Bayesian Nonparametric Learning for Point Processes with Spatial Homogeneity: A Spatial Analysis of NBA Shot Locations

Fan Yin
Jieying Jiao
Jun Yan
Guanyu Hu

39th International Conference on Machine Learning
July 2022

Microsoft

UCONN
UNIVERSITY OF CONNECTICUT

Modeling shot location data

- In professional basketball, shot location data remains to be a fundamental metric for evaluating players and has aroused great research interests (e.g., Reich et al., 2006; Miller et al., 2014; Jiao et al., 2021).

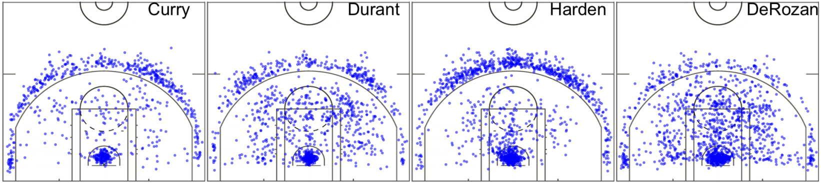


Figure: Shot data Display (2017-2018 season). On half court image, each point represents one shot.

Key Questions and Our Contributions

- Key questions
  - random nature of shot location data
  - identifying clusters of regions on which a given player has similar shooting behavior while incorporating the prior belief that spatially contiguous regions are similar
- Our contributions
  - A novel nonparametric Bayesian method for point process
  - A Gibbs sampler that enables efficient Bayesian inference across models of different dimensions without RJMCMC or samplers allocation

Key Questions and Our Contributions

- Key questions
  - random nature of shot location data
  - identifying clusters of regions on which a given player has similar shooting behavior while incorporating the prior belief that spatially contiguous regions are similar
- Our contributions
  - A novel nonparametric Bayesian method for point process
  - A Gibbs sampler that enables efficient Bayesian inference across models of different dimensions without RJMCMC or samplers allocation

Point Process Data

- A spatial point pattern is a data set $\boldsymbol{y} = (s_1, s_2, ..., s_\ell)$ consists of locations $(s_1, s_2, ..., s_\ell)$ of points that are observed in a bounded region $\mathcal{B} \subseteq \mathcal{R}^2$, which is a realization of spatial point process $\boldsymbol{Y}$.
- $N_{\boldsymbol{Y}}(A) = \sum_{i=1}^{\ell} 1(s_i \in A)$ is a counting process associated with the spatial point process $\boldsymbol{Y}$, which counts the number of points of $\boldsymbol{Y}$ for area $A \subseteq \mathcal{B}$.
- For the Poisson process $\boldsymbol{Y}$ over $\mathcal{B}$, which has the intensity function $\lambda(\boldsymbol{s})$, $N_{\boldsymbol{Y}}(A) \sim \text{Poisson}(\lambda(A))$, where $\lambda(A) = \int_A \lambda(s) ds$.
- We signify that a set of points $\boldsymbol{y} = (s_1, s_2, ..., s_\ell)$ follows a Poisson process as

$$\boldsymbol{y} \sim \mathcal{PP}(\lambda(\cdot)). \tag{1}$$

Poisson Process

- Let $A_1, A_2, \cdots, A_n$ be a partition of $\mathcal{B}$, i.e., disjoint subsets such that $\bigcup_{i=1}^{n} A_i = \mathcal{B}$.
- For each region $A_i, i = 1, \cdots, n$, we have constant intensity λ_i over region A_i. The likelihood is written as:

$$L = \prod_{i=1}^{n} f_{\text{poisson}}(N_{\boldsymbol{Y}}(A_i)|\lambda_i), \tag{2}$$

where f_{poisson} is the probability density of Poisson distribution. In later sections, we use $N(A_i)$ to denote $N_{\boldsymbol{Y}}(A_i)$

Dirichlet Process Mixture Model

$$\begin{aligned}
&y_i|Z_i,\beta_{Z_i} \sim F(\beta_{Z_i})\\
&Z_i|p \sim Discrete(p_1,\cdots,p_K)\\
&\beta_{Z_i} \sim G_0 \quad p \sim Dirichlet_K(\alpha/K,\cdots,\alpha/K)
\end{aligned} \tag{3}$$

- Z_i stands for the cluster of ith observation
- β_{c_i} means the parameter of c_ith cluster
- α stands for the precision parameter
- G_0 is the base measure in Dirichlet process
- Issue: Incorporating spatial information?

Dirichlet Process Mixture Model

$$
\begin{aligned}
& y_i | Z_i, \beta_{Z_i} \sim F(\beta_{Z_i}) \\
& Z_i | p \sim Discrete(p_1, \cdots, p_K) \\
& \beta_{Z_i} \sim G_0 \quad p \sim Dirichlet_K(\alpha/K, \cdots, \alpha/K)
\end{aligned}
\tag{3}
$$

- Z_i stands for the cluster of ith observation
- β_{c_i} means the parameter of c_ith cluster
- α stands for the precision parameter
- G_0 is the base measure in Dirichlet process
- Issue: Incorporating spatial information?

Incorporating Spatial Homogeneity

- We impose a Markov random field constraint (Besag et al., 1995; Orbanz and Buhmann, 2008) $M(\lambda_1, \ldots, \lambda_n) := \frac{1}{Z_H} \exp\{-H(\lambda_1, \ldots, \lambda_n)\}$ on $\boldsymbol{\lambda}$ to encourage rate parameters in nearby regions to be similar

$$
\begin{aligned}
G &\sim \mathrm{DP}(\alpha, G_0) \\
(\lambda_1, \ldots, \lambda_n) &\sim M(\lambda_1, \ldots, \lambda_n) \prod_{i=1}^{n} G(\lambda_i) \\
N(A_i) \mid \lambda_1, \ldots, \lambda_n &\sim \mathit{Poisson}(\lambda_i \mu(A_i)) \; i = 1, \ldots, n,
\end{aligned}
\tag{4}
$$

Cost Function

- $H(\lambda_1, \ldots, \lambda_n) := \sum_{C \in \mathcal{C}} H_C(\lambda_C)$
- $\mathcal{C}$ denotes the set of all cliques, or completely connected subsets in the underlying *neighborhood graph* $\mathcal{N} = (V_{\mathcal{N}}, E_{\mathcal{N}}, W_{\mathcal{N}})$
- Pairwise interactions

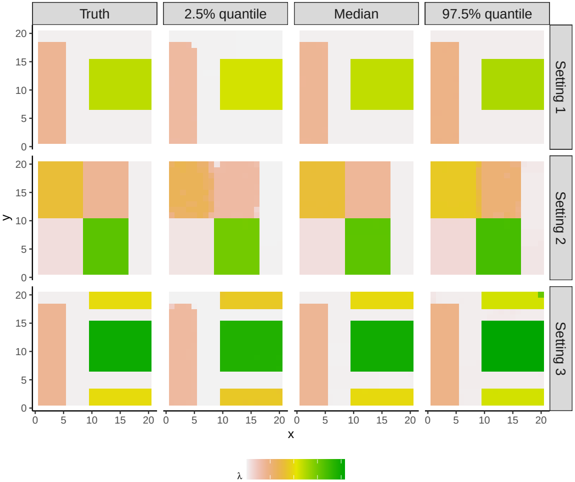
$$H(\lambda_i | \boldsymbol{\lambda}_{-i}) := -\eta \sum_{j \in \partial(i)} I(\lambda_i = \lambda_j) = -\eta \sum_{j \in \partial(i)} I(z_j = z_i) \quad (5)$$

- η: a parameter controlling the extent of spatial homogeneity with larger values dictating higher degree of spatial homogeneity

Algorithm Sketch

- A collapsed Gibbs sampler that enables the iterative updating of Poisson rate parameters and latent variables
- Dahl's method (Dahl, 2006) for post MCMC inference - identifying the *best* post burn-in iteration for estimation
- Selecting the optimal tuning parameter via Bayesian information criteria

Simulation study: Ground truth



Comparison to other methods

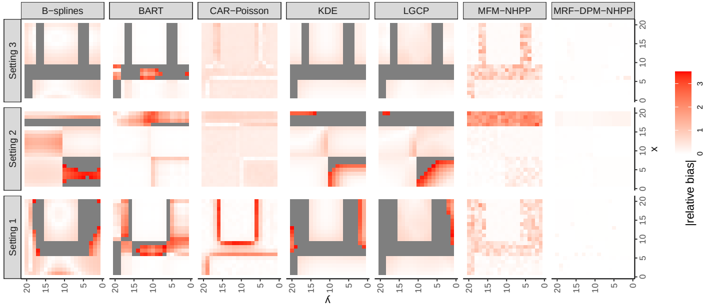
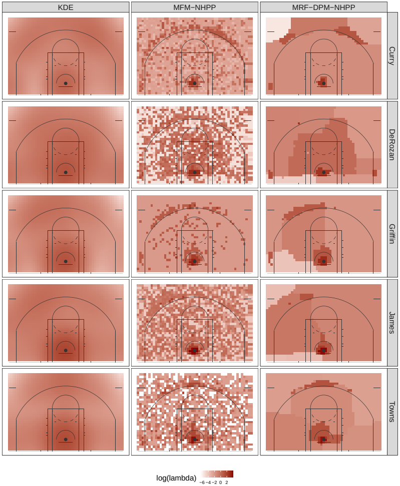


Figure: Absolute relative bias of posterior mean estimates. Dark grey: regions with large absolute relative bias (> 3.5).

NBA Data Analysis

- 20 top players in the 2017-2018 NBA regular season.
- The attacking half of the basketball court is divided into $50 \times 35 = 1750$ equally-sized grid boxes of approximately 1ft $\times$ 1ft following Miller et al. (2014).
- Parallel MCMC chains with $\eta \in \{0, 0.5, \ldots, 6\}$ for 5000 iterations using random initial values
- **Assessing predictive performance** via p-thinning approach (Illian et al., 2008) mean absolute error (MAE) defined as $\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1-p}{p}\widehat{\lambda}(A_i) - N(A_i)\right|$, where $\widehat{\lambda}(A_i)$ is the estimated intensity of region A_i based on a random subset of points (training data) and $N(A_i)$ is the number of observed points falling into region A_i in the hold-out data.

NBA Data Results



MAE Comparison for predictive assessment

Table 6: MAE results. The MAE for B-splines fitted to the training data from DeAndre Jordan is very large due to its poor fit at the boundary.

	position	MRF-DPM-NHPP	MFM-NHPP	LGCP	CAR-Poisson	BART	KDE	B-splines	Miller
DeAndre Jordan	C	0.041	0.044	0.077	0.300	0.038	0.100	$>> 1$	0.109
Joel Embiid	C	0.147	0.143	0.150	0.304	0.137	0.167	0.181	0.181
Karl-Anthony Towns	C	0.162	0.156	0.193	0.317	0.149	0.215	0.241	0.245
Dwight Howard	C	0.077	0.083	0.117	0.300	0.075	0.144	0.171	0.172
Giannis Antetokounmpo	PF	0.158	0.153	0.201	0.321	0.153	0.228	0.252	0.255
Blake Griffin	PF	0.141	0.140	0.151	0.312	0.137	0.171	0.182	0.186
LaMarcus Aldridge	PF	0.181	0.199	0.207	0.317	0.176	0.226	0.232	0.236
Kristaps Porziņģis	PF	0.169	0.175	0.174	0.318	0.168	0.183	0.183	0.183
Stephen Curry	PG	0.137	0.134	0.146	0.302	0.131	0.158	0.160	0.152
Damian Lillard	PG	0.192	0.201	0.229	0.331	0.190	0.251	0.269	0.269
Chris Paul	PG	0.128	0.126	0.130	0.296	0.129	0.138	0.138	0.138
Kyrie Irving	PG	0.167	0.177	0.186	0.315	0.167	0.200	0.209	0.208
Kevin Durant	SF	0.190	0.192	0.204	0.327	0.185	0.218	0.215	0.213
LeBron James	SF	0.200	0.199	0.241	0.329	0.181	0.273	0.292	0.298
Paul George	SF	0.200	0.200	0.233	0.334	0.194	0.250	0.255	0.249
Jimmy Butler	SF	0.149	0.156	0.162	0.311	0.150	0.174	0.181	0.181
James Harden	SG	0.192	0.196	0.237	0.330	0.182	0.264	0.287	0.280
DeMar DeRozan	SG	0.215	0.218	0.221	0.330	0.206	0.237	0.247	0.249
Russell Westbrook	SG	0.200	0.209	0.261	0.340	0.207	0.294	0.300	0.305
Klay Thompson	SG	0.188	0.190	0.191	0.321	0.182	0.204	0.203	0.200

Thank you

Welcome to reading our paper!

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statistical science*, 3–41.

Dahl, D. B. (2006). Model-based clustering for expression data via a dirichlet process mixture model. *Bayesian inference for gene expression and proteomics 4*, 201–218.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical analysis and modelling of spatial point patterns.*, Volume 70. John Wiley & Sons.

Jiao, J., G. Hu, and J. Yan (2021). A bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports 17*(2), 77–90.

Miller, A., L. Bornn, R. Adams, and K. Goldsberry (2014). Factorized point process intensities: A spatial analysis of professional basketball. In *International conference on machine learning*, pp. 235–243.

Orbanz, P. and J. M. Buhmann (2008). Nonparametric bayesian image segmentation. *International Journal of Computer Vision 77*(1-3), 25–45.

Reich, B. J., J. S. Hodges, B. P. Carlin, and A. M. Reich (2006). A spatial analysis of basketball shot chart data. *The American Statistician 60*(1), 3–12.