




LÓGICA ESTADÍSTICA DE LA INTERACCIÓN DE MARCA EN LOS PLAYOFFS DE LA NBA (2014-2019): FACTORES ESPACIALES Y TEMPORALES

Statistical logic of scoring interaction in NBA playoffs (2014-2019): The influence of time and space

Raúl Martínez-Santos ¹, * , Asier Oiarbide ¹, , Mario Enjuanes ¹, 


¹ Universidad del País Vasco, UPV/EHU, España.

* Correspondencia: raul.martinezdesantos@ehu.es

Recibido: 26/01/2021; Aceptado: 16/06/2021; Publicado: 20/12/2021

OPEN ACCESS

Sección / Section:
Análisis del rendimiento deportivo /
Performance analysis in sport

 Editor de Sección / Edited by:
Sebastián Feu, Universidad de
Extremadura, España

Citación / Citation:
Martínez-Santos, R., Oiarbide, A., &
Enjuanes, M. (2021). Lógica
estadística de la interacción de
marca en los playoffs de la NBA
(2014-2019): factores espaciales y
temporales. *E-balonmano Com*,
17(3), 233-240.

Fuentes de Financiación / Funding:
-

Agradecimientos/
Acknowledgments:
-

Conflicto de intereses / Conflicts of
Interest: NO

Resumen

El baloncesto es un deporte de engaños, pero también de números, y el conocimiento de su lógica interna se ha beneficiado desde hace décadas con las aportaciones de colegas como Sampaio e Ibáñez-Godoy. Aceptando que los playoffs de la NBA son la competición más exigente de nuestro deporte, nos planteamos dos objetivos complementarios: aportar las probabilidades asociadas a los lanzamientos de campo en función del espacio y el tiempo, y mostrar el camino seguido en el entorno R.

Nuestros análisis nos permiten aportar las distribuciones de probabilidad asociadas a los lanzamientos de campo en función del momento de juego y de la distancia al aro, como muestra la tabla 1. Además, mediante sencillos modelos *logit* podemos constatar y mensurar el impacto que las distancias y los momentos del partido (periodo y minuto) tienen sobre la probabilidad de acierto.

Los mejores jugadores de baloncesto del mundo maximizan la utilidad de sus lanzamientos sin poder escapar de la lógica espacio-temporal del juego: la relación de fuerzas entre el ataque y la defensa sale a la luz en forma de funciones probabilísticas con una alta linealidad, sobre todo en lo que respecta al espacio, y apunta hacia interesantes cuestiones prospectivas. En este sentido, nuestro estudio proporciona bases estadísticas generales para la comprensión de esta lógica, permitiendo en un futuro avanzar en modelos de estudio bayesianos mejor informados.

Palabras clave: baloncesto, lógica interna, probabilidad, lanzamiento

Abstract

Basketball is a sport of deceptions, but also of numbers, and the knowledge of its internal logic has benefited for decades from the contributions of colleagues such as Sampaio and Ibáñez-Godoy. Accepting that NBA playoffs are the most demanding competition in our sport, we set two complementary goals for this study: to provide the a priori probabilities associated with field shots based on space and time, and to show the way we followed to calculate them in the R environment.

Our analyses allow us to provide the probability distributions associated with field shots based on clock-time and court position, as shown in Table 1. In addition, by simple logit models we can see and measure the impact that distances and moments during the match (period and minute) have on the probability of success.

The best basketball players in the world maximize the utility of their shots without being able to escape the space-time logic of the game: the balance between attack and defense comes to light in the form of probabilistic functions with high linearity, especially in terms of space, and points to interesting forward-looking questions. In this sense, our study provides general statistical basis for understanding this logic, allowing in the future to advance in better informed, bayesian study models.

Key words: basketball, internal logic, probability, shooting

Introducción

Como todo el mundo sabe, hay una relación inversa entre la distancia al aro desde la que se consigue lanzar y la probabilidad de acierto: a mayor distancia menor acierto. Se puede afirmar también que hay una fuerte relación, aunque directa en este caso, entre la probabilidad de lanzar desmarcado y la distancia al aro, y que ambas relaciones se articulan para dar forma a la lucha por el control del espacio que lleva a los entrenadores a organizar sistemas defensivos (individuales, zonales, mixtos...) e inventar infinitas formas de saltárselos. En este sentido, podemos intentar sacar a la luz *los códigos secretos* (Oboeuf, Buffet, Collard, Martínez-Santos, & Lecroisey, 2019; Oboeuf, Collard, & Gerard, 2008) de estos juegos tomando como indicador privilegiado del contenido de las decisiones motrices de los jugadores *la distancia de tiro* desde la que se producen los lanzamientos.

Este estudio sobre la selección de lanzamiento en los playoffs de la NBA de las últimas cinco temporadas pre-COVID forma parte de la ingente cantidad de aportaciones que tienen por objeto la *lógica estadística* del baloncesto. Es un abordaje ciertamente alejado del estudio del *principio del engaño* que vincula a los adversarios, esa «esencia del juego» que Koppet (1973) identificó hace ya tanto, pero no tanto como para olvidar que las probabilidades asociadas a los comportamientos posibles es la clave de las creencias y las conductas de los jugadores de máximo nivel. El baloncesto es un deporte de «imputación colectiva» (Martínez-Santos, 2007): las marcas obtenidas a partir de los actos individuales (interacciones de marca cuyo valor en caso de acierto es 1, 2 o 3 puntos) se asigna al equipo de pertenencia. Aunque hay otros contadores fundamentales para la dinámica sociomotriz (la faltas o el mismo reloj), los tantos obtenidos son la *memoria fundamental* del juego, y todas las decisiones se articulan alrededor del espacio privilegiado de la canasta, asediado y continuamente atacado por el equipo en posesión del balón, numantamente defendido por el equipo que lo quiere recuperar.

Este tipo de indagación conlleva un riesgo cierto de pérdida de validez ecológica: ¿hasta qué punto existe el objeto *baloncesto estadístico*? ¿Hasta qué punto los signos comportamentales y situacionales registrados (posiciones «x» e «y», estados de marcador y relojes...) permiten explicar suficientemente las dinámicas de afectivo-cognitivas implicadas en los procesos de exploración, decisión y hasta intimidación que se ponen en marcha en la competición del más alto nivel? ¿En qué medida podemos tomar decisiones didácticas o estratégicas a partir de las regularidades numéricas que, sin duda, salen continuamente a la luz gracias a ellos? Cualquier investigación asume que sus modelos, representaciones matematizadas de la realidad, son simplificaciones, y nuestro caso no es diferente, aunque contamos con el respaldo de una amplia base de datos que, sin llegar al *big data*, nos permite confiar en que los resultados obtenidos en forma de funciones de distribución de probabilidades tengan una alta capacidad interpretativa. Es más: la estadística existe, precisamente, para encontrar similitudes *significativas desde el punto de vista humano* en sucesos y eventos que son, *empíricamente únicos*, y los trabajos de colegas como Jaime Sampaio (Sampaio, Drinkwater, & Leite, 2010; Sampaio, Godoy, & Feu, 2004; Sampaio, Gonçalves, Mateus, Shaoliang, & Leite, 2018; Sampaio et al., 2015) y Sergio Ibáñez-Godoy (García, Ibáñez, De Santos, Leite, & Sampaio, 2013; Ibáñez, Feu, García, Parejo, & Cañadas, 2009; Ibáñez, García, Feu, Lorenzo, & Sampaio, 2009; Ibáñez et al., 2008) son una muestra de excelente de este proceder. La cuestión es, por tanto, si los jugadores en su conjunto, conocedores de estas funciones de probabilidad individuales y colectivas, muestran en sus decisiones de lanzar, en un momento dado y desde un lugar concreto, alguna tendencia *estadísticamente significativa* que nos permita describir *la lengua que hablan* cuando juegan al máximo nivel (Wallace, Caudill, & Mixon, 2013).

Nuestro objetivo es doble, y ambos objetivos son igual de importantes para nosotros: por un lado, nos proponemos poner a prueba la hipótesis de que las interacciones de marca de campo están moduladas espacialmente por la probabilidad de acierto asociada a la distancia de tiro, y temporalmente por el tiempo restante de partido; por otro lado, queremos compartir el camino seguido para la obtención y explotación de los datos en R (R Core Team, 2020), un entorno poco conocido y utilizado en nuestros estudios a pesar de su inmenso potencial.

Método

Muestra

Para este estudio se ha completado una muestra de 69.252 lanzamientos de campo registrados por el servicio oficial de estadísticas de la NBA en 410 partidos de la fase de playoff de cinco temporadas (de la 2014-15 a la 2018-19).

Obtención y preprocesamiento de los datos

El objetivo de esta fase fue validar los datos y contextualizarlos. Entendemos por *dato contextualizado* cualquier evento del juego registrado en el espacio y el tiempo internos de cada partido. Cada evento es único, aunque el análisis estadístico los agrupará en función de las preguntas pertinentes. La validación busca asegurar que los registros empleados satisfacen las condiciones de inclusión, y la calidad observacional se presupone.

Los registros fueron obtenidos el día 9/19/2020 desde la página oficial <https://stats.nba.com> por medio del paquete de *nbstatR* (Bresler, 2020). Fueron empleados además los paquetes *tidyverse* (Wickham et al., 2019) y *magrittr* (Bache & Wickham, 2014). Para confeccionar la tabla de datos final fue necesario manejar varias tablas primarias:

- *Equipos de la NBA*, para lo que se empleó la función *nbstatR:: nba_teams()*, dando como resultado los 30 equipos actuales en activo tras ser filtrados según *isNonNBATeam == 0*.
- *Partidos por temporadas*: mediante la función *nbstatR:: seasons_schedule()*, con los parámetros *seasons=2015:2019* y *season_types= c("Pre Season", "Regular Season", "Playoffs", "All Star")*, dando como resultado un total de 7.050, de los que 410 (5,81%) son de los playoffs.
- *Lanzamientos de campo*; mediante la función (*nbstatR::teams_shots()*, con los parámetros *all_active_teams = T*, *season_types = "Playoffs"*, *seasons = 2015:2019*, *nest_data = F*), dando como resultado un total de 69.252 lanzamientos de campo caracterizados por 27 variables.

De cara al análisis estadístico, se decidió trabajar con variables discretas de espacio y tiempo (Grubis; Papalexakis & Pelechrinis, 2018; Reich, Hodges, Carlin, & Reich, 2006; Shortridge, Goldsberry, & Adams, 2014): con respecto a la distancia de lanzamiento reservamos la variable *zoneRange* que la divide en rangos de 8 pies (255 cm): *Less Than 8 ft.*, *8-16 ft.*, *16-24 ft.*, *24+ ft.* y *Back Court Shot*, aunque decidimos crear un nivel más para diferenciar los tiros en campo de ataque más alejados de 32 pies: *24-32 ft.* y *32 ft-HalfCourt*.

El último paso de esta fase consistió en: seleccionar las variables espaciales de los lanzamientos (*nameZone* y *zoneRange*) y las variables temporales del reloj (*numberPeriod* y *minutesRemaining*). El resultado fue un marco de datos con 69.252 registros y 14 variables.

Procesamiento de los datos

Los análisis exploratorio y predictivo fueron realizados con las funciones estadísticas básicas de R 4.0.3 (10/10/2020). El paquete *visreg* (Breheny & Burchett, 2017) fue empleado para la elaboración de los modelos *logit*.

Resultados y discusión

El objetivo empírico de este estudio fue sacar a la luz las funciones de distribución de probabilidad asociadas a los lanzamientos de campo en el baloncesto. Las «interacciones de marca» (Parlebas, 2001, p. 267), con independencia de su resultado, son el signo externo más importante de la acción de juego, siempre y cuando se pongan en relación con los ejes espacio-temporales de la propia situación: el punto de lanzamiento, el estado del marcador y el tiempo de juego restante. En este sentido, tomamos como objeto de estudio la máxima expresión del juego, la NBA, y como máxima expresión de esta competición la fase de *playoff*, más concretamente todos los partidos de la fase final de las últimas cinco temporadas previas a la COVID.

El número total de lanzamientos de campo fue 69.252, 66,15% de 2 puntos y 33,85% de 3 puntos. La tasa de acierto global fue de 44,66%: 49,55% para los tiros de 2 y 35,10% para los tiros de tres. La tabla 1 muestra los porcentajes de acierto desde cada tramo de distancia en cada temporada. Como sabemos, los tiros laterales de tres se hacen desde 22 pies, por lo que en esta tabla aparecen mezclados ambas categorías. Sin embargo, es sencillo calcular el impacto del valor del «paso de la marca» (Parlebas, 2001, p. 346) sobre el acierto de los lanzadores desde ese tramo de 22-24 pies (6,71-7,31 m): la probabilidad de acierto de los tiros de 2 es de .4112, y de .388 de los tiros de 3, una diferencia no significativa ($\chi^2(1, N = 5.307) = 0,765, p = 0,381$).

Tabla 1. probabilidades de acierto en cada tramo de distancia de lanzamiento

temporada	Distancia de tiro (pies)*						N
	0-8ft	8-16ft	16-24ft	24-32ft	32-HC	BCShot	
2014/15	.539	.380	.376	.339	.108	.074	13.689
2015/16	.535	.409	.390	.350	.047	.000	14.285
2016/17	.578	.407	.404	.357	.033	.000	13.203
2017/18	.583	.405	.396	.341	.149	.060	13.816
2018/19	.557	.411	.390	.345	.138	.000	14.259
	.558	.403	.391	.347	.102	.025	69.252

* Los puntos de corte son los siguientes: 2,44 m, 4,88 m, 7,32m, 9,75, medio campo y campo propio.

La figura 1 muestra en detalle esta relación en cualquier momento del partido, sin tener en cuenta, claro está, que recibir falta durante el tiro también es un resultado positivo. El modelo lineal general muestra una tendencia dentro de los márgenes de tiro habituales (menos de 9 metros), y asintótica a medida que las distancias quedan fuera de la lógica del entrenamiento y la competición (lanzamientos excesivamente lejanos). El modelo logit toma como variable continua independiente la distancia de tiro en metros y como dependiente la binomial «acierto/fallo», resultando una constante para la variable dependiente de 0.317 y un valor de $\beta = -0.135$ para la independiente: las razones de probabilidad indican que por cada acierto hay 1,373 fallos y por cada metro de separación del aro la probabilidad de acierto se ve afectada a razón de 0,873 por unidad. Con todo, lo más sorprendente es que intervalo de confianza ($p = .05$) resulta inapreciable, lo que indica la tremenda fortaleza de la relación.

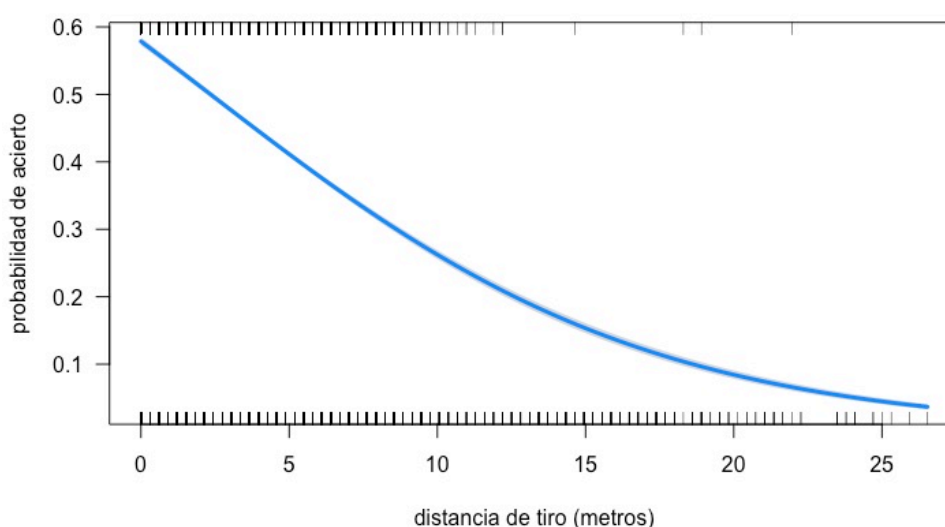


Figura 1. Estimaciones de probabilidad de acierto en función de la distancia al aro.

Con respecto al tiempo, disponemos de dos variables que nos permiten contrastar el impacto que sobre las decisiones y aciertos tiene la proximidad del final del partido en estos encuentros al máximo nivel: el periodo de juego (1-4 y prórrogas) y el minuto de cada periodo. La tabla 2 muestra las distribuciones de tiros y aciertos por periodos. El contraste estadístico ($\chi^2(7, N = 69.252) = 69,998, p = 0.000$) indica diferencias en la distribución espacial de los tiros en los periodos registrado, prórrogas incluidas, una diferencia que se mantiene para los cuatro cuartos ($\chi^2(3, N = 69.252) = 62,234, p = 0.000$), aunque la V de Cramer resulta ser $V=0.03$. Los resultados de esta tabla 2 permiten dos apuntes: que la ingente cantidad de datos muestra diferencias estadísticas con un tamaño de efecto mínimo, y que en los periodos con pocas ocurrencias la distribución es muy particular, alejada de la función de distribución general.

Tabla 2. Probabilidades de ocurrencia y acierto en los lanzamientos de campo en función del rango de distancia por cada periodo de juego registrado.

periodo	Distancia de tiro (pies)*						N
	0-8ft	8-16ft	16-24ft	24-32ft	32-HC	BCShot	
1	.263/.546	.271/.392	.266/.403	.245/.356	.252/.037	.333/.188	18.206
2	.259/.568	.253/.427	.266/.386	.240.349	.243/.077	.289/.021	17.424
3	.240/.562	.242/.390	.239/.402	.252/.357	.290/.145	.333/.037	16.860
4	.231/.558	.226/.401	.238/.372	.255/.328	.210/.155	.038/.000	16.450
5	.001/.506	.006/.423	.006/.381	.006/.241	.000/NaN	.006.000	408
6	.0002/.555	.0007/.000	.0005/.500	.0006/.272	.0047/.000	.000/NaN	52
7	.0002/.600	.0003/.333	.0002/.666	.0002/.500	.000/NaN	.000/NaN	15
8	.0003/.428	.0003/.666	.0003/.250	.0002/.333	.000/NaN	.000/NaN	17
acierto	1.0 /.558	1.0 /.403	1.0 /.391	1.0 /.347	1.0 /.102	1.0 /.025	
N	27.872	9.762	13.079	18.166	214	159	69.252
prob.	(.4024)	(.1409)	(.1888)	(.2623)	(.0030)	(.0022)	

Las celdas muestran la distribución empírica de los lanzamientos en cada tramo de distancia por periodos y la probabilidad empírica de acierto asociada.

* Los puntos de corte son los siguientes: 2,44 m, 4,88 m, 7,32m, 9,75, medio campo y campo propio.

La tabla 2 nos proporciona dos funciones espaciales de distribución de las interacciones de marca: la de probabilidad de ocurrencia (.4024, .1409, .1888, .2623, .0030, .0022) y la de probabilidad de acierto (.558, .403, .391, .347, .103, .025). estos vectores son las *probabilidades a priori* del mejor baloncesto del mundo, un rasgo de su lógica interna que podemos emplear como punto de partida para futuras investigaciones de corte estadístico. Estos vectores muestran, además, una organización de las decisiones particular: aunque la probabilidad de acierto disminuye linealmente en función de la distancia (del .558 al .347), la distribución de los lanzamientos no es lineal, ya que la zona central (de 8 a 24 pies) resulta menos atractiva que los extremos: si queremos lanzar de cerca, cuanto más cerca mejor porque se acierta más (.558 contra 4.03), y si queremos lanzar de lejos, cuanto más lejos mejor porque la reducción de acierto (de .391 a .347) se ve más que compensada con el punto extra en caso de acierto (una esperanza de 0,782 puntos contra una de 1,041).

Sin embargo, el espacio no nos cuenta la historia completa de los partidos de baloncesto. Como muestra la figura 2, hay una relación lineal inversa entre el tiempo restante en cada periodo y la probabilidad de acierto. La constante de la variable dependiente -0,260 y una $\beta = 0,00873$ permiten estimaciones con un margen de error mayor, como muestra el área correspondiente a los intervalos de confianza para $\alpha = ,05$. La probabilidad general de acierto disminuye a medida que nos acercamos al final del periodo (0 minutos restantes), aunque el gradiente acumulado (estadísticamente significativo) sea de un dos por ciento: aparentemente pequeño, pero definitivo al máximo nivel.

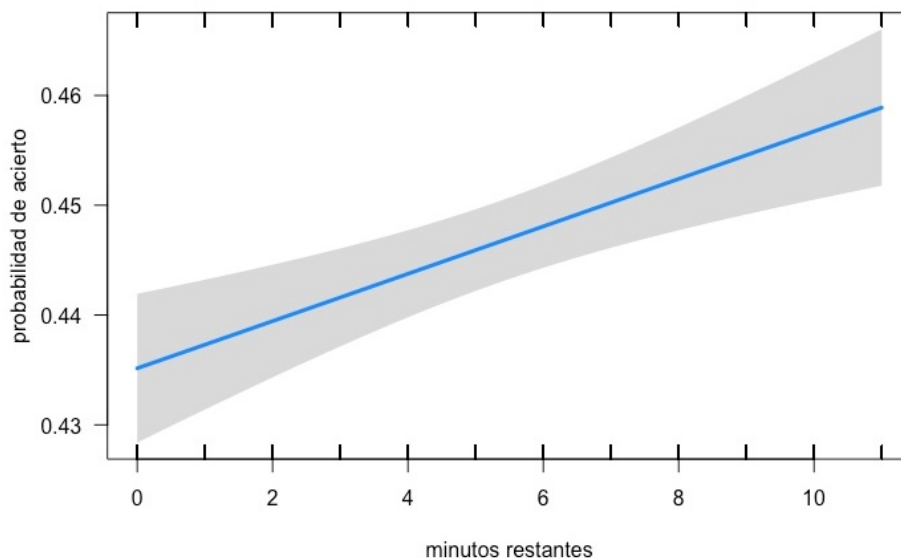


Figura 2. Estimaciones de probabilidad de acierto de los lanzamientos de en función de los minutos restantes para el fin del periodo.

Evidentemente, los lanzamientos tienen distintos valores, aunque la tendencia mostrada se mantiene, ajustada a la relación espacial ya descubierta en la figura 1, e incrementada por la mayor intensidad de los últimos cinco minutos. La tabla 3 y la figura 3 muestran este interesante impacto temporal en los últimos cinco minutos de los periodos de juego.

Tabla 3: Probabilidades de acierto en los últimos cinco minutos de cada periodo de juego para cada paso de marca.

<i>minutos restantes</i>	<i>valor de la marca</i>		<i>N</i>
	<i>2 puntos</i>	<i>3 puntos</i>	
0	.475 (4.080)	.283 (2.930)	7.010
1	.490 (3.778)	.334 (2.042)	5.820
2	.500 (3.694)	.362 (1.915)	5.609
3	.498 (3.829)	.375 (1.910)	5.730
4	.510 (3.868)	.377 (1.888)	5.756
	.494 (19.249)	.340 (10.685)	29.934

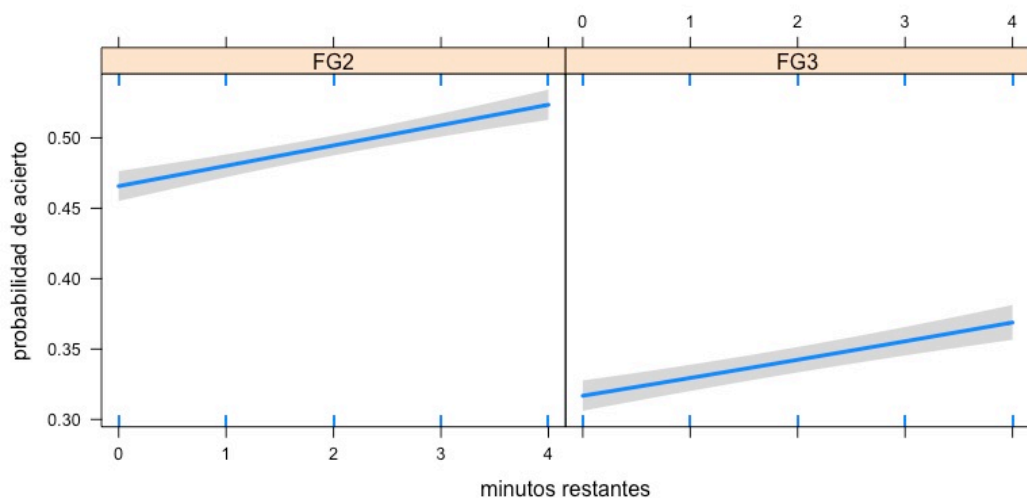


Figura 3. Estimaciones de probabilidad de acierto en los últimos cinco minutos en función del valor del lanzamiento.

El «soporte de marca» a tiempo límite (Parlebas, 2001, p. 435) se materializa en las decisiones de los jugadores, sobre todo en el último minuto, que acumula más interacciones de marca que el resto. Además, la decreciente probabilidad de acierto indica que la lucha de equipos se ve modificada por la cercanía del final del encuentro, como era de esperar. A modo de curiosidad final, cabe que decir que los porcentajes totales de acierto en los últimos cinco minutos de los últimos periodos de los partidos, prórrogas incluidas, son ligeramente inferiores que en los tres primeros cuartos: 48,51% en los tiros de 2 puntos y 31,96% en los tiros de tres por 49,72% y 34,75% respectivamente.

Conclusión

«R» es mucho más que un lenguaje de programación o un entorno de producción: R es una comunidad de investigadores en acción que comparten sin cesar herramientas avanzadas de gestión y análisis de datos de todo tipo. Este estudio muestra cómo operar eficientemente con la infinidad de datos estadísticos que las ligas profesionales de baloncesto producen cada día. Por lo que respecta al espacio, se prefieren las zonas más próxima y más alejada al aro, aunque por razones diferentes: acercarse al aro aumenta la probabilidad relativa de acierto en los tiros de dos puntos; desmarcarse lejos del aro puede resultar más sencillo porque la probabilidad de acierto es menor, pero la esperanza matemática del pago recibido en caso de acierto puede explicar por qué la proporción de interacciones de marca de tres puntos sigue aumentando (Caporale & Collier, 2015; Fichman & O'Brien, 2018; Smith, 2017). Por lo que respecta al tiempo, el sistema de tanteo se manifiesta a través de una consecuencia práctica que forma parte de la dinámica propia de todas las competiciones deportivas: la probabilidad de acierto disminuye a medida que se llega al momento final del encuentro.

Los datos han permitido mostrar que los mejores equipos de baloncesto del mundo maximizan la utilidad de los lanzamientos que sus jugadores realizan. Sin embargo, este modesto estudio descriptivo oculta una ambiciosa invitación: los vectores de probabilidades son los *a priori* con los que alimentar modelos bayesianos equivalentes (Yin, Hu, & Shen, 2020), que aspiren a predecir resultados (Alameda-Basora, 2019), a entrenar la selección de tiros (Hu, Yang, & Xue, 2020) o a reducir los sesgos arbitrarios (Constantinou, Fenton, & Hunter Pollock, 2014). «La teoría que nunca murió» (McGrayne, 2012) puede dar nueva vida a la investigación estadística del baloncesto.

La lógica interna del baloncesto es probabilística. La presencia de adversarios hace de la acción de juego del baloncesto un «proceso estocástico» (Collard, 1998) en el que la probabilidad de aparición de cada comportamiento (Obœuf, Collard, Pruvost, & Lech, 2009) debe ser matizada, además, por el grado de verosimilitud que se le conceden a las intenciones ajenas. La inteligencia táctica y la toma de riesgos propias de nuestro juego son bellos ejemplos de pensamiento bayesiano (Bach, 2014), por lo que no parece descabellado pensar que su investigación pueda adoptar también un esquema de pensamiento similar, si no más ajustado también a la propia indagación de máxima calidad que llamamos ciencia.

Referencias

- Alameda-Basora, E. M. (2019). A dynamic Bayesian network to predict the total points scored in national basketball association games.
- Bach, D. R. (2014). El cerebro bayesiano. *Mente y cerebro*(69), 70.
- Bache, S. M., & Wickham, H. (2014). *magrittr: A Forward-Pipe Operator for R*. Retrieved from <https://CRAN.R-project.org/package=magrittr>
- Breheny, P., & Burchett, W. (2017). Visualization of Regression Models Using visreg. *The R Journal*, 9(2), 56-71. Retrieved from <https://journal.r-project.org/archive/2017/RJ-2017-046/index.html>
- Bresler, A. (2020). R's interface to NBA data (Version 0.1.1503). Retrieved from <https://github.com/abresler/nbastatR>
- Caporale, T., & Collier, T. C. (2015). To Three or Not to Three? *Journal of Labor Research*, 36(1), 1-8.
- Collard, L. (1998). *Sports, enjeux et accidents*. Paris: PUF.
- Constantinou, A. C., Fenton, N. E., & Hunter Pollock, L. J. (2014). Bayesian networks for unbiased assessment of referee bias in Association Football. *Psychology of Sport and Exercise*, 15(5), 538-547. doi:<https://doi.org/10.1016/j.psychsport.2014.05.009>

- Fichman, M., & O'Brien, J. (2018). Three point shooting and efficient mixed strategies: A portfolio management approach. *Journal of Sports Analytics*, 4(2), 107-120.
- García, J., Ibáñez, S. J., De Santos, R. M., Leite, N., & Sampaio, J. (2013). Identifying basketball performance indicators in regular season and playoff games. *Journal of human kinetics*, 36(1), 161-168.
- Grubis, D. Dimensionality Reduction on NBA Shot Data.
- Hu, G., Yang, H. C., & Xue, Y. (2020). Bayesian group learning for shot selection of professional basketball players. *Stat*, e324.
- Ibáñez, S. J., Feu, S., García, J., Parejo, I., & Cañadas, M. (2009). Shot differences between professional (ACB) and amateur (EBA) basketball teams. Multifactorial study. *Revista de psicología del deporte*, 18(3), 313-317.
- Ibáñez, S. J., García, J., Feu, S., Lorenzo, A., & Sampaio, J. (2009). Effects of consecutive basketball games on the game-related statistics that discriminate winner and losing teams. *Journal of sports science & medicine*, 8(3), 458.
- Ibáñez, S. J., Sampaio, J., Feu, S., Lorenzo, A., Gómez, M. A., & Ortega, E. (2008). Basketball game-related statistics that discriminate between teams' season-long success. *European journal of sport science*, 8(6), 369-372.
- Koppet, L. (1973). *The essence of the game is deception*. Boston: Little, Brown & Company.
- Martínez-Santos, R. (2007). *La praxeología motriz aplicada al fútbol*. (PhD doctoral thesis). University of the Basque country, Vitoria-Gasteiz. Retrieved from http://www.ehu.es/argitalpenak/images/stories/tesis/Ciencias_Sociales/8779%20MtZDeSantosTH.pdf
- McGrayne, S. B. (2012). *La teoría que nunca murió: de cómo la Regla de Bayes permitió descifrar el código Enigma, perseguir los submarinos rusos y emerger triunfante de dos siglos de controversia*. Barcelona: Planeta.
- Oboeuf, A., Buffet, J., Collard, L., Martínez-Santos, R., & Lecroisey, L. (2019). Influence de l'apprentissage du « code secret » du jeu pour développer la créativité motrice. Une illustration en football chez des enfants de 8-9 ans. *Intellectia*, 71, 79-100.
- Oboeuf, A., Collard, L., & Gerard, B. (2008). Le jeu de la « balle assise »: un substitut au questionnaire sociométrique? *Les Cahiers internationaux de psychologie sociale*(1), 87-100.
- Oboeuf, A., Collard, L., Pruvost, A., & Lech, A. (2009). La prévisibilité au service de l'imprévisibilité. À la recherche du «code secret» du football. *Réseaux*(4), 241-270.
- Papalexakis, E., & Pelechrinis, K. (2018). *thoops: A multi-aspect analytical framework for spatio-temporal basketball data*. Paper presented at the Proceedings of the 27th ACM International Conference on Information and Knowledge Management.
- Parlebas, P. (2001). *Juegos, deporte y sociedad. Léxico de praxeología motriz*. Barcelona: Paidotribo.
- R Core Team. (2020). R: A language and environment for statistical computing
Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.
- Reich, B. J., Hodges, J. S., Carlin, B. P., & Reich, A. M. (2006). A Spatial Analysis of Basketball Shot Chart Data. *The American Statistician*, 60(1), 3-12. doi:10.1198/000313006X90305
- Sampaio, J., Drinkwater, E. J., & Leite, N. M. (2010). Effects of season period, team quality, and playing time on basketball players' game-related statistics. *European journal of sport science*, 10(2), 141-149.
- Sampaio, J., Godoy, S. I., & Feu, S. (2004). Discriminative power of basketball game-related statistics by level of competition and sex. *Perceptual and motor Skills*, 99(3_suppl), 1231-1238.
- Sampaio, J., Gonçalves, B., Mateus, N., Shaoliang, Z., & Leite, N. (2018). 6 Basketball. *Modelling and simulation in sport and exercise*.
- Sampaio, J., McGarry, T., Calleja-González, J., Jiménez Sáiz, S., Schelling i del Alcázar, X., & Balciunas, M. (2015). Exploring game performance in the National Basketball Association using player tracking data. *PloS one*, 10(7), e0132894.
- Shorridge, A., Goldsberry, K., & Adams, M. (2014). Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. *Journal of Quantitative Analysis in Sports*, 10(3), 303-313.
- Smith, J. M. (2017). Does Three Point Shooting Percentage Affect Winning Percentage in the NBA?
- Wallace, S., Caudill, S. B., & Mixon, F. G. (2013). Homo certus in professional basketball? Empirical evidence from the 2011 NBA Playoffs. *Applied Economics Letters*, 20(7), 642-648. doi:10.1080/13504851.2012.727965
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686
- Yin, F., Hu, G., & Shen, W. (2020). Analysis of professional basketball field goal attempts via a Bayesian matrix clustering approach. *arXiv preprint arXiv:2010.08495*.