Probabilistic Depth Charts: Attention-Based Probabilistic Substitution Model

Aaron John Danielson

July 5, 2025

Contents

1 Theory of Depth Charts

Depth charts represent the hierarchy or ordering of playing time. Player A outranks another in the depth chart if the player higher in the depth chart is preferred to the player lower on the depth. This preference means that, all else equal, the team would prefer to play A rather than B given both are available to play. Importantly, player B should have a substitution effect on player A and vice versa. Typically this occurs when two players occupy the similar roles or, more simply, play the same position.

For example, a point guard and a center may not share the court regularly. But, this is probably not a substitution effect. This is more likely due to the fact that one of the values is much better than the other. The rookie center does not compete with the veteran point guard for minutes.

The depth chart is a combination of preferences revealed over the course of a season. These include

- **Starter status:** Presence in the starting lineup given an availability context for the other rostered players.

- **Co-occurrence** Players who play with one another across different availability contexts are less likely to compete with one another for playing time.

- **Active status:** If the availability of one player implies an inactive status for another, then these players are in competition.

- **Substitution patterns:** If a player replaces another in games, this is direct evidence that the players are competing for the same role on the team.

- **Minutes:** Players who play more minutes across different availability contexts are higher in the depth chart.

Notation

Status Indicators Let $\mathbf{h}_g \in \{0,1\}^{|\mathcal{R}_g|}$ be the vector of player availabilities for game g. Players are available when they are not injured. While injured players are always inactive, healthy players can be set to inactive status. This is important data as it reflects their ranking in the hierarchy for playing time. To this end, let's define their active statuses as $\mathbf{a}_g \in \{0,1\}^{|\mathcal{R}_g|}$. Adding to our formal setup, define the vector $\mathbf{s}_g \in \{0,1\}^{|\mathcal{R}_g|}$ to be the vector of starting lineup assignments.

Minutes Next, let $\mathbf{m}_g \in \mathbb{R}^{|\mathcal{R}_g|}$ be the vector of minutes played by each player on the roster during game g. And, let $\mathbf{M}_g = (M_{ij,g})$ be the $|\mathcal{R}_g| \times |\mathcal{R}_g|$ matrix with entry $M_{ij,g}$ equal to the number of minutes that player i and player j share the court.

Substitution Events Let $\mathcal{D}_t$ denote the set of substitution events observed during game t. Each substitution event occurs at a game time τ and is represented as a pair of disjoint sets

$$\mathcal{S}_t = (\text{Out}_t, \text{In}_t), \quad \text{with } \text{Out}_t, \text{In}_t \subset \mathcal{R}_g, \quad \text{Out}_t \cap \text{In}_t = \emptyset.$$

Here, Out_t is the set of players who exit the court at time t, and In_t is the set of players who enter the game at that same moment. We assume substitutions are instantaneous and synchronized.

The complete set of substitution events in game g is

$$\mathcal{D}_g = \{\mathcal{S}_t\}_{t \in T_g},$$

where T_g is the set of all timepoints in game g at which substitutions occur. Importantly, we do not observe which specific player in In_t substitutes for which player in Out_t. That is, the substitution mapping is latent. However, we can define the set of all plausible substitution pairs at time t as the Cartesian product:

$$\mathcal{P}_t = \text{Out}_t \times \text{In}_t,$$

which we will later use to model substitution likelihoods based on attention and depth scores.

Top of the Depth Chart

The top line of the depth chart is defined by the starting lineup: these are the players who play against the opponent's best players. Hence, the roles on the team are defined relative to the starting lineup. Non-starters compete with starters to occupy these roles. Hence, identification of the top line of the depth chart amounts to observing the lineup chosen to play when all players are available.

Player i is at the top of the depth chart for game g if for all $\mathbf{h}_{-i,g} \in \{0,1\}^{|\mathcal{R}_g|-1}$

$$s_{i,g} = 1.$$

The probability that player i is at the top of the depth chart is

$$\sum_{\mathbf{h}_{-i,g} \in \{0,1\}^{|\mathcal{R}_g|-1}} p(s_{i,g} = 1 \mid \mathbf{h}_{-i,g}) p(\mathbf{h}_{-i,g}).$$

Then the depth chart manifests relative to the starting lineup. One way to formalize this is as a directed graph between players on the same team. The strength of the substitutive (versus complementary) intensity of B on A can be represented as a directed edge between from B to A.

Substitutes or Complements?

But, how do we identify substitutive versus complementary roles? If two players frequently share the court, their roles are complementary. They do not compete with one another for minutes. However, the lack of overlap in playing time does not imply that the players are substitutive; one of the two players can be much better than the other. For that reason, we do not see them player together. In general,

Active Status

Let's be more precise. Let $\mathbf{h}_t \in \{0,1\}^{|\mathcal{R}_t|}$ be the vector of player availabilities. Players are available when they are not injured. Healthy players can be set to inactive status, but this is important data as it reflects their ranking in the hierarchy for playing time. To this end, let's define their active statuses as $\mathbf{a}_t \in \{0,1\}^{|\mathcal{R}_t|}$. Then, aside from playing time, players compete for active status (given they are available). A player is globally preferred to another in active status if

$$h_{t,A} = h_{t,B} = 1 \implies a_{t,A} = 1 \text{ and } a_{t,B} = 0.$$

Of course, the availability of the other players can influence this. Let's consider the perfect health scenario. Suppose $h_{t,i} = 1$ for all $i \in \mathcal{R}_t$. Then A is preferred to B in active status given perfect health when

$$\mathbf{h}_t = \mathbf{1} \implies a_{t,A} = 1 \text{ and } a_{t,B} = 0.$$

Hence, an ordering appears between active status.

Substitution Network

On-Court Propensity

Adding to our formal setup, define the vector $\mathbf{s}_t$ to be the vector of starting lineup assignments. We say that player A is preferred to player B in starting status in a globally sense when

$$h_{t,A} = h_{t,B} = 1 \implies a_{t,A} = 1 \text{ and } a_{t,B} = 0.$$

This is a strong condition. It is likely that there exists an availability context vector such that the two could start together. For example, many players are injured or disqualified from play for some reason. The strength of this relation is the extent to which it holds across availability contexts.

We could see this in empirical data if there exists a vector of availabilities

3

2 Formal Model

Define a depth chart to be a directed graph over the set of nodes $\mathcal{R}_g$ representing players on the roster for game g. The directed edge from player i to player j indicates that player i is preferred to player j in the depth chart; that is,

$$E_{ij} = 1 \iff i \succ j,$$

where $E_{ij} \in \{0, 1\}$ is the directed edge from player i to player j.

Player i is preferred to player j in the depth chart if player i is more likely to play than player j and player j competes for playing time with player i. We model this using a latent probabilistic structure that combines multiple sources of playing-time evidence: substitutions, co-occurrence, starter status, and minutes played. In notation,

$$i \succ j \iff (m_i > m_j) \text{ and } (j \sim_c i)$$

where m_i, m_j are the minute shares for players i and j, and $j \sim_c i$ the relationship that player j competes with player i for playing time.

The minutes condition is straightforward to evaluate because it is directly observable. The competition relation, however, is a latent variable we learn through several phenomena. For this reason, we must estimate it using evidence from random variables like substitutions, co-occurrence and starter status. As such, rather than infer a hard binary relationship, we posit a probabilistic model of competition between players.

Distribution over Depth Chart Graphs

Assume that the edges are independent given a set of parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ where $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{R}|}$ is a vector of oncourt propensity scores and $\boldsymbol{\Omega} \in [0, 1]^{|\mathcal{R}| \times |\mathcal{R}|}$ is a matrix of competition strengths. Under these conditions the expected graph defined as

$$\mathbb{E}[E] = \sum_{e \in \{0,1\}^{|\mathcal{R}_g| \times |\mathcal{R}_g|}} e \cdot p(E = e)$$

is equal to the matrix of expected edges weights given by

$$\mathbb{E}[E] = (\mathbb{E}[E_{ij}])_{i,j \in \mathcal{R}_g} = (\Pr(i \succ j))_{i,j \in \mathcal{R}_g}.$$

This holds because the expectation operator is linear and the individual edges are independent given the parameters.

Hence, it is sufficient to define the probability of a directed edge from player i to player j

$$\begin{aligned}
\Pr(i \succ j) &= \Pr(i \sim_c j) \cdot \Pr(m_i > m_j) \\
&= \omega_{ji} \sigma(\theta_i - \theta_j) = \sigma(\alpha_{ji}) \sigma(\theta_i - \theta_j)
\end{aligned}$$

where $\sigma(x) = \frac{1}{1 + e^{-x}}$ is the sigmoid function, θ_i is the latent on-court propensity (depth score) for player i, $\omega_{ij} \in [0, 1]$ is the latent competition strength between players i and j, and α_{ji} is a real-valued substitution parameter indicating how likely player j is to replace player i given i subs out of the game and j is on the bench. Alternative expressions for ω_{ji} include

$$\omega_{ji} = \frac{\exp(\alpha_{ji})}{\sum_{k \neq j} \exp(\alpha_{ki})}.$$

The advantage of this formulation is that the αji scores are re-scaled relative to the other players who may sub-in when i exits the game. Then, for each player, this enforces a hierarchy and ensure that

$$\sum_{j \in \mathcal{R}_g} \omega_{ji} = 1$$

for all $i \in \mathcal{R}_g$. We can show that there exist coefficients $\{\alpha_{ji}\}_{j \neq i}$ such that

$$\sum_{j \in \mathcal{R}_g} \sigma(\alpha_{ji}) \neq 1.$$

Hence the choice of depends on whether we want the columns of $\boldsymbol{\Omega}$ to sum to one.

We cast the estimation of values for $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ as a multi-task model in which models for mulitple phenomena such as substitutions, co-occurrence, starter status, minutes played, and active status contribute information. The paremters $\{\alpha\}$ and θ are shared across the tasks. The next sections detail each of the component tasks.

2.1 Substitution Event Likelihood

Substitution events occur as unordered pairs of sets: a group of players exiting the game and another group entering the game simultaneously. We do not observe exact one-to-one mappings between players, but we aim to model the likelihood of such events based on latent substitution compatibility between players.

Let:

- $\mathrm{OnCourt}_{t,g} \subseteq \mathcal{A}_g$: players currently on the court at time t in game g.

- $\mathrm{Out}_{t,g} \subseteq \mathrm{OnCourt}_{t,g}$: players subbing out at time t.

- $\mathrm{Bench}_{t,g} = \mathcal{A}_g \setminus \mathrm{OnCourt}_{t,g}$: eligible substitution candidates (not currently on court).

- $\mathrm{In}_{t,g} \subseteq \mathrm{Bench}_{t,g}$: players subbing in, with $|\mathrm{In}_{t,g}| = |\mathrm{Out}_{t,g}| = k$.

We define the conditional probability of observing the specific substitution-in set $\mathrm{In}_{t,g}$ given the players subbing out $\mathrm{Out}_{t,g}$ as:

$$\Pr(\mathrm{In}_{t,g} \mid \mathrm{Out}_{t,g}) = \frac{\tilde{p}(\mathrm{In}_{t,g} \mid \mathrm{Out}_{t,g})}{\sum_{\mathrm{In}' \subseteq \mathrm{Bench}_{t,g}, |\mathrm{In}'|=k} \tilde{p}(\mathrm{In}' \mid \mathrm{Out}_{t,g})}$$

where the unnormalized score is:

$$\tilde{p}(\mathrm{In}_{t,g} \mid \mathrm{Out}_{t,g}) = \exp\left(\sum_{j \in \mathrm{In}_{t,g}} \sum_{i \in \mathrm{Out}_{t,g}} \alpha_{ji} \right)$$

Here, $\alpha_{ji} \in \mathbb{R}$ is a learned substitution affinity representing how likely player j is to replace player i when i exits the game and j is available. This formulation defines a discrete conditional exponential family over subsets of players.

Log-Likelihood

Given a dataset of substitution events $\mathcal{D}_{\mathrm{sub}} = \{(\mathrm{In}_{t,g}, \mathrm{Out}_{t,g})\}$ across all games and time steps, the total log-likelihood of the observed substitution events is:

$$\mathcal{L}_{\mathrm{sub}} = \sum_{(\mathrm{In},\mathrm{Out}) \in \mathcal{D}_{\mathrm{sub}}} \left[\sum_{j \in \mathrm{In}} \sum_{i \in \mathrm{Out}} \alpha_{ji} - \log \sum_{\mathrm{In}' \subseteq \mathrm{Bench}, |\mathrm{In}'|=|\mathrm{Out}|} \exp\left(\sum_{j' \in \mathrm{In}'} \sum_{i \in \mathrm{Out}} \alpha_{j'i} \right) \right]$$

Interpretation

This log-likelihood encourages the model to assign high scores α_{ji} to player pairs (j, i) that frequently appear in substitution events (i.e., j subs in when i subs out). The softmax-like normalization over all candidate substitution sets ensures that these scores are interpreted probabilistically.

Discussion

- The outer summation defines a probability distribution over all k-subsets of the bench players.

- This formulation is equivalent to a **log-linear model over sets**, a special case of exponential family models for structured outputs.

- To ensure numerical stability, we parameterize the unnormalized compatibility scores via additive affinities α_{ji} rather than multiplicative ω_{ji}.

- Gradients of this loss push high-affinity players j to appear more often as substitutes for i, and suppress others.

- Sampling approximations (e.g., Gumbel tricks, contrastive subsets) can be used for tractable training.

2.2 Co-Occurrence Likelihood via Shared Minutes

We model player competition based on the proportion of time players share the court during a game. Intuitively, if either of the two on-court propensity parameters for a player in a dyad is very small, then their joint minutes share will be very small. Hence low co-occurrence can be explained by at least one of the two players having a lower importance in the depth chart. But, when both player have higher on-court propensities, then the presence or absence of co-occurrence depeonds on whether the players are substitutes or complements; i.e., it depends upon whether they compete for the same role on the team. Hence, the co-occurrence likelihood can learn about the substitutive or complementary nature of the players' roles, improving the quality of the depth chart. For this second point, if either player has replaces the other in frequently in substitution events, then at least one of the two paramters, α_{ij} or α_{ji}, is larger. Hence, the minimum of oncourt propensities modulates the effect of the maximum of the substitution affinities on co-occurrence.

We summarize our intuition in the following points.

- Players with high on-court propensities, $\min\{\theta_i, \theta_j\}$, are more likely to play significant minutes.

- Players with large substitution affinities, $\max\{\alpha_{ij}, \alpha_{ji}\}$, are more likely to be substitutes.

- Players who frequently share those minutes are more likely to be complements than substitutes.

- Players who play regularly and rarely overlap are more likely to be in competition.

Probabilistic Model with Strength and Competition

We define:

- m_{ij}: minutes players i and j share the court,

- M_{game}: total game duration (e.g., 48 minutes),

- $\rho_{ij} = \frac{m_{ij}}{M_{\text{game}}} \in [0,1]$: percentage of game time players i and j share the court.

We model ρ_{ij} as a Beta-distributed random variable. The two shape parameters are defined to reflect both on court propensity via θ_i and θ_j and substitution affinities via α_{ij} and α_{ji}. As stated above, it is the mininum between propensities and the maximum between substitution affinities that modulate the co-occurrence likelihood. Hence, we define the expected value of ρ_{ij} as a function of these parameters. But, the standard `min` and `max` functions are nondifferentiable at $a = b$. For gradient-based optimization, we use smooth approximations:

$$\text{SmoothMin}_\beta(a, b) = -\frac{1}{\beta} \log\left(e^{-\beta a} + e^{-\beta b}\right)$$

$$\text{SmoothMax}_\beta(a, b) = \frac{1}{\beta} \log\left(e^{\beta a} + e^{\beta b}\right)$$

As $\beta \to \infty$, these converge to $\min(a, b)$ and $\max(a, b)$, respectively.

The Beta distribution is parameterized by two shape parameters, a and b such that the expected value is given by

$$\mathbb{E}[\rho_{ij}] = \frac{a}{a + b}.$$

We want the expected value of ρ_{ij} to reflect the intuition that players with high on-court propensities are more likely to share the court, while players with high substitution affinities are less likely to do so. Two players with very large values for θ and very small values for α shoudl have a very large mean. But, if either θ is small, then the expected value should be small even if the substitution effect is zero. Similarly, if the substition effect is very large, then the expeced value should be small even if the θ parameters are both large. Hence a target function for the expected value of ρ_{ij} is: We want the expected value of ρ_{ij} to reflect the intuition that players with high on-court propensities are more likely to share the court, while players with high substitution affinities are less likely to do so. To implement this, we define:

$$T_{ij} = \sigma(\text{SmoothMin}_\beta(\theta_i, \theta_j))$$

$$A_{ij} = \sigma(\text{SmoothMax}_\beta(\alpha_{ij}, \alpha_{ji}))$$

Then the expected value of ρ_{ij} is:

$$\mathbb{E}[\rho_{ij}] = (1 - A_{ij}) \cdot T_{ij}$$

This expression captures the idea that co-occurrence is only expected to be high when both players are important (via θ) and not strong substitutes (via α). To model ρ_{ij} as a Beta-distributed variable, we scale this expected value by a concentration parameter $\gamma_{m_{ij}} > 0$

$$\text{Shape}_1 = \gamma_{m_{ij}} \cdot \mathbb{E}[\rho_{ij}]$$
$$\text{Shape}_2 = \gamma_{m_{ij}} \cdot (1 - \mathbb{E}[\rho_{ij}])$$

$$\rho_{ij} \sim \text{Beta}(\text{Shape}_1, \text{Shape}_2)$$

If necessary, we can define T_{ij} using a two-step smooth, positive transformation of the latent propensities. To ensure that the on-court propensities are positive, we can use a softplus transformation to produce

$$\tilde{\theta}_i = \log(1 + e^{\theta_i}) \quad \text{(softplus)}.$$

Then, these transformed values are fed into SmoothMin_β to get

$$T_{ij} = \sigma(\text{SmoothMin}_\beta(\tilde{\theta}_i, \tilde{\theta}_j)).$$

This expression is symmetric, fully differentiable, and softly encourages co-occurrence among players with high on-court propensity. It serves as a stable approximation of $\min(\theta_i, \theta_j)$ without requiring bounded θ values.

Log-Likelihood

Let ρ_{ij}^{obs} be the observed co-occurrence proportion for players i and j in a given game. The negative log-likelihood of the observed data under our model is:

$$\mathcal{L}_{\text{cooc}} = -\sum_{i,j} \log \left[\text{BetaPDF} \left(\rho_{ij}^{\text{obs}} \mid a_{ij} = \gamma_{\text{co}} \cdot \mathbb{E}[\rho_{ij}], \quad b_{ij} = \gamma_{\text{co}} \cdot (1 - \mathbb{E}[\rho_{ij}]) \right) \right]$$

where the expected value is defined as:

$$\mathbb{E}[\rho_{ij}] = (1 - A_{ij}) \cdot T_{ij}$$

$$T_{ij} = \sigma \left(\text{SmoothMin}_\beta(\tilde{\theta}_i, \tilde{\theta}_j) \right), \qquad A_{ij} = \sigma \left(\text{SmoothMax}_\beta(\alpha_{ij}, \alpha_{ji}) \right)$$

and

$$\tilde{\theta}_i = \log(1 + e^{\theta_i}) \quad \text{(softplus)}.$$

2.3   Unified Probabilistic Model for Starter Selection

Since the starter parameter is highly informed by the share of minutes played by each player, players who frequently starter will tend to have high latent on-court propensity θ_i. Additionally, starters generally do not compete with one another for playing time. The 10 pairs of players in the starting lineup should have low substitution affinities α_{ij} and α_{ji}. Let $\mathcal{R}_{\text{start}}$ denote the set of players who start when the full roster is healthy. For each starter $i \in \mathcal{R}_{\text{start}}$, we define the following probabilistic model. We define a probabilistic model over starter subsets $S \subseteq \mathcal{R}$ with the following parameters.

- **Main effects:** players with high on-court propensity θ_i are more likely to be selected.

- **Bias term:** a global scalar b_{start} acts as a threshold for inclusion.

- **Pairwise penalties:** players with high substitution affinity α_{ij} and α_{ji} are discouraged from being selected together.

Let $\mathcal{S}$ denote the set of valid size-k starter subsets from the roster $\mathcal{R}$. Define:

$$\Pr(S = s) = \frac{1}{Z} \cdot \exp\left(\sum_{i \in s}(\theta_i - b_{\text{start}}) - \gamma_{\text{starter}} \sum_{\substack{i,j \in s \\ i \neq j}} \psi_{ij} \right),$$

where

$$\psi_{ij} = \sigma(\alpha_{ij} + \alpha_{ji}) \quad \text{and} \quad Z = \sum_{S' \in \mathcal{S}} \exp\left(\sum_{i \in S'}(\theta_i - b_{\text{start}}) - \gamma_{\text{starter}} \sum_{\substack{i,j \in S' \\ i \neq j}} \psi_{ij} \right)$$

Log-Likelihood

Given an observed starter subset $S \in \mathcal{S}$, the log-likelihood becomes:

$$\mathcal{L}_{\text{start}} = \sum_{i \in S}(\theta_i - b_{\text{start}}) - \gamma_{\text{starter}} \sum_{\substack{i,j \in S \\ i \neq j}} \psi_{ij} - \log \sum_{S' \in \mathcal{S}} \exp\left(\sum_{i \in S'}(\theta_i - b_{\text{start}}) - \gamma_{\text{starter}} \sum_{\substack{i,j \in S' \\ i \neq j}} \psi_{ij} \right)$$

Interpretation

The bias term b_{start} controls the global threshold of starter selection. It acts as a regularizer that reduces the total "attractiveness" of including any player, and can absorb systemic variation in the scale of θ. This allows us to learn a more stable ranking signal and facilitates top-k decoding.

Alternative Inference Method

Starter Propensity Model   vspace1em We treat $s_i = 1$ as a binary variable indicating that player i is selected as a starter. This is modeled using a logistic function of the latent court propensity:

$$\Pr(s_i = 1) = \sigma(\theta_i - b_{\text{start}}),$$

where $b_{\text{start}} \in \mathbb{R}$ is a learned threshold that governs the overall propensity required to be a starter, This yields the log-likelihood term:

$$\log \Pr(s_i = 1) = \log \sigma(\theta_i - b_{\text{start}}).$$

Starter Competition Suppression   vspace1em To reflect the assumption that starters are not direct substitutes, we introduce a penalization term expresses the idea that starters are not in direct competition with one another. This affinity is shared with previous models via the pairwise latent variable α_{ij}. More specifically, the penalty term is

$$Z_{ij} \sim \text{Bernoulli}(1 - \omega_{ij}) \quad \text{for } i, j \in \mathcal{R}_{\text{start}}, \ i \neq j,$$

with log-likelihood:

$$\log \Pr(Z_{ij} = 1) = \log(1 - \sigma(\alpha_{ij} + \alpha_{ji})).$$

Starter Log-Likelihood   vspace1em Combining both components, the full log-likelihood over starters is:

$$\mathcal{L}_{\text{start}} = \sum_{i \in \mathcal{R}_{\text{start}}} \log \sigma(\theta_i - b_{\text{start}}) + \sum_{\substack{i,j \in \mathcal{R}_{\text{start}} \\ i \neq j}} \log\left(1 - \sigma(\alpha_{ij} + \alpha_{ji})\right).$$

Comparison of the Two Starter Modeling Approaches

We present two probabilistic formulations for modeling the selection of starters:

1. A **unified log-linear subset model** that defines a joint distribution over valid starter sets of fixed size k.

2. A **factorized likelihood model** that treats each starter's inclusion and each pairwise non-competition relation independently.

Unified Log-Linear Subset Model   The joint probability of a starter set S is modeled as:

$$\Pr(S) = \frac{1}{Z} \exp\left(\sum_{i \in S}(\theta_i - b_{\text{start}}) - \gamma_{\text{starter}} \sum_{\substack{i,j \in S \\ i \neq j}} \psi_{ij}\right),$$

with $\psi_{ij} = \sigma(\alpha_{ij} + \alpha_{ji})$ and Z a normalizing constant over all size-k subsets.
   This model captures:

- **Subset-level structure** — the score for a subset depends on both who is in it and how they interact.

- **Global normalization** — probabilities are defined over the entire space of valid starter subsets.

- **Joint inference** — decisions about one player depend on all others selected.

Factorized Likelihood Model   This alternative treats inclusion and competition as conditionally independent events:

$$\mathcal{L}_{\text{start}} = \sum_{i \in \mathcal{R}_{\text{start}}} \log \sigma(\theta_i - b_{\text{start}}) + \sum_{\substack{i,j \in \mathcal{R}_{\text{start}} \\ i \neq j}} \log\left(1 - \sigma(\alpha_{ij} + \alpha_{ji})\right).$$

This model captures:

- **Independent inclusion** — each player's probability of starting is determined solely by their own θ_i.

- **Pairwise competition suppression** — penalizes selection of pairs with high substitution affinity.

- **Simplified optimization** — no partition function required.

Comparison Summary

Aspect	Unified Subset Model	Factorized Likelihood Model
Structure	Joint probability over full starter sets	Independent probabilities for each player and pair
Normalization	Globally normalized over all valid subsets	Local (per-player and per-pair)
Captures Interaction	Yes — pairwise penalties affect joint score	Yes — but only via additive terms
Inclusion Coupling	Fully coupled: changing one player affects the rest	Independent: each player's score is separable
Computational Cost	High (requires subset sampling or approximation)	Low (exact and efficient)
Use Case	When exact subset structure matters, e.g., ranked decoding or MAP inference over lineups	When inference speed or modular training is more important

Inference

Computing the normalizing constant Z involves summing over $\binom{|\mathcal{R}|}{k}$ possible subsets and may be intractable for large rosters. In practice, this can be approximated using:

- Gumbel-softmax sampling with top-k constraint,

- Contrastive estimation with negative samples,

- Variational approximations or MAP decoding with learned temperature.

- Caching and reusing subset scores across epochs.

Conclusion The unified subset model is more expressive and allows for holistic inference across lineup configurations, but is computationally intensive. The factorized model is more scalable and interpretable, and serves as a useful approximation or pretraining target. Both share parameters (θ, α) and can be used in multi-task learning to regularize one another.

2.4   Minutes Played Likelihood

We model playing time as the **fraction of game minutes** that player i plays. This makes the scale consistent across games and interpretable in the unit interval.

Step 1: Define Minute Share

Let:

- $m_i \in [0, M_{\text{game}}]$: total minutes player i plays,

- M_{game}: total duration of the game (e.g., $M_{\text{game}} = 48$ minutes for an NBA game),

- $r_i = \frac{m_i}{M_{\text{game}}} \in [0, 1]$: fraction of game minutes played by player i.

Step 2: Probabilistic Model

We model r_i using a Beta distribution:

$$r_i \sim \text{Beta}\left(\gamma_{\min} \cdot \sigma(\theta_i),\ \gamma_{\min} \cdot (1 - \sigma(\theta_i))\right),$$

where:

- $\theta_i \in \mathbb{R}$ is a latent on-court propensity for player i,

- $\gamma_{\min} > 0$ is a learnable concentration parameter.

This implies:
$$\mathbb{E}[r_i] = \sigma(\theta_i), \quad \mathrm{Var}(r_i) = \frac{\sigma(\theta_i)(1 - \sigma(\theta_i))}{1 + \gamma_{\min}}.$$

Step 3: Log-Likelihood

The contribution to the total loss from minutes played is:
$$\mathcal{L}_{\mathrm{minutes}} = -\sum_i \log \left[\mathrm{BetaPDF} \left(r_i^{\mathrm{obs}} \mid \gamma_{\min} \cdot \sigma(\theta_i), \; \gamma_{\min} \cdot (1 - \sigma(\theta_i)) \right) \right].$$

Interpretation

- Players with high θ_i are expected to play more minutes, up to the full game.

- This shares parameters with the starter model and substitution model.

- $\gamma_{\min}$ controls how tightly actual playing time is expected to concentrate around the learned propensity.

2.5 Active Status Likelihood

We model whether a player is marked active (dressed and eligible to play) in game g using two complementary signals:

- Their latent on-court propensity θ_i, which reflects global ordering of playing strength.

- Their local competition with teammates, via learned substitution affinities α_{ij}.

Individual Activation Likelihood

Let $a_{ig} \in \{0, 1\}$ indicate whether player i is active (not a DNP) in game g, conditional on being healthy.
  We model this with a logistic function of player propensity:
$$\Pr(a_{ig} = 1) = \sigma(\theta_i - b_{\mathrm{active}}),$$
where $b_{\mathrm{active}} \in \mathbb{R}$ is a learnable threshold.
  The likelihood of observed activation statuses is:
$$\mathcal{L}_{\mathrm{active,indiv}} = \sum_g \sum_{i \in \mathcal{R}_g} a_{ig} \cdot \log \sigma(\theta_i - b_{\mathrm{active}}) + (1 - a_{ig}) \cdot \log(1 - \sigma(\theta_i - b_{\mathrm{active}})).$$

Pairwise Competition Likelihood

We assume that activation choices among similarly ranked players carry additional signal. Let $\pi_{ij} = \sigma(\alpha_{ji} + \delta \cdot \alpha_{ij}) \in (0, 1)$ represent the probability that player j is preferred over player i based on substitution competition.
  Let:

- α_{ij}: substitution affinity from player i to j,

- δ: asymmetry parameter shared with the co-occurrence model.

Then, for all player pairs (i, j) on the roster for game g where one is active and the other is not, we define:
$$\mathcal{L}_{\mathrm{active,pairwise}} = \sum_g \sum_{\substack{i,j \in \mathcal{R}_g \\ a_{ig} \neq a_{jg}}} w_{ijg} \cdot \log \left[\pi_{ij}^{\mathbb{1}[a_{jg}=1]} \cdot (1 - \pi_{ij})^{\mathbb{1}[a_{ig}=1]} \right],$$
where the weight $w_{ijg} = \exp\left(-|\theta_i - \theta_j|\right)$ ensures that competition between players of similar rank is emphasized.

Full Active Likelihood

The total log-likelihood is:
$$\mathcal{L}_{\text{active}} = \mathcal{L}_{\text{active,indiv}} + \mathcal{L}_{\text{active,pairwise}}.$$

Interpretation

- The individual term promotes global ordering: high θ_i increases activation likelihood.

- The pairwise term captures local tie-breaking behavior when multiple bench players are plausible activations.

- The use of $\pi_{ij} = \sigma(\alpha_{ji} + \delta \cdot \alpha_{ij})$ links this likelihood to our substitution and co-occurrence models.

2.6 Active Status Likelihood: Exponential Family Formulation

Alternatively, we define a fully generative exponential family model over the binary vector of activations $\mathbf{a}_g = (a_{ig})_{i \in \mathcal{R}_g} \in \{0,1\}^{|\mathcal{R}_g|}$ for each game g.

We define:
$$\Pr(\mathbf{a}_g) = \frac{1}{Z_g} \exp\left(\sum_{i \in \mathcal{R}_g} a_{ig}(\theta_i - b_{\text{active}}) - \sum_{\substack{i,j \in \mathcal{R}_g \\ i < j}} \psi_{ij} \cdot a_{ig} a_{jg} \right),$$

where:
$$\psi_{ij} = \gamma_{\text{active}} \cdot \sigma(\alpha_{ij} + \alpha_{ji}),$$

and Z_g is the normalizing constant over all binary vectors $\mathbf{a}_g \in \{0,1\}^{|\mathcal{R}_g|}$.

Interpretation

- The main term $a_{ig} \cdot \theta_i$ increases the likelihood of activating strong players.

- The pairwise penalty term $a_{ig} a_{jg} \cdot \psi_{ij}$ discourages simultaneous activation of players with high substitution affinity.

- This structure allows activation decisions to be modeled jointly across all players.

Log-Likelihood For an observed activation vector $\mathbf{a}_g$, the log-likelihood is:
$$\mathcal{L}_{\text{active}}^{(g)} = \sum_{i \in \mathcal{R}_g} a_{ig}(\theta_i - b_{\text{active}}) - \sum_{\substack{i,j \in \mathcal{R}_g \\ i < j}} \psi_{ij} \cdot a_{ig} a_{jg} - \log Z_g$$

The full log-likelihood is summed over all games:
$$\mathcal{L}_{\text{active}} = \sum_g \mathcal{L}_{\text{active}}^{(g)}.$$

2.7 Final Objective

The full model combines multiple signal-specific likelihoods, each contributing information about latent depth structure, competition, and activation.

$$\begin{aligned}
\mathcal{L}_{\text{total}} =& \lambda_{\text{sub}} \cdot \mathcal{L}_{\text{sub}}(\alpha_{ji}) + \lambda_{\text{cooc}} \cdot \mathcal{L}_{\text{cooc}}(\tilde{\theta}_i, \tilde{\theta}_j, \alpha_{ij}, \alpha_{ji}) + \\
& \lambda_{\text{start}} \cdot \mathcal{L}_{\text{start}}(\theta_i, \alpha_{ij}, \alpha_{ji}, b_{\text{start}}) + \lambda_{\text{minutes}} \cdot \mathcal{L}_{\text{minutes}}(\theta_i) + \\
& \lambda_{\text{active}} \cdot \mathcal{L}_{\text{active}}(\theta_i, \alpha_{ij}, \alpha_{ji}, b_{\text{active}}, \delta) - \lambda_{\text{reg}} \cdot \|\boldsymbol{\lambda}\|_2^2
\end{aligned}$$

where $\lambda_{\text{sub}}, \lambda_{\text{cooc}}, \lambda_{\text{start}}, \lambda_{\text{minutes}}, \lambda_{\text{active}}, \lambda_{\text{reg}}$ are tunable coefficients that balance the contributions of each signal.

Each likelihood term promotes consistency across different observable behaviors (e.g., substitution, co-occurrence, starter status), while sharing latent player-level parameters.

Parameter Dependency Table

Observation	Signal Type	Parameter Dependency
Substitution $(i \rightarrow j)$	Replacement likelihood	α_{ji}
Minutes overlap ρ_{ij}	Competition/complementarity	$\tilde{\theta}_i, \tilde{\theta}_j, \alpha_{ij}, \alpha_{ji}$
Starting status s_i	Priority and mutual exclusivity	$\theta_i, \alpha_{ij}, \alpha_{ji}, b_{\text{start}}$
Total minutes t_i	Global strength	θ_i
Activation status a_{ig}	Binary depth activity	$\theta_i, \alpha_{ij}, \alpha_{ji}, b_{\text{active}}, \delta$

Model Interpretation

- Learns latent on-court propensities θ_i that govern overall playing time and participation.

- Learns pairwise substitution affinities α_{ij} encoding directional competition structure.

- Encodes both global order and local competition via exponential family models over subsets.

- Supports differentiable inference over subsets (e.g., substitution events, starter sets).

- Modular: each signal can be removed or replaced without rearchitecting the full model.

3 Accounting for Health Context

Recall the definition of the health context vector $\mathbf{h}_g \in \{0, 1\}^{|\mathcal{R}_g|}$, which indicates the subset of the roster $\mathcal{R}_g$ not out with an injury. The model as presented in the prior section assumes full health. But, this is far from being a safe assumption in practice. The set of active players, the starting lineup, substitution patterns throughout the game, oncourt occurrence, and minutes depend on the health status of the roster.

Hence, every model we build for the 5 forces in the prior section take the form

$$\Pr(\text{event} \mid \mathbf{h}_g = \mathbf{1})$$

where $\mathbf{h}_g = \mathbf{1}$ indicates that all players in the roster are healthy. In practice, we can condition on the health context vector $\mathbf{h}_g$ to account for the health status of the team. This is particularly important when modeling substitutions, as the set of available players may change throughout the game due to injuries or other factors.

For example, two players may be actual competitors for a role, but injuries to a subset of the team lead to the two players playing in the same lineups. This will change the value of the model parameters.

Our hope is that we can estimate model parameters θ_i and ω_{ij} that are robust to the health context of the team. To do this, we can condition on the health context vector $\mathbf{h}_g$ in our model. This allows us to learn depth scores and substitution probabilities that are specific to the health context of the team.

Alternatively, we can assume that the model parameters θ_i and ω_{ij} are functions of the health context vector $\mathbf{h}_g$. The model can be extended to condition on this vector, allowing it to learn depth scores and substitution probabilities that are specific to the health context of the team.

Conditional Models

To accurately learn robust and generalizable depth chart parameters, we must either:

1. **Condition on the health context $\mathbf{h}_g$** when modeling each signal, or

2. **Model θ_i and ω_{ij} as functions of $\mathbf{h}_g$**, thereby allowing depth and competition to adapt to roster availability.

Formally, each probabilistic component of our model becomes:

$$\Pr(\text{event}_g \mid \mathbf{h}_g),$$

where the "event" might be a substitution $(i \to j)$, co-occurrence c_{ij}, active status a_{ig}, starter designation s_{ig}, or total minutes t_{ig}.

For example, if i and j share the court often, but only in injury-depleted contexts, we should not infer that ω_{ij} is low (i.e., that they are complements). Rather, this co-occurrence may be forced by limited availability. Similarly, a player with low total minutes might be a reliable backup whose minutes are masked by the full-health dominance of others.

Robust Parameter Estimation

Our goal is to estimate global parameters θ_i and ω_{ij} that reflect stable, context-independent preferences and competition. To achieve this, we propose two strategies:

- **Stratified Conditioning:** Use only full-strength games (where $\mathbf{h}_g = \mathbf{1}$) to directly estimate θ_i and ω_{ij}.

- **Health-Aware Likelihoods:** In each term of the log-likelihood, include $\mathbf{h}_g$ as a conditioning variable or weight, filtering or attenuating signals that occur in irregular availability contexts.

Alternative: Health-Dependent Parameters

A more flexible extension would allow player parameters to depend directly on the health context:

$$\theta_i^{(g)} = f_\theta(i, \mathbf{h}_g), \quad \omega_{ij}^{(g)} = f_\omega(i, j, \mathbf{h}_g),$$

where f_θ and f_ω are learned functions (e.g., neural networks or attention mechanisms) that encode how a player's role and competitive dynamics change based on teammate availability.

While this approach is more expressive, it may require more data and complexity to generalize well, especially for rare or unusual availability configurations.

Summary

Incorporating the health context vector $\mathbf{h}_g$ into our modeling pipeline is critical for accurate inference of the underlying depth chart. Ignoring this structure can lead to spurious inferences about preference and competition due to confounding from roster availability. Our goal is to learn stable player depth scores and competition strengths that generalize across availability contexts while properly handling the combinatorial variation introduced by partial rosters.