer T should get to a fixed value.

4. Lastly, in Table 3 on page 204, the authors give a correlation matrix for the parameter MLEs. They say that the correlations are computed with a quadratic fit to their estimated log-likelihood surface, but I'm not quite sure how they would go about

doing something like that. By quadratic fit do they mean that once they get their MLEs, they just sample points around it and estimate the Hessian?

References

- Peter Diggle, Raquel Menezes, and Ting-li Su. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2): 191–232, 2010.
- Peter J Diggle, Paula Moraga, Barry Rowlingson, Benjamin M Taylor, et al. Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, 28(4):542–563, 2013.
- Jesper Møller. *Properties of spatial Cox process models*.