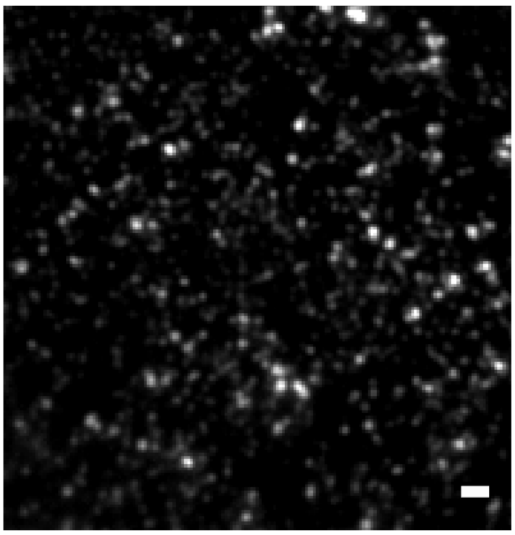
Exercise Sheet 2

1. An important question in cell biology is whether or not structures are organized in a regular way or do not have particular relations amongst them. Clathrin-coated pits play a major role in endocytosis. Whether clathrin-coated pits are positioned in an ordered fashion is of major interest in cell biology. Translated into the language of spatial statistics, we are therefore interested in whether or not clathrin-coated pits are distributed in a completely spatially random fashion.

   Below is a fluorescence microscopy image of clathrin-coated pits on the membrane of an HMEC-1 cell. Scale bar = 1μm.



   This image was then processed with the QuickPALM algorithm to extract pit locations as a spatial point pattern. The ROI is the square $[0, 18500] \times [0, 18500]$ (nm). The output of this algorithm is stored in `clat_qpalm_locs.csv`.

   Your task is to analyse these data to test whether Clathrin coated pits exist in a competely spatially random manner, or if their spatial distribution is not consistent with complete spatial randomness.

   You should first extract and plot the localizations in the file. These are in the columns labelled x and y. You should then analyse the nearest neighbour distribution, point to event distribution, quadrat counts and Ripley's K-function to test for compete spatial randomness.

   Aim to go beyond just envelope plots to run a formal Monte Carlo hypothesis test for Complete Spatial Randomness. For this, consider Ripley's K-function, and one of the two test statistics proposed in the notes. `SpatStat` does not have an inbuilt function for performing this test, so you will need to consider coding this up yourself. Can you provide an estimate for the p-value of your test statistic?

   Do all the tests give consistent results with one another? What are your scientific conclusions about the distribution of Clathrin-coated pits on the membrane of HMEC-1?

   Acknowledgement: data is courtesy of the Ward-Ober lab, Centre of Cancer Immunology, Faculty of Medicine, University of Southampton, U.K..

2. As discussed in the notes, Ripley's K-function can be used to fit the Thomas process model. The data in `clustered_data.csv` is (simulated) localizations of CD3ζ, a subunit of the CD3 T cell receptor complex, on a human T cell. The ROI is $[0, 10000] \times [0, 10000]$ (in nm). Using the `thomas.estK` function, estimate *all* parameters (σ, λ_p and ξ) of the Thomas process and hence determine the size of clusters.

3. As discussed in the notes, the cross K-function can be used to analyse dependency between between two point processes.

   In the files `channel1.csv` and `channel2.csv` and two spatial point patterns of two types of protein on a T Cell membrane take from the paper Pageon, S. V., et al. "Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination." Proceedings of the National Academy of Sciences (2016).

   Compute the cross-K-function using `Kcross`. You will notice from the function description that you will need to put this into a marked point pattern format. You'll also note that the datasets are very large! You could try cropping the point pattern to a much smaller ROI where the number of events is manageable. This can be achieved using the `subset.ppp` function.

   Compute envelopes for the cross-K-function. Why should these envelopes be treated with caution?