Shota Gugushvili

Fast and scalable non-parametric Bayesian inference for Poisson point processes

In this talk I consider the problem of non-parametric Bayesian estimation of the intensity function of a Poisson point process. The observations are assumed to be n independent realisations of a Poisson point process on the interval $[0, T]$. I propose two related approaches, in both of which the intensity function is modelled a priori as piecewise constant on N bins forming a partition of the interval $[0, T]$. In the first approach the coefficients of the intensity function are assigned independent Gamma priors. This leads to a closed form posterior distribution, for which posterior inference is straightforward to perform in practice. On the theoretical side, the approach is consistent: as $n \to \infty$, the posterior distribution asymptotically concentrates around the "true", data-generating intensity function at the rate that is optimal for estimating h-Hölder regular intensity functions ($0 < h \leq 1$), provided the number of coefficients N of the intensity function grows at a suitable rate depending on the sample size n.

In the second approach it is assumed that the prior distribution on the coefficients of the intensity function forms a Gamma Markov chain. The posterior distribution is no longer available in closed form, but inference can be performed using a straightforward version of the Gibbs sampler.

Practical performance of our methods is first demonstrated via synthetic data examples. Both methods scale well with data, but the second one depends in a less sensitive way on the choice of the number of bins N and outperforms the first one in practice. Finally, I show analyses of three real datasets: the UK coal mining disasters data, the US mass shootings data and Donald Trump's Twitter data.