

**Quem será o próximo campeão
brasileiro?**

Modelos de previsão em esportes

Problema

- Estimar probabilidades de campeonato, classificação para Libertadores e rebaixamento com base no retrospecto

Idéia básica

- Resultados dos jogos modelados por variáveis aleatórias, com distribuição inferida dos resultados anteriores
- Probabilidades estimadas via simulação.

Pergunta inicial

- Que distribuição é apropriada para modelar a ocorrência de gols nas partidas?
 - Colwell and Gillett (**The Mathematical Gazette**, 1981): distribuição de Poisson

$$P(k \text{ gols}) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

onde λ é a média de gols por partida

Brasileiro 2011

- 320 jogos, 640 amostras, 859 gols
- Número médio de gols por equipe por partida:
1,3422

Brasileiro 2011

- 320 jogos, 640 amostras, 859 gols
- Número médio de gols por equipe por partida: 1,3422
- Frequências observadas

Gols	0	1	2	3	4	5	>5
Freq. abs.	167	221	154	68	24	6	0
Freq. rel.	0.2609	0.3453	0.2406	0.1063	0.0375	0.0094	0
Poisson	0.2613	0.3507	0.2353	0.1053	0.0353	0.0095	0.0026

Incorporando retrospecto

- Desempenho de cada time descrito por 4 parâmetros (AC, AF, DC, DF) que descrevem poder de ataque e defesa, em casa e fora de casa
- No confronto $T_i \times T_k$
Gols de $T_i \sim \text{Poisson}(AC_i/DF_k)$
Gols de $T_k \sim \text{Poisson}(AF_k/DC_i)$

Incorporando retrospecto

- Valores de (AC_i, AF_i, DC_i, DF_i) estimados por máxima verossimilhança, a partir dos resultados das partidas, com os valores de DC e DF normalizados de modo a ter média 1

Função de verossimilhança

$$L = \prod_{\text{jogos}} (L_{\text{jogo}})^{w_{\text{jogo}}}$$

$$L_{i \times k} = (AC_i / DF_k)^{g_i} e^{-AC_i / DF_k} (AF_k / DC_i)^{g_i} e^{-AF_k / DC_i}$$

w_{jogo} = importância (jogos mais recentes são mais importantes)

Função de verossimilhança

$$\log L = \sum_{\text{jogos}} w_{\text{jogo}} \log(L_{\text{jogo}})$$

$$\log(L_{i \times k}) = g_i \log(AC_i) - g_i \log(DF_k) + g_k \log(AF_k) - g_k \log(DC_i) \\ - \frac{AC_i}{DF_k} - \frac{AF_k}{DC_i}$$

Maximizando a verossimilhança

- Anular derivadas parciais em relação a cada parâmetro
- Processo iterativo: estimar, alternadamente, parâmetros de defesa e ataque.
- Iniciar com $DC_i = DF_i = 1$, para todo i .

Mais detalhes...

$$L = \sum_{i \times k} w_{i \times k} \left(g_i \log(AC_i) - g_i \log(DF_k) + g_k \log(AF_k) - g_k \log(DC_i) - \frac{AC_i}{DF_k} - \frac{AF_k}{DC_i} \right)$$

$$\frac{\partial L}{\partial AC_i} = \sum_k w_{i \times k} \left(\frac{g_i}{AC_i} - \frac{1}{DF_k} \right)$$

$$\frac{\partial L}{\partial AC_i} = 0 \Leftrightarrow AC_i = \frac{\sum_k w_{i \times k} g_i}{\sum_k w_{i \times k} / DF_k}$$

$$\frac{\partial L}{\partial DF_k} = \sum_i w_{i \times k} \left(-\frac{g_i}{DF_k} + \frac{AC_i}{(DF_k)^2} \right)$$

$$\frac{\partial L}{\partial DF_k} = 0 \Leftrightarrow DF_k = \frac{\sum_i w_{i \times k} AC_i}{\sum_i w_{i \times k} g_i}$$

Resultados

- Aplicativo em <http://simulfutebol.appspot.com>

O que mais?

- Modelos mais apropriados (por exemplo, limitar a vantagem em gols)
- Estudo histórico dos modelos
- Intervalos de confiança
- Dissertação(ões) de mestrado?
- Manter página na Internet