Bayesian Group Learning for Shot Selection of Professional Basketball Players

Hou-Cheng Yang (Joint Work with: G. Hu, Y. Xue)[1]
[1] Florida State University

Abbreviated abstract: In this paper, we develop a group learning approach to analyze the underlying heterogeneity structure of shot selection among professional basketball players in the NBA. We propose a mixture of finite mixtures (MFM) model to capture the heterogeneity of shot selection among different players based on Log Gaussian Cox process (LGCP). Our proposed method can simultaneously estimate the number of groups and group configurations. Ultimately, our proposed learning approach is further illustrated in analyzing shot charts of selected players in the NBA's 2017–2018 regular season.

Data: We focus on players that have made more than 400 field goal attempts (FTA). $D \in [0, 47] \times [0, 50]$. Indexing the players with $i \in \{1, \dots, 191\}$, the locations of shots, both made and missed, for player~i are denoted as $X_i = \{x_{i,1}, \dots, x_{i,T_i}\}$, $\forall x_{i,T_i} \in D$, where T_i is the total number of attempts made by player~i on the offensive half court.

Chanllenge: Considering the random nature of shot locations; the number and estimation of groups; Computational algorithm

Contribution: A novel methodology based on mixtures of finite mixture model and spatial point process; A Gibbs sampler that enables efficient Bayesian inference; Consistently estimates the number of groups

Model and Methods

The LGCP can be written hierarchically as

$$\begin{aligned} \boldsymbol{y} &\sim \mathcal{PP}(\lambda(\cdot)), \\ \lambda(\cdot) &= \exp(Z(\cdot)), \\ Z(\cdot) &\sim \mathcal{GP}(0, k(\cdot,\cdot)), \end{aligned}$$

where $k(\cdot,\cdot)$ is the covariance function of the Gaussian process, $Z(\cdot)$.

Let the matrix $\mathbf{C}$ be that $\mathbf{C} \equiv (\widehat{\lambda}^{(1)}, \widehat{\lambda}^{(2)}, \ldots, \widehat{\lambda}^{(n)})$, and denote $\mathbf{C}^{(i)} = \lambda^{(i)}$. Then, following the approach in Cervone et al. (2016), we compute the players' similarity matrix $\mathbf{H}$ as:

$$\mathrm{H}_{i,j} = \exp\left\{ - \left\| \frac{\mathbf{C}^{(i)}}{\sum \mathbf{C}^{(i)}} - \frac{\mathbf{C}^{(j)}}{\sum \mathbf{C}^{(j)}} \right\| \right\},$$

where $i, j \in \{1, \ldots, n\}$ and $\|\cdot\|$ is L_2 norm. It can be seen that $\mathbf{H}$ is symmetric, and $\mathbf{H} \in \mathcal{R}^{n \times n}$.

Model and Methods

$k \sim p(\cdot)$, where $p(\cdot)$ is a p.m.f on $\{1,2, \ldots\}$

$$T_{rs} = T_{sr} \overset{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta), \quad r, s = 1, \ldots, k,$$

$$U_{rs} = U_{sr} \overset{\text{ind}}{\sim} \text{N}(\mu_0, k_0^{-1} T_{rs}^{-1}), \quad r, s = 1, \ldots, k,$$

$$\text{pr}(z_i = j \mid \pi, k) = \pi_j, \quad j = 1, \ldots, k, \; i = 1, \ldots, n,$$

$$\pi \mid k \sim \text{Dirichlet}(\gamma, \ldots, \gamma),$$

$$\mathscr{S}_{ij} \mid z, U, T, k \overset{\text{ind}}{\sim} \text{N}(\mu_{ij}, \tau_{ij}^{-1}), \quad \mu_{ij} = U_{z_i z_j}, \tau_{ij} = T_{z_i z_j} \quad 1 \leq i < j \leq n,$$

A default choice of $p(\cdot)$ is a Poisson(1) distribution truncated to be positive. $\mathscr{S}$ is Fisher-Transformation of $\boldsymbol{H}$. The larger $\mathscr{S}_{ij}$ indicates the closer $\boldsymbol{\lambda}_i$ and $\boldsymbol{\lambda}_j$. $U = (U_{rs}) \in (-\infty, +\infty)^{k \times k}$ and $T = (T_{rs}) \in (0, +\infty)^{k \times k}$ are symmetric matrices. $U_{rs} = U_{sr}$ indicating the mean closeness of any function $\boldsymbol{\lambda}_i$ in cluster r and any function $\boldsymbol{\lambda}_j$ in cluster s. $T_{rs} = T_{sr}$ indicating the precision of closeness between any intensity $\boldsymbol{\lambda}_i$ in cluster r and any function $\boldsymbol{\lambda}_j$ in cluster s.

Algorithm 2 Bayesian Group Learning Procedure for Basketball Players

1: Fit LGCPs for n different players $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(n)}$ via **inlabru** and get n underlying intensity surface $\widehat{\lambda}^{(1)}(\cdot), \widehat{\lambda}^{(2)}(\cdot), \ldots, \widehat{\lambda}^{(n)}(\cdot)$,

2: Use (4) and (5) to construct matrix $\boldsymbol{S}$ and matrix $\mathscr{S}$ and based on $\widehat{\lambda}^{(1)}(\cdot), \widehat{\lambda}^{(2)}(\cdot), \ldots, \widehat{\lambda}^{(n)}(\cdot)$,

3: Get B posterior samples of $\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \ldots, \boldsymbol{z}^{(B)}$ from $\mathscr{S}$ via Algorithm 1,

4: Summary posterior samples by Dahl's method.

Real Data Analysis

We run 1,000 MCMC iterations and the first 500 iterations as burn-in period. The sizes of the nine groups are 10, 59, 8, 19, 24, 8, 10, 51, and 2 respectively.

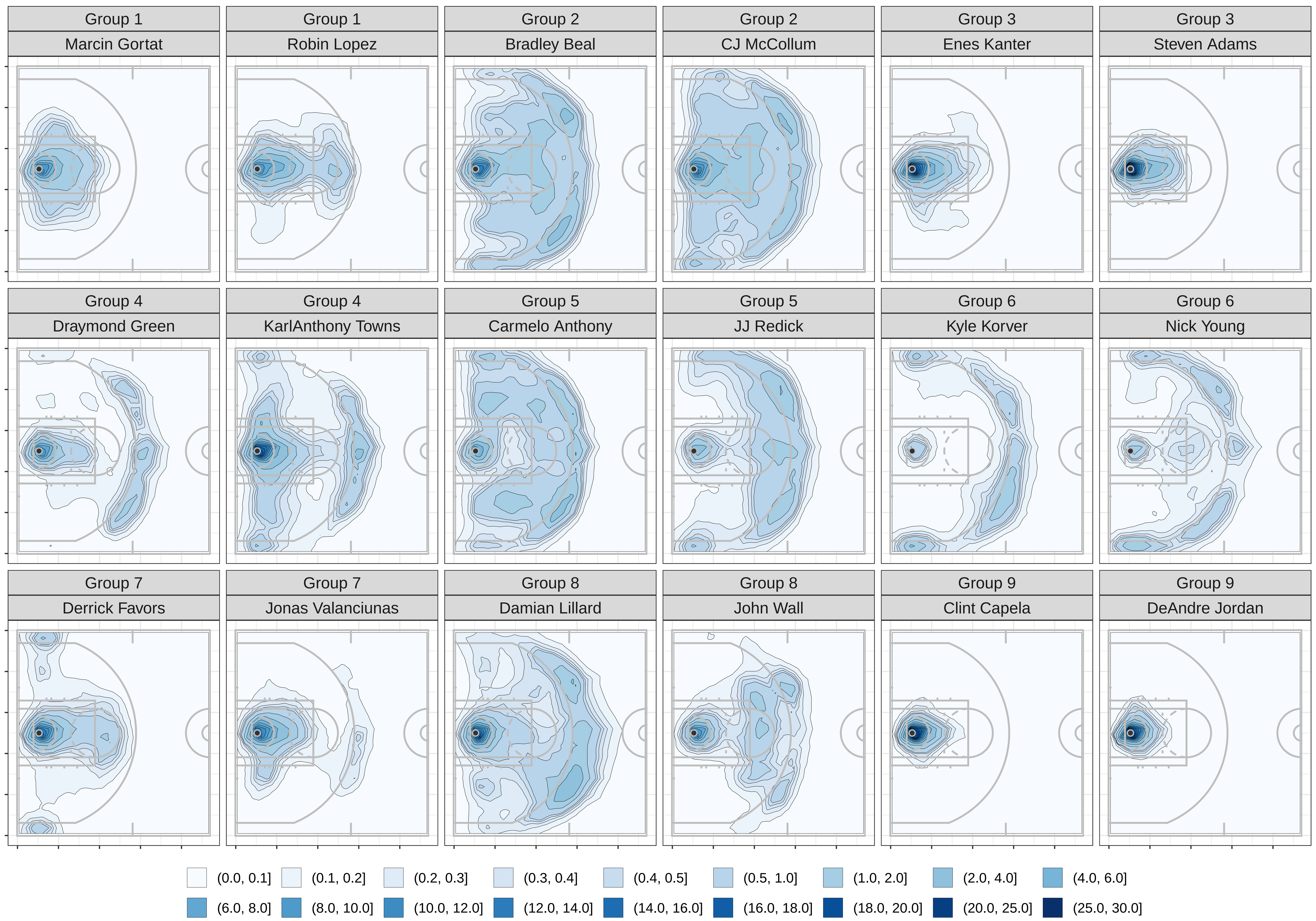

Figure 1: Fitted intensities with contour lines for two selected players in each of the nine identified groups.