



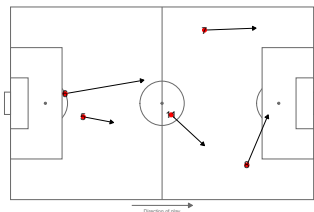
Statistics in the football analytics: a composite model for the probability of scoring a goal

Danny Paganin University of Padua
Bruno Scarpa, Univeristy of Padua

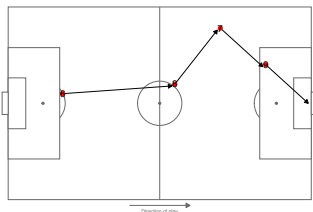
Introduction

- ▶ During a football match, through computer vision, we can obtain two types of data.

Tracking data
example



On the ball data
example



- ▶ In this work we analyze all the shots, which are part the “on the ball data”, in the five major European championships of the one season 2020-21

Offensive action

- ▶ The objective of this analysis is the study of the probability of scoring a goal;
- ▶ To study this probability we use two random variables:
 - ▶ $Y_i \in \{0, 1\}$, 1 if the i -th shot hits the goal, 0 otherwise;
 - ▶ $Z_i \in \{0, 1\}$, 1 if the i -th shot turns into a goal, 0 otherwise.
- ▶ we decide to analyze the probability of scoring goal, $Pr(Z_i)$, like the product between the probability that a shot hits the goal and the probability that the shot turns into a score since it is on goal:

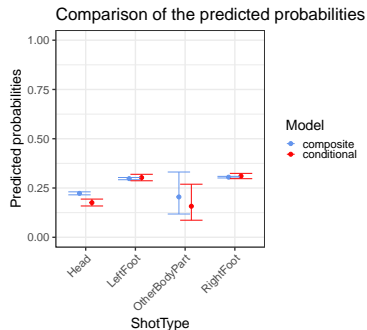
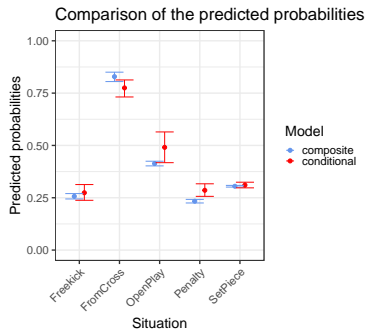
$$\begin{aligned}Pr(Z_i) &= \sum_j Pr(Z_i | Y_i = j), \\ &= \underbrace{Pr(Y_i = 0)Pr(Z_i | Y_i = 0)}_{\text{by hypothesis} = 0} + Pr(Y_i = 1)Pr(Z_i | Y_i = 1), \\ &= Pr(Y_i = 1)Pr(Z_i | Y_i = 1).\end{aligned}$$

Composite likelihood




- ▶ We study then two probabilities simultaneously through the use of composite likelihood, a pseudo-likelihood:

$$L_C(\theta; z, y) = L_1(\beta, \delta; y)^a L_2(\gamma, \delta; z)^b.$$

- ▶ We can assume the same effect for the variables shared by the two probabilities, obtaining an increase in precision on the parameter estimates of the shared variables.



References

-  A. Azzalini e B. Scarpa. Data analysis and data mining: An introduction. pages 40–135. OUP USA. 2012.
-  A. Gelman e J.Hill. Data analysis using regression and multilevel/hierarchical models, pages 301-345. Cambridge university press, 2006.
-  B. G. Lindsay. Composite likelihood methods. Contemporary mathematics, 80(1):221-239, 1994