Flexible multivariate processes for modelling football matches

Santhosh Narayanan

Supervisors: Dr. Ioannis Kosmidis and Prof. Petros Dellaportas
Department of Statistics
University of Warwick

MathSport International 2019

Outline

Data description

Data from the English Premier League for two consecutive seasons recording all touch-ball events, i.e., events where a player has acted on the ball by touching it with some part of their body. In total, we have approximately 1.1 million events recorded over the 760 games.

second	minute	team_id	player_id	type	outcome	x	y	end_x	end_y
0	0	665	68312	Pass	Successful	49.1	51.0	52.5	44.8
2	0	665	14036	Pass	Successful	52.2	44.5	36.7	60.6
3	0	665	79050	Pass	Successful	36.7	60.6	24.9	39.1
5	0	665	14107	Pass	Unsuccessful	25.0	37.9	97.0	22.9
9	0	660	73379	Tackle	Successful	1.9	73.7	1.9	73.7
15	0	660	73379	Pass	Successful	5.5	65.3	20.9	21.5
17	0	660	6292	Pass	Successful	20.9	21.5	29.0	38.5
19	0	660	26820	Foul	Successful	25.8	37.4	25.8	37.4

Table: A snapshot of the dataset.

Data source: Stratagem Technologies, London.

Football as a spatio-temporal point process

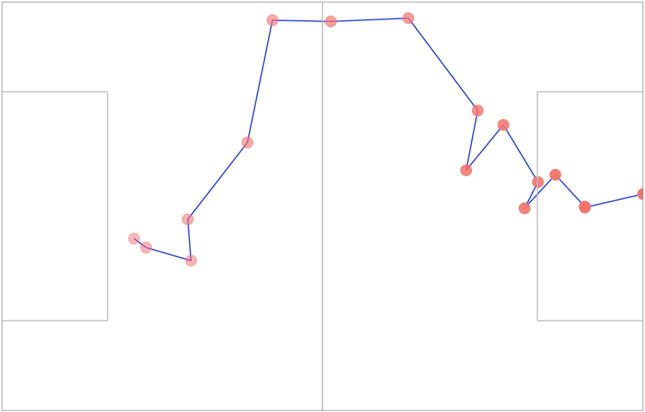


Figure: Goal of the season 2013-14, Jack Wilshere vs Norwich City.

Heat maps: Home advantage

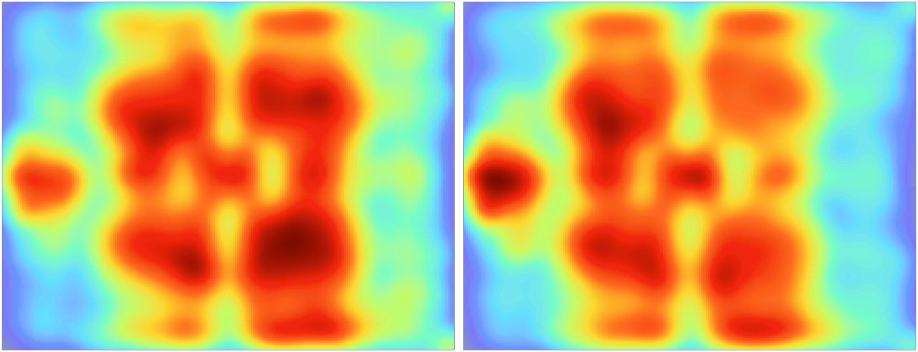


Figure: Ball touches for Arsenal, Home (left) and Away (right).

Heat maps: Shots vs Goals

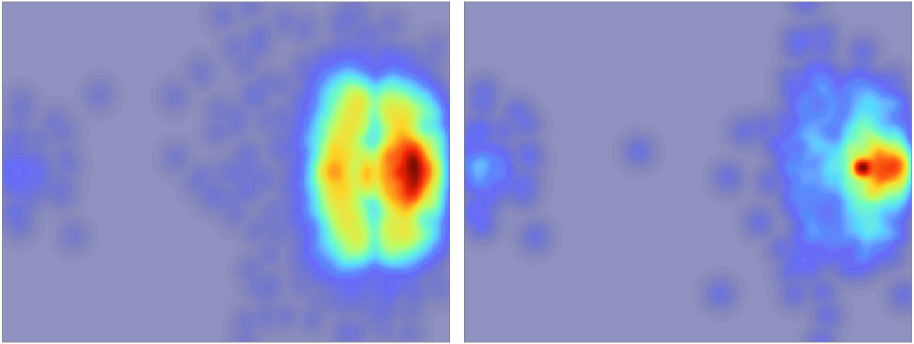


Figure: All shots on goal (left) and goals (right).

Research goal

Simulation

The primary goal is to develop machinery to simulate the process in $(T, T + \Delta)$ where T is the current time and Δ is the time resolution of prediction, properly accounting for the dependence between events, dependence on past processes, and process-specific characteristics.

- Predict match outcome probabilities in real-time.
- Predict team-specific probabilities of any event, e.g. goal, in the next t minutes

Poisson processes

Homogeneous Poisson process

Let $\lambda \in \mathbb{R}_+$. A Poisson process with constant rate λ is a point process defined by

$$\mathbf{P}[N(t+\epsilon) - N(t) = 1 | \mathcal{F}_t] = \lambda\epsilon + o(\epsilon),$$
$$\mathbf{P}[N(t+\epsilon) - N(t) > 1 | \mathcal{F}_t] = o(\epsilon).$$

- The intensity of the process $N(t)$ is time-invariant, and the probability of occurrence of an event in $(t, t+\epsilon]$ is independent of the history of the process at time t given by $\mathcal{F}_t$.

Self-exciting processes

Linear self-exciting process

$$\lambda^*(t) = \mu(t) + \alpha \sum_{t_i < t} g(t - t_i),$$

where $\mu : \mathbb{R} \mapsto \mathbb{R}_+$ is a deterministic base intensity, α is the excitation from a past event and $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ expresses how quickly this excitation dies over time.

Hawkes process

Hawkes (1971) proposed an exponential kernel $g(t) = \beta \mathrm{e}^{-\beta t}$, so that the intensity of the model becomes

$$\lambda^*(t) = \mu(t) + \alpha \sum_{t_i < t} \beta \mathrm{e}^{-\beta(t - t_i)}.$$

Conditional intensity

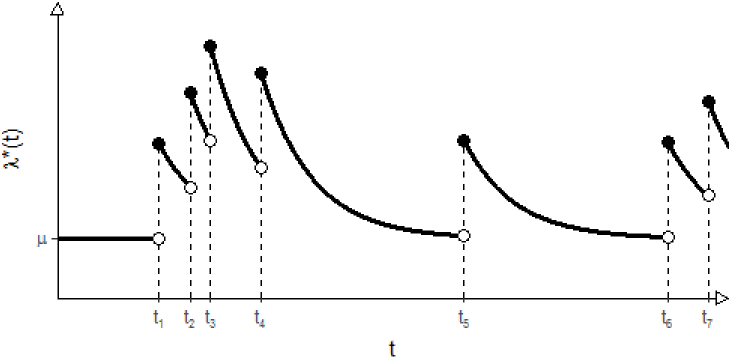


Figure: Conditional intensity function of a Hawkes process.

Marked Hawkes processes

A marked Hawkes process $\boldsymbol{X}$, consists of event times $\boldsymbol{t} = \{t_i : t_i \in \mathbb{R} \text{ and } t_i > t_{i-1}\}$ and marks $\boldsymbol{m} = \{m_i : m_i \in 1, \ldots, M\} \ \forall\ i = 1, \ldots, n$.

$$\lambda(t, m \mid \mathcal{F}_t) = \mu\delta_m + \sum_{t_j < t} \alpha\beta \mathrm{e}^{-\beta(t - t_j)} \gamma_{m_j \to m},$$

where $\delta_m \in [0, 1]$ is the base mark probability for mark m and $\gamma_{m_j \to m} \in [0, 1]$ is the probability a event with mark m_j produces an offspring of mark m.

Constraints

$$\sum_{m=1}^{M} \delta_m = 1, \qquad \sum_{m=1}^{M} \gamma_{m_j, m} = 1 \ \forall\ m_j = 1, \ldots, M.$$

Clustering of times is a problem!

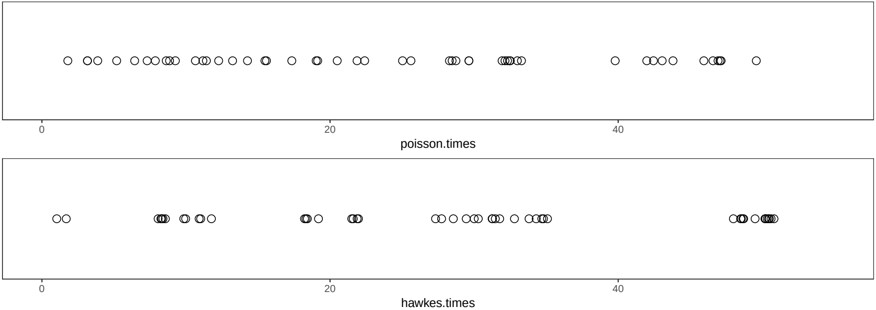


Figure: Simulations from a Poisson process (top) of unit rate and a Hawkes process (bottom) with ($\mu = 0.2$, $\alpha = 0.8$, $\beta = 1$).

Splitting times and marks

To restrict the self-exciting property of the process to the mark dimension, we use the decomposition that motivated *partial likelihood* Cox (1975). We can factorize the full likelihood of a marked point process as follows,

$$\prod_{i=1}^{n} g(t_i \mid \mathcal{F}_{t_{i-1}}; \zeta) \prod_{i=1}^{n} f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta}).$$

where g and f are the probability density functions for times and marks respectively and $\zeta, \boldsymbol{\eta}$ are the unknown parameter vectors. The second product above is called the *partial likelihood* based on the mark sequence $\boldsymbol{m} = \{m_i\}_{i=1}^{n}$.

Partial likelihood

The log-likelihood for a marked point process, Daley et al.(2003),

$$\sum_{i=1}^{n} \log\left(\lambda(t_i, m_i \mid \mathcal{F}_{t_{i-1}})\right) - \int_0^T \int_{\mathcal{X}} \lambda(u, v \mid \mathcal{F}_u)\,\mathrm{d}u\,\mathrm{d}v.$$

We can calculate the contribution of the i-th event to the partial likelihood from the log-likelihood, Diggle (2013),

$$f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta}) = \frac{\lambda(t_i, m_i \mid \mathcal{F}_{t_{i-1}})}{\sum_{\mathcal{X}_i} \lambda(t_i, m_i \mid \mathcal{F}_{t_{i-1}})}.$$

where $\mathcal{X}_i$ is the sample space for marks of the i-th event or also referred to as the *risk set*.

Model specification

Marks

$$f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta}) = \frac{\sum_{t_j < t_i} \mathrm{e}^{-\beta_{m_j}(t_i - t_j)} \gamma_{m_j \to m_i}}{\sum_{t_j < t_i} \mathrm{e}^{-\beta_{m_j}(t_i - t_j)}} \,,$$

where m_i is the mark of the i-th event in the match, β the decay rate and γ the mark probability or conversion rate.

Times

$$g(t_i \mid \mathcal{F}_{t_{i-1}}; \zeta) \sim \textbf{Gamma}[a(m_{i-1}), b(m_{i-1})] \,,$$

The shape and rate parameters of the gamma distribution depend on the mark of the last observed event.

Model specification

Team dependent conversion rates

$$\log\left(\frac{\gamma_{m_j \to m}}{\gamma_{m_j \to M}}\right) = \theta_{m_j \to m} + \mu_{t,m} \qquad \forall\ m \in 1, \dots, M-1,$$

where θ is the baseline conversion parameter and t is the team in possession of the ball attempting the event conversion. The parameter μ is the relative ability of a team to complete a conversion to an event of mark m.

Priors

$$\begin{aligned} \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b} &\sim \textbf{Exp}(0.01) \\ \boldsymbol{\theta}, \boldsymbol{\mu} &\sim \textbf{N}(0, \sigma_\gamma) \\ \sigma_\gamma &\sim \textbf{half-Cauchy}(0, 1) \end{aligned}$$

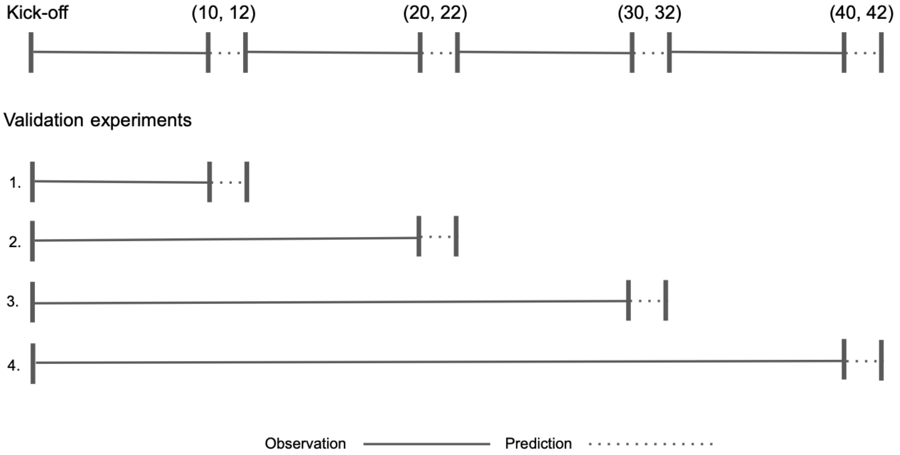
Model training

- First 20 games of the 2013/14 season used as training data, with all teams playing 1 game each at home and away.
- Each game half is modelled as a single process, i.e., history is reset at the beginning of the second period.
- Tracking a total of 15 event types for both the home and away teams separately, i.e, total marks $M = 30$.
- Bayesian posterior sampling via HMC using Stan.
- Samples obtained from 3 chains run in parallel with 1000 iterations each after burn-in.

Prediction Framework

- Collect R samples of the posterior parameter vector, $\boldsymbol{p}_k = \{\zeta_k, \boldsymbol{\Theta}_k\}$ for $k = 1, \ldots, R$.
- For each game period, for each $\boldsymbol{p}_k$, generate S simulations of the game in the interval $(T, T + d)$.
- Iteratively simulate the occurrence time of next event given history and then its mark given time and history.
- Add the generated pair of (time, mark) to the history as the most recent event.
- Stop simulation when the time exceeds $T + d$.
- Finally, for each game period in the test set, calculate event counts from the $R \times S$ simulations and validate against the observed counts.

Prediction Framework



Gamma distribution parameters $a(m_{i-1}), b(m_{i-1})$

Mark	shape	rate	mean
Win	2.85	1.56	1.83
Dribble	2.82	1.58	1.78
Pass_S	3.30	1.18	2.79
Pass_U	2.60	1.03	2.51
Shot	9.68	8.74	1.11
Keeper	1.34	0.16	8.55
Save	3.06	0.91	3.37
Clear	2.60	1.18	2.21
Lose	3.55	2.44	1.45
Goal	74.19	1.44	51.58
Foul	3.10	0.11	27.67
Out_Throw	2.82	0.19	14.85
Out_GK	9.95	0.34	29.35
Out_Corner	8.97	0.38	23.92
Pass_O	6.77	0.24	28.36

Table: Posterior parameter means of $g(t_i \mid m_{t_{i-1}}; \zeta)$.

Decay rates β_{m_i}

Mark	beta
Win	2.96
Dribble	4.03
Pass_S	3.06
Pass_U	2.93
Shot	3.88
Keeper	2.66
Save	2.59
Clear	3.10
Lose	3.52

Mark	beta
Goal	1.37
Foul	2.04
Out_Throw	2.13
Out_GK	1.68
Out_Corner	1.90
Pass_O	1.57

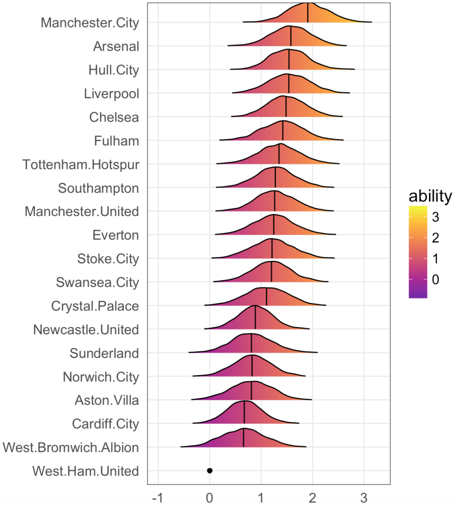
Table: Posterior parameter means for in-play events (left) and out-of-play events (right).

Conversion rates $\gamma_{m_j \to m_i}$

	Home_Pass_S	Home_Pass_U	Home_Shot
Home_Win	0.35	0.16	0.03
Home_Dribble	0.17	0.10	0.07
Home_Pass_S	0.58	0.25	0.02
Home_Foul	0.41	0.44	0.07
Away_Save	0.08	0.03	0.09
Away_Clear	0.09	0.11	0.05
	Away_Pass_S	Away_Pass_U	Away_Shot
Away_Win	0.25	0.16	0.02
Away_Dribble	0.12	0.14	0.01
Away_Pass_S	0.75	0.10	0.02
Away_Foul	0.56	0.34	0.06
Home_Save	0.10	0.01	0.03
Home_Clear	0.09	0.11	0.03

Table: Posterior means of conversion rates for selected events.

Team parameters



1	Manchester City
2	Liverpool FC
3	Chelsea FC
4	Arsenal FC
5	Everton FC
6	Tottenham Hotspur
7	Manchester United
8	Southampton FC
9	Stoke City
10	Newcastle United
11	Crystal Palace
12	Swansea City
13	West Ham United
14	Sunderland AFC
15	Aston Villa
16	Hull City
17	West Bromwich Albion
18	Norwich City
19	Fulham FC
20	Cardiff City

Figure: (Left) $\mu_{t,3} + \mu_{t,18}$ cumulative (home + away) team ability relative to West Ham (baseline) to retain possession by converting to a successful pass. (Right) Final league table for the 2013/14 season.

Validation

For 20 game periods in the test set, events are simulated in each prediction interval 500 times each for 1000 samples from the posterior. Model evaluated by using scoring rules, validating event counts in the interval, aggregated over all event types.

- Baseline: Homogeneous Poisson process for each event type.

Scoring Rule \ Game minutes	(10, 12)	(20, 22)	(30, 32)	(40, 42)
Logarithmic	12	17	15	18
Brier	8	12	12	13
Spherical	10	14	12	16
Ranked Probability	10	13	13	15
Squared Error	9	14	13	14
Dawid Sebastiani	6	14	13	18

Table: Counts of experiments (out of 20) where the model outperforms the baseline in each prediction interval.

Future work

- Location dependent conversion rates.
- Game states as covariates, e.g., score.
- Latent auto-regressive structure for parameters over games.
- Matrix parameterisation for decay rates.
- Alternative applications, e.g. cyber security.

Selected references

Daley, D. J. and Vere-Jones, D.
An introduction to the theory of point processes. Vol. I.
Springer-Verlag, New York 2003.

Diggle, Peter J.
Statistical analysis of spatial and spatio-temporal point patterns.
Chapman and Hall/CRC 2013.

Cox, David R.
Partial likelihood.
Biometrika, 62(2):269–276, 1975.

Hawkes, Alan G and Oakes, David
A cluster process representation of a self-exciting process.
Journal of Applied Probability, 11(3):493–503, 1974.