Sections and local group meeting reports.

Quantifying uncertainty in a predictive model for popularity dynamics

Written by **Gilbert MacKenzie** on March 26th. 2022. Posted in Section and local group meeting reports April 7th, 2023.

The Northern Ireland local group of the RSS held an online meeting on Wednesday, January 26th 2022, at 2pm (GMT), using MS Teams. .

The speaker was Dr. Joseph O'Brien, Department of Mathematics & Statistics, University of Limerick, Republic of Ireland.

Dr. O'Brien explained that the Hawkes Process (HP) is a form of point process which is *self-exciting* in nature, implying that the occurrence of current events increases the likelihood of future events. This process has attracted much attention in recent years as it is capable of describing the behaviour of online information cascades, which was his area of application. The intensity, of a general HP is defined as

$$\lambda(t) = \mu(t) + \xi \sum_{\tau_i < t} \phi(t - \tau_i) \tag{1}$$

where $\lambda(t)dt = E[N(t + dt) - N(t)]$ is the expected number of events in the interval [t, t + dt]. In (1), $\mu(t)$ is known as the background intensity, describing the likelihood of an exogenous event occurring independently of other events. The non-negative function $\phi(t)$ term is known as the memory kernel or the *excitation function*, as it denotes the increase in intensity of future events, due to the occurrence of previous events at times $\tau_i < t$. Normalizing $\phi(t)$ as $\int_0^\infty \phi(t)dt = 1$ permits one to interpret $\phi(t)$ as a as a probability density function or a *memory-time* distribution. A common model choice for $\phi(t)$ is the Exponential distribution with constant background intensity, whence (1) becomes:

$$\lambda(t) = \lambda_0 + \xi \sum_{\tau_i < t} \beta \exp(-\beta(t - \tau_i)) \tag{2}$$

which has decreasing excitations. Finally, ξ is a ('*Fitness'* or *'Popularity'*) parameter controlling the boundedness of the subsequent cascade size associated with an initiating event. When $\xi = 0$ we recover the the inhomogeneous Poisson Process, when $\xi < 1$ the cascade size is finite and when $\xi > 1$ it is unbounded.

Dr. O'Brien glossed over the extensive mathematical derivation, based on a differential-equation approach (see reference for details) used to obtain the governing equations of a general branching process from which the properties of the HP were deduced. He explained that the paper referenced dealt with formal mathematical derivations, for example, on prediction, rather than on estimation - an area which required further research.

He focussed on the use of the sub-critical HP ($\xi < 1$) by considering the various possible patterns of posting behaviour arising in a discussion tree on a social media platform. These patterns arose from a combination of seed postings and their off-spring (postings). The seeds arrived with intensity $\mu(t)$ and the offspring, subsequently, with intensities $\phi_k(t)$, where k=1st, 2nd, ...etc. offspring

generation. These two processes are mutually independent, have no parameters in common, and thus the likelihood separates, conveniently.

The first step was to check that the algebriac derivations were correct and he conducted a series of simulations to re-assure the audience that this was the case. The Figure is representative of the good agreement obtained in such simulations.

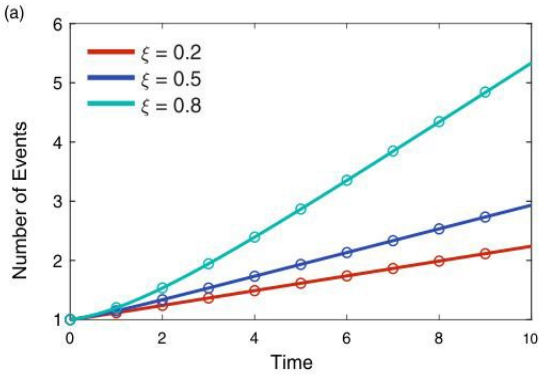

Figure 1: Mean number of events for a HP with exponential memory kernel (β= 3) and constant background intensity $\lambda_0 = 0.1$ for a range of ξ values.

Next, Dr. O'Brien demonstrated a generalization of the HP incorporating seasonality effects within the model and showed its usefulness in describing the diurnal variations of content popularity observed within empirical social media platforms. He hoped that taken together, these theoretical results and their application, would provide a firm basis for the analysis of the other self-exciting processes observed on communication platforms, including the potential to predict cascade dynamics within confidence limits.

This insightful talk was well received. The audience numbering c40 (on-line) showed their appreciation in the usual way. A short discussion ensued and Dr. O'Brien responded to questions about the balance between the background intensity and the memory kernel and how covariates might be included in these components.

The Chair thanked the speaker on a very stimulating talk and concluded the meeting by thanking everyone for their attendance and support.

References

J. D. O'Brien (2020). Quantifying uncertainty in a predictive model for popularity dynamics. Physical Review E 101 062311 [American Physical Society].