Time-interaction heterogeneous point processes

Alessio Farcomeni[1]
Department of Economics and Finance
University of Rome "Tor Vergata"

alessio.farcomeni@uniroma2.it

[1]with L. Altieri, D. Alunni, R. Barone, and M. Mezzetti

Outline

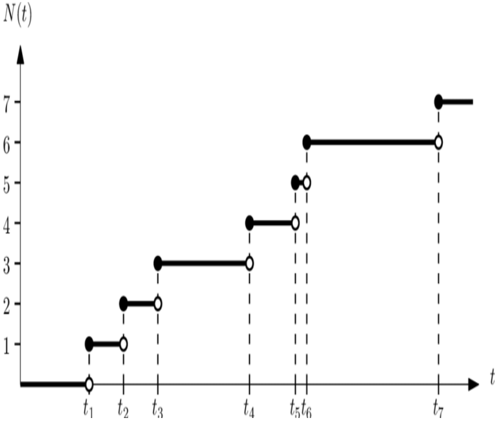
- Time-interaction point processes
- Unobserved heterogeneity: a latent Markov formulation
- Application to terrorist events in Europe between 2001 and 2017
- Inference under biased sampling, and estimation of population size
- Application to the Italian drug dealer population in 2005-2006.

Extensive simulation studies can be found in the accompanying papers Altieri *et al.* (2022) *Biometrics* and Barone *et al.* (2024) *Journal of Computational and Graphical Statistics*.

Setup

- Interest in continuous-time recurrent event processes
- Examples include occurrence of earthquakes, financial crises, system failures, soccer goals, and similar
- Real data often present time-inhomogeneity, dependence on covariates, unobserved heterogeneity, and complex dependence on past event history

Example of a realisation



Data

- $(t_{i1}, \ldots, t_{iK_i})$ occurrence times for $i = 1, \ldots, n$; in an observation period $(0, T)$
- Covariates X_{it}
- We allow $n \geq 1$, that is, we allow for time series or panel data
- If $n > 1$, we assume observations are i.i.d. conditionally on covariates

Poisson point processes

- Let $N(t)$ denote the number of events in the interval $(0, t)$.
- $$\lambda(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t)|F(t)]}{h}$$
- Under reasonable assumptions, $N(t) \sim \mathrm{Poi}\left(\int_0^t \lambda(s)\, ds\right)$

Homogeneous Poisson point process

- The assumption that risk of occurrence is time constant leads to
- $N(t) \sim \mathrm{Poi}(t\lambda)$
- Simple but inadequate for many real scenarios

Self-exciting processes

- Self-exciting processes (Hawkes, 1971)

$$\lambda(t) = \lambda + \sum_{k:t_k<t} g(t - t_k)$$

  for some smooth positive decreasing kernel $g(\cdot)$
- Often, $g(t) = \sum_{j=1}^{J} \alpha_j e^{-\beta_j t}$ for fixed J
- Occurrence of an event provides a transient (additive) increase in the future hazard

Università di Roma Tor Vergata

Self-correcting processes

- One istance of self-correcting processes (Isham and Westcott, 1979)

$$\lambda(t) = \lambda \exp\{-\theta(N(t) - \eta t)\}$$

- Hazard is continuously corrected towards the "target" mean number of events

Time interaction processes

Università di Roma Tor Vergata

- $N_i(t) \sim \mathrm{Poi}(\int_0^t \lambda_i(s)\ ds)$
- $$\lambda_i(s) = \lambda_0(s) \exp\left\{ \mathbf{X}_{it}\gamma - \theta \left(N_i(s) - \int_0^s \lambda_0(s)\ ds \right) \right\} + +\alpha \sum_{k: t_{ik} < s} \exp(-\beta(s - t_{ik}))$$

Baseline hazard

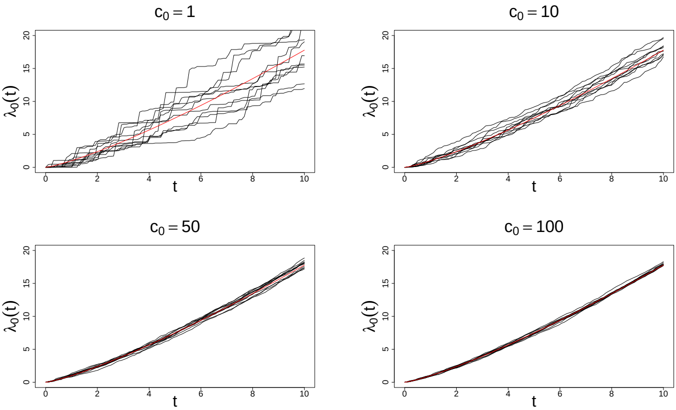
$$\lambda_i(s) = \lambda_0(s) \exp\left\{\mathbf{X}_{it}\gamma - \theta\left(N_i(s) - \int_0^s \lambda_0(s)\ ds\right)\right\} + +\alpha \sum_{k:t_{ik}<s} \exp(-\beta(s - t_{ik}))$$

- Baseline hazard is not time-homogeneous
- Parametric assumptions on $\lambda_0(s)$ (e.g., $\lambda_0(s) = \eta s^{\eta-1}$)
- Nonparametric assumptions through Gamma process prior

$$\int_0^t \lambda_0(s) \sim GP(c_0 H(t), c_0),$$

where we use $H(t) = t^\eta$

Gamma process with $\eta = 1.25$



Covariates

$$\lambda_i(s) = \lambda_0(s) \exp\left\{ \mathbf{X}_{it}\gamma - \theta \left(N_i(s) - \int_0^s \lambda_0(s)\ ds \right) \right\} + \\ +\alpha \sum_{k:t_{ik}<s} \exp(-\beta(s - t_{ik}))$$

- Under assumptions of proportionality of hazards γ provides slopes for multiplicative effects on the baseline hazard
- It can be shown that interpretation follows that of hazard ratios in Cox regression

Self-exciting and self-correcting parts

$$\lambda_i(s) = \lambda_0(s) \exp\left\{\mathbf{X}_{it}\gamma - \theta\left(N_i(s) - \int_0^s \lambda_0(s)\ ds\right)\right\} + \\ +\alpha \sum_{k:t_{ik}<s} \exp(-\beta(s - t_{ik}))$$

- For the self-exciting part we fix $J = 1$ and assume $\beta > \alpha > 0$
- For the self-correcting part, $\theta > 0$ and more importantly we use the time-inhomogeneous mean as point of attraction. Note: it makes sense to have zero-centered covariates
- The two time-interacting parts can cohexist at different time-scales, see Altieri *et al.* (2022) on this point

Unobserved heterogeneity

- When $n > 1$ the assumption that observations are identically distributed might be restrictive
- We take care of unobserved heterogeneity assuming the existence of a latent continuous-time Markov chain U_{it} with state-space $\{1, \ldots, k\}$ for some known k
- This leads to a continuous time latent Markov model (e.g., Bokenholt, 2005; Bartolucci and Farcomeni, 2019)

Unobserved heterogeneity

- The manifest model becomes $\lim_{h \to 0} \frac{E[N(t+h)-N(t)|F(t), U_{i,t+h=c}]}{h} =$

$$
\begin{aligned}
\lambda_{ic}(s) &= \lambda_0(s) \exp\left\{\mu_c + \mathbf{X}_{it}\gamma - \theta_c\left(N_i(s) - \Lambda_0(s)\right)\right\} + \\
&+ \alpha_c \sum_{k: t_{ik} < s} \exp(-\beta_c(s - t_{ik}))
\end{aligned}
$$

- We now have a latent state specific multiplicative effect for the baseline hazard μ_c, and state-specific behavioural effects $\alpha_c, \beta_c, \theta_c$, with $\alpha_1 = 0$.

The latent model

- $\Pr(U_{i0} = c) = \pi_c$
- Transition function is assumed to satisfy Chapman-Kolmogorov equations
- Transition intensity Q is a k by k matrix whose rows sum to zero and off-diagonal elements are non-negative
- Time-dependent transition matrix

$$\Pi_s = e^{sQ} = \sum_{j=0}^{\infty} \frac{s^j Q^j}{j!}$$

Inference

When $k = 1$ (that is, no latent process)

- Log-likelihood

$$l(\lambda) = \sum_{i=1}^{n} \sum_{l=1}^{K_i} \log(\lambda_i(t_{ik})) - \int_0^T \lambda_i(t)\ dt$$

- Both with parametric and non-parametric assumptions one can set up adaptive MCMC sampling schemes
- Details can be found in the accompanying paper

When $k > 1$

For computational reasons we proceed as follows:

- We set up an augmented likelihood approach
- The latent trajectory is first discretized at a small number of (equally spaced) time points
- Discrete-time forward-backward recursions (Zucchini *et al.*, 2006; Bartolucci *et al.*, 2013) are used to sample latent states at those time points
- The latent trajectory is then made continuous by sampling the rest of the process (uniformization)

For a similar strategy in a different context see Hobolth and Stone (2009)

When $k > 1$

- Update U_{it} through Forward-Backward Uniformization
- Update Q
- Update $(\mu_c, \alpha_c, \beta_c, \theta_c)$
- Update $\lambda_0(s)$ and γ

To recap

We have proposed four formulations:

1. A Time-Interaction Process (TIP) with parametric baseline and no unobserved heterogeneity
2. A Time-Interaction Process (TIP) with non-parametric baseline and no unobserved heterogeneity
3. A Time-Interaction Process (TIP) with parametric baseline and unobserved heterogeneity
4. A Time-Interaction Process (TIP) with non-parametric baseline and unobserved heterogeneity

For model selection we use WAIC, which requires little overhead after MCMC sampling.

Terrorist events in Europe

Università di Roma Tor Vergata

- We use data from the Global Terrorism Database, which for each event reports date and location, weapons used, nature of the target, number of casualties, and some more information
- We restrict to thirty European countries (excluding Russia and Ukraine), from 2001 to 2017, for a total of $n = 114$ attacks with at least one casualty
- Covariates (lagged one year): GDP growth, GDP, unemployment rate, bank or currency crisis, and scores of: judicial independence, impartiality of courts, military interference, legal system integrity, police reliability

Model fitting

- We fit models with parametric and non parametric baselines
- We let $k = 1, 2, 3$
- I am going to fit and show WAIC for the 54 univariate models resulting. Multivariate models in the accompanying paper.

WAIC

Cov.	(P)	(NP)	(P) $k=2$	(NP) $k=2$	(P) $k=3$	(NP) $k=3$
ΔGDP	-467.022	-302.608	**-554.138**	-520.591	-98.145	-300.074
GDP	-466.832	-372.077	**-548.830**	-436.262	306.377	190.127
Unemp.	-468.324	-343.193	**-562.701**	-475.520	1978.626	-320.101
Crisis	-463.070	-247.779	**-560.508**	-494.027	491.935	-44.920
Judges	-462.125	-407.130	**-557.093**	-520.011	-354.995	229.899
Courts	-462.508	-442.416	-549.765	**-638.864**	-317.617	110.853
Military	-459.774	-391.263	**-523.314**	-499.849	-218.323	241.559
Legal	-459.407	-408.970	**-557.700**	-544.447	746.341	-54.53246
Police	-464.500	-414.093	**-543.319**	-503.715	-292.372	-237.998

Posterior summaries (homogeneous pars)

ΔGDP	γ	η	q_1	q_2
$E(\cdot\|Y)$	-0.131	1.009	0.043	0.250
$SD(\cdot\|Y)$	0.053	0.132	0.007	0.054
$q_{0.025}(\cdot\|Y)$	-0.236	0.723	0.054	0.007
$q_{0.975}(\cdot\|Y)$	-0.015	1.236	0.109	0.650

Posterior summaries (state-specific pars)

ΔGDP	μ_c	θ_c	α_c	β_c
$E(\cdot \mid Y, c=1)$	-2.180	0.009	0.113	0.133
$SD(\cdot \mid Y, c=1)$	0.478	< 0.0001	0.051	0.066
$q_{0.025}(\cdot \mid Y, c=1)$	-3.191	0.009	0.044	0.050
$q_{0.975}(\cdot \mid Y, c=1)$	-1.271	0.010	0.221	0.312
$E(\cdot \mid Y, c=2)$	-0.796	0.041	2.502	3.944
$SD(\cdot \mid Y, c=2)$	0.782	0.001	1.205	2.054
$q_{0.025}(\cdot \mid Y, c=2)$	-2.360	0.040	0.860	1.334
$q_{0.975}(\cdot \mid Y, c=2)$	0.717	0.042	5.565	9.376

Biased sampling

- A common scenario in real data is the case in which $n > 1$, but a unit is recoded only whenever $N(T) > 1$
- E.g., (timing of) hospital accesses for MS patients sampled in hospitals.
- Two consequences: the sample size n is random, and all parameter estimates would be biased.
- We first propose a correction, based on conditional likelihoods, to obtain unbiased parameter estimates.
- We then assume the population is finite with size N, and show how to estimate N. This leads a method for continuous-time capture-recapture designs.

Conditional likelihood

- A common approach (Sanathanan, 1972) is to optimise the conditional likelihood
- When $k = 1$,

$$l_C(\lambda) = \sum_{i=1}^{n} \sum_{l=1}^{K_i} \log(\lambda_i(t_{ik})) - \int_0^T \lambda_i(t)\, dt - \log(\Pr(N_i(T) > 0$$

- Expression for $\Pr(N_i(T) > 0)$ is a little cumbersome, but available in closed form.
- For inference in the Bayesian framework we can simply substitute (augmented) likelihoods with (augmented) conditional likelihoods
- In Altieri *et al.* (2022) we also set up an Expectation-Maximisation algorithm to optimise the conditional likelihood

Estimation of population size

Università di Roma Tor Vergata

- MLE based on the conditional likelihood is consistent and unbiased in the presence of a biased sample
- Suppose now the sample comes from a finite population of size N
- Horvitz-Thompson estimator

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{\Pr(\widehat{N_i(T)} > 0)}$$

Drug dealing in Italy

Università di Roma Tor Vergata

- We have a database of drug dealers (repeatedly) identified by the Italian police in the years 2005 and 2006. Time zero is the onset of the (unconstitutional) 309/90 law.
- For population size estimation, a strong assumption: population closed in the period (common to other works)
- $n = 4271$, 4000 captured only once. About 10% female, median age 32 with IQR 13.
- We restrict to parametric models with Weibull baseline.

See Farcomeni and Scacciatelli (2013) *Annals of Applied Statistics* for the estimation of the number and features of cannabis users in Italy in the same period.

Model choice and population size

	$\widehat{N}$	95% CI	AIC
M_{h2tb}^{constr}	91387	(75149; 107625)	602.00
M_{h2t}	91387	(74929; 107845)	606.00
M_{h2ot} (sex, age)	93063	(72882; 113245)	608.24
M_{h2tb}	91332	(73137; 109528)	614.00
M_{h2otb} (sex, age)	93100	(75537; 110662)	616.25
M_{h3otb} (sex, age)	93123	(75060; 111186)	624.25
M_{1bt}	91397	(75027; 107768)	719.45
M_{1otb} (sex, age)	93124	(73224; 113024)	721.69
M_{h2b}	91373	(75136; 107609)	727.45
Chao	92093	(73981; 110205)	717.35
M_t	91168	(74641; 107694)	732.92
M_{th}	110963	(68315; 153611)	760.55
GC (sex,age)	93986	(91190; 96781)	737.21

Parameter estimates

	estimate	95% CI	*p*-value
θ_2	1.09	(0.99; 1.77)	<0.001
μ	-1.89	(-2.06; -1.71)	<0.001
η	1.17	(1.13; 1.21)	<0.001
	$\pi_1 = 0.394$	$\pi_2 = 0.606$	

Concluding remarks

- We have proposed a model for recurrent event-time processes which can accommodate (i) time-dependence, (ii) self-exciting and self-correcting effects, (iii) observed heterogeneity, (iv) unobserved heterogeneity
- The model generalises various proposals, including Wu et al. (2022).
- The inferential strategy overcomes issues of Wu et al. (2022), which had to restrict their practical examples to two latent states.