

# Métodos Estatísticos para Análises de Dados do Esporte - Aula 2

Daniel Takata Gomes

IX Workshop on Probabilistic and Statistical Methods

3 de fevereiro de 2022

## Métodos para análise de dados de futebol

# Sumário

## 1 Introdução

# Sumário

- ① Introdução
- ② Modelo baseado na distribuição de Poisson

# Sumário

- ① Introdução
- ② Modelo baseado na distribuição de Poisson
- ③ Probabilidades e as apostas esportivas

# Sumário

- ① Introdução
- ② Modelo baseado na distribuição de Poisson
- ③ Probabilidades e as apostas esportivas
- ④ Redes neurais

# Sumário

- ① Introdução
- ② Modelo baseado na distribuição de Poisson
- ③ Probabilidades e as apostas esportivas
- ④ Redes neurais
- ⑤ Considerações finais



## Após vitória do Flamengo sobre o Atlético-MG, probabilidades de título do Brasileirão mudam; veja chances

Segundo a projeção do Five Thirty Eight, site especializado em cálculos de probabilidade, o Flamengo tem 22% de chance de ser campeão brasileiro...

30 de out. de 2021





Parece óbvio que tais probabilidades são calculadas através de simulações, baseadas nas probabilidades dos resultados dos jogos.

Parece óbvio que tais probabilidades são calculadas através de simulações, baseadas nas probabilidades dos resultados dos jogos.

O que não se comenta muito é: como as probabilidades dos resultados dos jogos são calculadas?

Encontre times, competições, corridas e ...

e-mail/nome de us

senha

☐ Lembrar senha
 ☐ Esqueceu seu N

Exchange

Esportes

Cassino

Slots

Cassino ao vivo

Pôquer

Jogos Exchange

Bingo

Esportes Virtuais

Início

Ao vivo

Cash Out

Promoções

Futebol

Tênis

Cavalos

Galgos

Basquete

Segure esta Bolada

Quero Aprender

Tutoriais

Últimas Notic

Ocultar menu

★ Meus Favoritos

Esportes

Apostas Especiais

Automobilismo

Basquete

Beisebol

Boxe

Ciclismo

Corrida de Cavalos

Corrida de Galgos

Corridas de Cavalos - Hoje

Corridas de Galgos - Hoje

Crquete

Dardos

Desportos de Inverno

E-Sports

Futebol

Futebol Americano

Futebol Australiano

Golfe

Handebol

Hóquei no Gelo

Jogos de Mesa

Próximas Corridas de Cavalos

Guia Betfair TV

Destaques Esportivos

Futebol

Correspondido

1

X

2

jan 29 14:30

Peterborough Sheff Utd

R\$954,470

3.75 R\$82

3.95 R\$171

3.55 R\$322

3.8 R\$143

2.14 R\$236

2.18 R\$2373

71' 1 0

Envigado Deportivo Pasto

R\$294,594

1.25 R\$2638

1.28 R\$2492

5.2 R\$453

5.8 R\$940

36 R\$127

44 R\$132

jan 29 12:00

Barnsley Bournemouth

R\$246,559

4.8 R\$81

5.2 R\$207

3.6 R\$113

3.85 R\$147

1.88 R\$107

1.92 R\$1105

jan 28 16:45

Huddersfield Stoke

R\$237,423

2.56 R\$111

2.66 R\$634

3.35 R\$46427

3.4 R\$39795

3.05 R\$837

3.2 R\$317

Tênis

Correspondido

1

2

qua 05:30

Felix Auger-Aliassime Daniil Medvedev

R\$1,457,598

7.2 R\$18143

7.4 R\$3469

1.15 R\$398523

1.16 R\$111410

1 2 1 2

Jenson Brooksby Mitchell Krueger

R\$1,273,276

1.38 R\$3430

1.39 R\$5513

3.55 R\$2158

3.65 R\$1327

Hoje às 23:00

Iga Swiatek Kaia Kanepi

R\$670,725

1.27 R\$128370

1.28 R\$20429

4.5 R\$9721

4.6 R\$114

Basquete

Correspondido

1

2

Hoje às 21:10

New Orleans Pelicans Philadelphia 76ers

R\$508,303

5.8 R\$684

6 R\$742

1.2 R\$3712

1.21 R\$6838

Hoje às 21:30

Los Angeles Lakers Brooklyn Nets

R\$152,323

1.67 R\$488

1.68 R\$4213

2.48 R\$1382

2.5 R\$2713

Hoje às 21:10

Charlotte Hornets Toronto Raptors

R\$232,715

2.04 R\$1479

2.14 R\$314

1.88 R\$232

1.96 R\$1070

Nesta aula, serão utilizados como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

Nesta aula, serão utilizados como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.

Nesta aula, serão utilizados como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.
- Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.

Nesta aula, serão utilizados como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.
- Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.
- Baker, R. D., McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is... *Journal of the Royal Statistical Society Series A* 178, n. 2, pp. 481-492.

O modelo proposto por Maher (1982) tem sido utilizado, com adaptações e variações, como um dos principais para o cálculo de probabilidades de resultados de jogos de futebol.



O modelo proposto por Maher (1982) tem sido utilizado, com adaptações e variações, como um dos principais para o cálculo de probabilidades de resultados de jogos de futebol.

Pode ser utilizado, também, para o cálculo de probabilidades de resultados de jogos de outros esportes, como basquete, com as devidas adaptações relativas às distribuições de probabilidade das pontuações.



# Bem-vindo ao CHANCE DE GOL

Última atualização: 24/01/2022 às 11:16



## CAMPEONATOS EM ANDAMENTO

### Futebol Brasileiro:

- [Copa do Nordeste 2022](#)

### Futebol Internacional:

- [Liga dos Campeões 2021/2022](#)

### Futebol de Seleções:

- [Eliminatórias da Copa do Mundo 2022 - América do Sul](#)
- [Eliminatórias da Copa do Mundo 2022 - Europa](#)

## PRÓXIMOS CAMPEONATOS

### Futebol Brasileiro:

- Campeonato Paulista 2022
- Campeonato Carioca 2022
- Campeonato Mineiro 2022
- Campeonato Gaúcho 2022

### Futebol Internacional:

- Mundial de Clubes 2021

### Futebol de Seleções:

## RANKINGS CHANCE DE GOL (janeiro/2022)

[Ranking de Seleções](#)



[Ranking Mundial de Clubes](#)



[Ranking de Clubes Brasileiros](#)





Se a probabilidade de um ataque de um time resultar em um gol é  $p$  (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Se a probabilidade de um ataque de um time resultar em um gol é  $p$  (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Nessas circunstâncias, pode-se pensar que a distribuição de Poisson é uma boa aproximação para o número de gols marcados por um time em uma partida.

Se a probabilidade de um ataque de um time resultar em um gol é  $p$  (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Nessas circunstâncias, pode-se pensar que a distribuição de Poisson é uma boa aproximação para o número de gols marcados por um time em uma partida.

Esse foi o ponto de partida para o modelo sugerido por Maher (1987).

Obviamente, cada time terá um parâmetro da Poisson (média do número de gols em uma partida) próprio, a depender da qualidade do time.

Obviamente, cada time terá um parâmetro da Poisson (média do número de gols em uma partida) próprio, a depender da qualidade do time.

Maher considerou que, ao avaliar confrontos entre dois times, o número de gols marcados de um time é independente do número de gols marcados pelo outro time.



Obviamente, cada time terá um parâmetro da Poisson (média do número de gols em uma partida) próprio, a depender da qualidade do time.

Maher considerou que, ao avaliar confrontos entre dois times, o número de gols marcados de um time é independente do número de gols marcados pelo outro time.

Problema?

*"All models are wrong, but some are useful"* (George Box)

$X_{ij}$ : v.a. número de gols do time  $i$  (mandante) jogando contra o time  $j$ .

$X_{ij}$ : v.a. número de gols do time  $i$  (mandante) jogando contra o time  $j$ .

$Y_{ij}$ : v.a. número de gols do time  $j$  jogando contra o time  $i$  (mandante).

$X_{ij}$ : v.a. número de gols do time  $i$  (mandante) jogando contra o time  $j$ .

$Y_{ij}$ : v.a. número de gols do time  $j$  jogando contra o time  $i$  (mandante).

Placar observado:  $(x_{ij}, y_{ij})$ .

$X_{ij}$ : v.a. número de gols do time  $i$  (mandante) jogando contra o time  $j$ .

$Y_{ij}$ : v.a. número de gols do time  $j$  jogando contra o time  $i$  (mandante).

Placar observado:  $(x_{ij}, y_{ij})$ .

Por suposição,  $X_{ij}$  e  $Y_{ij}$  são independentes.

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$$



$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$$

$\alpha_i$ : parâmetro de força do ataque do time  $i$  quando é mandante.

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$$

$\alpha_i$ : parâmetro de força do ataque do time  $i$  quando é mandante.

$\beta_j$ : parâmetro de fragilidade da defesa do time  $j$  quando não é mandante.

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$$

$\alpha_i$ : parâmetro de força do ataque do time  $i$  quando é mandante.

$\beta_j$ : parâmetro de fragilidade da defesa do time  $j$  quando não é mandante.

$\gamma_i$ : parâmetro de fragilidade da defesa do time  $i$  quando é mandante.

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$$

$\alpha_i$ : parâmetro de força do ataque do time  $i$  quando é mandante.

$\beta_j$ : parâmetro de fragilidade da defesa do time  $j$  quando não é mandante.

$\gamma_i$ : parâmetro de fragilidade da defesa do time  $i$  quando é mandante.

$\delta_j$ : parâmetro de força do ataque do time  $j$  quando não é mandante.

Obviamente, a média da distribuição  $Poisson(\alpha_i \beta_j)$  continua sendo a mesma se  $\alpha_i$  for multiplicado por uma constante  $k$ , e se  $\beta_j$  for dividido pela mesma constante  $k$ .

Obviamente, a média da distribuição  $Poisson(\alpha_i \beta_j)$  continua sendo a mesma se  $\alpha_i$  for multiplicado por uma constante  $k$ , e se  $\beta_j$  for dividido pela mesma constante  $k$ .

Logo, há infinitos valores possíveis para os parâmetros que levam ao mesmo modelo, o que gera um problema de identificabilidade.

Obviamente, a média da distribuição  $Poisson(\alpha_i \beta_j)$  continua sendo a mesma se  $\alpha_i$  for multiplicado por uma constante  $k$ , e se  $\beta_j$  for dividido pela mesma constante  $k$ .

Logo, há infinitos valores possíveis para os parâmetros que levam ao mesmo modelo, o que gera um problema de identificabilidade.

Sugestão: aplicar as restrições  $\sum \alpha_i = \sum \beta_j$  e  $\sum \gamma_i = \sum \delta_j$ .

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será



Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underline{\alpha}, \underline{\beta}) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!))).$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underline{\alpha}, \underline{\beta}) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!))).$$

Sua derivada será

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\alpha, \beta) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!))).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left( -\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\alpha, \beta) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!))).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left( -\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \text{ e } \hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\alpha, \beta) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!))).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left( -\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \text{ e } \hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.$$

De maneira similar,  $\gamma$  e  $\delta$  podem ser determinados a partir de  $y_{ij}$ .

	mandante	alpha	beta_emv\$beta	gamma_emv\$gamma	delta_emv\$delta	posicao
4	AtleticoMG	1.9979328	0.9945181	0.7374696	1.3013590	1
10	Flamengo	1.4746647	0.9704690	0.8853876	1.9934145	2
16	Palmeiras	1.5563296	0.7885939	1.4749392	1.3533105	3
12	Fortaleza	1.2503475	1.1892053	1.0557144	0.9527619	4
8	Corinthians	1.2288361	0.8224867	0.9885545	0.7384121	5
17	RedBullBragantino	1.3465280	1.1944599	1.1414033	1.4356545	6
11	Fluminense	1.1491378	1.0926698	0.7688758	0.7299544	7
1	AmericaMG	1.0579863	1.1788335	0.6130007	0.9826660	8
3	AtleticoGO	0.8104413	0.9864035	0.7733590	0.8344286	9
19	SaoPaulo	0.8618534	1.0785628	0.8214145	0.6796766	10
6	Ceara	1.1491378	1.0926698	0.7711109	0.7821851	11
14	Internacional	1.3765455	0.9200842	1.2117456	0.8005802	12
18	Santos	1.0098960	1.1762688	0.7688758	0.7299544	13
2	AthleticoPR	1.1012571	1.0902931	1.1668422	0.9583994	14
9	Cuiaba	0.8489071	0.7635377	1.1047986	0.8491055	15
15	Juventude	0.9197532	1.3066633	0.8310217	0.8892546	16
13	Gremio	1.4069328	1.3820583	1.1566663	0.7982336	17
5	Bahia	1.3070177	1.3298684	1.2117456	0.8005802	18
20	Sport	0.6224497	1.0670822	0.7078013	0.5717150	19
7	Chapecoense	0.6377330	1.5572197	1.7574303	0.7696229	20

Obviamente, os parâmetros são estáticos, e representam a força “média” dos times ao longo de um período (no caso, ao longo de um campeonato completo).

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Palmeiras (mandante) e Corinthians.



Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Palmeiras (mandante) e Corinthians.

$$\lambda_{PAL} = 1.56 \times 0.82 = 1.28$$

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Palmeiras (mandante) e Corinthians.

$$\lambda_{PAL} = 1.56 \times 0.82 = 1.28$$

$$\lambda_{COR} = 1.47 \times 0.74 = 1.09$$

Para o cálculo de probabilidades:

Para o cálculo de probabilidades:

$$\begin{aligned} P(A > B) &= \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k) \\ &= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left( \sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}. \end{aligned}$$

Para o cálculo de probabilidades:

$$\begin{aligned} P(A > B) &= \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k) \\ &= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left( \sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}. \end{aligned}$$

Prob. de vitória do Palmeiras: 0.41

Para o cálculo de probabilidades:

$$\begin{aligned} P(A > B) &= \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k) \\ &= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left( \sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}. \end{aligned}$$

Prob. de vitória do Palmeiras: 0.41

Prob. de empate: 0.28

Para o cálculo de probabilidades:

$$\begin{aligned} P(A > B) &= \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k) \\ &= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left( \sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}. \end{aligned}$$

Prob. de vitória do Palmeiras: 0.41

Prob. de empate: 0.28

Prob. de vitória do Corinthians: 0.31

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.



Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

Na parametrização anterior, se  $n$  times são considerados, o modelo possui  $2(2n - 1)$  parâmetros.

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

Na parametrização anterior, se  $n$  times são considerados, o modelo possui  $2(2n - 1)$  parâmetros.

Agora, o modelo possuirá  $2n$  parâmetros.

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos  $\eta$ .

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos  $\eta$ .

O mesmo vale para a fragilidade da defesa, e a mesma constante  $\eta$  é considerada.

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos  $\eta$ .

O mesmo vale para a fragilidade da defesa, e a mesma constante  $\eta$  é considerada.

Ou seja, considera-se que  $\delta_i = \eta\alpha_i$  e  $\gamma_j = \eta\beta_j$ .

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos  $\eta$ .

O mesmo vale para a fragilidade da defesa, e a mesma constante  $\eta$  é considerada.

Ou seja, considera-se que  $\delta_i = \eta\alpha_i$  e  $\gamma_j = \eta\beta_j$ .

Nessa parametrização,  $\eta$  tipicamente tem valor menor que um.

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$



$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\eta^2 \alpha_j \beta_i)$$

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\eta^2 \alpha_j \beta_i)$$

Estimadores de máxima verossimilhança:

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\eta^2 \alpha_j \beta_i)$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} (x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{j \neq i} \hat{\beta}_j},$$

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\eta^2 \alpha_j \beta_i)$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} (x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{j \neq i} \hat{\beta}_j},$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j} (x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{i \neq j} \hat{\alpha}_i}, \text{ e}$$

$$X_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$$

$$Y_{ij} \sim \text{Poisson}(\eta^2 \alpha_j \beta_i)$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} (x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{j \neq i} \hat{\beta}_j},$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j} (x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{i \neq j} \hat{\alpha}_i}, \text{ e}$$

$$\hat{\eta}^2 = \frac{\sum_i \sum_{j \neq i} y_{ij}}{\sum_i \sum_{j \neq i} x_{ij}}.$$

O ajuste é surpreendentemente bom para um modelo tão simples!

O ajuste é surpreendentemente bom para um modelo tão simples!

Principalmente levando-se em conta que o número de gols de uma equipe em uma partida não deve ser independente do número de gols da outra equipe.

O ajuste é surpreendentemente bom para um modelo tão simples!

Principalmente levando-se em conta que o número de gols de uma equipe em uma partida não deve ser independente do número de gols da outra equipe.

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.



$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$

$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$

$$Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$$

$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$
$$Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$$

(aqui, a parametrização considera  $\eta$  como o efeito de se jogar como mandante, e não como visitante).

$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$
$$Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$$

(aqui, a parametrização considera  $\eta$  como o efeito de se jogar como mandante, e não como visitante).

Para simplificação de notação, considerando  $\lambda$  e  $\mu$  como os parâmetros das distribuições de tal modo que  $\log \lambda = \alpha_i\beta_j\eta$  e  $\log \mu = \alpha_j\beta_i$ , tem-se:

$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$
$$Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$$

(aqui, a parametrização considera  $\eta$  como o efeito de se jogar como mandante, e não como visitante).

Para simplificação de notação, considerando  $\lambda$  e  $\mu$  como os parâmetros das distribuições de tal modo que  $\log \lambda = \alpha_i\beta_j\eta$  e  $\log \mu = \alpha_j\beta_i$ , tem-se:

$$P(X_{ij} = x, Y_{ij} = y) = \tau_{\lambda,\mu}(x, y) \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^y e^{-\mu}}{y!}$$

$$X_{ij} \sim \text{Poisson}(\eta\alpha_i\beta_j),$$
$$Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$$

(aqui, a parametrização considera  $\eta$  como o efeito de se jogar como mandante, e não como visitante).

Para simplificação de notação, considerando  $\lambda$  e  $\mu$  como os parâmetros das distribuições de tal modo que  $\log \lambda = \alpha_i\beta_j\eta$  e  $\log \mu = \alpha_j\beta_i$ , tem-se:

$$P(X_{ij} = x, Y_{ij} = y) = \tau_{\lambda,\mu}(x, y) \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^y e^{-\mu}}{y!}$$

Condição de identificabilidade:  $\sum \alpha_i = n$ .

O modelo tem o acréscimo do parâmetro  $\tau$ , que é justamente o que regula a dependência entre os times.

O modelo tem o acréscimo do parâmetro  $\tau$ , que é justamente o que regula a dependência entre os times.

$$\tau_{\lambda,\mu}(x,y) = \begin{cases} 1 - \lambda\mu\rho, & \text{se } x = y = 0 \\ 1 + \lambda\rho, & \text{se } x = 0, y = 1 \\ 1 + \mu\rho, & \text{se } x = 1, y = 0 \\ 1 - \rho, & \text{se } x = y = 1 \\ 1, & \text{caso contrário.} \end{cases}$$



O modelo tem o acréscimo do parâmetro  $\tau$ , que é justamente o que regula a dependência entre os times.

$$\tau_{\lambda,\mu}(x,y) = \begin{cases} 1 - \lambda\mu\rho, & \text{se } x = y = 0 \\ 1 + \lambda\rho, & \text{se } x = 0, y = 1 \\ 1 + \mu\rho, & \text{se } x = 1, y = 0 \\ 1 - \rho, & \text{se } x = y = 1 \\ 1, & \text{caso contrário.} \end{cases}$$

Quando  $\rho = 0$ , há independência. Considera-se uma estrutura de dependência quando  $x \leq 1$  e  $y \leq 1$ .

Como  $\log \lambda = \alpha_i \beta_j \eta$  e  $\log \mu = \alpha_j \beta_i$ , então

Como  $\log \lambda = \alpha_i \beta_j \eta$  e  $\log \mu = \alpha_j \beta_i$ , então

$\lambda = \exp(\alpha_i + \beta_j + \eta)$  e  $\mu = \exp(\alpha_j + \beta_i)$ .

Como  $\log \lambda = \alpha_i \beta_j \eta$  e  $\log \mu = \alpha_j \beta_i$ , então

$$\lambda = \exp(\alpha_i + \beta_j + \eta) \text{ e } \mu = \exp(\alpha_j + \beta_i).$$

Excluindo-se o efeito de se jogar em casa ( $\eta$ ), se  $\lambda = \mu$ , os números esperados de gols dos times  $i$  e  $j$  são iguais. Logo, pode-se pensar que os times têm forças iguais se  $\alpha_i + \beta_j = \alpha_j + \beta_i$ .

Como  $\log \lambda = \alpha_i \beta_j \eta$  e  $\log \mu = \alpha_j \beta_i$ , então

$\lambda = \exp(\alpha_i + \beta_j + \eta)$  e  $\mu = \exp(\alpha_j + \beta_i)$ .

Excluindo-se o efeito de se jogar em casa ( $\eta$ ), se  $\lambda = \mu$ , os números esperados de gols dos times  $i$  e  $j$  são iguais. Logo, pode-se pensar que os times têm forças iguais se  $\alpha_i + \beta_j = \alpha_j + \beta_i$ .

Se  $\alpha_i + \beta_j > \alpha_j + \beta_i$ , considera-se que o time  $i$  é mais forte que o time  $j$ , ou, equivalentemente, se  $\alpha_i - \beta_i > \alpha_j - \beta_j$ .

Como  $\log \lambda = \alpha_i \beta_j \eta$  e  $\log \mu = \alpha_j \beta_i$ , então

$\lambda = \exp(\alpha_i + \beta_j + \eta)$  e  $\mu = \exp(\alpha_j + \beta_i)$ .

Excluindo-se o efeito de se jogar em casa ( $\eta$ ), se  $\lambda = \mu$ , os números esperados de gols dos times  $i$  e  $j$  são iguais. Logo, pode-se pensar que os times têm forças iguais se  $\alpha_i + \beta_j = \alpha_j + \beta_i$ .

Se  $\alpha_i + \beta_j > \alpha_j + \beta_i$ , considera-se que o time  $i$  é mais forte que o time  $j$ , ou, equivalentemente, se  $\alpha_i - \beta_i > \alpha_j - \beta_j$ .

Logo,  $\alpha_i - \beta_i$  pode ser usado como medida de força do time  $i$ .

Time	Ataque	Defesa	Força	Classif. Força	Classif. Campeonato
AtleticoMG	1.50	-1.30	2.80	1	1
Flamengo	1.53	-1.24	2.77	2	2
Palmeiras	1.37	-1.08	2.44	3	3
RedBullBragantino	1.31	-1.01	2.33	4	6
Corinthians	0.98	-1.28	2.26	5	5
AmericaMG	1.01	-1.25	2.25	6	8
Internacional	1.08	-1.12	2.20	7	12
Ceara	0.96	-1.22	2.18	8	11
Fluminense	0.93	-1.23	2.16	9	7
Fortaleza	1.09	-1.05	2.14	10	4
Cuiaba	0.82	-1.26	2.08	11	15
AtleticoGO	0.79	-1.29	2.07	12	9
AthleticoPR	1.02	-1.05	2.07	13	14
Santos	0.85	-1.18	2.03	14	10
Gremio	1.09	-0.92	2.02	15	17
Bahia	1.05	-0.92	1.97	16	18
Juventude	0.88	-1.08	1.97	17	16
SaoPaulo	0.73	-1.21	1.93	18	13
Sport	0.39	-1.26	1.65	19	19
Chapecoense	0.62	-0.67	1.29	20	20

Por exemplo, Atlético-MG jogando em casa contra o São Paulo.



Por exemplo, Atlético-MG jogando em casa contra o São Paulo.

$$\lambda_{ATL} = \exp(0.32 + 1.50 - 1.21) = 1.84$$

Por exemplo, Atlético-MG jogando em casa contra o São Paulo.

$$\lambda_{ATL} = \exp(0.32 + 1.50 - 1.21) = 1.84$$

$$\mu_{SP} = \exp(0.73 - 1.32) = 0.56$$

Por exemplo, Atlético-MG jogando em casa contra o São Paulo.

$$\lambda_{ATL} = \exp(0.32 + 1.50 - 1.21) = 1.84$$

$$\mu_{SP} = \exp(0.73 - 1.32) = 0.56$$

O valor esperado de gols do Atlético-MG é maior que o do São Paulo.

Para calcular as probabilidades de vitória de cada time e de empate, é preciso utilizar a distribuição de Poisson, além de considerar o parâmetro de dependência.

Para calcular as probabilidades de vitória de cada time e de empate, é preciso utilizar a distribuição de Poisson, além de considerar o parâmetro de dependência.

Probabilidade de vitória do Atlético-MG: 0.68

Probabilidade de vitória do São Paulo: 0.11

Probabilidade de empate: 0.21

Por exemplo, Santos jogando em casa contra o Palmeiras.

Por exemplo, Santos jogando em casa contra o Palmeiras.

$$\lambda_{SAN} = \exp(0.32 + 0.85 - 1.08) = 1.10$$

Por exemplo, Santos jogando em casa contra o Palmeiras.

$$\lambda_{SAN} = \exp(0.32 + 0.85 - 1.08) = 1.10$$

$$\mu_{PAL} = \exp(1.37 - 1.18) = 1.21$$



Por exemplo, Santos jogando em casa contra o Palmeiras.

$$\lambda_{SAN} = \exp(0.32 + 0.85 - 1.08) = 1.10$$

$$\mu_{PAL} = \exp(1.37 - 1.18) = 1.21$$

O valor esperado de gols do Palmeiras é maior que o do Santos, mesmo o Santos sendo o time mandante, devido à diferença dos parâmetros em favor do Palmeiras.

Probabilidade de vitória do Santos: 0.33  
Probabilidade de vitória do Palmeiras: 0.38  
Probabilidade de empate: 0.29

Dixon e Coles sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Dixon e Coles sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Essa abordagem é especificamente útil quando se deseja calcular as probabilidades de um determinado jogo que efetivamente irá se realizar.

Dixon e Coles sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Essa abordagem é especificamente útil quando se deseja calcular as probabilidades de um determinado jogo que efetivamente irá se realizar.

Afinal, faz sentido levar em consideração o momento dos times, ao invés de somente uma média de desempenho ao longo de todo o campeonato até aquele momento.

Seja  $\phi(t)$  o peso em função do tempo - quanto mais próximo  $t$  é de zero, maior é o valor da função, que varia entre 0 e 1, indicando que resultados mais recentes têm peso maior.

Seja  $\phi(t)$  o peso em função do tempo - quanto mais próximo  $t$  é de zero, maior é o valor da função, que varia entre 0 e 1, indicando que resultados mais recentes têm peso maior.

Uma escolha possível é:

$$\phi(t) = \begin{cases} 1, & \text{se } t \leq t_0 \\ 0, & \text{se } t > t_0, \end{cases}$$

Seja  $\phi(t)$  o peso em função do tempo - quanto mais próximo  $t$  é de zero, maior é o valor da função, que varia entre 0 e 1, indicando que resultados mais recentes têm peso maior.

Uma escolha possível é:

$$\phi(t) = \begin{cases} 1, & \text{se } t \leq t_0 \\ 0, & \text{se } t > t_0, \end{cases}$$

o que indica que, em um determinado instante  $t$ , todos os resultados a partir do instante  $t_0$  têm pesos iguais no processo de estimação.



Uma abordagem mais interessante é:

Uma abordagem mais interessante é:

$$\phi(t) = \exp(-\xi t),$$

Uma abordagem mais interessante é:

$$\phi(t) = \exp(-\xi t),$$

em que todos os resultados anteriores, com pesos que decaem exponencialmente de acordo com o parâmetro  $\xi > 0$ , são considerados no processo de estimação.

Uma abordagem mais interessante é:

$$\phi(t) = \exp(-\xi t),$$

em que todos os resultados anteriores, com pesos que decaem exponencialmente de acordo com o parâmetro  $\xi > 0$ , são considerados no processo de estimação.

Se  $\xi = 0$ , todos os resultados anteriores têm peso 1, que é justamente a abordagem que vínhamos utilizando.

Uma abordagem mais interessante é:

$$\phi(t) = \exp(-\xi t),$$

em que todos os resultados anteriores, com pesos que decaem exponencialmente de acordo com o parâmetro  $\xi > 0$ , são considerados no processo de estimação.

Se  $\xi = 0$ , todos os resultados anteriores têm peso 1, que é justamente a abordagem que vínhamos utilizando.

Quanto maior é o valor de  $\xi$ , maior é o peso das partidas mais recentes em relação às partidas anteriores.

A determinação de  $\xi$  pode ser feita em termos de otimização de previsão de resultados. É um processo trabalhoso e não será discutido aqui.

A determinação de  $\xi$  pode ser feita em termos de otimização de previsão de resultados. É um processo trabalhoso e não será discutido aqui.

Na abordagem utilizada neste curso, em que  $t$  é definido em dias, valores razoáveis para  $\xi$  variam entre 0.0015 e 0.0025, dependendo das características dos times envolvidos.

Utilizando  $\xi = 0.0018$  para o Campeonato Brasileiro, temos o seguinte cenário na última rodada:



Utilizando  $\xi = 0.0018$  para o Campeonato Brasileiro, temos o seguinte cenário na última rodada:

Time	Ataque	Defesa	Força	Classif. Força	Classif. Campeonato
Flamengo	1.56	-1.47	3.03	1	2
AtleticoMG	1.55	-1.39	2.93	2	1
AmericaMG	1.12	-1.48	2.60	3	8
Corinthians	1.13	-1.32	2.45	4	5
Palmeiras	1.32	-1.05	2.38	5	3
Internacional	1.15	-1.18	2.32	6	12
Gremio	1.30	-0.98	2.27	7	17
RedBullBragantino	1.24	-1.01	2.25	8	6
Ceara	0.99	-1.25	2.24	9	11
Fluminense	1.00	-1.24	2.24	10	7
SaoPaulo	0.82	-1.40	2.22	11	13
Cuiaba	0.68	-1.49	2.17	12	15
Juventude	1.05	-1.11	2.16	13	16
Bahia	0.95	-1.16	2.11	14	18
AtleticoGO	0.77	-1.32	2.09	15	9
Santos	0.75	-1.31	2.06	16	10
AthleticoPR	0.71	-1.17	1.87	17	14
Sport	0.68	-1.18	1.86	18	19
Fortaleza	0.70	-0.95	1.65	19	4
Chapecoense	0.55	-0.72	1.27	20	20

Com essa abordagem, é possível até mesmo elaborar um modelo em que os parâmetros variam ao longo do tempo

Com essa abordagem, é possível até mesmo elaborar um modelo em que os parâmetros variam ao longo do tempo

Por exemplo, a partir da segunda rodada do segundo turno, considerar os últimos 20 jogos de cada time para a estimação dos parâmetros, para que se calcule as probabilidades dos resultados da rodada seguinte.

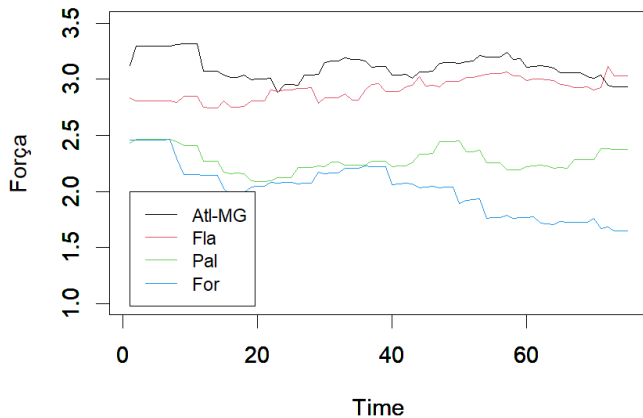
Para o Campeonato Brasileiro, temos o seguinte cenário na última rodada:

Para o Campeonato Brasileiro, temos o seguinte cenário na última rodada:

Time	Ataque	Defesa	Força	Classif. Força	Classif. Campeonato
Flamengo	1.56	-1.47	3.03	1	2
AtleticoMG	1.55	-1.39	2.93	2	1
AmericaMG	1.12	-1.48	2.60	3	8
Corinthians	1.13	-1.32	2.45	4	5
Palmeiras	1.32	-1.05	2.38	5	3
Internacional	1.15	-1.18	2.32	6	12
Gremio	1.30	-0.98	2.27	7	17
RedBullBragantino	1.24	-1.01	2.25	8	6
Ceara	0.99	-1.25	2.24	9	11
Fluminense	1.00	-1.24	2.24	10	7
SaoPaulo	0.82	-1.40	2.22	11	13
Cuiaba	0.68	-1.49	2.17	12	15
Juventude	1.05	-1.11	2.16	13	16
Bahia	0.95	-1.16	2.11	14	18
AtleticoGO	0.77	-1.32	2.09	15	9
Santos	0.75	-1.31	2.06	16	10
AthleticoPR	0.71	-1.17	1.87	17	14
Sport	0.68	-1.18	1.86	18	19
Fortaleza	0.70	-0.95	1.65	19	4
Chapecoense	0.55	-0.72	1.27	20	20

Comparação da evolução de quatro times:

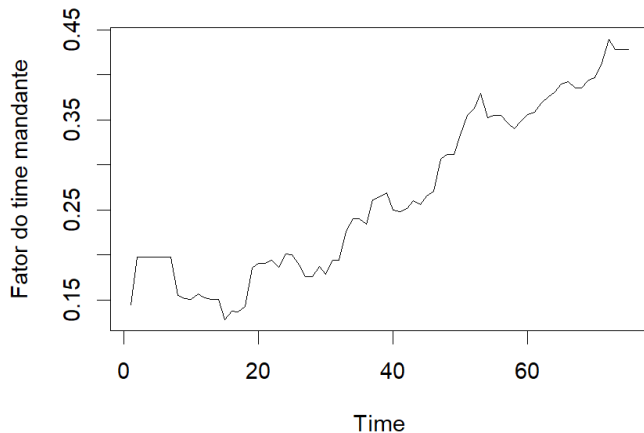
## Comparação da evolução de quatro times:



## Evolução do parâmetro que indica a força do time mandante



## Evolução do parâmetro que indica a força do time mandante



1o. turno

Vitórias de times mandantes: 68 (37,2%)

Empates: 60 (32,8%)

Vitórias de times visitantes: 55 (30,0%)

### 1o. turno

Vitórias de times mandantes: 68 (37,2%)

Empates: 60 (32,8%)

Vitórias de times visitantes: 55 (30,0%)

### 2o. turno

Vitórias de times mandantes: 106 (53,8%)

Empates: 53 (26,9%)

Vitórias de times visitantes: 38 (19,3%)

Métodos para cálculos de probabilidade envolvendo a distribuição de Poisson em jogos de futebol são amplamente utilizados.

Métodos para cálculos de probabilidade envolvendo a distribuição de Poisson em jogos de futebol são amplamente utilizados.

Eventualmente, a distribuição binomial negativa também é utilizada.

mostrar artigos mchale, premier league e selecoes

*Appl. Statist.* (1997)  
**46**, No. 2, pp. 265–280

# Modelling Association Football Scores and Inefficiencies in the Football Betting Market

By MARK J. DIXON<sup>†</sup> and STUART G. COLES

*Lancaster University, UK*

[Received November 1995. Revised September 1996]

## SUMMARY

A parametric model is developed and fitted to English league and cup football data from 1992 to 1995. The model is motivated by an aim to exploit potential inefficiencies in the association football betting market, and this is examined using bookmakers' odds from 1995 to 1996. The technique is based on a Poisson regression model but is complicated by the data structure and the dynamic nature of teams' performances. Maximum likelihood estimates are shown to be computationally obtainable, and the model is shown to have a positive return when used as the basis of a betting strategy.

**Keywords:** Betting strategy; Expected return; Football (soccer); Maximum likelihood; Poisson distribution

*International Journal of Sport Finance*, 2008, 3, 156-166, © 2008 West Virginia University

# “Say It Ain’t So”: Betting-Related Malpractice in Sport

David Forrest<sup>1</sup>, Ian McHale<sup>1</sup>, and Kevin McAuley<sup>1</sup>

<sup>1</sup>University of Salford

David Forrest is a professor of economics at the Centre for the Study of Gambling. His research interests include the economics of sport and the economics of gambling.

Ian McHale is a senior lecturer in statistics at the Centre for the Study of Gambling and the Centre for Operational Research and Applied Statistics. His research interests include statistics in sport and the statistical analysis of gambling issues.

Kevin McAuley is a graduate student at the Centre for the Study of Gambling. His research interests include the efficiency of wagering markets.

## Abstract

The paper identifies key changes in the betting environment that have raised risks to the integrity of sport. The risks are discussed in the context of a model where potential fixers evaluate the costs and benefits to them of engaging in manipulation of events on the field for betting gain. Using this framework, particular markets and situations are indicated as especially susceptible to corruption and these predictions appear to be consistent with the set of cases that have been exposed in contemporary sport. Possible policy responses are discussed.



*The Statistician* (2002)  
51, Part 2, pp. 157–168

## Dynamic modelling and prediction of English Football League matches for betting

Martin Crowder,

*Imperial College of Science, Technology and Medicine, London, UK*

Mark Dixon

*City University, London, UK*

and Anthony Ledford and Mike Robinson

*Man Investment Products Ltd, London, UK*

[Received July 2001. Revised November 2001]

**Summary.** We focus on modelling the 92 soccer teams in the English Football Association League over the years 1992–1997 using refinements of the independent Poisson model of Dixon and Coles. Our framework assumes that each team has attack and defence strengths that evolve through time (rather than remaining constant) according to some unobserved bivariate stochastic process. Estimation of the teams' attack and defence capabilities is undertaken via a novel approach involving an approximation that is computationally convenient and fast. The results of this approximation compare very favourably with results obtained through the Dixon and Coles approach. We note that the full model (i.e. the model before the above approximation is made) may be implemented using Markov chain Monte Carlo procedures, and that this approach is vastly more computationally expensive. We focus on the probabilities of home win, draw or away win because these outcomes constitute the primary betting market. These probabilities are estimated for games played between any two of the 92 teams and the predictions are compared with the actual results.

**Keywords:** Attack and defence strengths; Football results; Poisson models; State space modelling

# Optimal betting under parameter uncertainty: improving the Kelly criterion

Rose D. Baker\* and Ian G. McHale†

Centre for Sports Business, Salford Business School, University of  
Salford, M5 4WT, UK.

March 21, 2013

## Abstract

The Kelly betting criterion ignores uncertainty in the probability of winning the bet, and uses an estimated probability. In general, such replacement of population parameters by sample estimates gives poorer out-of-sample than in-sample performance. We show that to improve out-of-sample performance the size of the bet should be shrunk in the presence of this parameter uncertainty, and compare some estimates of the shrinkage factor. From a simulation study and from an analysis of some tennis betting data we show that the shrunk Kelly approaches developed here offer an improvement over the ‘raw’ Kelly criterion. One approximate estimate of the shrinkage factor gives a ‘back of envelope’ correction to the Kelly criterion

*The European Journal of Finance*  
Vol. 13, No. 8, 751–768, December 2007



## Anyone for Tennis (Betting)?

DAVID FORREST\* & IAN McHALE\*\*

*\*Centre for the Study of Gambling, University of Salford, Salford, Greater Manchester, UK, \*\*Centre for Operational Research and Applied Statistics, University of Salford, Salford, Greater Manchester, UK*

**ABSTRACT** *The most robust anomaly noted in the literature on wagering markets is (positive) longshot bias: over a period of 50 years, it has been well documented in horse betting that higher expected returns accrue to short- than to long-odds bets. However, a few examples of betting markets with zero or negative bias have been found, for example in certain American sports. The understanding of longshot bias is likely to be informed by comparing and contrasting conditions in markets displaying positive, zero, and negative bias but, to date, relatively few markets have been examined. This paper employs a large data set on professional men's tennis matches and a new econometric approach to the estimation of the relationship between returns and odds. It finds positive bias throughout the range of odds. It discusses this finding in the context of the debate on why biases exist and persist in wagering markets, focusing particularly on bettors' attitudes towards risk and skewness.*

**KEY WORDS:** Sports betting, longshot bias, risk preference

### 1. Introduction

Wagering markets are specialist financial markets where bettors purchase state-contingent assets. As in other financial markets, most studies (Sauer (1998) and Vaughan Williams (1999)) provide

O cálculo de probabilidades de resultados de jogos é de muito interesse em diversos países especialmente europeus, principalmente no que se refere ao universo das apostas esportivas.

Afinal de contas, as cotações das apostas são calculadas baseadas em tais probabilidades.

Encontre times, competições, corridas e ...

e-mail/nome de us

senha

☐ Lembrar senha
 ☐ Esqueceu seu N

Exchange

Esportes

Cassino

Slots

Cassino ao vivo

Pôquer

Jogos Exchange

Bingo

Esportes Virtuais

Início

Ao vivo

Cash Out

Promoções

Futebol

Tênis

Cavalos

Galgos

Basquete

Segure esta Bolada

Quero Aprender

Tutoriais

Últimas Notic

Ocultar menu

★ Meus Favoritos

Esportes

Apostas Especiais

Automobilismo

Basquete

Beisebol

Boxe

Ciclismo

Corrida de Cavalos

Corrida de Galgos

Corridas de Cavalos - Hoje

Corridas de Galgos - Hoje

Crquete

Dardos

Desportos de Inverno

E-Sports

Futebol

Futebol Americano

Futebol Australiano

Golfe

Handebol

Hóquei no Gelo

Hóquei no Gelo

Proximas Corridas de Cavalos

Guia Betfair TV

Destaques Esportivos

Futebol

Correspondido

1

X

2

jan 29 14:30

Peterborough Sheff Utd

R\$954,470

3.75 R\$82

3.95 R\$171

3.55 R\$322

3.8 R\$143

2.14 R\$236

2.18 R\$2373

71' 1 0

Envigado Deportivo Pasto

R\$294,594

1.25 R\$2638

1.28 R\$2492

5.2 R\$453

5.8 R\$940

36 R\$127

44 R\$132

jan 29 12:00

Barnsley Bournemouth

R\$246,559

4.8 R\$81

5.2 R\$207

3.6 R\$113

3.85 R\$147

1.88 R\$107

1.92 R\$1105

jan 28 16:45

Huddersfield Stoke

R\$237,423

2.56 R\$111

2.66 R\$634

3.35 R\$46427

3.4 R\$39795

3.05 R\$837

3.2 R\$317

Tênis

Correspondido

1

2

qua 05:30

Felix Auger-Aliassime Daniil Medvedev

R\$1,457,598

7.2 R\$18143

7.4 R\$3469

1.15 R\$398523

1.16 R\$11410

1 2 1 2

Jenson Brooksby Mitchell Krueger

R\$1,273,276

1.38 R\$3430

1.39 R\$5513

3.55 R\$2158

3.65 R\$1327

Hoje às 23:00

Iga Swiatek Kaia Kanepi

R\$670,725

1.27 R\$128370

1.28 R\$20429

4.5 R\$9721

4.6 R\$114

Basquete

Correspondido

1

2

Hoje às 21:10

New Orleans Pelicans Philadelphia 76ers

R\$508,303

5.8 R\$684

6 R\$742

1.2 R\$3712

1.21 R\$6838

Hoje às 21:30

Los Angeles Lakers Brooklyn Nets

R\$152,323

1.67 R\$488

1.68 R\$4213

2.48 R\$1382

2.5 R\$2713

Hoje às 21:10

Charlotte Hornets Toronto Raptors

R\$232,715

2.04 R\$1479

2.14 R\$314

1.88 R\$232

1.96 R\$1070

Peterborough x Sheffield United:

Peterborough x Sheffield United:

Probabilidade de vitória do Peterborough: 0.265

Peterborough x Sheffield United:

Probabilidade de vitória do Peterborough: 0.265

Probabilidade de empate: 0.28



Peterborough x Sheffield United:

Probabilidade de vitória do Peterborough: 0.265

Probabilidade de empate: 0.28

Probabilidade de vitória do Sheffield United: 0.455

$$1/0.265 =$$

$$1/0.265 = 3.77$$

$$1/0.265 = 3.77$$

$$1/0.28 =$$

$$1/0.265 = 3.77$$

$$1/0.28 = 3.57$$

$$1/0.265 = 3.77$$

$$1/0.28 = 3.57$$

$$1/0.455 =$$

$$1/0.265 = 3.77$$

$$1/0.28 = 3.57$$

$$1/0.455 = 2.20$$

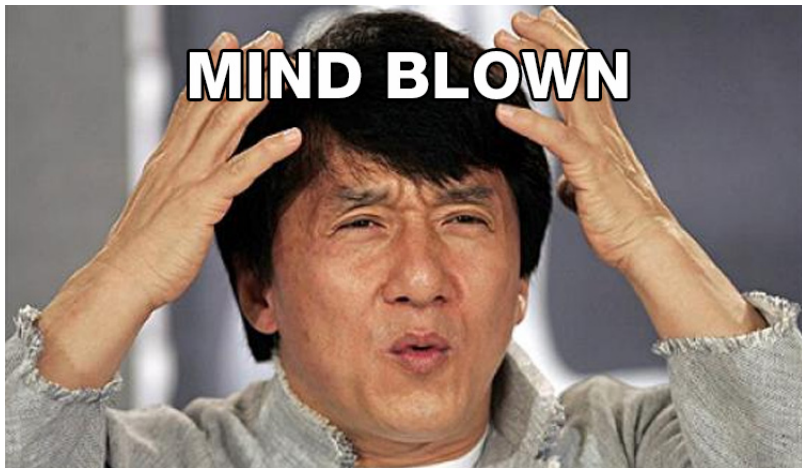
$$1/0.265 = 3.77$$

$$1/0.28 = 3.57$$

$$1/0.455 = 2.20$$

Futebol	Correspondido	1		X		2	
jan 29 14:30	Peterborough Sheff Utd R\$954,470	3.75 R\$82	3.95 R\$171	3.55 R\$322	3.8 R\$143	2.14 R\$236	2.18 R\$2373





O apostador tem como objetivo calcular as probabilidades de maneira mais precisa do que as casas de apostas.

O apostador tem como objetivo calcular as probabilidades de maneira mais precisa do que as casas de apostas.

Dessa maneira, ele pode encontrar cotações que o favoreçam e, em longo prazo, obter lucro.

Suponha que irão jogar Bayern de Munique x Ituano.

Suponha que irão jogar Bayern de Munique x Ituano.

Suponha que o apostador, em sua análise, obteve que:

Prob. vitória do Bayern: 90%

Prob. empate: 5%

Prob. vitória do Ituano: 5%

Suponha que irão jogar Bayern de Munique x Ituano.

Suponha que o apostador, em sua análise, obteve que:

Prob. vitória do Bayern: 90%

Prob. empate: 5%

Prob. vitória do Ituano: 5%

Segundo essa análise, as cotações deveriam ser: Bayern 1.11, empate 20, Ituano 20.

Mas suponha que, ao entrar no site da casa de apostas, ele encontre o seguinte:

Mas suponha que, ao entrar no site da casa de apostas, ele encontre o seguinte:

Bayern 1.04, empate 33, Ituano 100.



Mas suponha que, ao entrar no site da casa de apostas, ele encontre o seguinte:

Bayern 1.04, empate 33, Ituano 100.

Como há uma diferença entre as cotações que ele avalia serem “justas” e as cotações que efetivamente estão disponíveis, ele enxerga uma oportunidade.

De acordo com seus cálculos, em 100 jogos, haverá, em média, 5 vitórias do Ituano.

De acordo com seus cálculos, em 100 jogos, haverá, em média, 5 vitórias do Ituano.

Se, nesses 100 jogos, ele apostar em cada jogo 1 real na vitória do Ituano, ocorrerá o seguinte:

De acordo com seus cálculos, em 100 jogos, haverá, em média, 5 vitórias do Ituano.

Se, nesses 100 jogos, ele apostar em cada jogo 1 real na vitória do Ituano, ocorrerá o seguinte:

Em 90 jogos, o Bayern vencerá e em 5 ocorrerá empate, ou seja, ele perderá 95 reais.

De acordo com seus cálculos, em 100 jogos, haverá, em média, 5 vitórias do Ituano.

Se, nesses 100 jogos, ele apostar em cada jogo 1 real na vitória do Ituano, ocorrerá o seguinte:

Em 90 jogos, o Bayern vencerá e em 5 ocorrerá empate, ou seja, ele perderá 95 reais.

Em 5 jogos, o Ituano irá vencer, e ele ganhará 100 reais.

De acordo com seus cálculos, em 100 jogos, haverá, em média, 5 vitórias do Ituano.

Se, nesses 100 jogos, ele apostar em cada jogo 1 real na vitória do Ituano, ocorrerá o seguinte:

Em 90 jogos, o Bayern vencerá e em 5 ocorrerá empate, ou seja, ele perderá 95 reais.

Em 5 jogos, o Ituano irá vencer, e ele ganhará 100 reais.

Ou seja, ele sairá com lucro de 5 reais.

Perceba que, em apenas um jogo, a chance de ele ter prejuízo é grande, pois ele estará apostando no time com probabilidade pequena de vencer.

Perceba que, em apenas um jogo, a chance de ele ter prejuízo é grande, pois ele estará apostando no time com probabilidade pequena de vencer.

Mas, se seus cálculos estiverem corretos, e ele conseguir identificar essas oportunidades, em longo prazo ele conseguirá lucro.



Perceba que, em apenas um jogo, a chance de ele ter prejuízo é grande, pois ele estará apostando no time com probabilidade pequena de vencer.

Mas, se seus cálculos estiverem corretos, e ele conseguir identificar essas oportunidades, em longo prazo ele conseguirá lucro.

Como se vê, ser lucrativo em apostas esportivas nada tem a ver em apostar no favorito ou na zebra, e sim em saber calcular probabilidades.

Não é fácil identificar essas oportunidades, afinal as casas de apostas trabalham com modelos refinados e complexos para seus cálculos.

Não é fácil identificar essas oportunidades, afinal as casas de apostas trabalham com modelos refinados e complexos para seus cálculos.

Como as probabilidades dependem dos momentos dos times, desfalques, contusões etc, modelos computacionalmente complexos são amplamente empregados

Não é fácil identificar essas oportunidades, afinal as casas de apostas trabalham com modelos refinados e complexos para seus cálculos.

Como as probabilidades dependem dos momentos dos times, desfalques, contusões etc, modelos computacionalmente complexos são amplamente empregados

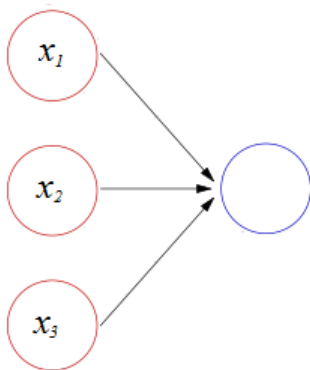
Exemplo: redes neurais.

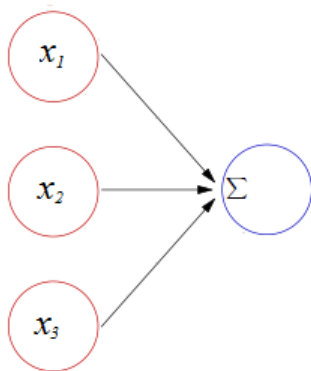
Modelo linear generalizado: exemplo simples de rede neural.

$$x_1$$

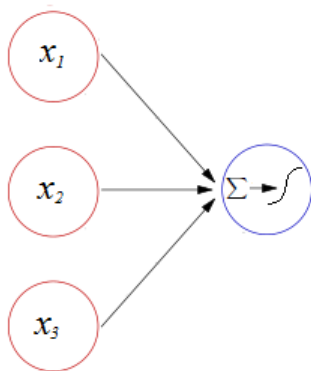
$$x_2$$

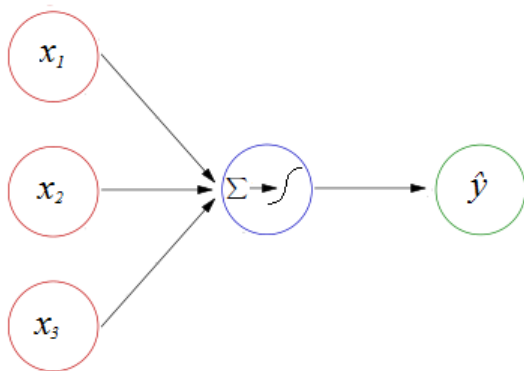
$$x_3$$



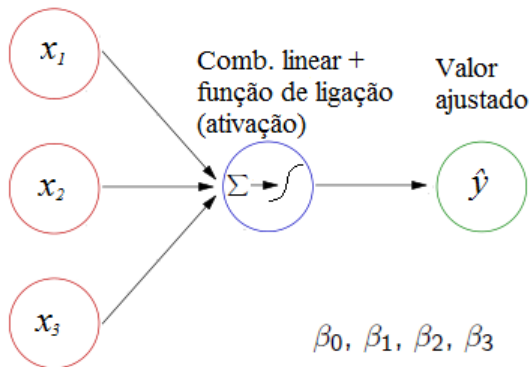








Covariáveis

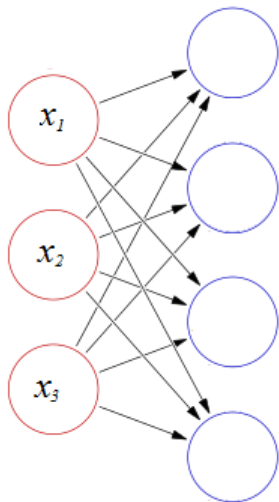


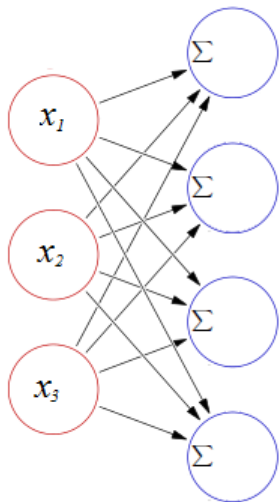
Rede neural: modelo linear generalizado turbinado!

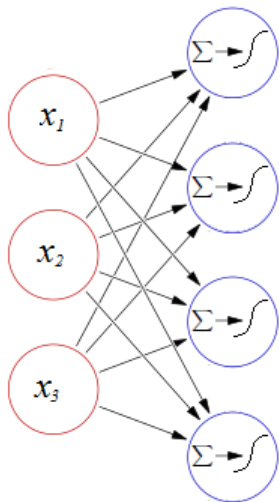
$$x_1$$

$$x_2$$

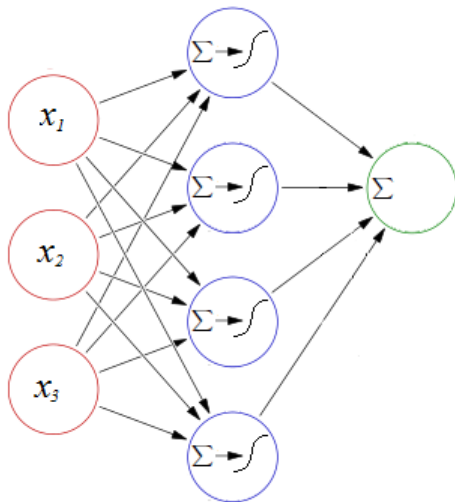
$$x_3$$

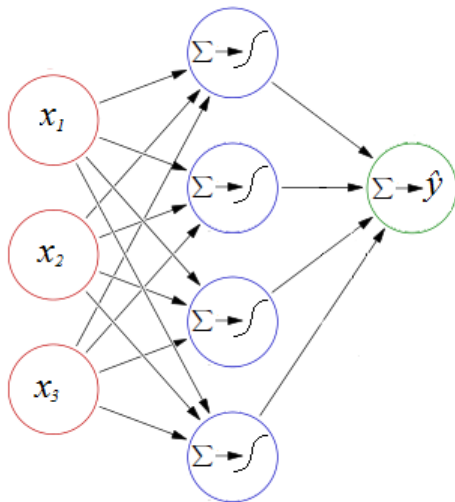


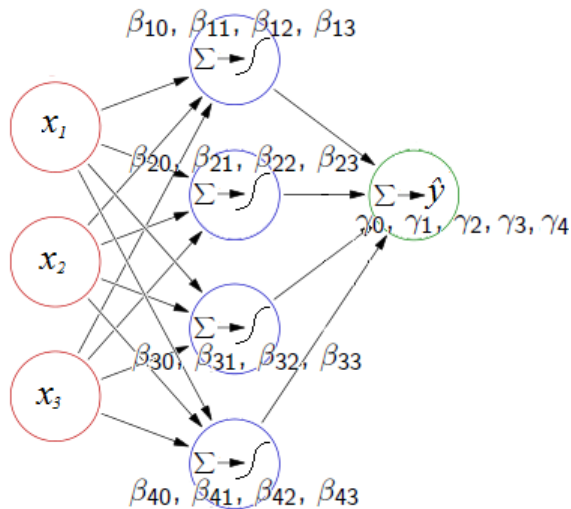


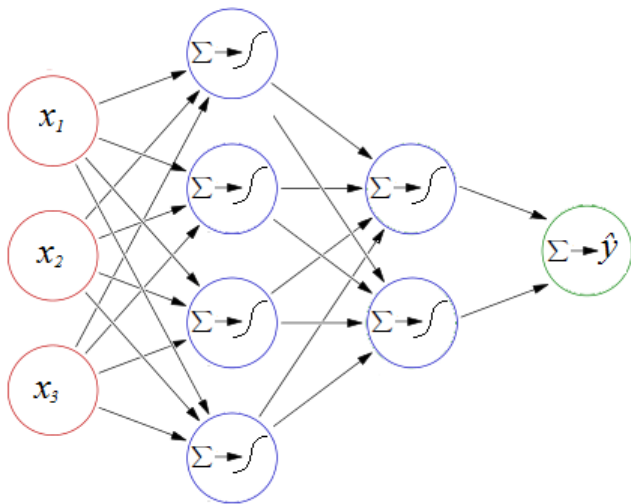






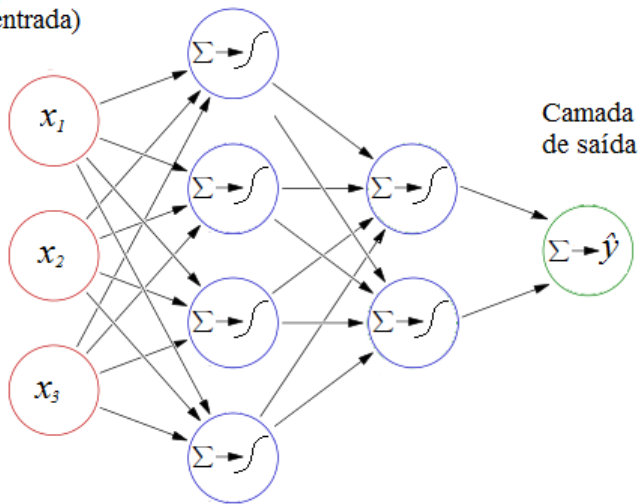






Covariáveis  
(camada de  
entrada)

Camadas intermediárias, seus  
neurônios e funções de ativação



Teorema de Kolmogorov-Nielsen: Dada uma função  $f : [0; 1]^n \rightarrow \mathbb{R}^m$ ,  $f(x) = y$ , existe sempre para  $f$  uma implementação exata com uma rede neural de três camadas com  $n$  entradas, sendo a primeira camada intermediária composta por  $n(2n + 1)$  neurônios, a segunda camada intermediária com  $2n + 1$  neurônios e a camada de saída com  $m$  neurônios representando as  $m$  componentes do vetor  $y$  (Bishop, 1995).

Redes neurais com muitas camadas intermediárias: *deep neural networks*, ou redes neurais profundas.

Redes neurais com muitas camadas intermediárias: *deep neural networks*, ou redes neurais profundas.

Modelos altamente não lineares.



Redes neurais com muitas camadas intermediárias: *deep neural networks*, ou redes neurais profundas.

Modelos altamente não lineares.

Utilizada para classificação, previsão, visão computacional, reconhecimento de imagens, de fala, processamento de linguagem natural etc.

Redes neurais com muitas camadas intermediárias: *deep neural networks*, ou redes neurais profundas.

Modelos altamente não lineares.

Utilizada para classificação, previsão, visão computacional, reconhecimento de imagens, de fala, processamento de linguagem natural etc.

Alta complexidade devido ao grande número de parâmetros.

Redes neurais com muitas camadas intermediárias: *deep neural networks*, ou redes neurais profundas.

Modelos altamente não lineares.

Utilizada para classificação, previsão, visão computacional, reconhecimento de imagens, de fala, processamento de linguagem natural etc.

Alta complexidade devido ao grande número de parâmetros.

Não se sabe exatamente a razão teórica, mas redes com muitas camadas apresentam melhor desempenho do que redes com poucas camadas.

Redes neurais profundas são usadas para selecionar os posts que aparecem no Facebook, para sugerir amigos no Instagram, para escolher as ofertas para você no aplicativo do supermercado.

Redes neurais profundas são usadas para selecionar os posts que aparecem no Facebook, para sugerir amigos no Instagram, para escolher as ofertas para você no aplicativo do supermercado.

Você realmente achou que elas não estariam no esporte?

Redes neurais profundas são usadas para selecionar os posts que aparecem no Facebook, para sugerir amigos no Instagram, para escolher as ofertas para você no aplicativo do supermercado.

Você realmente achou que elas não estariam no esporte?

Casas de aposta alimentam suas redes neurais a todo o tempo com uma imensidão de informações, de modo a terem suas cotações calculadas de modo a não darem margem para os apostadores.



## Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer

Javier Fernández

F.C. Barcelona

javier.fernandezr@fcbarcelona.cat

Luke Bornn

Simon Fraser University, Sacramento

Kings

lbornn@sfu.ca

Dan Cervone

Los Angeles Dodgers

dcervone@gmail.com

### 1. Introduction

What is the right way to think about analytics in soccer? Is the sport about measured events such as passes and goals, possession percentages and traveled distance, or even more abstract notions such as mistakes (to quote Cruyff, "Soccer is a game of mistakes, whoever makes the fewer wins")? Analytical work to date has focused primarily on these more isolated aspects of the sport, while coaches tend to focus on the tactical interplay of all 22 players on the pitch. Soccer analytics is lacking from a comprehensive approach that can start to address performance-related questions that are closer to the language of the game. Questions such as: who adds more value? How and where is this value added? Are the teammates creating spaces of value? When and how should a backward pass be taken? How risky is a team attacking strategy? What is a player's decision-making profile?

In order to make an impact on key decision-makers within the sport, soccer analytics



# Using Deep Learning to Understand Patterns of Player Movement in the NBA

Akhil Nistala, John Gutttag

Track: Basketball

## Abstract

In 2011, SportVU fundamentally changed the way that basketball can be analyzed. STATS SportVU utilized a six-camera system installed in basketball arenas to track the real-time positions of players, at 25 frames per second. In this paper, we demonstrate how we can apply deep learning techniques to this data to produce a queryable database of basketball possessions.

We trained an unsupervised machine learning pipeline that generates a representation, called a *trajectory embedding*, of how individual players move on offense. The representation is a 32-dimensional vector of floating-point numbers that captures the semantics of a single player's movement, such as locations of the endpoints, screen actions, court coverage, and other spatial features. We generated nearly 3 million trajectory-embeddings from three seasons of data (2013-2014, 2014-2015, 2015-2016).

We found that the Euclidean distance between trajectory-embeddings is an excellent indicator of the visual similarity of the movements they encode. For example, two different movements of a post-up in the right block will have nearby embeddings; a post-up in the right block and a screen action above the left wing will have distant embeddings. This result led to the Similar Possessions Finder, a queryable database of basketball possessions.

The Similar Possessions Finder can be used to quickly answer queries such as "How much more frequently did Andre Drummond establish position on the right block than on the left block during the 2015-2016 regular season?" and "Find all possessions from the 2014 playoffs in which Chris Paul ran a screen action in the high post that ended with DeAndre Jordan scoring."

## 1. Introduction

Our goal was to develop an automated framework to quantitatively examine and compare patterns of individual player movements on offense. This process currently entails a film analyst watching hundreds of hours of game footage, carefully examining each possession and taking notes on





## DeepQB: Deep Learning with Player Tracking to Quantify Quarterback Decision-Making & Performance

Brian Burke, ESPN Analytics, [brian.j.burke@espn.com](mailto:brian.j.burke@espn.com)

DeepQB is a proposed application of deep neural networks to player tracking data from over two full seasons of American professional football. This novel approach demonstrates the ability to successfully understand complex aspects of the passing game, most notably quarterback decision-making. It can assess and compare individual quarterback pass target selection based on a snapshot presented to the passer by the receivers and defenders. Assessments of quarterback decision-making are made by comparing actual target selection to that predicted by our model. The model performs well, correctly identifying the targeted receiver in 60% of cross-validated cases. When passers target the predicted receiver, passes are completed 74% of the time, compared to 55% when the QB targets any other receiver. This performance is surprisingly strong, given that the offense often conceals its intent by design, while defenses try not to allow any single receiver to be open. Further, quarterback passing skills separate and apart from his receivers and defense are isolated and assessed by comparing metrics of actual play success to the metrics of success predicted by the situation presented to the passer. This approach represents a new way for teams, media, and fans to understand and quantitatively assess quarterback decision-making, an aspect of the sport which has previously been opaque and inaccessible.

### 1. Introduction

Perhaps the most enigmatic and yet most important player attribute in all of American football is the decision-making abilities of quarterbacks. Although mental abilities are important for every position, quarterback is unique in that psycho-cognitive abilities rival physical abilities in regards to successful performance. Measurable and physical attributes are easily observed through the scouting process, but a professional quarterback's ability to process and exploit highly dynamic information during the course of a play is not well understood. Previously only relatively crude aggregate statistical methods - that cannot fully separate the quarterback's individual impact from those of his teammates, play design, and opponents - have existed to quantify this skill.

Early efforts to exploit football tracking data only scratched the surface of what is possible with such rich information. Typical applications of tracking data merely involved measuring the



## You Cannot Do That Ben Stokes: Dynamically Predicting Shot Type in Cricket Using a Personalized Deep Neural Network

Will Gürpınar-Morgan, Daniel Dinsdale, Joe Gallagher,  
Aditya Cherukumudi & Patrick Lucey

Track: Other Sports  
Paper ID: 1548748

### 1. Introduction

The ability to predict what shot a batsman will attempt given the type of ball and match situation is both one of the most challenging and strategically important tasks in cricket.

The goal of each batsman is to score as many runs as possible without being dismissed. Batsmen can be dismissed in several ways, including being caught by fielders or having their wickets knocked over. While simple in principle, the type of shots and style of a batsman is greatly influenced by the format of the game. In short forms of the game such as T20 and One Day Internationals (the focus of this paper), batsmen are typically more aggressive since their team have a limited number of balls from which to score their runs (120 and 300 balls respectively).

Getting the right batsman vs bowler match-up is of paramount importance. For example, for the fielding team, the choice of bowler against the opposition star batsman could be the key difference between winning or losing. Therefore, the ability to have a predefined playbook (as in the NFL) which would allow a team to predict how best to set their fielders given the context of the game, the batsman they are bowling to and bowlers at their disposal would give them a significant strategic advantage.

***In this paper, we present a personalized deep neural network approach which can predict the probabilities of where a specific batsman will hit a specific bowler and bowl type, in a specific game-scenario.***

As a motivating example let us consider the 2019 Cricket World Cup Final between England and New Zealand, with England leading 0 runs from 2 balls to win. The ball was an attempted "scooper"



## Baseball Predictions and Strategies Using Explainable AI

Joshua Silver (Singularity, [joshua.silver@singularity.com](mailto:joshua.silver@singularity.com))

Tate Huffman (Harvard University, [thuffman@college.harvard.edu](mailto:thuffman@college.harvard.edu))

### 1. Abstract

Over the last decade, Major League Baseball has dramatically increased the amount of data it captures and makes available to the public. Meanwhile, there is a growing technology trend that applies AI to analyze massive amounts of data to build complex models that find relationships and meaning in the data. By marrying these two concepts together, we built a neural-network-based AI model called Singularity-PA (pronounced single-arity-P-A) to solve one of the most fundamental questions in baseball: *How can we predict the outcome of a batter vs. pitcher plate appearance (PA)?*

We show how our model learned even the most subtle rules and strategies of the game, and is able to answer our batter vs. pitcher question with accurate and precise predictions. We demonstrate how to apply techniques to interpret and visualize Singularity-PA's predictions to make the model's rationale understandable, and sharable for humans.

Finally, we provide open-source tools that allow for the readers' own experimentation and analyses.

### 2. Introduction and Motivation

Recently, new technologies and databases have rapidly increased the amount of available data in baseball. We now have many new statistics at our disposal. For instance, a single pitch at the major league level contains measurements for approximately 90 different parameters, with such varied data points as pitch information (including velocity, spin rates, and ball movement), batted-ball information (including exit velocity and launch angle) and fielder-position information. Baseball domain experts wishing to improve their abilities to forecast outcomes have pored over this data, looking to answer questions such as "How does a pitched ball spin rate affect pitcher performance?"



## Making Offensive Play Predictable - Using a Graph Convolutional Network to Understand Defensive Performance in Soccer

Michael Stöckl, Thomas Seidl, Daniel Marley & Paul Power | Stats Perform

### 1. Introduction

#### 1.1 Measuring defensive quality in soccer

The art of good defending is to prevent something from happening before it has even happened. Virgil Van Dijk is considered one of the best defenders in world soccer as he has the ability to prevent a pass being made to an open attacker to shoot by forcing the ball carrier to pass somewhere else less dangerous. However, while we know this is great defending, in today's stats, Van Dijk would not receive any acknowledgement. A defender's contribution is simply measured by the number of tackles or interceptions they make. But what if we were able to measure actions that have been prevented before they were made?

The aim of a defense and a defender is to make offensive play predictable. For example, Jürgen Klopp's Liverpool, press the opposition with the aim of forcing them to give the ball away in specific areas of the pitch by limiting the number of passing options available in dangerous areas. If the art of good defending is to make play predictable, then it should be measurable. Given enough data, we should be able to predict where a player will pass the ball, the likelihood of that pass being completed and whether this pass will result in a scoring opportunity. It therefore stands that we should be able to measure if a defender forces an attacker to change their mind or to prevent an attacker from even becoming an option.

Figure 1 shows a situation from a match between Liverpool vs Bayern Munich in the 2018/19 UEFA Champions League that leads to Mané (red 10) scoring. Our model identifies that Milner (red 7) is the primary target for Van Dijk (red 4) in the first instance. However, due to the combination of Gnabry (blue 22) closing down Milner, Lewandowski (blue 9) closing down Van Dijk and Mané making an *active run*, behind the defence, Mané becomes both the most likely receiver and a high threat for scoring. This demonstrates our ability to model how players decision making is influenced and how a situation can move from low threat to high threat by the off-ball actions of attackers and defenders. [\[LINK TO VIDEO\]](#).



Alan Couzens, M.Sc. (Sports Science)

*"Devoted to the science of Maximal Athletic Development"*

## Why Neural Networks are better than the old Banister/TSS model at predicting athletic performance.

Alan Couzens, M.Sc.(Sports Science)

July 26, 2018

I received a lot of follow up qu's/discussion from this tweet on how I'm seeing [Neural Networks](#) consistently out-perform the [Banister model](#) (the model behind [Training Peaks' Performance Management Chart](#)) as a performance predictor for the vast majority of athletes....

Nesta aula, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Nesta aula, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Tais cálculos podem ser utilizados para mensurar as chances de seu time ser campeão ou ser rebaixado, que tanto aparecem na imprensa ao final dos campeonatos.

Nesta aula, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Tais cálculos podem ser utilizados para mensurar as chances de seu time ser campeão ou ser rebaixado, que tanto aparecem na imprensa ao final dos campeonatos.

Também são úteis no contexto de apostas esportivas, nas quais redes neurais profundas também têm um papel muito importante.





Baker, R. D., McHale, I. G. (2013). Optimal betting under parameter uncertainty: Improving the Kelly criterion. *Decision Analysis*, 10(3), pp. 189-199.



Baker, R. D., McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is... *Journal of the Royal Statistical Society Series A* 178, n. 2, pp. 481-492.



Baker, R. D., McHale, I. G. (2018). Time-varying ratings for international football teams. *European Journal of Operational Research*, 267(2), pp. 659-666.



Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press



Crowder, M., Dixon, M., Ledford, A., Robinson, M. (2002). Dynamic modelling and prediction of English Football League matches for betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2), pp. 157-168.



Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.



Forrest, D., McHale, I. (2007). Anyone for tennis (betting)? *The European Journal of Finance*, 13(8), pp. 751-768.



Forrest, D., McHale, I., McAuley, K. (2008). "Say it ain't so": Betting-related malpractice in sport. *International Journal of Sport Finance*, 3(3), pp. 156-166.



Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.