Estimating Player Skills in Real-World Communities using Variational Inference

Logan Blaine | MIT 6.435 | 5-5-2022

Statistical modeling of paired comparison data

Let $y \in \{0,1\}$ be the result of a *binary paired comparison* between two entities (i, j)

$$y = \begin{cases} 0 & \text{if } j \text{ is preferred over } i \\ 1 & \text{if } i \text{ is preferred over } j \end{cases}$$

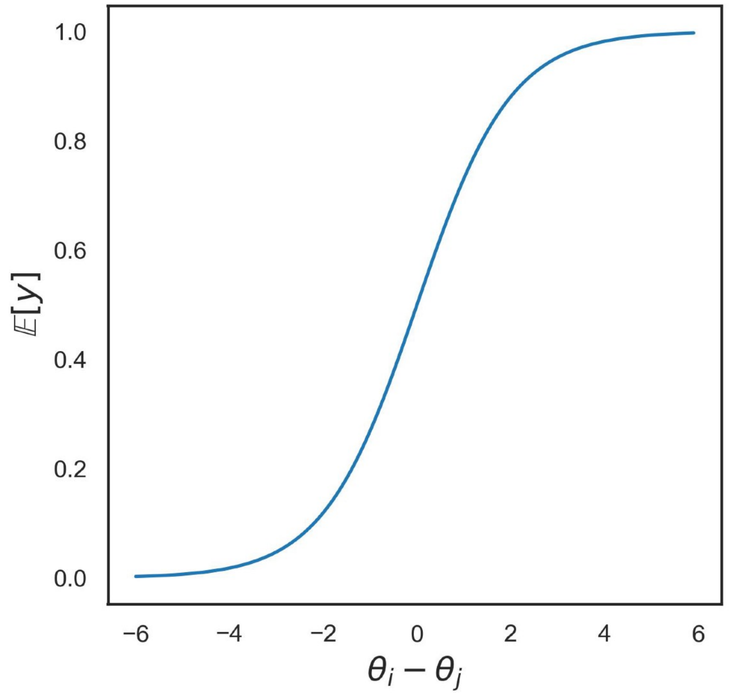
Bradley-Terry Model (a.k.a. logistic regression)

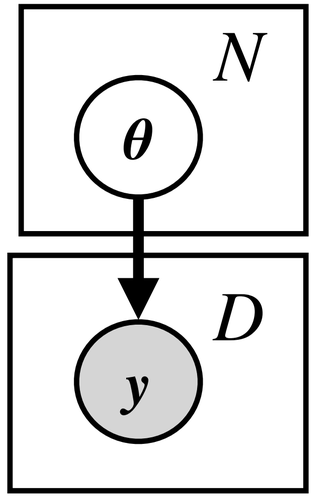
Assume each entity has a "merit"/"skill" parameter

$$\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^\top \in \mathbb{R}^N$$

Then the outcome y is distributed according to:

$$y \,|\, \theta_i, \theta_j \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-(\theta_i - \theta_j)}}\right)$$





Applications

iChess.com

NETFLIX

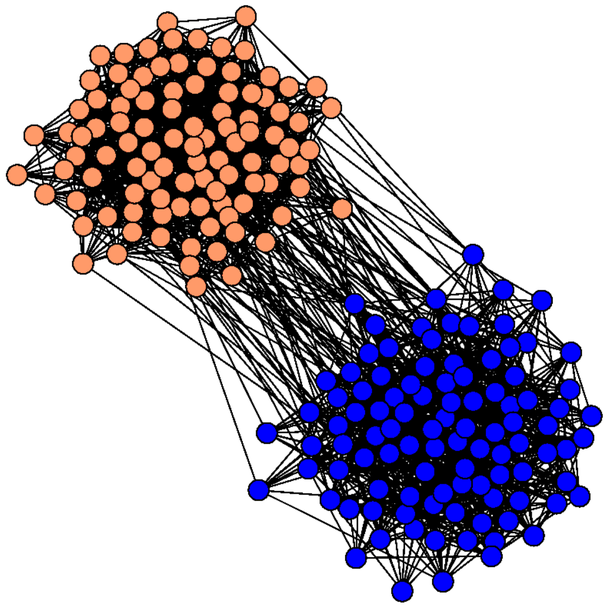
Sports & Games

Recommendations & Ads

And more...

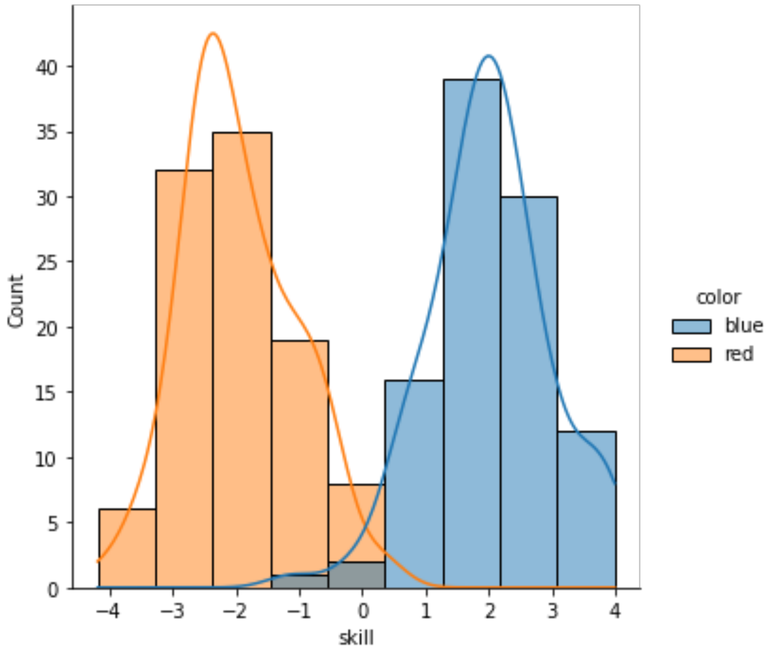
Simulated data model with regional match disparity and skill disparity

Simulated Matches $\boldsymbol{y}$:
Stochastic Block Model



2 clusters with 100 nodes each
p(within-group connections) = 0.2
p(between-group connections) = 0.01

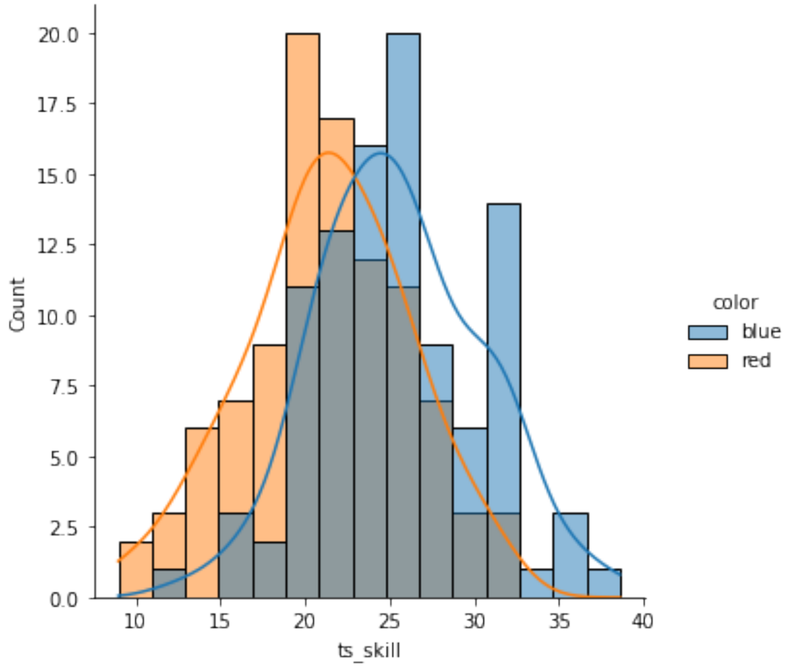
Simulated Skill Parameters $\boldsymbol{\theta}$



$$\theta_{red} \sim \mathcal{N}(-2,1)$$

$$\theta_{blue} \sim \mathcal{N}(2,1)$$

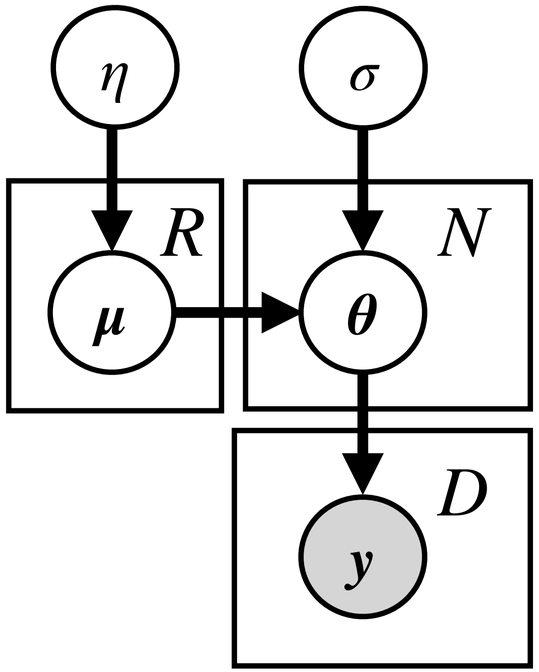
Microsoft TrueSkill Estimates
$E[\boldsymbol{\theta} \mid \boldsymbol{y}]$



Hierarchical Bayesian model of regional skill disparity

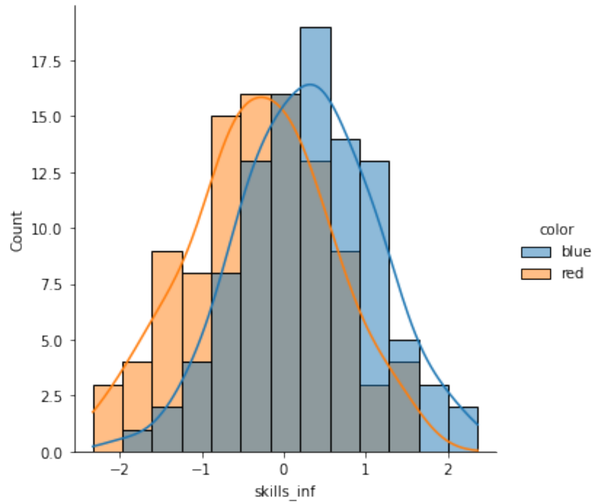
Regional Bradley-Terry (B-T)

Each player i is assigned to a region $r(i) \in \{1, \ldots, R\}$, their prior skill is Normal with mean $\mu_{r(i)}$ and variance σ^2

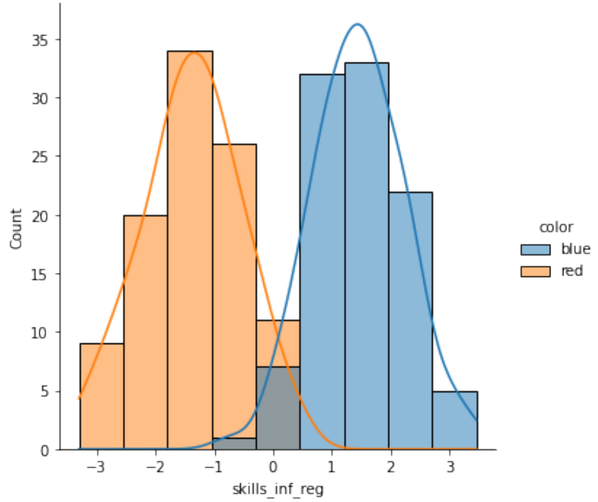


$$\sigma \sim \text{Cauchy}^+(1)$$
$$\eta \sim \text{Cauchy}^+(1)$$
$$\boldsymbol{\mu} \,|\, \eta \sim \mathcal{N}(0,\, \eta)$$
$$\theta_i \,|\, \boldsymbol{\mu}, \sigma \sim \mathcal{N}(\mu_{r(i)}\,,\, \sigma)$$
$$y_i \,|\, \boldsymbol{\theta} \sim \text{Bernoulli}(\,\cdot\,)$$

Pyro Bradley-Terry



Pyro Regional B-T



	Bradley-Terry	Regional B-T
Train ELBO	0.622	0.604
Test ELBO	0.673	0.651

Inference using regional model recalibrates rankings across states