Point processes (2)

1. **Maximum likelihood for Markov chain** Let us imagine a cognitive context where we are interested in sleeping patterns. Every 5 minutes, a check is performed on the participants and we note their state. We want to model this by a (discrete) Markov chain with 3 states : "sleeping", "(fully) awake", "drowsy".

   (a) Plot the diagrams and parametrizes the two following (informal) models with the least possible amount of parameters

   - You can only pass from "fully awake" to "sleeping" through the state "drowsy" and vice-versa: you can only go from sleeping to "fully awake" through "drowsy".
   - It is possible (but not necessary) to be instantly awake or instantly sleeping and to skip the "drowsy" state.

   (b) Make a function in R which simulates the first model if you give in entries the state x_0, the parameters of the model (only the ones parametrizing the transitions) and the length of the chain to simulate, n.

   (c) Write the (log) likelihood of this problem and give the maximum likelihood estimator of $p_{D \to S}$ the probabilty to fall asleep when you are drowsy. Would it be different if we were in the second model ?

   (d) Show on simulation that this estimator converges towards the true probability when n tends to infinity.

   (e) How would you decide if we are in the first or second model when you have the observations?

2. **Maximum likelihood for Poisson processes**
   We are observing an inhomogeneous Poisson process with intensity $\lambda_0(.)$ on $[0, T_{max}]$. We want to estimate $\lambda_0(.)$ by a piecewise constant function on a regular partition, which means that the candidate intensities $\lambda(.)$ are, for a certain integer d giving the number of bins, of the form

$$\lambda(t) = \sum_{i=1}^{d} a_i^d \mathbf{1}_{b_{i-1}^d \leq t < b_i^d}.$$

   where the $b_i^d = \frac{iT_{max}}{d}$.

   (a) When $d = 1$ write the likelihood of the process and provide the maximum likelihood estimator of a_1^1. Implement it and show on simulation that when T_{max} grows, the estimation is converging when the process is homogeneous.

   (b) For a more general d write the corresponding likelihood and give the maximum likelihood estimator of the different a_i^d.

(c) Simulate by thinning a Poisson process of intensity ne^{-x} and $n(1+\sin(x))$ on $[0, T_{max}]$ for various n. Superpose different MLE of the intensities for various choices of d and n.

(d) Apply AIC criterion to select for a given observed Poisson process a good d.

(e) Apply it on the data of the STAR package, for instance the data citronellal (after having aggregated all the trials into just one point process).

3. **How do we check that our simulation is correct ?**

(a) Build on Exercise 3 of tutorial 4 to provide a way to simulate a univariate Hawkes process with intensity

$$\lambda(t) = \nu + \sum_{T \in N, T < t} h(t - T),$$

with $h(u) = 1 - u^2$ if $u \in [0, 1]$, 0 elsewhere.

(b) By exchanging sums and integrals, show that $\Lambda(t) = \int_0^t \lambda(s) ds$ is equal to

$$\nu t + \sum_{T \in N, T < t} \int_T^{\min(T+1,t)} (1 - (s - T)^2) ds,$$

that is

$$\nu t + \frac{2}{3} N_{t-1} + \sum_{T \in N, t-1 < T < t} \left[t - T - \frac{(t - T)^3}{3} \right],$$

where N_{t-1} is the corresponding counting process at time $t - 1$.

(c) Program a function which for a given t computes $\Lambda(t)$.

(d) Simulate a Hawkes process thanks to a/ and transform its points with Λ.

(e) On the transformed points, perform Ogata's tests to verify that we have indeed simulated a Hawkes process.