

# Un estudio de goles “Solo Run” usando sistemas complejos

Cristian José Álvarez Bran

522462263

## 1. Introducción

El uso de sistemas complejos como herramienta para el estudio del fútbol es un acercamiento relativamente nuevo, siendo [1] uno de los primeros estudios realizados de esta manera. Sin embargo, para la fecha ya hay una considerable cantidad de artículos e investigación al respecto. Y es que hoy en día no sólo el poder computacional necesario para esto es más accesible, sino que ahora hay disponibles extensas bases de datos que guardan características espacio-temporales de cada juego.

Al ser el fútbol un deporte de equipo donde pasar la pelota es lo más fundamental, es bastante sencillo encontrar estudios en los que se enfatiza este aspecto. Pero, ciertamente existen situaciones, especialmente las llamadas “jugadas de peligro”, en las que una acción individual que realiza el jugador en posesión de la pelota es lo más relevante. En este trabajo se busca hacer un modelo usando sistemas complejos para estudiar probabilísticamente acciones individuales que culminan en gol. Para ello se usa un modelo de agentes que recrea la jugada del gol, lo cual permite tomar en cuenta factores dinámicos como la posición del atacante en el campo, su velocidad, su posición relativa a los defensas, etc. Y a partir de esto, se toman dos acercamientos para definir una función de probabilidad y se examinan los alcances de ellos.

## 2. El modelo de agentes

Los goles que estudiaremos son;

1. Lionel Messi, Barcelona vs Real Madrid Semifinal UEFA Champions League 2010-2011. [\[Video\]](#)
2. Diego Maradona, “Gol del Siglo”, Argentina vs Inglaterra Cuartos de Final Copa Mundial México 1986. [\[Video\]](#)

Debido a que estos no son tan recientes no se pudo hallar datos espaciales y temporales para construir el modelo, así que fueron recaudados personalmente viendo distintos videos de cada gol. Por simplicidad, se linealizó la trayectoria que siguió cada jugador a lo largo de la jugada y se midieron los tiempos que tardaban los mismos en llegar de un punto a otro en la linearización para calcular sus velocidades. El código para toda la simulación puede hallarse en un repositorio de github. [\[Código\]](#)

## 3. Primer Acercamiento

La primera aproximación que tomaremos para el estudio de los goles tiene como pilar un modelo de probabilidad de gol que depende de la posición en la que se realiza un tiro en el campo. David Sumpter en su página, [soccermetics.medium](#), consideró un modelo que dependía únicamente del ángulo  $\theta$  que se forma entre los dos postes de la portería y el punto en el que se realiza un tiro (ver [2]). Esta idea resulta bastante natural, puesto que este ángulo representa qué tan amplio es el rango de opciones a las que se puede dirigir un disparo; entre más grande es el ángulo, más fácil es que un tiro vaya en dirección a gol, ver figura 1. Como dato curioso, el teorema del ángulo central permite reconocer regiones “equivalentes” respecto al ángulo. Ver figura 2.

Luego, para crear el modelo tomó los datos de todos los tiros realizados en una edición de “La Liga” y utilizando regresión logística calculó la probabilidad como una función del ángulo de disparo. Al hacer esto obtuvo

$$P(\text{Gol}) = \frac{1}{1 + e^{-3,54\theta + 3,9}}.$$

Sumpter menciona que este modelo no es perfecto y que para tener una función que se ajuste mucho mejor a los datos también hay que tomar en cuenta al parámetro  $x$ , la distancia vertical del punto de disparo hacia la portería, con lo cual obtuvo

$$P(\text{Gol}) = \frac{1}{1 + \exp(0,0069x^2 - 0,11x\theta - 0,12x - 2,53\theta + 4,03)}.$$

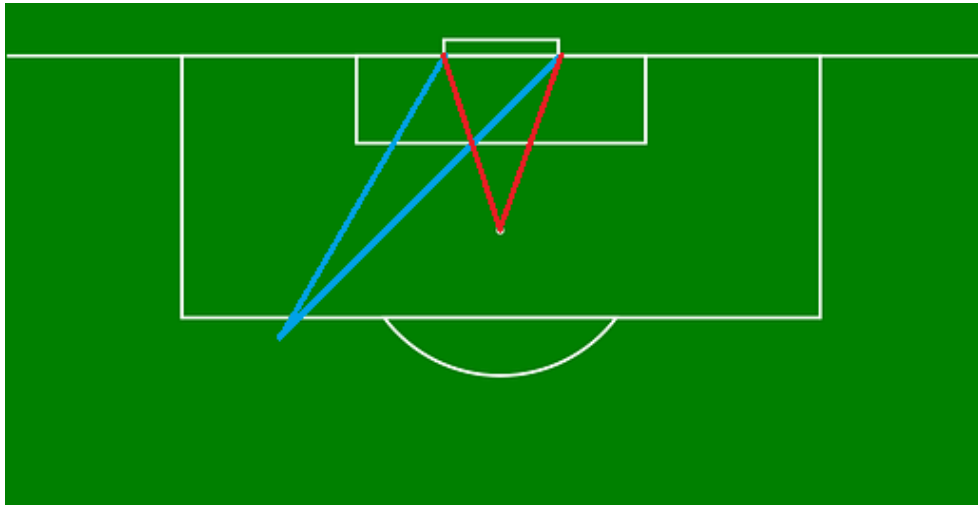


Figura 1: Dos ángulos de tiro  
Fuente: Elaborado con Microsoft Paint.

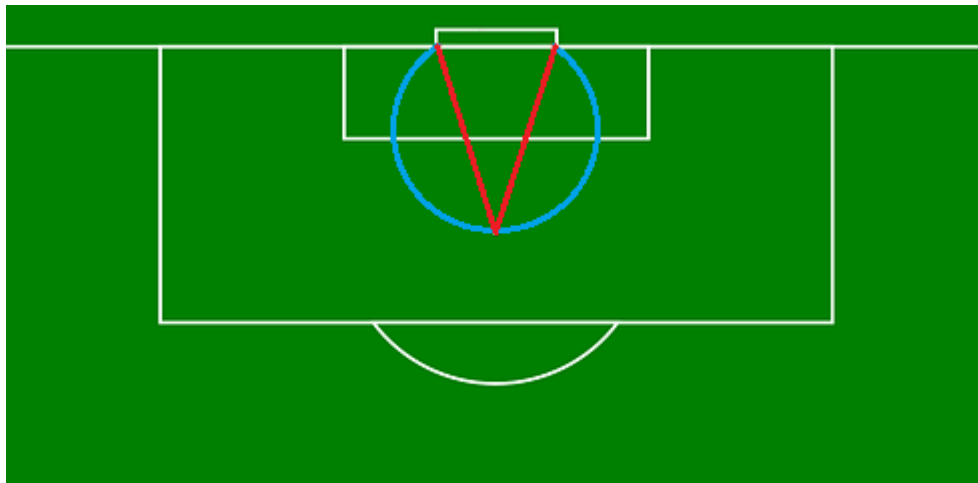


Figura 2: Dos ángulos de tiro  
Fuente: Elaborado con Microsoft Paint.

Con esta herramienta podemos tomar la primera aproximación a este estudio; la evolución de la probabilidad de gol a lo largo de la jugada. Ver figuras 3 y 4.

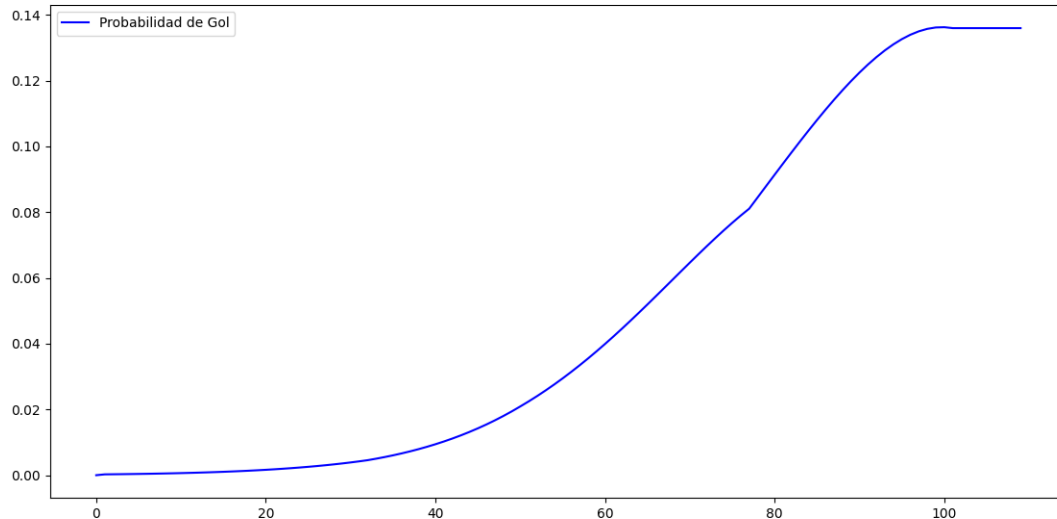


Figura 3: Evolución de la probabilidad de en Gol de Messi  
Fuente: Elaborado con Python.

A primera vista estas gráficas pueden ser decepcionantes, no parece como que haya habido un cambio significativo en la probabilidad durante las jugadas pero, realmente esto sucede a raíz de la función de probabilidad. Al realizar la regresión logística, los datos usados comprenden tiros de todo tipo de naturaleza. En particular, tiros en los que hay varios jugadores entre la pelota y la portería son varios, pero no suelen terminar en gol. Esto hace que la función de probabilidad no pueda distinguir situaciones “claras de gol”. De hecho bajo este modelo, ¡un tiro desde el punto de penal tiene asociada una probabilidad de apenas 34%! Sin embargo, algunas fuentes estiman que alrededor del 75 % de los penales terminan en gol.

Además de la intuición, herramientas como el “control de campo” (ver [3]) sugieren que si se tomara en cuenta la posición de los defensores la probabilidad de gol podría aumentar significativamente. Recordemos que Messi llega a tener un duelo uno contra uno con Casillas y que Maradona rebasa a Shilton.

## 4. Segundo Acercamiento

A continuación se presenta un acercamiento que tiene la esencia de un modelo de contagio por medio de agentes. La idea detrás de aplicar esto es que uno podría argumentar que ciertos defensas involucrados en la jugada pudieron haber tomado decisiones distintas y a lo mejor evitar el gol. Dicho esto, el setup de la simulación es el siguiente:

- Los defensas elegidos se movilizan siguiendo ciertas reglas que toman en consideración la posición del atacante en el campo, la posición del atacante relativa a ellos, el perfil nato del atacante y la posición de sus demás compañeros defensas en el campo y relativa al atacante. En cada fotograma la velocidad de un defensa es una clase de promedio ponderado de las velocidades que surgen de seguir estas reglas, donde los pesos del promedio siguen una distribución uniforme en  $(0,5,1)$  y son generados y fijados en el primer fotograma de la simulación.
- El movimiento de todos los demás jugadores involucrados, incluido el atacante, sigue el curso de la jugada que sí pasó en la vida real.

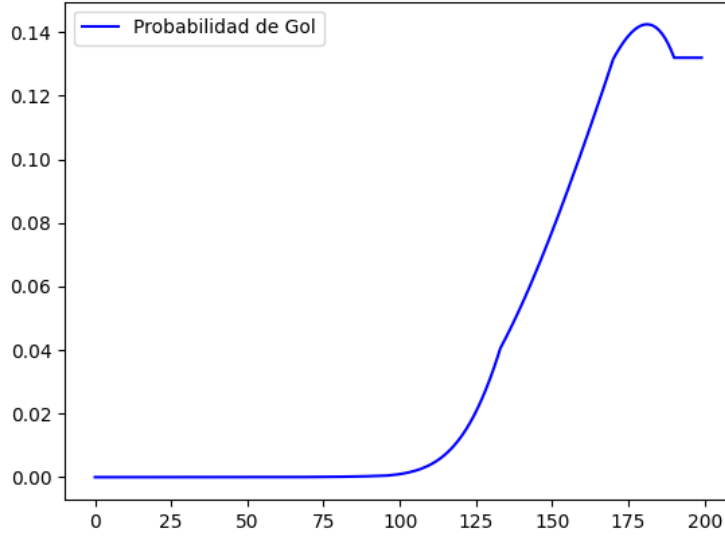


Figura 4: Evolución de la probabilidad de en Gol de Maradona  
Fuente: Elaborado con Python

- Similar a un modelo de contagio, si los defensas elegidos están dentro de una vecindad del atacante suficientemente pequeña (se eligió un radio de 1.5 m) entonces pueden tomar la decisión de intentar quitarle la pelota y lograrlo con cierta probabilidad.
- Se propuso que la probabilidad de robarle la pelota a un atacante tenga la forma

$$P(\text{Robo}) = (1 - E) \cdot f(d)$$

donde  $E$  es el porcentaje de regates completados exitosamente por el atacante en la temporada o campeonato en cuestión, 68 % para Messi y 59 % para Maradona (ver inciso 7 en Notas y Disclaimers). Y  $f : (0, 1.5) \rightarrow [0, 1]$  es una función de peso decreciente, es decir entre menor sea la distancia  $d$  que separa al defensa del atacante, más cercano es  $f(d)$  a 1.

- Usando la misma función  $f$  del punto anterior, la probabilidad de que un defensa tome la decisión de quitar la pelota o “meter el pie” se definió como

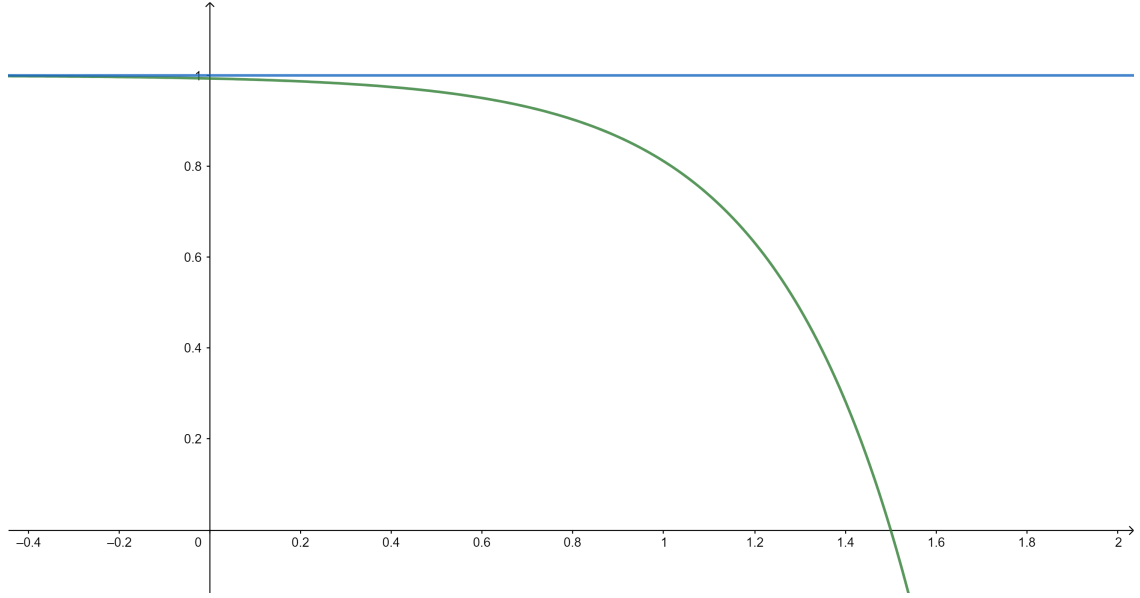
$$P(\text{Meter el pie}) = A \cdot \frac{x}{h} + (1 - A)f(d)$$

donde  $A$  es una constante que depende de cada jugador y representa su “agresividad”,  $h$  es el largo de la sección relevante del campo y  $x$  es la resta entre  $h$  y distancia vertical del atacante a la portería. Los coeficientes de agresividad siguen una distribución uniforme en  $(0,1)$  y son generados y fijados en el primer fotograma de la simulación.

- En las simulaciones realizadas se tomó

$$f(d) = 1 - e^{\frac{10d}{3} - 5}$$

restringida a  $(0, 1.5)$ , debido a que posee un decaimiento relativamente lento cerca de  $d = 0$  pero rápido cerca de  $d = 1.5$  como se puede apreciar en la gráfica.



El resultado de cada una de estas simulaciones es ya sea “Gol” o “No Gol”, posibilidades que pueden codificarse como 1 o 0 respectivamente. Esto quiere decir que podemos interpretar al resultado de una simulación como una variable aleatoria  $X \sim \text{Ber}(p)$ , donde el parámetro  $p$  puede ser estimado por la ley de los números grandes como la proporción de simulaciones que acaban en gol entre simulaciones realizadas.

Una vez estimado el parámetro  $p$  resulta de interés compararlo con  $E^n$  donde  $n$  es la cantidad de defensores elegidos, esto porque  $E^n$  puede interpretarse como la probabilidad de que el atacante logre rebasar a los  $n$  defensas en un modelo donde no se toma en cuenta el contexto espacio-temporal de la jugada. Para esto realizamos la prueba de hipótesis

$$H_0 : E^n \geq p, \quad H_A : E^n < p$$

con el test de Wald a un nivel de confianza  $1 - \alpha$

$$\mathbb{1} \left\{ \frac{\hat{p} - E^n}{\sqrt{\text{var}(\hat{p})}} > q_\alpha \right\} = \mathbb{1} \left\{ \frac{\bar{X}_N - E^n}{\sqrt{\frac{\bar{X}_N(1-\bar{X}_N)}{N}}} > q_\alpha \right\},$$

donde  $q_\alpha$  es tal que  $\Phi(q_\alpha) = 1 - \alpha$  donde  $\Phi$  es la función de densidad acumulada de una variable normal estándar. Se corrieron 200 simulaciones por cada gol y obtuvimos los siguientes resultados

Gol	$\hat{p}$	Estadístico	$p$ -value
Messi	0.575	3.2212	0.0006
Maradona	0.445	2.7421	0.00305

Recordemos que el  $p$ -value es el menor nivel  $\alpha$  para el cual rechazamos la hipótesis nula, o bien equivalentemente, la probabilidad de que bajo la hipótesis nula se obtenga un resultado igual o más extremo que el estadístico en la dirección de la hipótesis alterna. Lo anterior nos permite rechazar la hipótesis nula para ambos goles con una confianza del 99%. Más aún, repitiendo un ejercicio análogo cambiando  $E^n$  por 0,5 y 0,38 en los goles de Messi y Maradona respectivamente, rechazamos  $H_0$  con un nivel de confianza de 95%.

## 5. Notas y Disclaimers

Se hicieron varias suposiciones lo largo del estudio, en especial en el segundo acercamiento. Algunas de ellas claramente no se cumplen en una situación real, otras se adoptaron en búsqueda de simplificar el modelo y puede que se haya hecho de sobremanera. En cualquier caso, estas suposiciones probablemente sesgan los resultados y por lo tanto se enumeran a continuación para tomarlas en cuenta:

1. En la realidad, el delantero reacciona a su entorno para tomar decisiones acerca de su movimiento.
2. La toma de datos fue realizada de manera personal y agentpy, la biblioteca usada en python, parece discretizar el espacio en una retícula muy fina, lo cual puede alejar las simulaciones de copias fidedignas de las jugadas.
3. En el gol de Maradona habían más delanteros creando opciones de pase y por lo tanto influían en el movimiento de los demás defensas.
4. La función  $f$  considerada está basada únicamente en un argumento razonable, no en datos.
5. Durante la simulación y la evaluación del segundo acercamiento se usó  $E$  para encuentros 1v1. Sin embargo,  $E$  también toma en cuenta encuentros consecutivos de este tipo.
6. Para definir la probabilidad de robo se apeló principalmente al número  $E$ , sin tomar en cuenta factores como la velocidad de los jugadores, el tiempo de reacción humana, el centro de gravedad del atacante, etc.
7. La cifra de  $E$  para Messi puede ser hallada en [4]. En el caso de Maradona, es bien sabido que este completó 53 regates exitosamente (ver por ejemplo [5]) y algunas fuentes (ver [6]) afirman que intentó 90 regates en total.

## 6. Conclusiones

1. Intuitivamente, el primer acercamiento se ve fuertemente afectado porque el modelo de probabilidad no toma en consideración la posición de los demás jugadores. Sin embargo, con un modelo que pueda adaptarse a la posición de los defensas este acercamiento nos podría dar mucha información de cómo realmente evolucionó la jugada.
2. El segundo acercamiento promete ser mejor que un modelo probabilístico natural, aunque rudimentario, basado en una cifra estadística indiferente al contexto espacio temporal del futbol. Además, también sugiere (no evidencia) que la probabilidad de que la jugada de Messi culminara en gol es mayor al 50 % y la de Maradona mayor al 38 %.

## Referencias

- [1] Carlos Cotta, et al. (2011). FIFA World Cup 2010: A Network Analysis of the Champion Team Play. CoRR. abs/1108.0261. 10.1007/s11424-013-2291-2.
- [2] David Sumpter. (2017). The Geometry of Shooting. 2/06/2022, de Medium Sitio web: <https://soccermatics.medium.com/the-geometry-of-shooting-ae7a67fdf760>
- [3] Javier Fernandez. (2018). Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. de MIT Sloan Sports Analytics Conference.
- [4] Gabriele Marcotti (2020). Lionel Messi's evolution as a player: From Ronaldinho's Barcelona understudy to GOAT candidate. de ESPN. <https://www.espn.com/soccer/blog-marcottis-musings/story/4107181/lionel-messis-evolution-as-a-player-from-ronaldinhos-barcelona-understudy-to-goat-candidate>
- [5] Anirudh Menon (2020). Diego Maradona's stats: You can't measure genius in numbers, but here's a try. de ESPN. <https://www.espn.com/soccer/argentina-arg/story/4245614/diego-maradonas-career-stats-you-cant-measure-genius-in-numbers-but-heres-a-try>
- [6] James Darby (NA). Castrol World Cup Legends: Diego Maradona - 1986. de GOAL. <https://www.goal.com/en/news/1863/world-cup-2010/2010/06/10/1968349/castrol-world-cup-legends-diego-maradona-1986>.