



Original

Análisis de los goles conseguidos en 13 temporadas (2000/01-2012/13) correspondientes a la Primera División de la Liga Española de Fútbol Profesional



J. Sánchez-Flores^a, J.M. Martín-González^b, J.M. García-Manso^{a,*}, Y. de Saa^a, E.J. Arriaza-Ardiles^c y M.E. Da Silva-Griglotetto^d

^a Departamento de Educación Física, Universidad de Las Palmas de Gran Canaria, Las Palmas, Islas Canarias, España

^b Departamento de Física, Universidad de Las Palmas de Gran Canaria, Las Palmas, Islas Canarias, España

^c Centro de Estudios Avanzados, Universidad de Playa Ancha, Valparaíso, Chile

^d Centro de Ciencias Biológicas e da Saúde, Universidade Federal de Sergipe, Brasil/Scientific Sport, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 19 de marzo de 2014

Aceptado el 20 de mayo de 2015

Palabras clave:

Fútbol

Gol

Entropía normalizada de Shannon

Distribución de Poisson

Distribución binomial negativa

Efecto Mateo

R E S U M E N

Objetivo: El propósito del estudio es analizar, en las temporadas 2000/01-2012/13, la distribución de los goles y la entropía normalizada de Shannon para el cálculo del grado de competitividad de las ligas; además se ha analizado su relación con el grado de competitividad de la liga.

Método: Se utilizó la distribución de Poisson y la distribución binomial negativa para el análisis de los goles y la entropía normalizada de Shannon para el cálculo del grado de competitividad de las ligas.

Resultados: La liga española ha perdido competitividad en las temporadas evaluadas, ya que demuestran resultados y en el índice de dispersión entre equipo-partido, especialmente en las últimas temporadas evaluadas. La distribución de los goles por equipos deja de ser poissoniana, especialmente a partir de la temporada 2008-09, tal y como muestra el aumento del índice de dispersión (ajuste lineal: $a = 0.0162 \pm 0.009$; $b = 0.9952 \pm 0.0715$; $R^2 = 0.588$; $p = 0.002$). Sin embargo, no ocurre lo mismo si el análisis lo hacemos de tiempo entre gol, con independencia del número de partidos, el comportamiento es diferente a partir de los 200 minutos, donde el proceso sigue una exponencial y puede considerarse un proceso poissoniano. Este cambio parece indicar cierto efecto de memoria que se puede interpretar como un efecto Mateo que explica la incapacidad de los equipos más débiles para recuperarse de dinámicas perdedoras.

Conclusiones: La superioridad de los equipos más potentes parece clara, quizás excesiva, respecto al resto de equipos que participan en la principal liga española de fútbol. También se ha incrementado la probabilidad de que se consiga un número elevado de goles (> 5 goles) en un partido. Esto provoca que la distribución del número de goles por partido sea de la junta binomial negativa.

© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Analysis of goals (score) studied in thirteen seasons (2000/01 to 2012/13) for a league of professional spanish football League

A B S T R A C T

Objective: The aim of this study is to analyze, 2000/01season through 2012/13season, the goal scored distribution by game and team; as its behavior in time. We also analyze the relationship with the league competitiveness degree.

Method: We used the Poisson and the Negative Binomial distributions in order to study the goals distribution; and the Normalized Shannon Entropy for calculating the leagues uncertainty.

Keywords:

Soccer

Goal

Shannon entropy normalized

Poisson distribution

Negative binomial distribution

Mathews Effect

* Autor para correspondencia.

Correo electrónico: jgarciamanso@gmail.com (J.M. García-Manso).

<http://dx.doi.org/10.1016/j.ramd.2015.05.006>

1888-7546/© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Results: The Spaniard league has lost competitiveness in the seasons evaluated as the entropy and index of dispersion (team-game) display, especially in the last seasons analyzed. From the perspective of teams, it is not Poisson anymore, above all beyond 2008/09 season. From the perspective of games it does not take place the same phenomenon, specially the last seasons studied ($a = 0.0099 \pm 0.0097$; $b = 0.9622 \pm 0.077$; $R2 = 0.316$; $p = 0.045$ vs. $a = 0.0162 \pm 0.009$; $b = 0.9952 \pm 0.0715$; $R2 = 0.588$; $p = 0.002$). Regarding time difference between goals, the behavior is different from 200 minutes, where the process follows an exponential distribution, and can be considered as a Poissonian process. This modification points out a possible memory effect that can be understood as a Mathew effect, which explains that the less powerful teams are unable to overcome the situation.

Conclusions: The superiority of most powerful teams seems to be more clear, perhaps excessive, compared to the rest of participating teams, as well as the probability that a lot of goals take place (> 5 goals) in a single game.

© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Análise dos golos marcados em 13 temporadas (2000/1 a 2012/13) da primeira divisão da Liga Espanhola de Futebol Profissional

R E S U M O

Palavras-chave:

Futebol
Objetivo
Entropia normalizada de Shannon
Poisson
Distribuição binomial negativa
Efeito Mateus

Objetivo: O objetivo do estudo é analisar, nas temporadas 2000/01-2012/13, a distribuição de golos marcados por partida e equipas, e o seu comportamento ao longo do tempo. Além disso, analisou a sua relação com o grau de competitividade da liga.

Métodos: Foram utilizadas a distribuição de Poisson e a distribuição binomial negativa para análise dos golos; e a entropia normalizada de Shannon para o cálculo das incertezas da liga.

Resultados: A liga espanhola perdeu competitividade nas épocas avaliadas, como mostrado pela entropia e pelo índice de dispersão entre as equipas em jogo, especialmente nas últimas temporadas avaliadas (≈ 1.2). Do ponto de vista do equipamento, de Poisson das partes, especialmente uma vez que a estação 2008-09. Não é por isso, se a análise é feita do ponto de vista mais, especialmente nas últimas temporadas ($a = 0.0099 \pm 0.0097$, $b = 0.9622 \pm 0.077$; $R2 = 0.316$, $p = 0.045$ vs. $a = 0.0162 \pm 0.009$; $b = 0.9952 \pm 0.0715$; $R2 = 0.588$, $p = 0.002$). Quanto às diferenças de tempo entre cada gol, o comportamento é diferente a partir dos 200 minutos, em que o processo segue uma distribuição exponencial e pode ser considerado um processo Poissoniano. Esta modificação aponta um possível efeito de memória que pode ser entendido como um efeito Mathew, o que explica que as equipas menos potentes são incapazes de superar a situação.

Conclusões: A superioridade das equipas de futebol mais fortes parece clara, talvez excessiva, em comparação com outras equipas participantes na principal liga de futebol espanhola. Também tem aumentado a probabilidade de que seja alcançada um grande número de golos (> 5 golos) numa partida. Isto significa que, por partida, o número de golos é o tipo binomial negativa.

© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Published by Elsevier España, S.L.U. Este é um artigo Open Access sob a licença de CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

Como en cualquier modalidad deportiva, el objetivo principal del fútbol es establecer estrategias de juego que permitan a los equipos participantes superar a sus rivales para conseguir la victoria en cada encuentro o en la competición en que participan. El código de puntuación establecido viene supeditado a un reglamento en el que se establece la forma en la que se gana un partido y se consiguen los puntos que indican su posición en la competición final.

En el caso del fútbol el *score* es el gol (tanto). El número de goles, conseguidos o recibidos, es lo que determina que se consiga o no la victoria.

En el fútbol profesional, el número de goles que se marcan en cada partido, y también en una competición, es pequeño (alrededor de 2.5 goles/partido de media) si lo comparamos con la mayor parte de los deportes de cooperación-oposición (p. e. baloncesto, balonmano, waterpolo, etc.). Este aspecto incrementa el grado de incertidumbre en el resultado de los partidos y se justifica, en esta modalidad deportiva. Por este motivo, Anderson y Sally¹, en los años 1993-2011 tras analizar 43 000 partidos de las principales ligas europeas (España, Alemania, Italia e Inglaterra) concluyen que el fútbol profesional es un deporte en el que la suerte juega un papel relevante que hace que los equipos favoritos solo ganen el partido

algo más de la mitad de las veces (54.8%) que compiten. Aún más, plantean que para tener éxito en el fútbol hay 2 rutas: una es ser bueno, la otra ser afortunado. Según los autores, se necesitan las 2 para ganar un partido. Pero una para ganar un partido. La diferencia de puntos obtenidos por los equipos situados en los primeros puestos de la clasificación y el resto de equipos está influenciada más por el rendimiento que por la suerte. Para Lago², el azar puede ser un factor importante para explicar el resultado en un único partido o, a lo sumo, en un número muy limitado de ellos, pero que a partir de cierta cantidad de encuentros el rendimiento es determinante para dar cuenta de los puntos que alcanzan los conjuntos.

Esto hace especialmente interesante el estudio de los goles, su número y comportamiento, en la categoría profesional de esta modalidad deportiva. En general, en los deportes de equipo, la forma como se comporta el *score* ha sido considerada como un proceso de distribución de Poisson (DP), aunque con algunas restricciones³⁻⁷. La distribución de los goles conseguidos en un partido de fútbol sigue aproximadamente la mencionada distribución (varianza/media = 1.0) con pequeñas variaciones⁸⁻¹². Las DP son distribuciones de probabilidad discreta que expresan, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos en cierto periodo de tiempo. Es

decir, una DP parte de la distribución binomial en el que se expresa el comportamiento de un número elevado de ensayos (en nuestro caso partidos) con bajo índice de éxito (goles).

Las características que presenta el fútbol (número de jugadores, espacio de juego, duración del partido, etc.) son algunos de los principales factores que condicionan el número de goles que se pueden marcar. Todos ellos quedan claramente establecidos y normalizados en el reglamento que aprueba la *Fédération Internationale de Football Association* (FIFA). En el fútbol, y en todos los deportes en general, el reglamento busca la máxima espectacularidad en las competiciones para hacerlas más atractivas para los practicantes, aficionados y espectadores.

Sin embargo, el nivel de la liga (calidad y nivel de igualdad entre los equipos), los cambios de reglamento (p. e. sistema de puntuación) o las características de los equipos (p. e. sistema de juego) son parámetros que pueden alterar el *score* y, en consecuencia, la distribución de los goles conseguidos en cada partido o por cada equipo en cada partido o competición.

El objetivo de este trabajo es utilizar el marco de los procesos aleatorios de Poisson para comprobar hasta qué punto los datos reales siguen este comportamiento en el fútbol moderno y determinar si se han producido cambios en los últimos años. Para ello se analiza, en la Liga española de Primera División (Liga BBVA) en las temporadas 2000/01-2012/13, cuántos han sido los goles conseguidos, por partido y equipo, en cada temporada. Además, se evalúa el grado de competitividad que se detecta en cada temporada mediante el cálculo de la entropía normalizada de Shanon (S_N).

Método

Muestra

Se han analizado 13 287 goles conseguidos por los equipos (media: 2.69 goles/partido) de la Primera División española (Liga BBVA) durante las temporadas oficiales de 2000-01-2012-13.

Diseño experimental

Distribución de los goles por temporada. Para el estudio de los goles se analizó su distribución. Para ello se calcularon 2 tipos de distribuciones: DP y distribución Binomial Negativa (BN).

Distribución de Poisson. Como es conocido, la DP tiene la siguiente expresión matemática:

$$P(k, \lambda) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Donde $P(k; \lambda)$ es la probabilidad que existe de que, al analizar un fenómeno (partidos de fútbol), se den k eventos (goles) en un intervalo de tiempo y λ es la media de eventos por intervalo. Así, el modelo de Poisson depende de un solo parámetro λ que tiene un significado físico preciso que, en este caso, representa el promedio de tantos marcados en un determinado periodo de tiempo (por partido: 90 min). Además, el número de ceros (no eventos) queda también determinado por λ :

$$P(0, \lambda) = e^{-\lambda}$$

Una interesante propiedad de este tipo de distribución es que los valores de la media y la varianza son iguales o muy parecidos. El cociente entre la varianza y la media de eventos por intervalo es conocido como Índice de Dispersión (*ID*):

$$ID = \frac{\sigma^2}{\lambda}$$

Siendo σ su desviación estándar, en este tipo de distribución el *ID* toma el valor de 1. No obstante, este valor puede variar

sensiblemente aumentando o disminuyendo. Cuando $ID < 1$ se denomina subdispersión, y los datos tienden a agruparse más en torno al valor medio. En este caso la distribución del fenómeno es más predecible. Es decir el número de goles que se consiguen es más fácil de predecir. Si $ID > 1$ existe sobredispersión, por lo tanto los datos tienden a estar más dispersos, lo cual puede deberse a un mayor número de ceros de los que predeciría una DP, o a un problema más alejado de la media que los previstos por el modelo poissoniano. En este caso se propone ensayar la distribución BN.

Distribución Binomial Negativa. Una BN se define como una distribución de probabilidad discreta que trata de medir el número de éxitos en una secuencia para un número de eventos independientes entre sí. Esta distribución tiene más dispersión en la cola final con respecto a la de Poisson. La distribución BN depende de 2 parámetros r y q (parámetros del ajuste), y puede considerarse una generalización de la DP. Además, en el caso de la DP, los intervalos de tiempo entre eventos, este caso goles, siguen una distribución exponencial $P(dt) = e^{-\lambda dt}$ que solo depende del valor que tiene el parámetro λ y que es fácilmente detectable cuando la distribución del histograma se transforma en un gráfico semilogarítmico, donde el comportamiento exponencial se ve como una línea recta.

Análisis de la competitividad de la liga. La competitividad de una liga (nivel de igualdad entre los equipos) puede ser determinada por el grado de incertidumbre que exista en cada enfrentamiento. Una magnitud que se ha mostrado útil para el análisis de sistemas complejos es la entropía de Shannon (S) que, cuando el conjunto de probabilidades p_i , $i=1, \dots, N$; de un sistema es conocido, mide la incertidumbre promedio y, por tanto, hace referencia a la cantidad media de información que contiene una variable aleatoria. Se define como:

$$S = \sum_{i=1}^N p_i \log(p_i)$$

Siendo la incertidumbre máxima cuando todos los valores de p_i sean iguales. El valor de S cambia con el valor de N , en nuestro caso el número de equipos que participan en la liga cada temporada, y por tanto, si N cambia los valores de S no son comparables en cada temporada. En estos casos, es decir, cuando queremos comparar las diferentes temporadas es preferible utilizar la S_N .

$$S_N = \frac{S}{\log(N)}$$

De esta forma, el valor de S_N queda acotado entre 0-1, donde 1 corresponde a la situación de máxima incertidumbre, donde todos los valores de p_i son iguales.

Además, de variables (ID y S_N) fueron comparadas, calculando el coeficiente de correlación de Pearson, para conocer el grado de covariación entre ambos parámetros que tienen la peculiaridad de estar relacionados linealmente.

Análisis de los datos

Para el tratamiento de los datos se ha usado los programas SPSS (versión 17, SPSS, Chicago, Illinois) y MATLAB (versión R2008b, The MathWorks, Natick, MA).

Resultados

Goles totales, goles por partido y goles por equipo. En la [tabla 1](#) se muestra la estadística (total, media, varianza e *ID*) de los goles que se marcan, por temporada, de la Liga española de fútbol profesional durante las temporadas 2000/01-2012/13.

Nótese como el *ID* cambia sensiblemente cuando se compara entre cada equipo o partido ([tablas 2 y 3](#)). En el caso de los datos que muestran los equipos el *ID* es claramente mayor que 1 y, a su

Tabla 1
Estadística global de goles conseguidos en las 13 temporadas evaluadas y los estadísticos utilizados (media y varianza de goles e Índice de Dispersión). Estos datos se expresan en valores medios por cada equipo y por partido

Variables	13 287 goles	
	Por equipo	Por partido
Media	1.345	2.689
Varianza	1.494	2.779
Índice dispersión	1.109	1.032

Tabla 2
Datos promedio por equipo de cada una de las variables analizadas (goles totales, media, varianza e índice de dispersión) en las 13 temporadas. También se incluye el valor de entropía normalizada de Shannon (S_N) para cada temporada

Estadísticas globales por equipo					
Temporada	N.º goles	Media	Varianza	ID	S_N
2000-01	1090	1.434	1.511	1.053	0.991270
2001-02	961	1.265	1.312	1.038	0.993490
2002-03	1028	1.353	1.491	1.102	0.990621
2003-04	1015	1.336	1.412	1.057	0.990884
2004-05	976	1.284	1.327	1.033	0.988012
2005-06	936	1.232	1.222	0.992	0.987178
2006-07	942	1.240	1.341	1.082	0.989397
2007-08	1022	1.345	1.533	1.140	0.988368
2008-09	1101	1.449	1.623	1.120	0.988571
2009-10	1031	1.357	1.521	1.121	0.982032
2010-11	1044	1.374	1.654	1.204	0.985618
2011-12	1050	1.382	1.794	1.298	0.985470
2012-13	1091	1.436	1.688	1.176	0.983620
Media	1022	1.345	1.494	1.109	0.988041

ID: índice de dispersión; S_N : entropía normalizada de Shannon.

vez, mayor que el valor que muestra el análisis realizado por partidos. Esto indica que el modelo de Poisson se sigue mejor cuando consideramos los resultados obtenidos por partidos ($ID = 1032$) que por equipos ($ID = 1109$).

Tabla 3
Datos promedio por partido de cada una de las variables analizadas (goles totales, media, varianza e índice de dispersión) en las 13 temporadas. También se incluye el valor de entropía normalizada de Shannon (S_N) para cada temporada

Estadísticas globales por partido					
Temporada	N.º goles	Media	Varianza	ID	S_N
2000-01	1090	2.868	2.811	0.980	0.991270
2001-02	961	2.529	2.403	0.950	0.993490
2002-03	1028	2.674	2.891	1.081	0.990621
2003-04	1015	2.671	2.723	1.019	0.990884
2004-05	976	2.597	2.500	0.962	0.988012
2005-06	936	2.463	2.228	0.905	0.987178
2006-07	942	2.483	2.727	1.098	0.989397
2007-08	1022	2.687	2.833	1.054	0.988368
2008-09	1101	2.897	3.174	1.096	0.988571
2009-10	1031	2.713	2.701	0.996	0.982032
2010-11	1044	2.742	2.862	1.044	0.985618
2011-12	1050	2.763	3.094	1.120	0.985470
2012-13	1091	2.871	3.179	1.107	0.983620
Media	1022	2.700	2.779	1.032	0.988041

ID: índice de dispersión; S_N : entropía normalizada de Shannon.

Las variables ID y S_N , es decir la forma como se distribuyen los goles y el nivel de incertidumbre de las temporadas analizadas, muestran una moderada correlación entre ellas ($R^2 = -0.541$; P-value: 0.056) cuando se valoran los resultados por equipos.

En la figura 1 se representan los comportamientos de los parámetros evaluados (S_N de cada temporada; ID por cada equipo y partido y las medias y varianzas por equipo y partido) al analizar las 13 temporadas (2000/01-2012/13). En la parte superior izquierda (fig. 1a) se representa los valores obtenidos de la entropía normalizada de Shannon calculada al final de cada temporada. En ella se aprecia una tendencia decreciente de los valores de S_N (ajuste lineal: $a = -0.0007 \pm 0.0003$; $b = 0.9930 \pm 0.0022$; $R^2 = 0.7399$; p-value = 0.0002) donde el intervalo de confianza de la pendiente parece indicar que esta es negativa.

En la gráfica superior derecha (fig. 1b), se representan los valores del ID calculados para cada equipo (línea discontinua) y partido

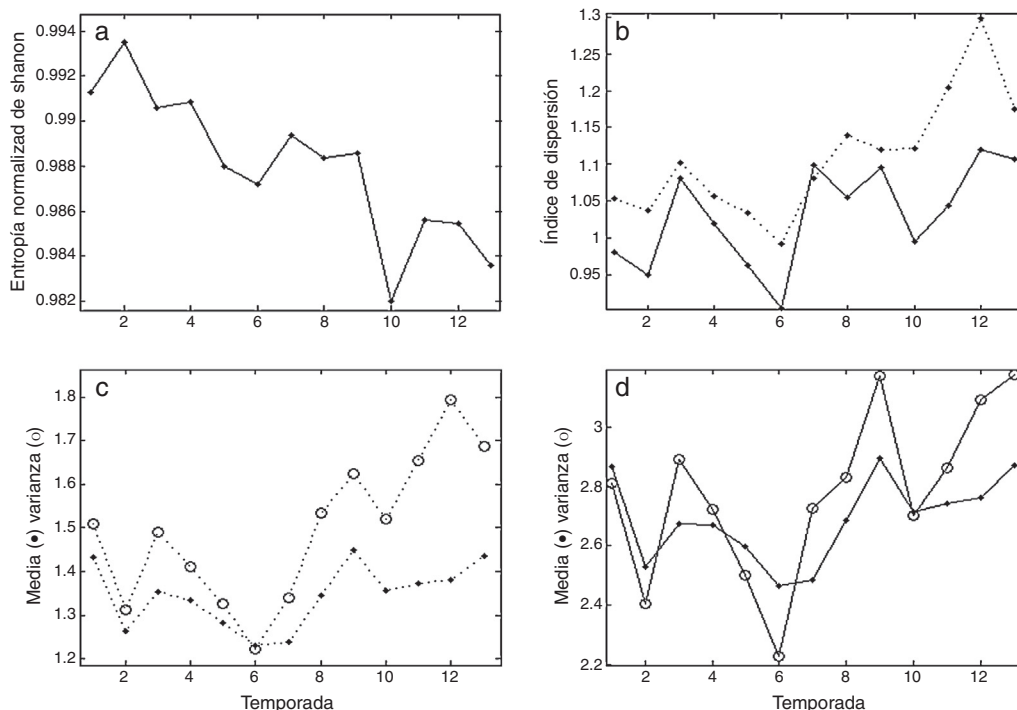


Figura 1. Representación gráfica de la entropía normalizada en cada temporada (1a), el índice de dispersión por partido (línea continua) y equipo (línea discontinua) en cada temporada (1b), las medias (●) y varianzas (○) por equipo y por partido en cada temporada (1c y 1d).

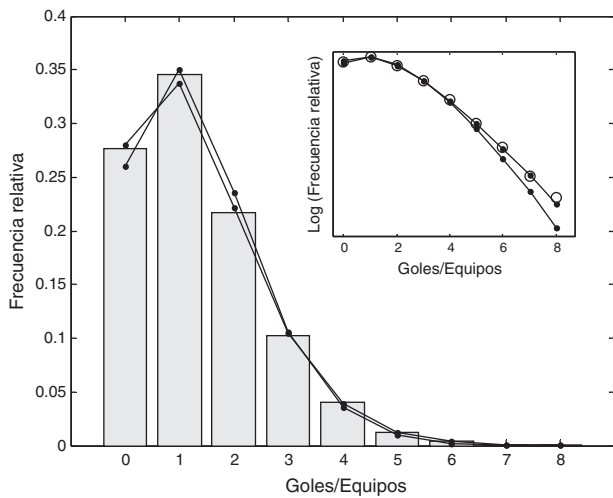


Figura 2. Número y frecuencia relativa de goles por equipos. Con línea continua se representa el ajuste de Poisson y con línea de puntos el ajuste de la distribución Binomial Negativa. En el recuadro la representación semilogarítmica, logaritmo de las frecuencias relativas, para ver con mayor precisión lo que sucede en la cola de la distribución.

(línea continua). En el caso de los equipos, se aprecia un aumento importante del *ID* a partir de la temporada 2005/06, con valores por encima de 1, (ajuste lineal: $a = 0.0162 \pm 0.009$; $b = 0.9952 \pm 0.0715$; $R^2 = 0.588$; $p\text{-value} = 0.002$). No se observa el mismo comportamiento en el caso de los partidos donde el índice se mantiene cerca del valor 1 (ajuste lineal: $a = 0.0099 \pm 0.0097$; $b = 0.9622 \pm 0.077$; $R^2 = 0.316$; $p\text{-value} = 0.045$).

En las gráficas inferiores (fig. 1c y fig. 1d), se muestran los valores de las medias (\cdot) y de las varianzas (\circ) calculadas para cada temporada, por equipo (1c) y por partido (1d). En el caso de los equipos, se observa como también a partir de la temporada 2005/06 la diferencia entre los valores de la varianza y la media aumenta durante la últimas temporadas.

Para analizar en detalle el promedio de goles que se consiguen durante las 13 temporadas en cada partido y por cada equipo se muestran las correspondientes distribuciones con sus correspondientes ajustes: Poisson y *BN*.

La figura 2 muestra las frecuencias relativas al número de goles anotados por cada equipo en las temporadas señaladas. Con línea continua, se muestra el ajuste de la *DP* y con línea de puntos la correspondiente a la *BN*. En el recuadro se muestra la misma distribución en modo semilogarítmico para mostrar mejor el comportamiento en la cola de la distribución. Si bien, las 2 distribuciones se ajustan bien entre los valores 0-4 goles, a partir del quinto gol la distribución *BN* es un mejor modelo de ajuste, ya que como se ve los últimos valores de la cola tienen probabilidades mayores de producirse que las predichas por el modelo de Poisson.

El resultado en el caso de los goles anotados en cada partido, es decir sumando los anotados por los 2 equipos rivales, se muestra en la figura 3. En ella se puede comprobar como, a diferencia del caso anterior, el modelo de Poisson parece representar bien esta distribución, no existiendo diferencias significativas entre esta distribución y la *BN* (recuadro interno).

Intervalos de tiempo transcurrido entre cada uno de los goles marcados en cada partido y los marcados por cada equipo en el total de temporadas. Una de las propiedades de la *DP* implica que los intervalos de tiempo de eventos sigue una distribución exponencial. Este tipo de distribución exige que los eventos sean independientes en el tiempo. Es decir, que la serie carezca de memoria y que no se pueda predecir un resultado a partir del anterior (memoriless).

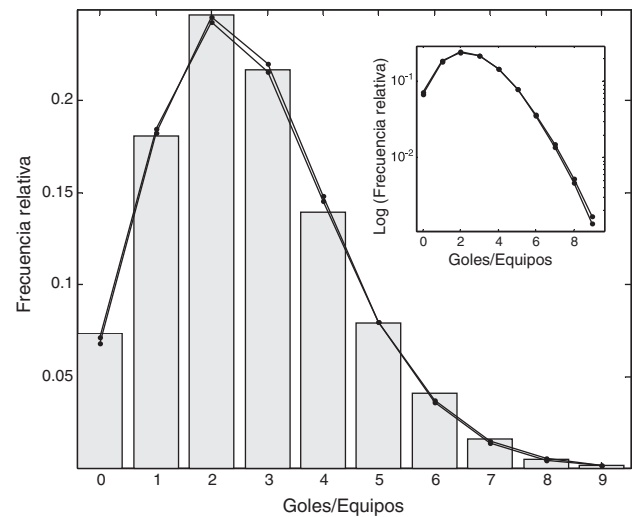


Figura 3. Número y frecuencia relativa de goles por partido. Con línea continua se muestra el ajuste de Poisson y con línea de puntos el ajuste de la distribución Binomial Negativa. En el recuadro aparece la representación semilogarítmica, logaritmo de las frecuencias relativas, para poder ver más claramente lo que sucede en la cola de la distribución.

Para ello, a partir de los minutos en los que se ha marcado cada tanto en las temporadas consideradas, hemos calculado las diferencias de tiempo (*dt*) en 3 casos: i) diferencias de tiempo de goles marcados por cada equipo en los 90 min de cada partido; ii) diferencias totales de tiempo por equipo en partidos sucesivos (es decir, tiempo transcurrido entre un tanto y el siguiente en encuentros sucesivos); iii) diferencias de tiempos en cada partido por cualquiera de los equipos.

En la figura 4 se muestran el histograma de las *dt* para el caso i). El recuadro representa el gráfico semilogarítmico del histograma. El comportamiento de este fenómeno deportivo es un indicador del carácter poissoniano de este fenómeno deportivo.

Sin embargo, hay 3 puntos en los minutos 3, 45 y 87 en los que el ajuste no es bueno. En los minutos 1-6, a diferencia de los que ocurre del 7-12, no es tan frecuente el marcarse un tanto. En los minutos 45 y 90 el fenómeno es más frecuente, aunque debe ser tenido en cuenta que en estos minutos (final de primer y segundo tiempo) se suman los tantos marcados en los correspondientes minutos de descuentos. No obstante, su incremento lo entendemos trascendente para comprender lo que realmente ocurre en los momentos

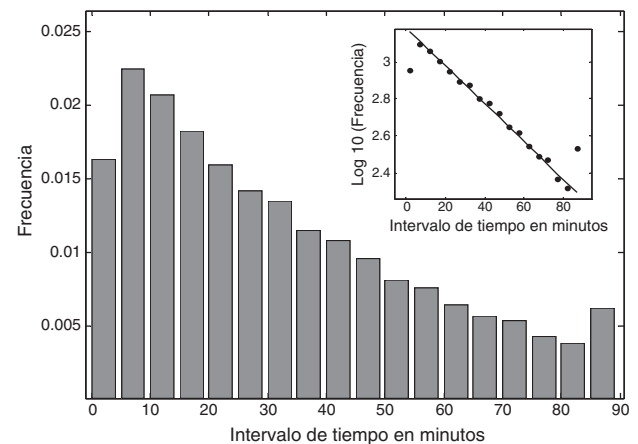


Figura 4. Histograma de las diferencias de tiempo entre goles marcados por cada equipo en cada partido, desde 0 hasta 90 minutos. El logaritmo de las frecuencias absolutas y el ajuste lineal en el recuadro. Nótese el comportamiento exponencial de la distribución.

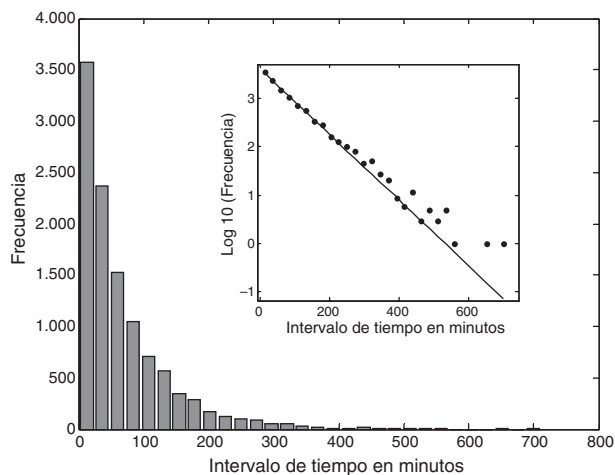


Figura 5. Histograma de las frecuencias absolutas (eje Y) respecto a las diferencias de tiempo entre goles marcados por cada equipo en partidos sucesivos (eje X). En el gráfico interior se representa el semilogarítmico de las frecuencias absolutas (eje Y) frente al tiempo que transcurre entre cada gol (eje X). El tiempo que transcurre entre goles se acumula también de un partido al siguiente.

finales de cada tiempo y, especialmente, en los últimos minutos del partido. Como vemos, el ajuste lineal ($a = -0.0114 \pm 0.0021$; $R^2 = 0.896$; $p\text{-value} = 0.0000$; error de la varianza = 0.0002) parece bastante adecuado, a simple vista, menos en el último valor donde se detecta un claro repunte.

La distribución de las diferencias de tiempo para el caso ii) se representan en la figura 5. Téngase en cuenta que el tiempo que transcurre entre los goles que un equipo marca se acumula también de un partido al siguiente. Es decir, si por ejemplo un equipo marca un tanto en el minuto 20 de un partido y el siguiente tanto lo marca en el minuto 80 del partido siguiente, se contabilizaría un tiempo total entre ambos tantos ($dt = 70 + 80 = 150$ min).

Nótese en estos casos que el ajuste lineal, es decir, el comportamiento exponencial, es bueno ($a = 0.0068 \pm 0.0005$; $b = 0.993 \pm 0.002$; $R^2 = 0.994$; $p\text{-value} = 0.000$; error de la varianza = 0.0012) hasta el valor 200 (minuto 200). Ya señalamos que una característica del modelo exponencial es la falta de memoria. Es decir, el tiempo que ha de transcurrir hasta que se produzca el evento siguiente, no depende del tiempo transcurrido para el evento anterior. Este fenómeno parece cumplirse para tiempos menores a 200 min, es decir, en torno a 2 partidos. Para diferencias de tiempos mayores, a medida que aumenta el valor de dt también aumenta la probabilidad del suceso, lo cual parece indicar que en estos casos sí existe un efecto de memoria. Esto nos está indicando que cuanto más tiempo se pasa sin anotar, mayor es la probabilidad de que no se anote en el siguiente partido. Este efecto, en la teoría de la complejidad, es conocido como *efecto Mateo* o *Preferential Attachment Process* (fig. 6).

El caso iii) es análogo al caso i), con un comportamiento también exponencial en las diferencias de tiempo ($a = 0.0167 \pm 0.0013$; $R^2 = 0.978$; $p\text{-value} = 0.000$; error de la varianza = 0.0046).

Discusión

El aspecto más destacable de este trabajo es observar como el valor de S_N decae desde el año 2000 (fig. 1a). Esto significa una pérdida de competitividad de la principal liga española durante el periodo evaluado que también fue constatado por Montes-Suay y Sala-Garrido¹³ entre las temporadas 2002/2003-2011/2012 con otro tipo de metodología (índice de Gini y test de Montecarlo).

El nivel de competitividad es un factor clave en la incertidumbre del resultado y, en consecuencia, en el interés que la modalidad

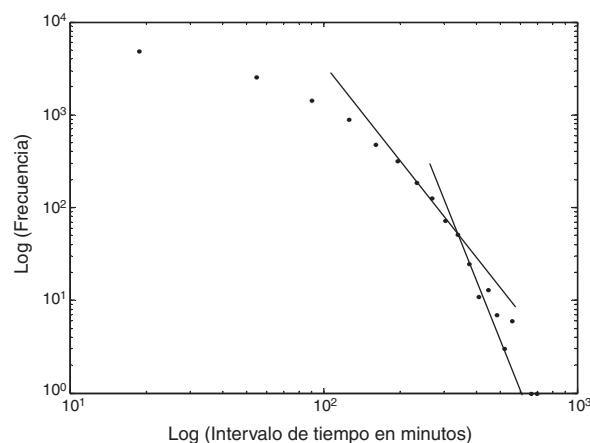


Figura 6. Valores doblemente-logarítmicos (log-log plot) de los valores de frecuencia y la justa de tiempo entre los goles marcados por cada equipo. Incluye los ajustes lineales que muestran la existencia de 2 Leyes de Potencia que se generan entre los minutos 200-400 y por encima de los 400 minutos.

deportiva provoca entre los diferentes actores (practicantes, espectadores, directivos, sponsors, medios de comunicación, etc.). La competitividad tiene un claro componente multidimensional^{14,15} pero sin duda, en el caso del deporte, se manifiesta sobre todo en la igualdad o desigualdad existente entre los competidores (equipos)¹⁶.

El caso de la BBVA, la pérdida de incertidumbre es especialmente significativa en la temporada 2009/10 donde se dieron 2 circunstancias: dominio absoluto de 2 equipos en la alcanza casi clasificación con menos de 42 puntos. Los desequilibrios económicos entre los clubes como consecuencia de una distribución poco equitativa de los recursos (i. e. ingresos por TV), el aumento de los costes de participación y formación de plantillas competitivas y la disminución progresiva de recursos, pueden ser 3 de las principales causas que subyacen detrás de este comportamiento^{17,18}. Estos desequilibrios afectan directamente a la calidad de las plantillas de cada equipo. Un ejemplo claro lo tenemos en la temporada 2013/2014 donde el valor de mercado de las plantillas del Real Madrid y F. C. Barcelona superaban los 580 millones de euros mientras que las plantillas de 8 equipos (Celta de Vigo, Getafe C. F., Elche C. F., Rayo Vallecano, C. A. Osasuna, Levante U. D., Real Valladolid y U. D. Almería) no llegaban a los 50 millones de euros¹⁹.

Un reflejo en la caída de competitividad lo vemos en el número de goles que se logran en las últimas temporadas²⁰. En ellas cada vez son más frecuentes los encuentros con un elevado número de goles. Esto queda reflejado en la evolución que muestra el *ID*. El *ID* (fig. 1b), tanto para equipo como por partido, normalmente estaba próximo a 1 hasta la temporada 2009/10. A partir de ese momento ambos índices se separan mostrando una tendencia incremental para el *ID* por equipo. Es decir, el *ID* por equipo se aleja de la *DP*, como consecuencia de que los equipos dominan en la tendencia a marcar más goles a su varianza. Esto también se refleja en las gráficas 1c y 1d, donde a la varianza en el número de tantos aumenta en el caso de goles marcados por equipo. Llama la atención que el *ID* por partido disminuye en las temporadas 2001/02, 2005/06 y 2009/10 ($ID < 1$). Curiosamente, estas temporadas finalizan con la celebración de la Copa del Mundo de selecciones nacionales (Corea-Japón'02; Alemania'06 y Sudáfrica'10). Es decir, el nivel de competitividad de la principal Liga española de fútbol aumenta. Es difícil, la relación causa-efecto, aunque se podría pensar que al ser una liga en la que participan muchos jugadores de diferentes países, el interés por ser convocados por sus selecciones nacionales aumenta su nivel juego y la igualdad entre los equipos

de la Liga. No obstante, esto sólo debe interpretarse como una mera especulación que debe ser contrastada en otros estudios.

Si bien la distribución del número de tantos por partido es siempre cercano a una DP^{3-7} , queda claro que se debe distinguir el comportamiento de los goles conseguidos en cada partido y en cada equipo. Respecto al ID por equipo, especialmente en las últimas temporadas, vemos como el valor de ID es mayor que 1 y, por lo tanto, la BN parece mostrarse como un mejor ajuste de la distribución (tabla 1). Los histogramas de tantos por equipo y partidos mostrados en las figuras 2 y 3 muestran que la distribución de tantos por equipo responde a una distribución BN que se hace especialmente evidente a partir de los 4 o 5 goles conseguidos. En esta situación, la probabilidad de que un determinado equipo supere los 5 goles en un partido es superior a la predicha por el modelo de Poisson. Sin embargo, el número de goles por partido parece seguir casi siempre este tipo de distribución.

No menos interesante, resulta analizar el tiempo que transcurre entre cada gol. Su valor nos muestra parte del nivel de efectividad de los equipos. En los intervalos de tiempo transcurrido entre los goles distinguimos 2 casos. Si atendemos a diferencias de tiempo en cada partido, esto es, entre los 90 min de juego que marca el reglamento, parecen seguir con claridad distribuciones exponenciales tanto por los goles conseguidos por equipos como los marcados en cada partido. Pero si consideramos el tiempo total transcurrido entre los goles marcados por los equipos, de partido a partido, la situación cambia. Hasta unos 200 min, es decir, algo más de 2 partidos, el proceso sigue una exponencial y puede considerarse un proceso poissoniano, pero a partir de este tiempo, la probabilidad de que un equipo siga sin marcar es mayor que la prevista por la distribución exponencial, lo que parece indicar cierto efecto de memoria. Es decir, la distribución exponencial se basa en el hecho de la independencia de sucesos seguidos, es decir, de no memoria. Sin embargo, lo que parece mostrar este estudio es que hay un cierto efecto de feedback positivo, en el sentido de que cuando a un equipo le va mal, y entra en una fase de bajo rendimiento, es muy probable que le siga yendo mal y conseguir marcar le resulte altamente complicado y entren en una dinámica perdedora. Si tenemos en cuenta que hay equipos que no ceden puntos (equipos muy dominantes), el resto de equipos deben buscarlos entre el resto de rivales, y, por lo tanto, aprovechar cualquier debilidad en los demás para conseguir anotar.

Además, es un hecho que cuando a un equipo le va mal, pueden darse una serie de factores (entrenadores, plantilla, ambiente, etc.) que dificultan que los equipos que entran una dinámica negativa puedan salir de ella.

En conclusión, la Liga profesional española de máxima categoría (Liga BBVA) ha perdido competitividad en las primeras 13 temporadas del siglo XXI, tal y como refleja la caída de la S_N y en el comportamiento del ID entre equipo-partido, especialmente en las últimas temporadas evaluadas. Desde el punto de vista de los equipos, deja de ser poissoniana, especialmente a partir de 2008/09,

y sobre todo en la temporada siguiente. No ocurre lo mismo si el análisis lo hacemos desde el punto de vista de partidos. La elevada diferencia entre equipos puede ser una de las principales causas de que exista una tendencia creciente a encontrar partidos con un elevado número de goles. El número de minutos, o partidos, que un equipo tarda en conseguir un gol parece ser un buen parámetro para caracterizar los resultados y el comportamiento de los goles en los enfrentamientos que tienen lugar en la Liga BBVA.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

- Anderson D, Sally C. The numbers game: Why everything you know about football is wrong. New York: Penguin Group; 2013.
- Lago C. Ganar o perder en el fútbol de alto nivel: ¿Una cuestión de suerte? Eur J Hum Mov. 2005;14:135–50.
- Karlis D, Ntzoufras I. Analysis of sports data by using bivariate Poisson models. Statistician. 2003;52:381–93.
- McHale I, Scarf P. Modelling the dependence of goals scored by opposing teams in international soccer matches. Stat Modelling. 2011;11(3):219–36.
- De Saá-Guerra R, Martín-González JM, Sarmiento-Montesdeoca S, Rodríguez-Ruiz D, García-Rodríguez A, García-Manso JM. A model for competitiveness level analysis in sports competitions: Application to basketball. Physica A. 2012;391:2997–3004.
- Gabel A, Redner S. Random walk picture of basketball scoring. J Quant Anal Sports. 2012;8:1–20.
- Sánchez-Flores J, García-Manso JM, Martín-González JM, Ramos-Verde E, Arriaza-Ardiles EJ, da Silva-Grigoletto ME. Análisis y evaluación del lanzamiento de esquina (córner) en el fútbol de alto nivel. Rev Andal Med Deporte. 2012;5:140–6.
- Dyte D, Clarke SR. A rating based Poisson model for World Cup soccer simulation. J Oper Res Soc. 2000;51:993–8.
- Greenhough J, Birch PC, Chapman A. Soccer goal distributions and extremal statistics. Physica A. 2002;316:615–24.
- Chu S. Using soccer goals to motivate the Poisson process. ITE. 2003;3(2):64–70.
- Bittner E, Nußbaumer A, Janke W, Weigel M. Football fever: Goal distributions and non-Gaussian statistics. Eur Phys J B. 2009;67:459–71.
- Heuer A, Mueller C, Rubner O. Soccer: Is scoring goals a predictable Poissonian process? Europhys Lett. 2010;89:38007–12.
- Montes-Suay F, Sala-Garrido R. ¿Fue alguna vez equilibrada la liga de la primera división española de fútbol? Anales de ASEPUMA. 2013;21:27–33.
- Houston J, Harris P, McIntire S, Francis D. Revising the competitiveness index using factor analysis. Psychol Rep. 2002;90:31–3.
- El-Hodiri M, Quirk J. An economic model of a professional sports League. J Polit Econ. 1971;79:1302–19.
- Levin RC, Mitchell GJ, Volcker PA, Will GF. The Report of the Independent Members of the Commissioner's Blue Ribbon Panel on Baseball Economics. New York: Major League Baseball; 2000.
- Gay de Liébana JM, Saludas JM, Mielgo FD, Moya DC. Liga de las Estrellas 2009/10: El fútbol español bajo las turbulencias de la crisis. Radiografía del estado patrimonial, situación financiera y posición económica de los clubes integrantes de la Liga BBVA. Temporada 2009/2010. Universidad de Barcelona. 2011.
- Gay de Liébana JM. Fútbol europeo, ¿finanzas inestables? ¿Modelo económico sostenible? Harv-Deusto Bus Rev. 2014;235:52–62.
- Transfermarkt. Es [sede Web]. Hamburg: Transfermarkt GmbH & Co. KG; [accedido 16 Sep 2015]. Disponible en: <http://www.transfermarkt.es/>
- Palacios-Huerta I. Structural changes during a century of the world's most popular sport. SMA. 2004;13:241–58.