STAT 468

Spring 2025

Bayesian and Spatiotemporal Methods in Sports: An Overview

Prof. Samuel Wong

Key references

- Gudmundsson, J. and Horton, M. (2017). *Spatio-Temporal Analysis of Team Sports*. ACM Computing Surveys, 50(2), article 22.
- Santos-Fernández, E., Wu, P., and Mengersen, K.L. (2019). *Bayesian statistics meets sports: A comprehensive review*. JQAS, 15(4), 289–312.

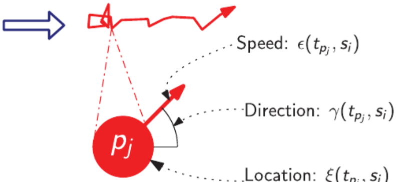
Sports and spatio-temporal data

- Rising availability of spatiotemporal data (tracking, cameras, etc.). These provide trajectories (of players, objects) and event logs (shot, pass, etc.).
- Leverage these data to provide quantitative evaluation of team and player performance; data-driven decision making in coaching, strategy, and training.

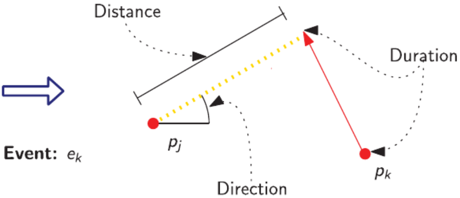
Football example

p	s_i	x_1	x_2
...			
p_j	10.1	3940	-2362
p_j	10.2	3948	-2346
p_j	10.3	3956	-2331
p_j	10.4	3923	-2392
p_j	10.5	3977	-2309
...			

Trajectory: t_{p_j}



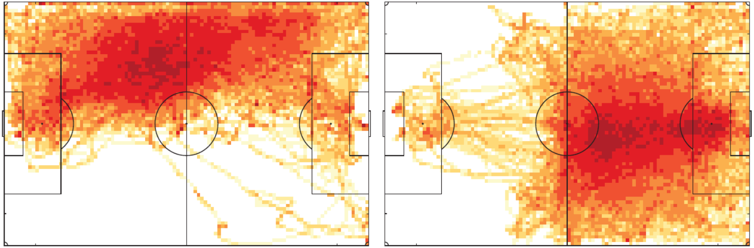
s_i	v	P
...		
10.1	*Touch*	$\{p_j\}$
10.2	*Pass*	$\{p_j\}$
10.6	*Touch*	$\{p_k\}$
11.0	*Tackle*	$\{p_l, p_k\}$
...		



Playing area subdivision

- Often discretize into regions or pixels
- Aggregate trajectories/events for analysis and visualization
- Point process representations, e.g., of where shots are taken
- Kernel-smoothed estimates
- Low-rank matrix factorizations $X \approx WB$

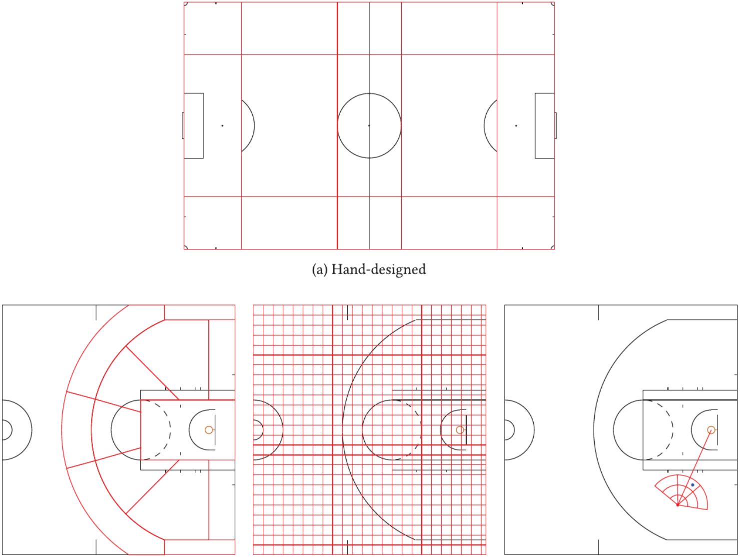
Football example



(a) Left-back

(b) Striker

Basketball example



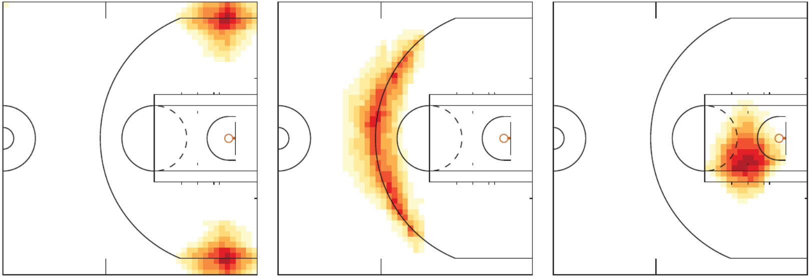
(a) Hand-designed

(b) Hand-designed

(c) Cartesian grid

(d) Polar grid

Basketball example: low-rank representation



(a) Corner three-point (b) Top-of-key three-point (c) Right low-post

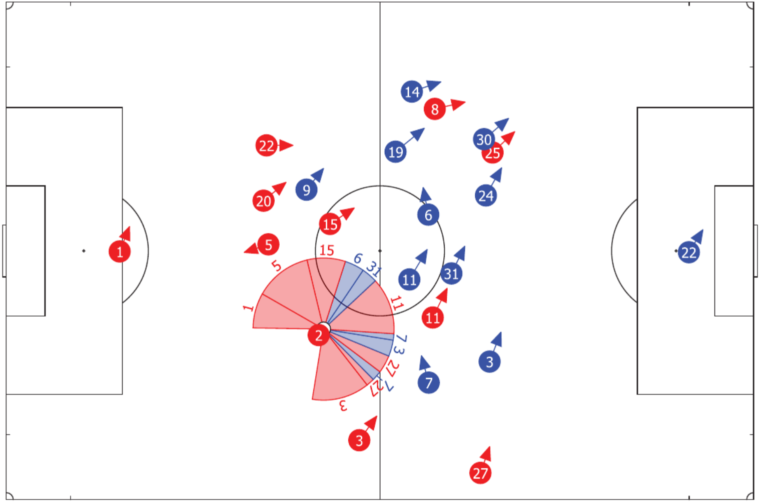
Player movement

- Basic equation of motion (ODE):

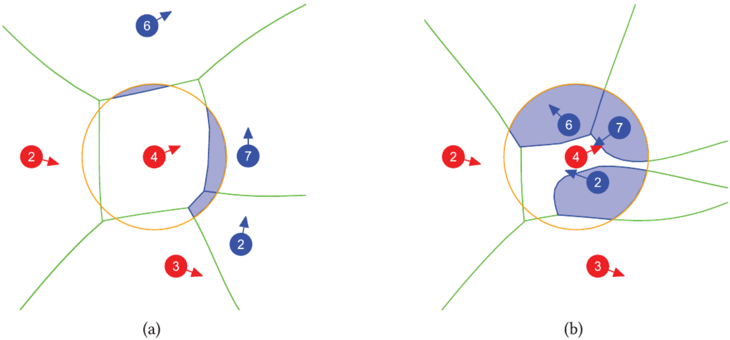
$$m\frac{d}{dt}v = F - kv$$

- Predict who can reach a location first
- The *dominant region* of a player p is the region of the playing area where p can arrive before any other player.
- Useful to identify pass availability and spatial pressure

Pass availability



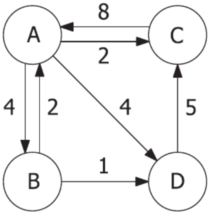
Spatial pressure



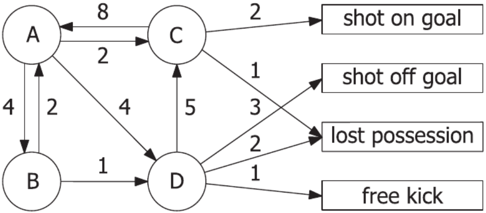
Player interaction networks

- Describe passing and transitions
- Graph $G = (V, E)$, each player is a vertex
- Edges represent, e.g., successful passes
- Measures of centrality
- Markov chains

Player interaction networks



(a)



(b)

What to do with all the data?

- Add labels (e.g., classify a pass as good, OK, or bad)
- Predict when/where the ball will go next (e.g., football, basketball)
- Identify patterns of team formations and tactical strategies
- Performance metrics and models
  - Attacking: shot quality (account for location, proximity to defenders, shooter, etc.)
  - Defending: e.g., effectiveness of basketball man-to-man defense based on hidden Markov models for defender assignment; rebound ability

Examples of common models

- Linear regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

- Logistic regression

$$y_i \sim Bern(p_i)$$

$$logit(p_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

- Mixed effects / multilevel models

$$y_{ij} = \beta_0 + \beta_1 x_{ij,1} + \cdots + \beta_k x_{ij,k} + m_i + \epsilon_{ij}$$

Space-Time Modelling

A general framework for modelling a space-time process:

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + e(\mathbf{s}, t)$$

- $\mathbf{s}$ refers to a spatial location
- t denotes time
- $Y(\mathbf{s}, t)$ a random process that depends on both spatial location and time
- $\mu(\mathbf{s}, t)$ the mean function of that process
  - E.g., via a linear regression
  - Some covariates might depend on location or time only
- $e(\mathbf{s}, t)$ a residual or error term
  - Not iid
  - Generally would be correlated over both space and time

Intro to Bayesian probability and inference

Figure: Let's start with the basics.

Why a Bayesian approach for sports analysis?

- Ability to encode prior information (including subjective expertise)
- Update models incrementally over time
- Handles missing data
- Fuse multiple data sources
- Regularizes small datasets (or rare players)
- Propagates uncertainty into downstream predictions and credible intervals

Bayesian Inference

- Recall Bayes' rule:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)}$$

- In the Bayesian approach to inference, we treat
  - $x_1, x_2, \ldots, x_n$: data/observations
  - θ: parameter(s)

  all as random variables, and thus they have a joint probability distribution; denote its joint PDF/PMF as $p(x_1, \ldots, x_n, \theta)$.
- In contrast, classical (frequentist) statistics treats θ as a fixed but unknown value.

Bayesian Inference

Using Bayes' rule, we can write the joint distribution of the data and parameters $p(x_1, \dots, x_n, \theta)$ in two equivalent ways:

$$
\begin{aligned}
p(x_1, \dots, x_n, \theta) &= \underbrace{f(x_1, \dots, x_n|\theta)}_{\text{likelihood function}} \times \underbrace{\pi(\theta)}_{\text{prior dist.}} \\
&= \underbrace{\pi(\theta|x_1, \dots, x_n)}_{\text{posterior dist.}} \times \underbrace{m(x_1, \dots, x_n)}_{\text{marginal dist. of data}}
\end{aligned}
$$

Or rearranging:

$$
\pi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{m(x_1, \dots, x_n)}
$$

Bayesian Inference

- The marginal data distribution $m(x_1, \ldots, x_n)$ does not depend on θ, i.e., it can be thought of as providing the normalizing constant for the posterior $\pi(\theta|x_1, \ldots, x_n)$.
- Thus, we often drop it and write the posterior up to a constant of proportionality:

$$\pi(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n|\theta)\pi(\theta)}{m(x_1, \ldots, x_n)} \propto f(x_1, \ldots, x_n|\theta)\pi(\theta)$$

- Thus, the fundamental relationship for Bayesian inference is

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Bayesian Inference

- Sample of observed data $(x_1, x_2, \ldots, x_n)$
- We assume a model for the data $f(x|\theta)$ which gives us the **likelihood** $L(\theta) = f(x_1, \ldots, x_n|\theta)$
- Parameters θ are unknown and characterized using probability distributions
- We assume a **prior distribution** for θ, namely $\pi(\theta)$, that encodes any information about θ before we observe the data
- Using Bayes theorem the **posterior distribution** of θ is

$$\pi(\theta|x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n|\theta)\pi(\theta)$$

up to a normalizing constant, which is the basis of inference for θ in a Bayesian framework.

Summarizing the posterior distribution

For simplicity, think of θ as a single continuous parameter.

- The **posterior mean** of θ is

$$E[\theta|x_1,\dots,x_n] = \int_{-\infty}^{\infty} \theta\, \pi(\theta|x_1,\dots x_n)d\theta$$

- The **maximum a posteriori** (MAP) estimate is the value of θ that maximizes $\pi(\theta|x_1,\dots x_n)$, that is, the mode of the posterior density.
- The posterior probability that $a \le \theta \le b$, for constants a, b is

$$P(a \le \theta \le b \mid x_1,\dots,x_n) = \int_a^b \pi(\theta|x_1,\dots x_n)d\theta$$

  These are known as **Bayesian credible intervals**. E.g., we might choose a, b to get a central 95% credible interval for θ, such that $P(\theta < a \mid \mathbf{x}) = 0.025 = P(\theta > b \mid \mathbf{x})$.

Bayesian hypothesis testing

Consider the problem of testing $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_A$.

Since θ is a r.v. in the Bayesian framework:

- Prior probability that H_0 is true is $P(\theta \in \Omega_0)$.
  - Calculate this using the prior $\pi(\theta)$.
- Posterior probability that H_0 is true is $P(\theta \in \Omega_0|\mathbf{x})$.
  - Calculate this using the posterior $\pi(\theta|\mathbf{x})$.

Then the **Bayes factor** in favour of H_0 is determined by

$$\underbrace{\frac{P(\theta \in \Omega_0|\mathbf{x})}{P(\theta \in \Omega_A|\mathbf{x})}}_{\text{posterior odds that } H_0 \text{ is true}} = \underbrace{\frac{P(\theta \in \Omega_0)}{P(\theta \in \Omega_A)}}_{\text{prior odds that } H_0 \text{ is true}} \times \text{Bayes Factor}$$