

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA



**ANÁLISIS DEL COMPORTAMIENTO DE LOS RESULTADOS DEL
FÚTBOL EUROPEO Y NACIONAL POR MEDIO DEL USO DE
HERRAMIENTAS DE ESTADÍSTICA**

Trabajo Especial de Grado presentado por
Luis Adalberto Hernández
ante la Facultad de Ciencias de la
Ilustre Universidad Central de Venezuela
como requisito parcial para optar al título
de: **Licenciado en Física**
Con la tutoría de: Prof. Esteban Álvarez
Prof. José Antonio López

Octubre-2017
Caracas-Venezuela

Escuela de Física

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA



**ANÁLISIS DEL COMPORTAMIENTO DE LOS RESULTADOS DEL
FÚTBOL EUROPEO Y NACIONAL POR MEDIO DEL USO DE
HERRAMIENTAS DE ESTADÍSTICA**

Trabajo Especial de Grado presentado por
Luis Adalberto Hernández
ante la Facultad de Ciencias de la
Ilustre Universidad Central de Venezuela
como requisito parcial para optar al título
de: **Licenciado en Física**

Con la tutoría de: Prof. Esteban Álvarez
Prof. José Antonio López

Octubre-2017
Caracas-Venezuela



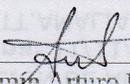
VEREDICTO

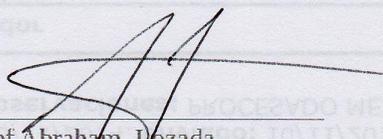
Quienes suscriben, miembros del Jurado designado por el Consejo de la Escuela de Física de la Facultad de Ciencias de la Universidad Central de Venezuela, para examinar el Trabajo Especial de Grado presentado por **Luis Hernández**, Cédula de Identidad **V-18363391**, bajo el título "**Análisis del comportamiento de los resultados del fútbol europeo y nacional por medio del uso de herramientas de estadística**", a los fines de cumplir con el requisito legal para optar al grado de **Licenciado en Física**, dejan constancia de lo siguiente:

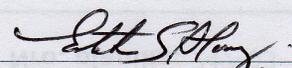
1. Leído como fue dicho trabajo por cada uno de los miembros del Jurado, éste fijó el día 25 de enero de 2018, a las 11:00 am, para que el autor lo defendiera en forma pública, lo que éste hizo en la Sala de Seminarios Guillermo Ruggeri de la Escuela de Física, mediante un resumen oral de su contenido, luego de lo cual respondió satisfactoriamente a las preguntas que le fueron formuladas por el jurado; todo ello conforme a los artículos 20, 21, 22, 25, 26 y 28 de la Normativa de Trabajo Especial de Grado de la Licenciatura en Física de la Facultad de Ciencias de la UCV vigente.
2. Finalizada la defensa pública del trabajo, el jurado decidió declararlo aprobado por considerar que se ajusta a lo dispuesto y exigido en la Normativa de Trabajo Especial de Grado de la Licenciatura en Física de la Facultad de Ciencias de la UCV vigente en sus artículos 1, 5 y 6.

Se levanta la presente acta a los 25 días del mes de enero de 2018, dejándose también constancia de que, conforme a la normativa jurídica vigente, actuó como coordinador del jurado el tutor del Trabajo Especial de Grado Prof. Esteban Segundo Álvarez .

Firma del jurado evaluador


Prof Fermín Arturo Dalmagro
Alonso
UCV


Prof Abraham Lozada
UCV


Prof. Esteban Segundo Álvarez
UCV



”Mis padres me enseñaron tres cosas fundamentales: que para poder estar orgulloso de tí mismo y ser alguien hace falta trabajar; que es preciso actuar con seriedad; y que debes respetar a los demás para recibir respeto a cambio. Trabajo, seriedad y respeto. -Si haces estas tres cosas, podrás ser alguien en la vida-, me dijeron”.

Zinedine Zidane

Es que puedes tener todas las virtudes del mundo, pero si no tienes suerte ni gente en el camino que te ayude, no te sirve de nada.

Zinedine Zidane

El juego del fútbol empieza en las calles, donde aprendes que hay muchos otros que comparten tu misma afición y persiguen el mismo objetivo. Si tú ayudas, ellos te ayudan. Eso es el fútbol

Zinedine Zidane

Agradecimientos

A las siguientes personas y organizaciones que colaboraron con la realización de este trabajo.

Escuela de Física de la Universidad Central de Venezuela, Laboratorio Docente de Computación.

Laboratorio de Fenómenos no lineales, Escuela de Física, UCV.

Al Prof. José Benito Hernández, al Prof. Nelson Bolívar, al Prof. Abraham Lozada, al Prof. Francisco Sans, al Prof. Andrés Pérez, al Prof. Jorge Peñuela, a la Profa. Anamaría Font, a Jesús Hernández, a Jedariana Belisario, a Wilfredo Montero, a Francisco Suárez, a Deyban Pérez, a July Marval, a Eliezer Pérez y a mis tutores Prof. Esteban Alvarez y Prof. José Antonio López.

A Alejandra Delgado por apoyarme incondicionalmente, al igual que mis amistades Hernán Pirela, Ronald Cuenca, Octavio Londoño, Familia Farias, Prof. Ernesto Fuenmayor, Familia Lira, Francisco Chacoa, Prof. Hely Cordero, Daniel Landi, Profa. Mary Lorena Araujo, Arley Larrota, Daniel Gomes, Yuary Pino, Profa. Mariela Castillo, Manuel Alvela, Fernando Machado, Francisco Marichal, Victor Briceño, Rodolfo Briceño, Gabriel Martínez, Miguel Galindo, María Piñero, Jhonny De Abreu y Profa. Marisol Ortega.

A Ciencias F.C., a U.C.V. F.C., al A.G.Q., a U.T.C., al grupo PHYSIS y a PRO-CIENCIAS.

A mi familia por apoyarme en el largo camino de la ciencia y el fútbol.

A todos, GRACIAS!

RESUMEN

En el presente trabajo se estudia por medio de herramientas estadísticas el comportamiento de las cuatro grandes ligas de fútbol de Europa, la liga femenina alemana y la liga profesional de Venezuela, siguiendo la metodología propuesta en la investigación realizada por Wolfhard Janke y colaboradores en el año 2005, los cuales implementaron técnicas de mecánica estadística para realizar estudios de comportamiento en el fútbol. Fueron evaluados los resultados finales de los partidos de fútbol de las ligas de España, Alemania, Italia e Inglaterra, la liga femenina alemana y la liga profesional de Venezuela, encontrando que existe una clara ventaja de la localía y una alta competitividad en todas las ligas profesionales incluyendo la venezolana. Además, analizando el modelo binomial negativo se encontró que los datos se ajustaron denotando un comportamiento de alta competitividad en las ligas europeas y venezolana. Adicionalmente, se analizó el comportamiento de las ligas mediante modelos microscópicos de probabilidad, para lo cual se encontró que en las ligas más profesionalizadas se fortalecen las tácticas defensivas cuando los equipos tienen ventaja en el marcador sobre el rival, además, en ligas menos profesionalizadas marcar goles ciertamente motiva a los equipos a seguir anotando.

Palabras claves: Estadística deportiva, comportamiento, fútbol.

Índice general

| | |
|--|-----------|
| Lista de figuras | 13 |
| Lista de tablas | 17 |
| 1. Introducción | 19 |
| 1.1. Planteamiento del Problema | 21 |
| 1.2. Objetivo General | 22 |
| 1.3. Objetivos Específicos | 22 |
| 1.4. Justificación | 23 |
| 2. Marco Teórico | 25 |
| 2.1. Historia del Fútbol | 25 |
| 2.1.1. Fútbol en el ámbito cultural | 27 |
| 2.1.2. Fútbol en el ámbito económico | 28 |
| 2.1.3. Fútbol en el ámbito estadístico | 29 |
| 2.2. Antecedentes de la Investigación | 30 |
| 2.3. Bases Teóricas | 32 |
| 2.3.1. Teoría de probabilidades y estadística | 32 |
| 2.3.2. Tabla de frecuencias de una variable discreta | 33 |
| 2.3.3. Representación gráfica de los datos | 36 |
| 2.3.4. Variables aleatorias discretas | 38 |
| 2.3.5. Variables aleatorias continuas | 40 |
| 2.3.6. Distribuciones discretas de probabilidad | 42 |
| 2.3.7. Distribuciones continuas de probabilidad | 51 |
| 3. Marco Metodológico | 59 |
| 3.1. Recolección y organización de datos | 59 |
| 3.2. Cálculo de Parámetros | 61 |
| 3.2.1. Modelos microscópicos de probabilidad | 64 |

| | |
|---|------------|
| 4. Resultados | 67 |
| 4.1. Frecuencia de goles | 67 |
| 4.2. Sumas y diferencias de goles | 70 |
| 4.3. Modelo de Probabilidad | 76 |
| 4.4. Modelos microscópicos | 85 |
| 4.4.1. Liga premier inglesa | 85 |
| 4.4.2. Liga femenina alemana (AFBL) | 86 |
| 4.4.3. Liga profesional de Venezuela FVF | 86 |
| 5. Conclusiones | 89 |
| A. Desarrollo Numérico A | 91 |
| A.1. Distribución de Poisson | 91 |
| B. Desarrollo Numérico B | 93 |
| B.1. Distribución Gamma | 93 |
| B.2. Distribución Binomial Negativa (DBN) | 94 |
| C. Frecuencias de goles | 99 |
| D. Sumas y diferencias de goles | 101 |
| E. Modelos de Probabilidad | 105 |

Índice de figuras

| | |
|--|----|
| 2.1. Diagrama de Barras y Polígono de frecuencias. | 36 |
| 2.2. Diagrama de Frecuencias acumuladas. | 37 |
| 2.3. Diagrama de rectángulos y de sectores. | 38 |
| 2.4. Función de probabilidad y función de distribución para una variable aleatoria discreta. | 39 |
| 2.5. Función de probabilidad y función de distribución para una variable aleatoria continua. | 42 |
| 2.6. Distribución Binomial de diversos parámetros | 45 |
| 2.7. Distribución de Poisson para distintos valores de λ | 48 |
| 2.8. Distribución binomial negativa para distintos valores de r y $p = 0,5$ | 51 |
| 2.9. $P(a \leq X \leq b) =$ el área bajo la curva de densidad a y b | 52 |
| 2.10. Función de densidad y función de distribución para una distribución continua uniforme. | 53 |
| 2.11. Función de densidad y función de distribución para una distribución normal. | 54 |
| 2.12. Distribución normal. | 55 |
| 2.13. Diferentes densidades de probabilidad gamma para diferentes grados de libertad. | 57 |
| 3.1. Ligas de fútbol | 62 |
| 4.1. Frecuencia de goles de local y visitante de la liga Premier inglesa. | 68 |
| 4.2. Frecuencia de goles de local y visitante de la Liga femenina alemana. | 69 |
| 4.3. Frecuencia de goles de local y visitante de la Liga profesional de Venezuela. | 70 |
| 4.4. Frecuencia de las sumas de goles de la liga Premier inglesa. | 71 |
| 4.5. Frecuencia de las sumas de goles de la Liga femenina alemana. | 72 |
| 4.6. Frecuencia de las sumas de goles de la Liga profesional de Venezuela. | 73 |
| 4.7. Frecuencia de las diferencias de goles de la liga Premier inglesa. | 73 |
| 4.8. Frecuencia de las diferencias de goles de la Liga femenina alemana. | 74 |
| 4.9. Frecuencia de las diferencias de goles de la Liga profesional de Venezuela. | 75 |

| | |
|---|-----|
| 4.10. Densidad de probabilidad para los goles de local en la liga premier inglesa. | 77 |
| 4.11. Densidad de probabilidad para los goles de visita en la liga Premier inglesa. | 77 |
| 4.12. Densidad de probabilidad para las sumas de goles de la liga premier inglesa. | 78 |
| 4.13. Densidad de probabilidad para las diferencias de goles de la liga Premier inglesa. | 78 |
| 4.14. Densidad de probabilidad para los goles de local en la liga femenina alemana. | 80 |
| 4.15. Densidad de probabilidad para los goles de visita en la liga femenina alemana. | 80 |
| 4.16. Densidad de probabilidad para las sumas de goles en la liga femenina alemana. | 81 |
| 4.17. Densidad de probabilidad para las diferencias de goles en la liga femenina alemana. | 81 |
| 4.18. Densidad de probabilidad para los goles de local en la liga profesional de Venezuela. | 83 |
| 4.19. Densidad de probabilidad para los goles de visita en la liga profesional de Venezuela. | 83 |
| 4.20. Densidad de probabilidad para las sumas de goles en la liga profesional de Venezuela. | 84 |
| 4.21. Densidad de probabilidad para las diferencias de goles en la liga profesional de Venezuela. | 84 |
| | |
| C.1. Frecuencia de goles de local y visitante de la Serie A italiana. | 99 |
| C.2. Frecuencia de goles de local y visitante de la liga española. | 100 |
| C.3. Frecuencia de goles de local y visitante de la Bundesliga alemana. | 100 |
| | |
| D.1. Frecuencia de la suma de goles de la Serie A italiana. | 101 |
| D.2. Frecuencia de la suma de goles de la liga española. | 102 |
| D.3. Frecuencia de la suma de goles de la Bundesliga alemana. | 102 |
| D.4. Frecuencia de las diferencias de goles de la Serie A italiana. | 103 |
| D.5. Frecuencia de las diferencias de goles de la liga española. | 103 |
| D.6. Frecuencia de las diferencias de goles de la Bundesliga alemana. | 104 |
| | |
| E.1. Ajuste a la distribución binomial negativa para los goles de local de la Serie A italiana. | 105 |
| E.2. Ajuste a la distribución binomial negativa para los goles de local de la liga española. | 106 |
| E.3. Ajuste a la distribución binomial negativa para los goles de local de la Bundesliga alemana. | 106 |
| E.4. Ajuste a la distribución binomial negativa para los goles de visita de la Serie A italiana. | 107 |

| | |
|--|-----|
| E.5. Ajuste a la distribución binomial negativa para los goles de visita de la liga española. | 107 |
| E.6. Ajuste a la distribución binomial negativa para los goles de visita de la Bundesliga alemana. | 108 |
| E.7. Ajuste a la distribución binomial negativa para la suma de goles de la Serie A italiana. | 108 |
| E.8. Ajuste a la distribución binomial negativa para la suma de goles de la liga española. | 109 |
| E.9. Ajuste a la distribución binomial negativa para la suma de goles de la Bundesliga alemana. | 109 |
| E.10. Ajuste a la distribución binomial negativa para la diferencia de goles de la Serie A italiana. | 110 |
| E.11. Ajuste a la distribución binomial negativa para la diferencia de goles de la liga española. | 110 |
| E.12. Ajuste a la distribución binomial negativa para la diferencia de goles de la Bundesliga alemana. | 111 |

Índice de tablas

| | |
|--|----|
| 2.1. Clubes de fútbol y su poder económico | 29 |
| 2.2. Tabla de frecuencias de variables discretas | 34 |
| 4.1. Parámetros de las distribuciones de probabilidad para los datos de la liga Premier inglesa 1946/47 - 2015/16 | 76 |
| 4.2. Parámetros de las distribuciones de probabilidad para los datos de la liga femenina alemana (AFBL) 1997/98 - 2015/16 | 79 |
| 4.3. Parámetros de las distribuciones de probabilidad para los datos de la liga Profesional de fútbol de Venezuela (FVF) 1957 - 2011 | 82 |
| 4.4. Parámetros de los modelos microscópicos para los datos de la liga Premier inglesa 1946/47 - 2015/16 | 85 |
| 4.5. Parámetros de los modelos microscópicos para los datos de la liga femenina alemana (AFBL) 1997/98 - 2015/16 | 86 |
| 4.6. Parámetros de los modelos microscópicos para los datos de la liga Profesional de fútbol de Venezuela (FVF) 1957 - 2011 | 87 |

Capítulo 1

Introducción

Es conocido hoy día que el fútbol representa un fenómeno socioeconómico, el cual surge en la Inglaterra del siglo XIX, convirtiéndose a través de los años en una pasión que va más allá de 22 jugadores en cancha, ya que mueve y promueve los más grandes valores del ser humano por un lado, y maneja una gran cantidad de factores que influyen en nuestra sociedad por el otro. Desde millones de jugadores organizados en cada país, pasando por las grandes audiencias frente al televisor y en los estadios, hasta innumerables empresarios gestionando a los equipos en pro de generar miles de ganancias.

Esta investigación responde al interés de conocer cómo un fenómeno de tal envergadura como el fútbol ha evolucionado a través del tiempo, la revisión bibliográfica deja ver que el mismo ha sido estudiado por más de 50 años en pro de entender su comportamiento, así como las tácticas de los entrenadores y por supuesto con el propósito de predecir los resultados de éste deporte.

En los años 70 Charles Reep y colaboradores [1], realizaron estudios sobre los deportes de balón a través del modelo de Poisson, tratando de predecir los resultados asumiendo que los mismos eran completamente aleatorios. Sin embargo, Wolfhard Janke y colaboradores [2], en el año 2009 realizaron estudios de comportamiento a través del modelo binomial negativo, usando también estadística de valores extremos y sugirieron ciertos modelos microscópicos de probabilidad, los cuales se utilizaron con el fin de entender de mejor manera el profesionalismo y la competitividad del fútbol europeo.

Siguiendo la misma línea de ideas, en este trabajo se realizarán los estudios de comportamiento de las cuatro grandes ligas de Europa (Alemania, Italia, España e Inglaterra), de la liga femenina alemana (Allianz Frauen-Bundesliga) y de la liga profesional de fútbol

de Venezuela (FVF), en una primera instancia mediante distribuciones de frecuencias de goles de local y de visitante, además de distribuciones de sumas y diferencias de goles, a fin de identificar la ventaja de la localía y parámetros que permitan comprender niveles de competitividad y profesionalismo.

Adicionalmente, se realizará el ajuste a la Distribución Binomial negativa y se estudiará el comportamiento de las ligas mencionadas a través de modelos microscópicos de probabilidad. Se implementará una metodología de cálculo con la ayuda de algoritmos escritos en lenguaje R y C.

La importancia de realizar estudios de comportamiento al fútbol europeo, femenino y nacional, se basa en la comprensión de las características fundamentales que poseen cada una de las ligas en estudio desde competitividad hasta profesionalismo y al entendimiento del progreso que ha tenido el fútbol nacional en los últimos años.

Este trabajo consta de cinco capítulos, en el capítulo 1 se presenta el planteamiento del problema, así como el objetivo general, los objetivos específicos y la justificación del mismo. En el capítulo 2 se desarrollará una reseña de los diversos aspectos relacionados con el fútbol, los antecedentes que sirven de base para ésta investigación y la teoría de probabilidades y estadística que dan pie a los métodos de análisis a implementar. En el capítulo 3 se introduce la metodología a usar, para entender el comportamiento de las ligas de fútbol, además de los parámetros a calcular que ayuden a entender la ventaja de la localía en el fútbol profesional y las características de profesionalismo y competitividad. En el capítulo 4 se presentan los resultados obtenidos basados en la metodología aplicada y finalmente en el capítulo 5 se exponen las conclusiones de éste trabajo de investigación.

1.1. Planteamiento del Problema

El fútbol, sin duda alguna, es una de las grandes actividades deportivas del mundo moderno. Nos encontramos con un fenómeno socioeconómico de alta envergadura donde se hallan envueltas diversas variables que influyen en su desarrollo. Los jugadores, el cuerpo técnico, la afición, las ventas, los derechos televisivos, el clima entre otros, vienen siendo los elementos que influyen en el mismo, lo cual, permite observarlo desde muchos puntos de vista.

El fútbol, a su vez, ha sido objeto de investigación por más de 50 años en todos los niveles, dándole mayor relevancia en este trabajo al estudio del fútbol a través de la estadística, entendiéndolo que en sí mismo es un deporte de alta complejidad desde la visión de muchas variables, es decir, nos encontramos más allá del juego entre dos equipos de once jugadores cada uno, cuyo objetivo es hacer entrar en la portería contraria un balón[3].

Sin embargo no fue hasta 1971 cuando Charles Reep y col. [1] realizaron su trabajo sobre deportes de balón entre ellos hockey, cricket y fútbol, para observar el mejor ajuste de los resultados finales a través de modelos de poisson con el fin de predecir los mismos en el futuro. Los primeros análisis realizados al fútbol envuelven una cantidad limitada de datos, lo cual permitió asumir, basado en las distribuciones de resultados, que los mismos eran simplemente aleatorios.

En la actualidad el fútbol goza de mayor influencia mediática y por ende se puede obtener mayor cantidad de datos sobre el mismo, permitiendo realizar estudios de mayor alcance sobre el comportamiento de las ligas a través de los resultados y así obtener mayor información. Es aquí donde Wolfhard Janke y col. [2] en el año 2009 analizaron resultados del fútbol europeo en casi su totalidad, haciendo énfasis en el modelo binomial negativo, y en estadística de valores extremos generalizados, ya que estos ofrecen la mejor descripción estadística de las distribuciones empíricas de goles observados y ciertos modelos microscópicos, los cuales sugieren una mejor justificación para la forma analítica de las densidades de probabilidad, donde se obtuvieron mayor información sobre las ligas de fútbol, su profesionalismo y competitividad.

Tomando como referencia lo expuesto anteriormente, se plantean varias interrogantes acerca del fútbol y las variables que influyen en el mismo. Por ejemplo, ¿son todas las variables ciertamente aleatorias?, ¿cómo influyen entre sí en la obtención de los resultados?, ¿cómo obtener información del comportamiento a través de la estadística?, pues

basados en el trabajo de W. Janke y col.[2], se dará respuesta a estas interrogantes y se confirmará la determinación del comportamiento de las ligas española, alemana, italiana e inglesa desde sus inicios profesionales, innovando en el estudio del comportamiento de la liga profesional venezolana.

El estudio del fútbol por medio de las distribuciones de probabilidad de goles marcados tanto por los equipos locales y visitantes, así como también desde el punto de vista de modelos de comportamiento, permitirá entender por ejemplo la ventaja de ser local en el fútbol, a su vez, la cantidad de goles marcados en total por partido y la diferencia de los mismos posibilita la comprensión sobre el profesionalismo de cada una de las ligas en estudio.

Se realizarán aportes sobre el comportamiento de la liga profesional de Venezuela lo cual dará una nueva visión del fútbol nacional, que significará un beneficio en la comprensión de la misma, así como una nueva manera de entender el fútbol de Venezuela aportando herramientas estadísticas para mejorar y contribuir en el crecimiento y desarrollo de nuestro fútbol.

1.2. Objetivo General

Estudiar el comportamiento de los resultados del fútbol europeo y fútbol nacional haciendo uso de herramientas de estadística a fin de identificar los efectos estocásticos.

1.3. Objetivos Específicos

1. Recopilar los resultados de los partidos de fútbol de las cuatro grandes ligas europeas y la liga profesional de fútbol de Venezuela.
 2. Estudiar las distribuciones de probabilidad a fin de observar el mejor ajuste de los datos recopilados.
 3. Estudiar los resultados finales de los partidos de fútbol para identificar la ventaja de la localía observando los ajustes de las distribuciones de probabilidad.
 4. Analizar las sumas y las diferencias de goles a través de los ajustes de las distribuciones de probabilidad.
-

-
5. Estudiar los factores de motivación y retroalimentación entre los equipos mediante modelos microscópicos de probabilidad.

1.4. Justificación

Actualmente se maneja información diversa relacionada al fútbol motivados por la pasión o fanatismo que ésta actividad genera, sin embargo, disponemos de un escaso dominio de las variables que repercuten de manera importante en los resultados de la misma, conllevando muchas veces a especulaciones alejadas de la realidad sobre algunos resultados observados. El estudio a realizar por medio del uso de herramientas estadísticas nos permitirá revisar trabajos ya elaborados sobre el fútbol europeo e innovar en el comportamiento del fútbol nacional, lo cual ayudará a sumar elementos que permitan evaluar la incidencia de algunas variables en consideración que muy bien puedan ser usadas a futuro para incidir en la formación de nuestra liga profesional y con ello enriquecer el fútbol nacional en pro de alcanzar mayor competitividad y profesionalismo, siendo este un elemento de peso que justifica el abordar la presente investigación.

Capítulo 2

Marco Teórico

A través de esta sección se desarrollará una breve reseña histórica sobre el fútbol, así como sus aspectos más importantes dentro de los ámbitos cultural, económico y estadístico. Por otro lado serán presentados diversos antecedentes a este trabajo de investigación, los cuales abordaron diferentes temáticas relacionadas con el fútbol desde el punto de vista estadístico y probabilístico y ayudarán a entrar en el contexto de los estudios realizados con anterioridad respecto al comportamiento del fútbol y las variables que influyen en el mismo. Finalmente se realizará un desarrollo de la teoría estadística y teoría de valores extremos, relacionado con las variables aleatorias discretas y continuas, en conjuntos con las diversas distribuciones de probabilidad que se ajustan de la mejor manera posible a la distribución de goles anotados en las ligas de fútbol.

“La historia de fútbol es un triste viaje del placer al deber. A medida que el deporte se ha hecho industrial, ha ido desterrando la belleza que nace de la alegría de jugar porque sí” [4], condenado lo que es útil e inútil, dándole un sentido mercantilista y empresarial al desarrollo del deporte, cambiando la sencillez del juego por una visión industrial globalizada, abriendo paso a nuevas tendencias dentro del marco socioeconómico.

2.1. Historia del Fútbol

La historia del fútbol asociación, conocido simplemente como fútbol, se considera que inicia a partir del año 1863, con la fundación de “The Football Association”, aunque sus orígenes se pueden remontar hacia la edad media, particularmente en las Islas Británicas. Si bien existían puntos en común entre diferentes juegos de pelota que se desarrollaron desde el siglo III A.C. y el fútbol actual, el deporte tal como se lo conoce hoy tiene sus

orígenes en las Islas Británicas [5].

La formación definitiva del fútbol asociación tuvo su momento culminante durante el Siglo XIX, debido a la carencia de códigos o reglas en el deporte, éste se caracterizó por mucha desorganización y alto grado de violencia en la práctica del mismo. No obstante, también existían otros códigos menos violentos y mejor organizados: quizás uno de los más conocidos fue el calcio florentino, deporte muy popular en Italia que tuvo incidencia en los códigos de algunas escuelas británicas. Por otro lado, la desorganización en el deporte llevó a que en 1848 diferentes colegios ingleses se reunieran en la Universidad de Cambridge, para darle forma a lo que se llamó código Cambridge, el cual sería la base para el reglamento moderno del fútbol. Finalmente en 1863 en Londres se oficializaron las primeras reglas del fútbol asociación [5].

Desde ese momento el fútbol ha experimentado un crecimiento ininterrumpido hasta la actualidad, obteniendo a millones de personas involucradas. En 1886 se llevó a cabo la primera reunión de la “International Football Association Board” para la consiguiente fundación de la FIFA en el año de 1904 alcanzando así los rincones más lejanos del planeta con el deporte, iniciando su globalización en el año de 1930 con la primera copa del mundo llevada a cabo en Uruguay, convirtiéndose en la actualidad en el evento deportivo con mayor audiencia del planeta [5].

En un censo realizado por la FIFA en el año 2006, se evidenció como más de 265 millones de jugadores y 26 millones de jugadoras estarían en activo en las 207 asociaciones de fútbol afiliadas a la FIFA. Este número crece progresivamente año tras año, motivado a la masificación y globalización del balompié mundial. Cabe destacar que el impacto socio cultural del fútbol es tan grande que la copa del mundo del año 2014 en Brasil fue vista por más de 3.000 millones de telespectadores acumulados, esto significa un 42 % de la población mundial, lo cual magnifica el alcance de éste deporte en el planeta [6].

Es importante destacar que en este trabajo de investigación se revelará información sobre el fútbol profesionalizado y competitivo, donde el profesionalismo implica una dedicación exclusiva del atleta hacia el deporte y la práctica del mismo como actividad laboral, devengando un sueldo y rigiéndose por la leyes del país donde juega. Por otra parte la competitividad se refiere a la paridad o igualdad en las ligas de fútbol la cual, se verá reflejada en los resultados de los partidos donde la ventaja no esta del todo marcada, no obstante la misma se observa en la igualdad económica donde los clubes pueden mantener

el nivel de competición.

2.1.1. Fútbol en el ámbito cultural

El fútbol en esencia es un fenómeno cultural que puede parecer inclusive superficial o vanidoso, sin embargo hay razones para creer todo lo contrario, ya que podrá entenderse a una sociedad, o un grupo de la misma, por sus matices culturales provenientes de este deporte, así como las dinámicas económicas, sociales y de globalización.

Desde una perspectiva más simbólica, que privilegia lo cultural y “procura comprender las cosas tal cual ellas se presentan y, si es posible, cómo llegaron a ser lo que son, que no se limita a las referencias objetivas sino a lo que ellos significan para las diferentes colectividades que crean, intercambian y modifican bienes culturales, en este caso específico, el fútbol” [7], se abre como una investigación y reflexión del fútbol, pero que permite brindar una serie de perspectivas a pensar ámbitos más amplios como la cultura, la globalización y las culturas locales.

Por otro lado se puede abordar el ámbito cultural del fútbol tal cual ha sido adaptado y configurado, así como, asumido en todas las comunidades tanto sociales, educativas e intelectuales, como de partidismo político, lo cual ha tenido cierto rechazo por algunos grupos y en parte esto ha permitido la ausencia de estudios profundos en el tema del deporte rey.

Es conveniente ver al fútbol como una corriente que moviliza a millones de personas desde aspectos sentimentales, físicos y pasionales, que evidencian el afecto que siente el ser humano por grupos que van desde lo local, pasando por lo nacional hasta llegar a lo mundial. Con esto se advierten varias cosas, como, la densidad cultural de las realidades que viven los individuos y también la manera en que lo local se conformó y se articuló a procesos globales [7].

“Por un lado, el estudio sobre la manera como el fútbol se hizo presente en nuestros territorios y la manera en la que actúa como fenómeno social, colectivo, nos hace ingresar en la zona de los *intersticios*, aquello que Michel Maffesoli (sociólogo francés) señala que es por donde se puede desplegar la mirada por una serie de estratos profundos a través de los cuales se vive ordinariamente la socialidad, que no son visibles a las miradas que intentan observar las dimensiones sociales mayores o estructurales, pero que es por donde han circulado los sentidos mediante los que se han conformado tanto los mundos simbóli-

cos y afectivos, como las maneras en que los individuos se han agrupado y actuado en territorios concretos[8]. Es decir, el fútbol, por un lado, ha sido una de las vías por las cuales las personas se han apropiado de un mundo simbólico por vía de todo lo que les representa ser un *hinch*a de un club de fútbol, pero que esto mismo será un mecanismo analógico de pertenecer a un espacio, un territorio. Pero también, es una manera de participar dentro de una comunidad a través de portar y ser leal a ese mundo simbólico, de ritos colectivos, estados de ánimo y sistemas de representaciones colectivas, que actuarán como mecanismos de integración, congregación y vivencia grupales” [8].

2.1.2. Fútbol en el ámbito económico

En la actualidad, el fútbol representa para la población mundial, una actividad recreativa y de rendimiento económico en todos los niveles y clases sociales, desde los más humildes hasta las élites económicas, motivados en construir espacios o en el simple disfrute de la práctica deportiva. El fútbol se convirtió en un activo donde los pequeños y grandes empresarios comenzaron a invertir, incentivado a la obtención de grandes ganancias económicas o por la simple difusión de valores e ideales. Lo anterior llevó a que se originaran múltiples redes de poder económico que decayó en la monopolización del deporte profesional, donde los empresarios y dirigentes políticos se convirtieron en dueños de clubes de fútbol [9].

El fútbol a su vez despierta pasiones a nivel social de gran envergadura y gracias a su repercusión en la economía, se ha derivado un efecto globalizador de este deporte. Hoy día se puede afirmar, como sostiene Pascal Boniface (geopolítico francés), que “el fútbol es el arquetipo de la globalización en mayor grado que la democracia, la economía de mercado o Internet. No existe actualmente un fenómeno más global. Su imperio no conoce fronteras ni límites y se ha ido extendiendo por todo el planeta de manera pacífica y sin necesidad de imponerse.” Además, “constituye por otra parte uno de los raros fenómenos de la globalización que escapa a la dominación estadounidense”. Fruto de la globalización del mismo, el fútbol ha ido progresivamente extendiéndose a nivel mundial, alcanzando principalmente a Norteamérica, luego poco a poco en el continente asiático, lo cual se observó en el mundial de Corea y Japón 2002, hasta llegar al Continente Africano por lo visto en el mundial de Sudáfrica 2010 [10].

Es importante destacar que en el aspecto económico, el fútbol, otorga una proyección social de alto nivel, ya sea por quienes buscan un reconocimiento social o para aquellos

empresarios que buscan expandir sus negocios a nivel mundial. Aunado a esto, existen también personajes que han usado al fútbol como impulso político, de hecho son cada vez más los casos de empresarios que se convierten en dueños de clubes de fútbol, como el ruso Roman Abramovich, el estadounidense Malcolm Glazer o el qatarí Abdullah Al Tani. Por ello hoy en día el fútbol es una plaza donde muchos intereses empresariales se cruzan, algo que el presidente de la FIFA en los años sesenta, Jules Rimet, ya advirtió: “Hay que tener mucho cuidado, porque el fútbol se está convirtiendo en un negocio” [10].

Para ejemplificar el nivel de manejo económico que poseen hoy en día los clubes de fútbol a nivel mundial, a continuación se mostraran el capital de los 10 clubes de fútbol más ricos del mundo para enero del año 2017 [11].:

Tabla 2.1: Clubes de fútbol y su poder económico

| Número | Club | Capital |
|--------|---------------------|-------------------------|
| 1 | Manchester United | 689 millones de euros |
| 2 | Barcelona | 620,2 millones de euros |
| 3 | Real Madrid | 620,1 millones de euros |
| 4 | Bayern de Múnich | 591 millones de euros |
| 5 | Manchester City | 524,9 millones de euros |
| 6 | Paris Saint Germain | 520,9 millones de euros |
| 7 | Arsenal | 468,5 millones de euros |
| 8 | Chelsea | 447,4 millones de euros |
| 9 | Liverpool | 403,8 millones de euros |
| 10 | Juventus de Turín | 341,1 millones de euros |

2.1.3. Fútbol en el ámbito estadístico

El centrocampista inglés Paul Gascoigne apodado “Gazza”, dijo en 1996: “Nunca he predicho nada y nunca lo haré”. Estas palabras demuestran que las predicciones en el mundo deportivo son más que inevitables, y el fútbol no está exento del mundo de las apuestas. Sin embargo Gazza dice algo muy importante “se pueden encontrar patrones en todo”, existen patrones en el tiempo que tardamos en llegar al trabajo día a día, en las redes de amistades de cada persona y la frecuencia de reuniones, en la alimentación, las compras del supermercado, entre otras. El fútbol ha sido estudiado con anterioridad, en

la búsqueda de los patrones que buscan entender los mismos y predecirlos. En cuanto se identifica un patrón, se puede realizar una predicción [12].

El poder predecir el resultado de un evento deportivo mediante el cálculo o la estimación de algunas variables, no sólo conlleva a que la gente apueste y tenga un beneficio económico, también ayuda a los directivos y entrenadores a ajustar estos parámetros y forzar un resultado beneficioso para el equipo o atleta. En la actualidad existen muchos métodos cuyo objetivo es obtener esas probabilidades. Este trabajo de investigación se revisarán herramientas a nivel estadístico y probabilístico para entender el fútbol desde el punto de vista del comportamiento, permitiendo observar al fútbol desde otra perspectiva mucho más científica, que a su vez con el enfoque deportivo concretar y unificar aspectos sociales y culturales del deporte con la visión científica del mismo.

2.2. Antecedentes de la Investigación

Una vez comprendido la magnitud, y el alcance que ha tenido el fútbol a través de los años, desde sus inicios hasta la actualidad, es importante resaltar los diversos estudios que se han realizado al mismo, desde el punto de vista científico, y estadístico de comportamiento, así como la importancia o ventaja de la localía en el balompié mundial.

Desde hace más de 50 años se han realizado estudios a nivel estadístico en este deporte, donde, figura el trabajo de Charles Reep et al.[2], en el cual se estudiaron los resultados de fútbol desde el punto de vista de la distribución binomial negativa, considerada como una Poisson compuesta, y observando los goles anotados a partir de los pases y tiros al arco, los autores consideran suficiente evidencia que muestra que la distribución binomial negativa puede ser aplicada en diversos deportes de balón, donde se encuentra involucrada la probabilidad de éxito con los jugadores y es relativamente invariante a los movimientos entre ellos.

En 1995, Stephen Clarke [13] realizó un estudio sobre la ventaja de ser local en clubes de la liga inglesa, donde ajustando a través de un modelo de mínimos cuadrados a los resultados de partidos entre 1981 y 1991, logró reproducir el efecto de ventaja de la localía de cada equipo así como una clasificación de los mismos. Finalmente encontrando que la ventaja de ser local viene dada a pares entre clubes, es decir la relación entre dos equipos, y está relacionada con la distancia entre los estadios de cada club de fútbol.

En el año 2000 Dyte y Clarke de la Universidad de Swinburne en Hawthorn, Australia [14] realizaron un estudio de la copa del mundo de fútbol, con una clasificación basada en un modelo de Poisson, donde se sugiere que para predecir la distribución de los goles de los partidos internacionales de una copa del mundo, se debe observar los goles marcados por cada equipo como variables independientes de la clasificación dada por la “Federation Internationale de Football Association” FIFA. Donde obtuvieron que la ventaja de ser local bajo este modelo se basa en un gran efecto defensivo de los equipos, demostrado en que los equipos locales marcan más goles que los visitantes.

Años más tarde Greenhough et al., en el año 2002 [15] analizaron la distribución del número de goles anotados por equipos locales y visitantes de más de 160 países. Donde observaron que la función de densidad de probabilidad de los goles anotados era muy pesada en las colas de la distribución para ser ajustada por los modelos de Poisson o binomial negativo, ya que se alejaba de los rangos de las mismas, lo cual se ajustaba a procesos no correlacionados entre sí. Sin embargo eran muy consistentes al ajuste de la estadística de valores extremos. No obstante, observaron los resultados de la “Football Association Challenge Cup” (FA Cup) de la primera división inglesa y éstos no se encontraban dentro del rango de la estadística de valores extremos dando un buen ajuste para los modelos de Poisson y binomial negativo.

Por otro lado el trabajo realizado por Wolfhard Janke et al., en 2009 [2], se basó en observar los resultados de los partidos de fútbol de la mayoría de los países de Europa, a través de diversos modelos de probabilidad, tomando en consideración que el resultado de un partido de fútbol depende directamente de una combinación de diversos factores externos y habilidades tanto de jugadores como de los entrenadores. Examinando las distribuciones de probabilidad para equipos locales y visitantes observaron que las colas de las distribuciones no estaban bien descritas por los modelos de Poisson o binomial, por el contrario se observa un buen ajuste a la estadística de valores extremos así como a la distribución binomial negativa con un efecto de autoafirmación del modelo. Sin embargo, para ir más allá en la investigación propusieron modelos microscópicos de probabilidad, para estudiar el comportamiento de los resultados locales y visitantes, así como la suma y diferencia de los goles, encontrando que dichos modelos permitieron observar muy bien las distribuciones antes mencionadas.

Con respecto al fútbol nacional, no se han encontrado hasta la fecha publicaciones o trabajos nacionales o internacionales relacionados a las herramientas estadísticas hasta ahora citadas.

2.3. Bases Teóricas

2.3.1. Teoría de probabilidades y estadística

La probabilidad y la estadística están relacionadas en una forma importante, la probabilidad se utiliza como una herramienta, que permite evaluar la confiabilidad de las conclusiones acerca de la población usada cuando tenga información de la muestra. En cambio, basándose en el ejemplo de lanzar una moneda al aire, la estadística puede ser utilizada de dos maneras, en primer lugar cuando la población o espacio muestral es conocida, se usa la probabilidad para describir la posibilidad de observar un resultado en particular, y dos, cuando la población es desconocida y sólo se tiene una muestra de dicha población, donde se usa la probabilidad para inferir sobre la composición de la población total [16].

Existen ciertas reglas que aplican a la teoría clásica de probabilidades:

1. La probabilidad de cualquier evento siempre será un número comprendido entre cero y uno.
2. Cuando un evento no puede ocurrir, la probabilidad será cero.
3. Cuando un evento es seguro que ocurrirá, la probabilidad será uno.
4. La suma de las probabilidades de todos los resultados en el espacio de muestras es uno.
5. La probabilidad de que un evento no ocurra es igual a uno menos la probabilidad de que ese evento ocurra [17].

Por otra parte, la estadística envuelve, la recolección, clasificación, resumen, organización, análisis e interpretación de los datos al igual que la construcción de modelos. Por ejemplo, se quiere estudiar el ingreso familiar en un vecindario, para ello se toman aleatoriamente un grupo de familias y se observan sus ingresos, o se quiere medir el largo de una barra de madera y para ello se realiza la medida 10 veces. En el primer caso la población (ingresos familiares) sí existe, en el segundo (medidas de longitud) sólo es conceptual. En cada caso puede visualizarse la totalidad de los valores de la población de las cuales las muestras son una pequeña parte. Así se define a una población, como la colección o grupo de todos los objetos o medidas que son de interés para el coleccionista. Y se define la muestra como un subconjunto de datos seleccionados de una población, dónde el tamaño

dependerá del número de elementos totales de la población [18].

Dentro de la estadística, es posible que los datos que se han recopilado deban ser resumidos y descritos a través de características importantes, para ello se utilizará la estadística descriptiva. La misma tiene diversos métodos para explicar la naturaleza de los datos recopilados, entre los cuales se tienen, la construcción de histogramas, diagramas de caja, y gráficas de puntos. Otros métodos descriptivos implican el desarrollo de algunos cálculos numéricos como: medias, desviaciones estándar y coeficientes de correlación[19].

Después de haber obtenido una muestra de una población, se puede utilizar la información muestral para obtener algún tipo de conclusión con respecto a la población, es decir, la muestra es un medio para llegar a un fin en lugar de un fin por sí misma. Las técnicas para generalizar desde una muestra hasta una población se reúnen dentro de la estadística inferencial [19].

2.3.2. Tabla de frecuencias de una variable discreta

Se comienza con una muestra de tamaño N , donde la variable estadística x toma los valores, $x_1, x_2, x_3, \dots, x_k$. En primer lugar se ordenan los valores de forma creciente, la diferencia entre el valor mayor y el menor que toma la variable será llamada rango. En el caso de variables discretas, usualmente un mismo valor de la variable aparecerá repetido enumeradas veces, es decir $k < N$. El siguiente paso es la construcción de la tabla donde se indicarán los valores posibles de la variable y su frecuencia de aparición, esta tabla será denominada tabla de frecuencias de una variable discreta[20]:

En la primera columna de esta tabla se escriben los distintos valores de la variable, x_i ordenados de mayor a menor. Es posible hacer también una tabla de frecuencias de una variable cualitativa. En ese caso, en la primera columna se escribirán las diferentes cualidades o atributos que puede tomar la variable. En las siguientes columnas se escriben para cada valor de la variable, donde:

Frecuencia absoluta n_i : Definida como el número de veces que aparece repetido el valor en cuestión de la variable estadística en el conjunto de las observaciones realizadas. Si N es el tamaño de la muestra, las frecuencias absolutas cumplen las propiedades.

$$0 \leq n_i \leq N \quad ; \quad \sum_{i=1}^k n_i = N \quad (2.1)$$

Tabla 2.2: Tabla de frecuencias de variables discretas

| Valores de la variable estadística x_i | Frecuencias absolutas n_i | Frecuencias relativas f_i | Frecuencias absolutas acumuladas N_i | Frecuencias relativas acumuladas F_i |
|---|--------------------------------|-----------------------------|---|---|
| x_1 | n_1 | f_1 | N_1 | F_1 |
| x_2 | n_2 | f_2 | N_2 | F_2 |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| x_k | n_k | f_k | N_k | F_k |

La frecuencia absoluta, aunque dice el número de veces que se repite un dato, no proporciona información de la importancia de este. Para ello se realiza la siguiente definición[20].

Frecuencia relativa f_i : Cociente entre la frecuencia absoluta y el número de observaciones realizadas N . Es decir,

$$f_i = \frac{n_i}{N} \quad (2.2)$$

cumpliéndose las propiedades.

$$0 \leq f_i \leq 1 \quad ; \quad \sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{N} = \frac{\sum_{i=1}^k n_i}{N} = 1 \quad (2.3)$$

Esta frecuencia relativa se puede expresar también en tantos por cientos del tamaño de la muestra, para lo cual basta con multiplicar por 100.

$$\%_{x_i} = 100x f_i \quad (2.4)$$

Por ejemplo, si $f_i = 0.25$, esto quiere decir que la variable x_i se repite en el 25 % de la muestra[20].

Frecuencia absoluta acumulada N_i : Suma de las frecuencias absolutas de los valores inferiores o igual a x_i , o número de medidas por debajo, o igual, que x_i . Evidentemente la frecuencia absoluta acumulada de un valor se puede calcular a partir de la correspon-

diente al anterior como,

$$N_i = N_{i-1} + n_i \quad y \quad N_1 = n_1. \quad (2.5)$$

Además la frecuencia absoluta acumulada del último valor será.

$$N_k = N \quad (2.6)$$

Frecuencia absoluta acumulada F_i : Cociente entre la frecuencia absoluta acumulada y el número de observaciones. Coincide además con la suma de las frecuencias relativas de los valores inferiores o iguales a x_i ,

$$F_i = \frac{N_i}{N} = \frac{\sum_{j=1}^i n_j}{N} = \sum_{j=1}^i \frac{n_j}{N} = \sum_{j=1}^i f_j \quad (2.7)$$

y la frecuencia relativa acumulada del último valor es 1

$$F_k = 1 \quad (2.8)$$

Se puede expresar asimismo como un porcentaje (multiplicando por 100) y su significado será el tanto por ciento de medidas con valores por debajo o igual que x_i [20].

Ejemplo 2.1: Tabla de frecuencias de una muestra de hijos de 20 familias.

Supongamos que el número de hijos de una muestra de 20 familias es el siguiente:

2 1 1 3 1 2 5 1 2 3
4 2 3 2 1 4 2 3 2 1

El tamaño de la muestra es $N = 20$, el número de valores posibles $k = 5$ y el recorrido es $5-1=4$.

| x_i | n_i | f_i , $n_i/20$ | N_i , $\sum_1^i n_j$ | F_i , $\sum_1^i f_i$ |
|-------|-------|---------------------|---------------------------|---------------------------|
| 1 | 6 | 0,30 | 6 | 0,30 |
| 2 | 7 | 0,35 | 13 | 0,65 |
| 3 | 4 | 0,20 | 17 | 0,85 |
| 4 | 2 | 0,10 | 19 | 0,95 |
| 5 | 1 | 0,05 | 20 | 1,00 |

2.3.3. Representación gráfica de los datos

Después de construir la tabla de frecuencias correspondiente es conveniente la representación gráfica de la distribución de los datos en un diagrama. Estas representaciones gráficas permiten una visualización rápida de la información recogida.

El diagrama principal para representar datos de variables discretas sin agrupar es el diagrama de barras. En este se representan en el eje de abscisas los distintos valores de la variable y sobre cada uno de ellos se levanta una barra de longitud igual a la frecuencia correspondiente. Pueden representarse tanto las frecuencias absolutas como las relativas f_i . En la práctica se puede graduar simultáneamente el eje de ordenadas tanto en frecuencias absolutas como en relativas en tantos por ciento. Un diagrama similar es el polígono de frecuencias. Este se obtiene uniendo con rectas los extremos superiores de las barras del diagrama anterior. De la misma forma, pueden representarse frecuencias absolutas, relativas, o ambas a la vez[20]. Ver Figura 2.2.

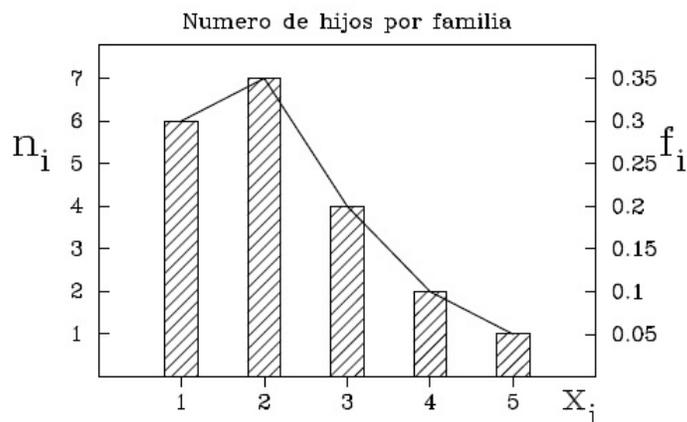


Figura 2.1: Diagrama de barras y polígono de frecuencias. Se han usado los datos del ejemplo 2.1. Tomado de [20].

Representaciones gráficas para variables cualitativas

Existe una gran variedad de representaciones para variables cualitativas, de las cuales se describirán las más usadas. El diagrama de rectángulos es similar al diagrama de barras y el histograma para las variables cuantitativas. Consiste en representar en el eje de abscisas los diferentes caracteres cualitativos y levantar sobre cada uno de ellos un

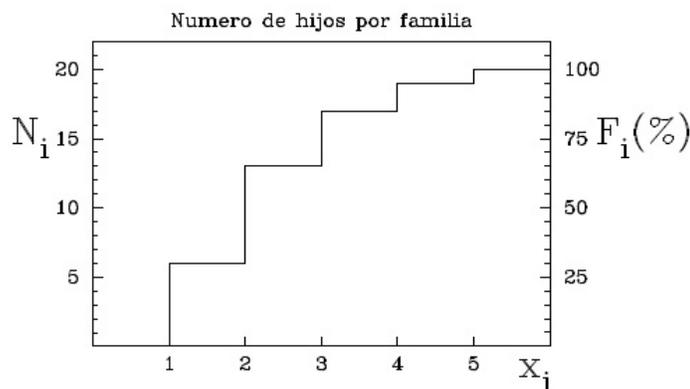


Figura 2.2: Diagrama de frecuencias acumuladas. Se han usado los datos del ejemplo 2.1. Tomado de [20].

rectángulo (de forma no solapada) cuya altura sea la frecuencia (absoluta o relativa) de dicho carácter [20].

Un diagrama muy usado es el diagrama de sectores, también llamado diagrama de torta. En él se representa el valor de cada carácter cualitativo como un sector de un círculo completo, siendo el área de cada sector, o, lo que es lo mismo, el arco subtendido, proporcional a la frecuencia del carácter en cuestión. De forma práctica, cada arco se calcula como 360° multiplicado por la frecuencia relativa. Es además costumbre escribir dentro, o a un lado, de cada sector la frecuencia correspondiente. Este tipo de diagrama proporciona una idea visual muy clara de cuáles son los caracteres que más se repiten[20].

Ejemplo 2.2: Tabla de frecuencias de notas de física, ver figura 2.3.

Las notas de una asignatura de física (en la UCM) del curso académico 95/96 se distribuyeron de acuerdo a la siguiente tabla para los alumnos presentados en junio:

| Nota | n_i | f_i | N_i | F_i | α_i |
|--------------------|-------|-------|-------|-------|------------|
| Suspense | 110 | 0,56 | 110 | 0,46 | 165,6 |
| Aprobado | 90 | 0,38 | 200 | 0,84 | 136,8 |
| Notable | 23 | 0,10 | 223 | 0,94 | 36 |
| Sobresaliente | 12 | 0,05 | 235 | 0,99 | 18 |
| Matrícula de Honor | 2 | 0,01 | 237 | 1,00 | 3,6 |

Los diagramas de rectángulos y de sectores correspondientes se muestran en la Figura 2.3.

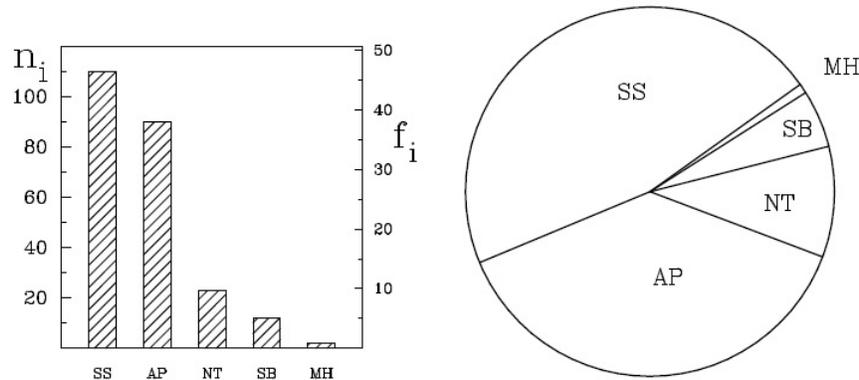


Figura 2.3: Diagrama de rectángulos (izquierda) y de sectores (derecha) para las notas del ejemplo 2.2. Las frecuencias relativas están dadas en tanto porcentaje. Los ángulos de cada sector circular se determinan como $\alpha_i = f_i \times 360$ (grados). Tomado de [20].

2.3.4. Variables aleatorias discretas

Dado un experimento aleatorio, definimos una variable aleatoria como una función definida sobre el espacio muestral que asigna un número real a cada uno de los puntos, o resultados posibles, de dicho espacio muestral. Por ejemplo en el lanzamiento de monedas podemos asignar 0 si sale cara y 1 si es cruz. De esta forma, la variable aleatoria toma valores (aleatorios) determinados por el resultado del experimento. Generalmente, la variable aleatoria se denota por una letra mayúscula (ej. X), reservándose las letras minúsculas (ej. x) para los distintos valores que puede tomar[20].

Una variable aleatoria que toma un número finito o infinito, pero numerable, de valores, se denomina variable aleatoria discreta. Por el contrario, cuando la variable puede tomar un número infinito no numerable de valores (o todos los valores posibles de un intervalo) se la denomina variable aleatoria continua. Un ejemplo sería la duración de un suceso, o el peso de una persona. En la mayoría de los casos, las variables aleatorias continuas representan datos medidos, mientras que las variables aleatorias discretas suelen representar datos que se cuentan (ej. número de veces que ha ocurrido un cierto suceso) [20].

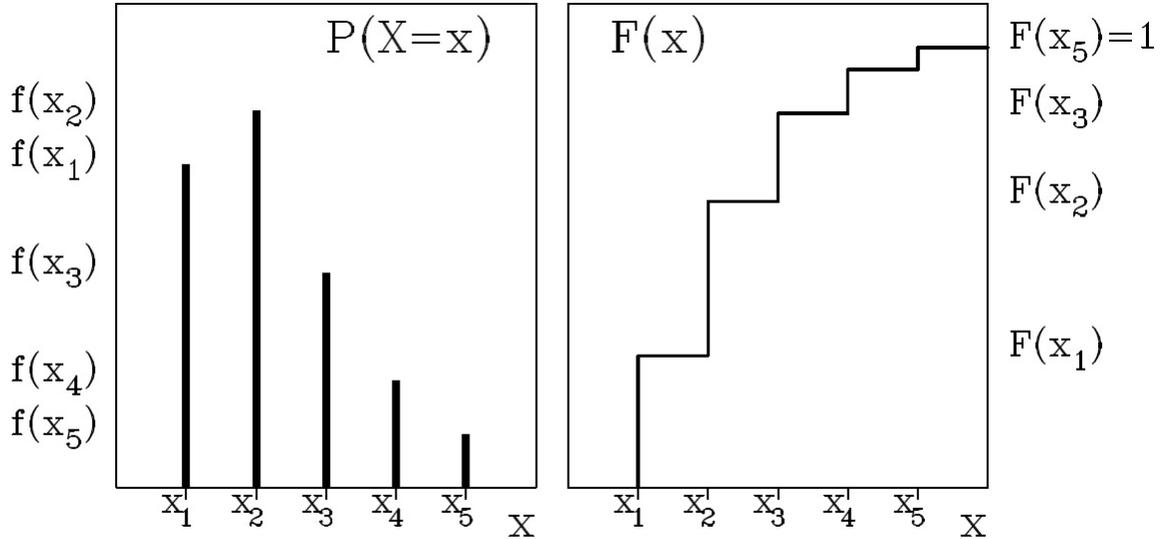


Figura 2.4: Función de probabilidad, $f(x)$, y función de distribución, $F(x)$, para una variable aleatoria discreta $X = x_1, x_2, x_3, x_4, x_5$. Tomado de [20].

Sea una variable aleatoria discreta X y supongamos que puede tomar los valores x_1, x_2, x_3, \dots . Como ya se ha indicado, para describir completamente la variable aleatoria hay que indicar las probabilidades de que tome cada uno de sus valores posibles. De esta forma a cada valor de la variable aleatoria se le asigna como probabilidad, la probabilidad de que ocurra el subconjunto del espacio muestral asociado con ese valor particular. Para esto se define una función $f(x)$ que indica la probabilidad de cada valor x de la variable aleatoria. Esta es la función de probabilidad, también llamada distribución de probabilidad, de la variable aleatoria discreta X

$$f(x) \equiv P(X = x) \quad (2.9)$$

En particular, para un valor x_i de la variable aleatoria: $f(x_i) = P(X = x_i)$. Además, por las propiedades de la probabilidad, la función de probabilidad cumple, para todo x_i .

$$f(x_i) \geq 0 \quad ; \quad \sum_i f(x_i) = 1 \quad (2.10)$$

Asimismo, gráficamente se suele representar usando un diagrama de barras donde en el eje de las abscisas se sitúan los diferentes valores de X y en las ordenadas las probabilidades correspondientes (Figura 2.4). Otra forma de caracterizar la distribución de una

variable aleatoria es mediante la función de distribución $F(x)$, o función de probabilidad acumulativa, definida para cada x como la probabilidad de que la variable aleatoria X tome un valor menor o igual que x . Es decir,

$$F(x) = P(X \leq x), \quad (2.11)$$

donde x no se restringe a los valores que puede tomar la variable aleatoria y es cualquier número real ($-\infty \leq x \leq \infty$). Es fácil ver que, por su definición, $F(x)$ es una función no decreciente y toma los valores extremos.

$$F(-\infty) = 0 \quad ; \quad F(\infty) = 1. \quad (2.12)$$

La función de distribución se puede evaluar a partir de la función de probabilidad, y al contrario, ya que,

$$F(x) = \sum_{x_i \leq x} f(x_i) = f(x_{i-1}) + f(x_i) \quad ; \quad f(x_i) = F(x_i) - F(x_{i-1}) \quad (2.13)$$

Si suponemos que la variable aleatoria puede tomar los valores $X = x_1, x_2, \dots, x_n$, ordenados de menor a mayor, entonces la función de distribución para cada punto estaría dada por.

$$F(x) = \begin{cases} 0 & x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ \sum_{i=1}^n f(x_i) = 1 & x_n \leq x \end{cases} \quad (2.14)$$

De modo que la representación gráfica de la función de distribución discreta tiene forma de escalera, con saltos en los valores aislados que toma la variable y con continuidad por la derecha es decir, en cada salto el valor que toma $F(x)$ es del escalón superior, ver Figura 2.4.

2.3.5. Variables aleatorias continuas

Respecto al caso de las variables aleatorias continuas, es decir, aquellas que pueden tomar cualquier valor en un intervalo (a, b) , o incluso $(-\infty, \infty)$. En este caso, la probabilidad de que la variable X tome un valor determinado dentro de ese intervalo es cero, ya que

existen infinitos valores posibles en cualquier intervalo, por pequeño que sea, alrededor del valor en cuestión. Por ejemplo, la probabilidad de que la altura de una persona sea exactamente 1.75 cm, con infinitos ceros en las cifras decimales, es cero. Por tanto no se puede definir una función de probabilidad igual que se hacía para las variables discretas, dando la probabilidad de cada valor de la variable[20]. Lo que se sí puede especificar es la probabilidad de que la variable esté en un cierto intervalo. Para ello se define una función $f(x)$ llamada función de densidad, o distribución de probabilidad, de la variable aleatoria continua X de forma que, para todo x , cumpla.

$$f(x) \geq 0 \quad ; \quad \int_{-\infty}^{\infty} f(x)dx = 1 \quad (2.15)$$

De forma que la probabilidad de que X se encuentre entre dos valores x_1 y x_2 se puede calcular como.

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx \quad (2.16)$$

Estas expresiones constituyen la definición de la función de densidad. Puede observarse que la representación gráfica de la función de densidad será la de una curva, normalmente continua, que toma siempre valores positivos o nulos, y con área comprendida entre la curva y el eje x . De igual forma por la expresión 2.16, la probabilidad de que la variable tome un valor entre x_1 y x_2 será el área bajo la función de densidad entre las abscisas x_1 y x_2 . Esta asociación de probabilidad área es sumamente útil para el estudio de las distribuciones continuas de probabilidad[20].

Al igual que para el caso discreto, se puede definir la función de distribución $F(x)$ en cada punto x de una variable aleatoria continua como la probabilidad de que la variable X tome un valor inferior a x .

$$F(x) = P(X < x) \quad (2.17)$$

Por la definición de función de densidad, esta se relaciona con la función de distribución por.

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (2.18)$$

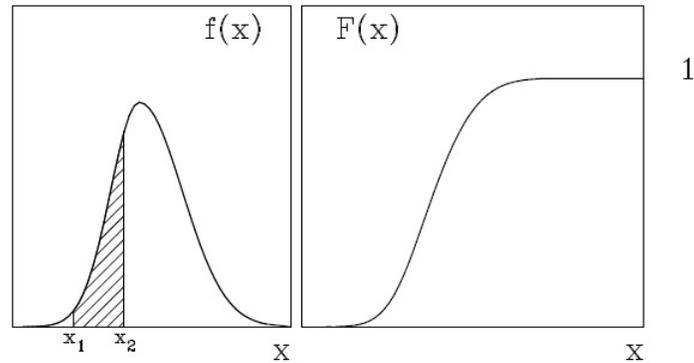


Figura 2.5: Función de densidad, $f(x)$, y función de distribución, $F(x)$, para una variable aleatoria continua. Tomado de [20].

2.3.6. Distribuciones discretas de probabilidad

Se realizará una reseña de las diversas distribuciones de probabilidad discreta, motivado a que en el fútbol la variable de los resultados se entiende que es aleatoria y de comportamiento discreto, por ende es necesario ahondar en la diversidad de distribuciones que describan el ajuste aleatorio de las variables que se estudiarán.

El comportamiento de una variable aleatoria queda, en general, descrito por su distribución de probabilidad, o función de probabilidad $f(x)$, que, en el caso de que la variable sea discreta, indica la probabilidad de que se de cada uno de los valores x posibles de la variable aleatoria ($f(x) = P(X = x)$). En la práctica esto indica que muchos experimentos aleatorios tienen comportamientos similares, de forma que sus resultados siguen la misma distribución de probabilidad[20].

Distribución discreta uniforme

La distribución uniforme es la más simple de todas las distribuciones discretas de probabilidad. Diremos que tenemos una distribución discreta uniforme cuando todos los posibles valores de la variable aleatoria sean igualmente probables. En este caso, si la variable aleatoria X puede tomar los valores x_1, x_2, \dots, x_n con probabilidades iguales, la función de probabilidad vendrá dada por.

$$F(x) = \int_{-\infty}^x f(t)dt \quad \text{donde} \quad x = x_1, x_2, \dots, x_n \quad (2.19)$$

Por la condición de normalización (2.10). Se ha utilizado la notación $f(x; n)$ puesto que, en este caso, la distribución de probabilidad depende (únicamente) del parámetro n , o número de valores posibles. Las expresiones para la media y varianza de esta distribución son, evidentemente.

$$\mu = \sum_{i=1}^n x_i f(x_i, n) = \sum_{i=1}^n \frac{x_i}{n} = \frac{\sum_{i=1}^n x_i}{n}. \quad (2.20)$$

$$\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 f(x_i, n) = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}. \quad (2.21)$$

Ejemplo 2.3: En este ejemplo se demostrará como el lanzamiento de un dado no trucado cumple con la distribución discreta uniforme, ya que cumple con las probabilidades de que cada lado tiene las mismas posibilidades de obtención.

$$x=1,2,3,4,5,6 \quad n=6 \quad f(x; 6) = \frac{1}{6}$$

$$\mu = \frac{\sum x_i}{n} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3,5$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n} = \frac{\sum (x_i - 3,5)^2}{6} = 2,92 \quad \Rightarrow \quad \sigma = 1,71$$

Distribución binomial

Se supone un experimento aleatorio consistente en realizar un número de ensayos o pruebas repetidas, cada una de ellas con únicamente dos posibles resultados mutuamente excluyentes, denominados éxito o fracaso. Se supone también que la probabilidad de obtener un éxito en un ensayo es siempre constante y que los diferentes ensayos son independientes, en el sentido de que el resultado de un ensayo no afecta a los otros. En este caso diremos que se tiene un proceso de Bernoulli[20]. En concreto, el proceso de Bernoulli debe tener las siguientes propiedades:

1. El experimento consiste en n ensayos.
2. El resultado de cada uno de los ensayos puede clasificarse en éxito o fracaso (excluyentes).
3. La probabilidad de éxito, que se denota por p , es constante en todos los ensayos.

4. Los diferentes ensayos son independientes.

Se define la variable aleatoria binomial como la función que da el número de éxitos en un proceso de Bernoulli. Evidentemente, la variable binomial X podrá tener valores en el rango $X = 0, 1, 2, \dots, n$, donde n es el número de veces que se repite el ensayo. La distribución de probabilidad asociada con esta variable aleatoria se denomina distribución binomial y vendrá representada por

$$f(x) = P(X = x) = b(x; n, p). \quad (2.22)$$

Ya que la distribución depende del número de ensayos n y la probabilidad de éxito p en un sólo ensayo. Para calcular una expresión para $b(x; n, p)$ se considera la probabilidad de que se obtengan x éxitos y $n - x$ fracasos en un orden determinado. Llamando q a la probabilidad de fracaso (que será evidentemente $q = 1 - p$) y teniendo en cuenta que los n ensayos son independientes, la probabilidad de esa disposición de resultados particular será el producto de las probabilidades de cada ensayo, es decir

$$\underbrace{p \cdots p}_n \underbrace{q \cdots q}_{n-x} = p^x q^{n-x} \quad (2.23)$$

Para calcular la probabilidad total de x éxitos, tenemos que sumar la probabilidad anterior para todas las disposiciones posibles de resultados en que se dan esos x éxitos. Ese número se puede expresar como.

$$P_n^{x, n-x} = \frac{n!}{x!(n-x)!} = \binom{n}{x} \quad (2.24)$$

De esta forma, la probabilidad de obtener x éxito, o la distribución de probabilidad binomial, viene dada por.

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad \text{donde } x = 0, 1, \dots, n \quad (2.25)$$

El término de distribución binomial viene del hecho de que los diversos valores de $b(x; n, p)$ con $x = 0, 1, 2, \dots, n$ corresponden a los $n+1$ términos de la expresión $(q + p)^n$ entonces.

$$(q + p)^n = \binom{n}{0} q^n + \binom{n}{1} p q^{n-1} + \binom{n}{1} p^n =$$

$$(q + p)^n = b(0; n, p) + b(1; n, p) + \cdots + b(n; n, p) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} \quad (2.26)$$

Es importante notar que, puesto que $(q + p) = 1$, la expresión anterior implica

$$\sum_{x=0}^n b(x; n, p) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = 1 \quad (2.27)$$

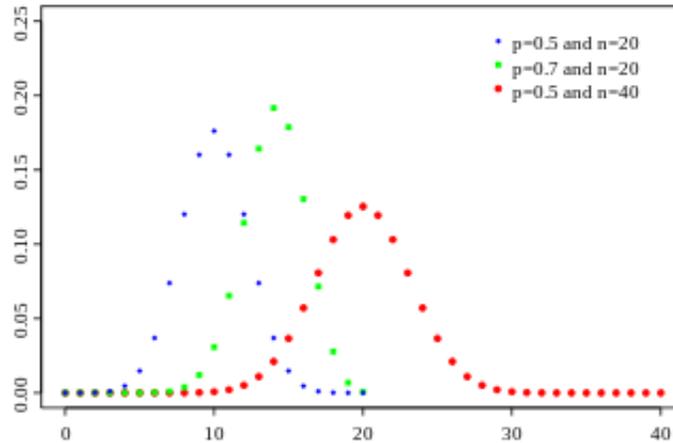


Figura 2.6: Distribución Binomial de diversos parámetros Fuente: https://es.wikipedia.org/wiki/Distribución_binomial. Consultado: 11/09/2017.

Ejemplo 2.4: basados en los lanzamientos de un jugador de baloncesto se observará la distribución binomial de manera analítica.

Sea un jugador de baloncesto que tiene que tirar 3 tiros libres. Se sabe que su promedio de acierto es del 80%. Determinar las probabilidades de que enceste 0, 1, 2 ó 3 canastas. Si se llama: Canasta S, Fallo N y x el número de puntos, se puede calcular la probabilidad de cada suceso como el producto de las probabilidades de cada tiro ya que son sucesos independientes.

$$\begin{aligned} P(S) &= 0,8 & P(N) &= 0,2 \\ P(SSS) &= 0,8 \times 0,8 \times 0,8 = 0,512 \\ P(SSN) &= 0,8 \times 0,8 \times 0,2 = 0,128 \\ P(SNN) &= 0,8 \times 0,2 \times 0,2 = 0,032 \\ P(NNN) &= 0,2 \times 0,2 \times 0,2 = 0,008 \end{aligned}$$

La probabilidad de que cada x se calcula sumando las probabilidades para cada disposición:

$$\begin{aligned} P(x = 0) &= 0,008 & P(x = 1) &= 3 \times 0,032 = 0,096 \\ P(x = 2) &= 3 \times 0,128 = 0,384 & P(x = 3) &= 0,512 \end{aligned}$$

| | x | P |
|-----|-----|-------|
| SSS | 3 | 0,512 |
| SSN | 2 | 0,128 |
| SNS | 2 | 0,128 |
| SNN | 1 | 0,032 |
| NSS | 2 | 0,128 |
| NSN | 1 | 0,032 |
| NNS | 1 | 0,032 |
| NNN | 0 | 0,008 |
| | | 1,000 |

Distribución de Poisson

Se considera un experimento aleatorio consistente en medir el número de resultados, o sucesos de un tipo dado, que se producen en un cierto intervalo continuo. Este intervalo puede ser un intervalo de tiempo, de espacio, una región dada, entre otros. Ejemplos de este experimento podrían ser: el número de partículas radiactivas emitidas por un material en un tiempo dado, el número de fotones que llegan a un detector en un tiempo fijado, el número de días al año en que llueve en un cierto lugar, el número de estrellas que se observan en el cielo en cuadrículas del mismo tamaño. Diremos que un experimento de este tipo sigue un proceso de Poisson cuando se cumplan las siguientes condiciones:

1. El número de resultados que ocurren en un intervalo es independiente del número que ocurre en otro intervalo disjunto. Es decir, los sucesos aparecen aleatoriamente de forma independiente. Se dice entonces que el proceso no tiene memoria.
2. La probabilidad de que un resultado sencillo ocurra en un intervalo pequeño es proporcional a la longitud de dicho intervalo. Además dicha probabilidad permanece constante, de forma que se puede definir un número medio de resultados por unidad de intervalo. Se dice que el proceso es estable.
3. La probabilidad de que ocurra más de un resultado en un intervalo suficientemente pequeño es despreciable

Se define entonces la variable aleatoria de Poisson como el número de resultados que aparecen en un experimento que sigue el proceso de Poisson. Nótese que el campo de variabilidad de la variable de Poisson será: $X = 0, 1, 2, \dots$. La distribución de probabilidad

asociada con esta variable se denomina distribución de Poisson y dependerá fundamentalmente del número medio de resultados (o sucesos) por intervalo, que se denotará por λ . De esta forma, la distribución de Poisson se escribe.

$$f(x) = P(X = x) = P(x; \lambda) \quad (2.28)$$

Para calcular una expresión para $p(x; \lambda)$ es importante relacionar la distribución de Poisson con la binomial. Efectivamente, la distribución de Poisson aparece como límite de la distribución binomial cuando el número de observaciones en ésta última es muy grande y la probabilidad de que en una observación se dé el suceso (se obtenga un éxito, en la nomenclatura de la distribución binomial) es muy pequeña. Para ello dividimos el intervalo de observación en n intervalos muy pequeños, con n suficientemente grande para que, por la tercera propiedad del proceso de Poisson, no se puedan dar dos sucesos en cada subintervalo, y la probabilidad p de que ocurra un suceso en un subintervalo sea muy pequeña. De esta forma, el experimento de observar cuantos sucesos aparecen en un intervalo se convierte en observar si ocurre o no un suceso en n subintervalos (proceso de Bernoulli). Se puede suponer entonces una distribución binomial con n ensayos y probabilidad de éxito en cada uno p , que podremos escribir.

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x} \quad (2.29)$$

Cabe destacar que, aunque $n \rightarrow \infty$ y $p \rightarrow 0$, el número medio esperado de sucesos en el intervalo total ha de permanecer constante, e igual a λ , es decir $\mu = np = \lambda$. Y haciendo tender n a infinito y sustituyendo por p por λ/n

$$\begin{aligned} \lim_{n \rightarrow \infty} b(x; n, p) &= \lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} = \frac{\lambda^x}{x!} e^{-\lambda} \end{aligned} \quad (2.30)$$

donde se ha introducido el valor de los siguientes límites.

$$\lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} = \lim_{n \rightarrow \infty} 1 \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) = 1 \quad (2.31)$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \left(\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n/(-\lambda)}\right)^{n/(-\lambda)}\right)^{-\lambda} = e^{-\lambda} \quad (2.32)$$

De esta forma, la distribución de probabilidad de Poisson, o probabilidad de que se den x sucesos en un proceso de Poisson con un valor promedio λ , vendrá dada por $p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, donde $x = 1, 2, \dots$

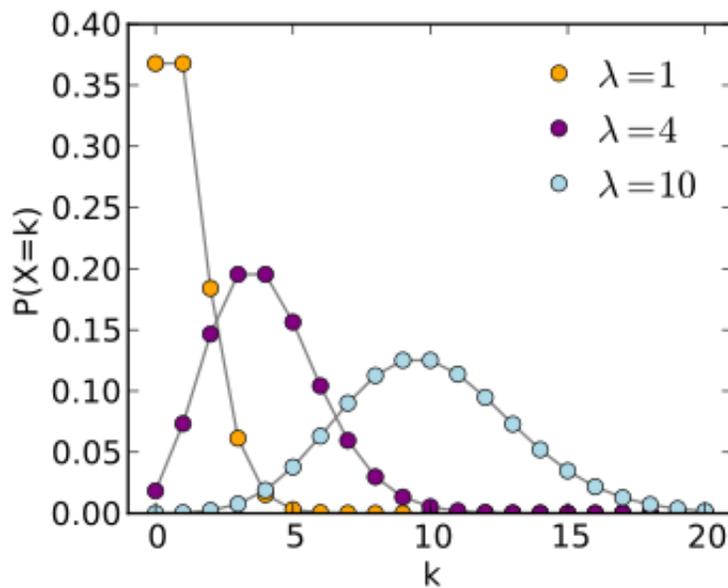


Figura 2.7: Distribución de Poisson para distintos valores de λ . Fuente: https://es.wikipedia.org/wiki/Distribución_de_Poisson. Consultado: 11/09/2017.

Ejemplo 2.5: Sea X el número de criaturas de un tipo particular capturadas en una trampa durante un período determinado. Se supone que X tiene una distribución de Poisson con $\lambda = 4,5$, así que en promedio las trampas contendrán 4,5 criaturas. La probabilidad de que una trampa contenga 5 criaturas es.

$$P(X = 5) = \frac{e^{-4,5}(4,5)^5}{5!} = 0,1708$$

La probabilidad de que una trampa contenga cuando mucho cinco criaturas es.

$$P(X \leq 5) = \sum_{x=0}^5 \frac{e^{-4,5}(4,5)^x}{x!} = e^{-4,5} \left[1 + 4,5 + \frac{(4,5)^2}{2!} + \dots + \frac{(4,5)^5}{5!} \right] = 0,7029$$

Distribución binomial negativa

La variable aleatoria y la distribución binomial negativa se basan en un experimento que satisface las siguientes condiciones:

1. El proceso consta de un número no definido de pruebas separadas o separables. El proceso concluirá cuando se obtenga un determinado número de resultados favorables r .
2. Cada prueba puede dar dos resultados posibles mutuamente excluyentes A y A^c .
3. La probabilidad de obtener un resultado A en cada una de las pruebas es p , siendo la probabilidad de A^c q . Lo que nos lleva a que $p + q = 1$.
4. Las probabilidades p y q son constantes en todas las pruebas. Todas las pruebas son independientes. Si se trata de un experimento de extracción este se llevará a cabo con devolución del individuo extraído, a no ser que se trate de una población en la que el número de individuos tenga carácter infinito[21].

Si, en estas circunstancias de manera aleatoria se hace que la variable X sea el número de pruebas necesarias para conseguir r éxitos o resultados A ; entonces la variable aleatoria X seguirá una distribución binomial negativa con parámetros p y r que se denotará $X \sim BN(p, r)$. La variable aleatoria X podrá tomar sólo valores superiores a $r \in r, r + 1, r + 2, \dots$ [21]. El suceso del que trata lo podemos ver como.

$$\underbrace{A^c, A^c, A^c, \dots, A^c}_{n \text{ veces}}, \overbrace{A, A, \dots, A}^{r \text{ veces}}. \tag{2.33}$$

O lo que es lo mismo,

$$\underbrace{A^c \cap A^c \cap A^c \cap \dots \cap A^c}_{n \text{ veces}} \cap \overbrace{A \cap A \cap \dots \cap A}^{r \text{ veces}}. \tag{2.34}$$

dado que las pruebas son independiente y además se sabe que $P(A) = p$ y $P(A^c) = q$, entonces se tiene

$$\underbrace{q \times q \times \dots \times q}_{n \text{ veces}} \overbrace{p \times \dots \times p}^{r \text{ veces}} \tag{2.35}$$

que sería la probabilidad de X si el suceso fuera precisamente con los resultados en ese orden. Dado que pueden darse otros ordenes, en concreto $\binom{n-1}{n-r} = \binom{n-1}{r-1}$ formas u órdenes distintos[21]. La función de probabilidad quedará como.

$$P(X = r) = \binom{n-1}{r-1} q^{n-r} p^r \quad (2.36)$$

Se define que una variable aleatoria X tiene Distribución Binomial Negativa de parámetros p, r , si su función de probabilidad está dada por.

$$g_{BN}(n; p, r) = P(X = r) = \binom{n-1}{r-1} q^{n-r} p^r \quad (2.37)$$

para $r = 1, 2, 3, \dots, n$

Ejemplo 2.6: en este ejemplo se observará como en un proceso de manufactura se involucra la distribución binomial negativa. Se sabe que un promedio de 1 de cada 10 productos es defectuoso en el proceso, ¿cuál es la probabilidad de que el artículo 16 examinado sea el segundo en estar defectuoso?

Tenemos $n = 16, r = 2, p = 1/10 = 0,1$, luego

$$P(X = 2) = BN(16; 2, 0, 1) = \binom{16-1}{2-1} (1 - 0,1)^{16-2} (0,1)^2 = \binom{15}{1} (0,9)^{14} (0,1)^2 = 0,0343$$

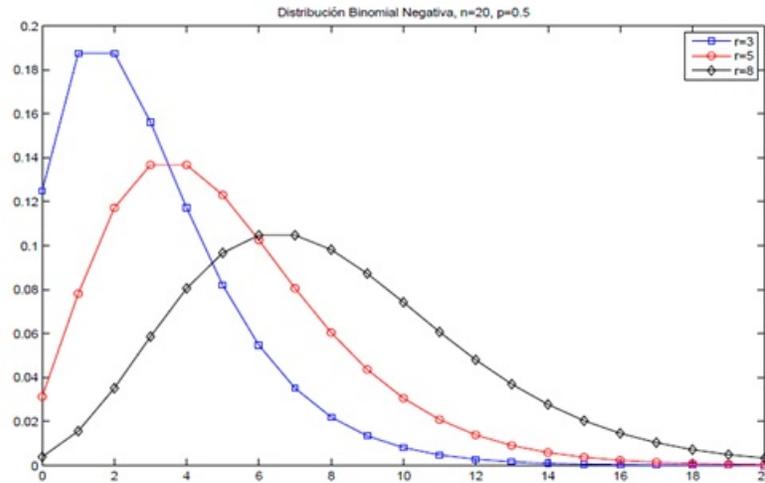


Figura 2.8: Distribución binomial negativa para distintos valores de r y $p = 0, 5$. Tomado de [21].

2.3.7. Distribuciones continuas de probabilidad

Sea X una variable aleatoria continua. Entonces, una distribución de probabilidad o función de densidad de probabilidad (FDP) de X es una función $f(x)$ tal que para dos números cualesquiera a y b con $a \leq b$.

$$P(a \leq X \leq b) = \int_a^b f(x)dx. \quad (2.38)$$

Es decir, la probabilidad de que X asuma un valor en el intervalo $[a, b]$ es el área sobre este intervalo y bajo la gráfica de la función de densidad, como se puede observar en la figura 2.9. La gráfica de $f(x)$ a menudo se conoce como curva de densidad[20].

Para que $f(x)$ sea una función de densidad de probabilidad legítima, se deben satisfacer las siguientes condiciones:

1. $f(x) \geq 0$ con todas las x .
2. $\int_{-\infty}^{\infty} f(x)dx = \text{área bajo la curva } f(x) = 1$

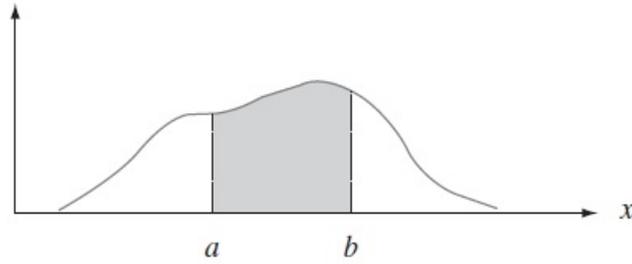


Figura 2.9: $P(a \leq X \leq b) =$ el área bajo la curva de densidad a y b . Tomado de [20].

Distribución continua uniforme

Se dice que una variable aleatoria X sigue una distribución continua uniforme cuando su función de densidad $f(x)$ toma valores constantes en el intervalo $[a, b]$. Es decir, $f(x) = K$ en ese intervalo y , por tanto, la probabilidad de que tome un valor en cualquier incremento dentro de ese intervalo es la misma. Para calcular esa constante se aplica la condición de normalización de la función de densidad.

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_a^b f(x)dx = \int_a^b Kdx = K(b-a) \quad \Rightarrow K = \frac{1}{b-a}. \quad (2.39)$$

Por lo tanto la función de densidad tiene la forma

$$F(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & x > b \end{cases} \quad (2.40)$$

Se puede calcular la función de distribución $F(x)$. Cuando x esté en el intervalo $[a, b]$

$$F(x) = P(X < x) = \int_{-\infty}^x f(t)dt = \int_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}. \quad (2.41)$$

y en general,

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases} \quad (2.42)$$

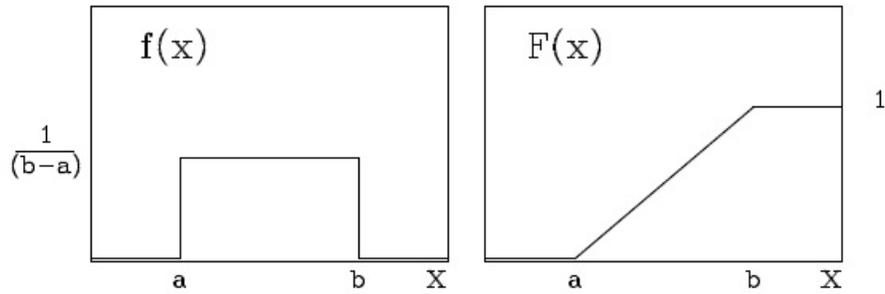


Figura 2.10: Función de densidad, $f(x)$, y función de distribución, $F(x)$, para una distribución continua uniforme. Tomado de [20]

Ejemplo 2.7: en este ejemplo se observará de manera numérica como una variable aleatoria distribuida uniformemente cumple con las características de la distribución. Si X es una variable aleatoria distribuida uniformemente en el intervalo $(0, 10)$, cuál sería la probabilidad de que $X < 3$, $X > 6$ y $3 < X < 8$.

$$P(X < 3) = \int_0^3 dx = \frac{3}{10}$$

$$P(X > 6) = \int_6^{10} dx = \frac{4}{10}$$

$$P(3 < X < 8) = \int_3^8 dx = \frac{1}{2}$$

Distribución normal

La distribución continua de probabilidad más importante de toda la estadística es, sin duda alguna, la distribución normal. La importancia de esta distribución se debe a que describe con gran aproximación la distribución de las variables asociadas con muchos fenómenos de la naturaleza. En particular, las medidas de magnitudes físicas suelen distribuirse según una distribución normal. Por ejemplo, la distribución de alturas de un grupo de población, las medidas de calidad de procesos industriales, o la distribución de temperaturas de una población, se pueden aproximar por distribuciones normales. Además, los errores en las medidas también se aproximan con mucha exactitud a la distribución normal. Por otra parte, bajo ciertas condiciones, la distribución normal constituye una buena aproximación a otras distribuciones de probabilidad, como la binomial y la de Poisson. Frecuentemente, a la distribución normal se la denomina también distribución Gaussiana[20].

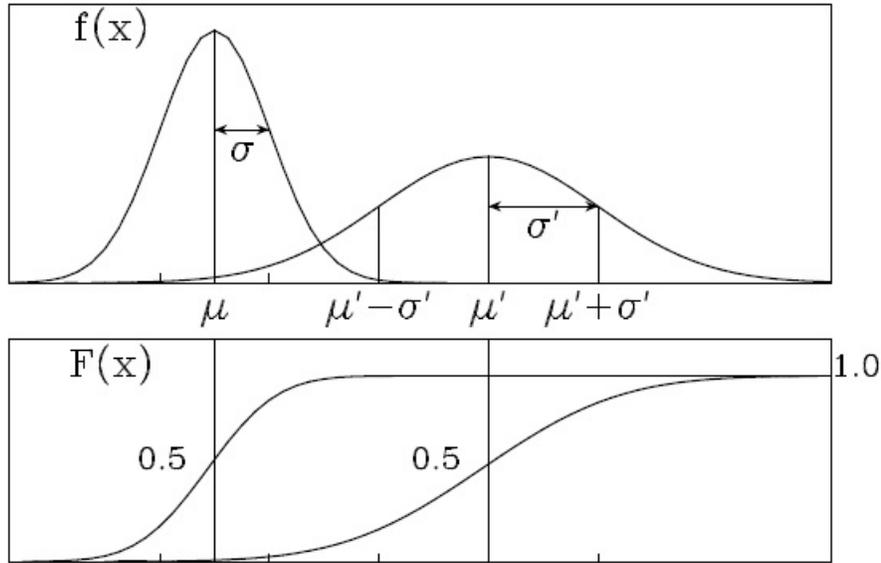


Figura 2.11: Función de densidad, $f(x)$, y función de distribución, $F(x)$, para una distribución normal. Se muestran las representaciones correspondientes a dos valores de la media μ y la desviación típica σ . Tomado de [20]

Por definición, se dice que una variable aleatoria continua X sigue una distribución normal de media μ y desviación típica σ si su función de densidad es

$$f(x) = N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; \quad -\infty < x < \infty. \quad (2.43)$$

De esta forma, una vez que se especifican μ y σ la distribución queda determinada completamente. Puede comprobarse que esta distribución de probabilidad cumple la condición de normalización dada en (2.15), ya que.

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1. \quad (2.44)$$

Donde se ha hecho el cambio de variable $z = (x - \mu)/\sigma$ (es decir $dx = \sigma dz$) y se ha aplicado el siguiente valor tabulado de la integral: $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a}$.

Gráficamente (Figura 2.11), la distribución de probabilidad normal tiene forma de campana (llamada campana de Gauss, o curva normal), simétrica (por depender de x a través del término $(x - \mu)^2$), centrada en μ y con anchura proporcional a σ (como es lógico

esperar del significado de la desviación típica). Evidentemente, el máximo de la función de densidad ocurre para $x = \mu$ y, por tanto, media, mediana y moda coinciden en ese punto. Se puede demostrar que los puntos de inflexión de la curva normal están situados en $\mu - \sigma$ y $\mu + \sigma$. La curva tiende asintóticamente a cero al alejarse del valor medio. Además, por (2.44), el área entre la curva normal y el eje X es la unidad[20].

La función de distribución normal, útil para el cálculo de probabilidades, vendrá dada por.

$$F(x) = P(X < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (2.45)$$

Es claro que la probabilidad de que X tome un valor entre x_1 y x_2 puede calcularse por.

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (2.46)$$

Ejemplo 2.8: Distribución normal. Haciendo uso de la figura 2.12 para una variable aleatoria normal estándar Z , conseguir el valor de z_0 que: (a) $P(Z > z_0) = 0,25$. (b) $P(Z < z_0) = 0,95$. (c) $P(Z < z_0) = 0,12$. (d) $P(Z > z_0) = 0,68$.

(a) Para la tabla normal y usando el hecho de que el área sombreada en la figura es 0.25 se obtiene que $z_0 \approx 0,675$

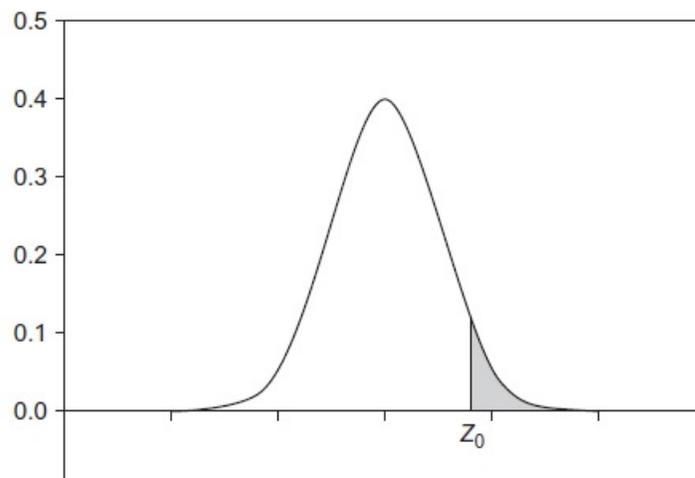


Figura 2.12: Ejemplo 2.8 Distribución normal. Tomado de [18].

(b) Ya que $P(Z < z_0) = 1 - P(Z \geq z_0) = 0,95 = 0,5 + 0,45$. para la tabla normal $Z_0 = 1,645$

(c) $Z_0 = -1,175$

(d) Usando la tabla, se tiene que $P(Z > z_0) = 0,5 + P(0 < Z < z_0) = 0,68$. Esto implica que $P(z_0 < Z < 0) = 0,18$ y de la tabla $Z_0 = -0,465$

Distribución gamma

La distribución de probabilidad gamma tiene diversas aplicaciones en varios campos, por ejemplo en la ingeniería, la distribución gamma ha sido empleada para fiabilidad de sistemas. Se describe la función gamma antes de introducir la distribución de probabilidad gamma[18]. La función gamma, denotada por $\Gamma(a)$ está definida como.

$$\Gamma(a) = \int_0^{\infty} e^{-x} x^{a-1} dx, \quad a > 0 \quad (2.47)$$

Se puede observar que integrando por partes que $a > 1$, $\Gamma(a) = (a-1)\Gamma(a-1)$ y en particular si n es un entero positivo, $\Gamma(n) = (n-1)!$. Entonces se dice que una variable aleatoria X posee una distribución de probabilidad gamma con parámetros $\alpha > 0$ y $\beta > 0$ si posee una función de densidad de probabilidad dada por.

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \quad (2.48)$$

La densidad gamma tiene dos parámetros, α y β , la cual se denotará $\text{gamma}(\alpha, \beta)$. El parámetro α es llamado parámetro de forma y β el parámetro de escala. Cambiando α cambia la forma de la densidad, mientras que al variar β se corresponde a un cambio en las unidades de medida. Variando ambos parámetros, se podrá generar diferentes miembros de la familia de gammas[18].

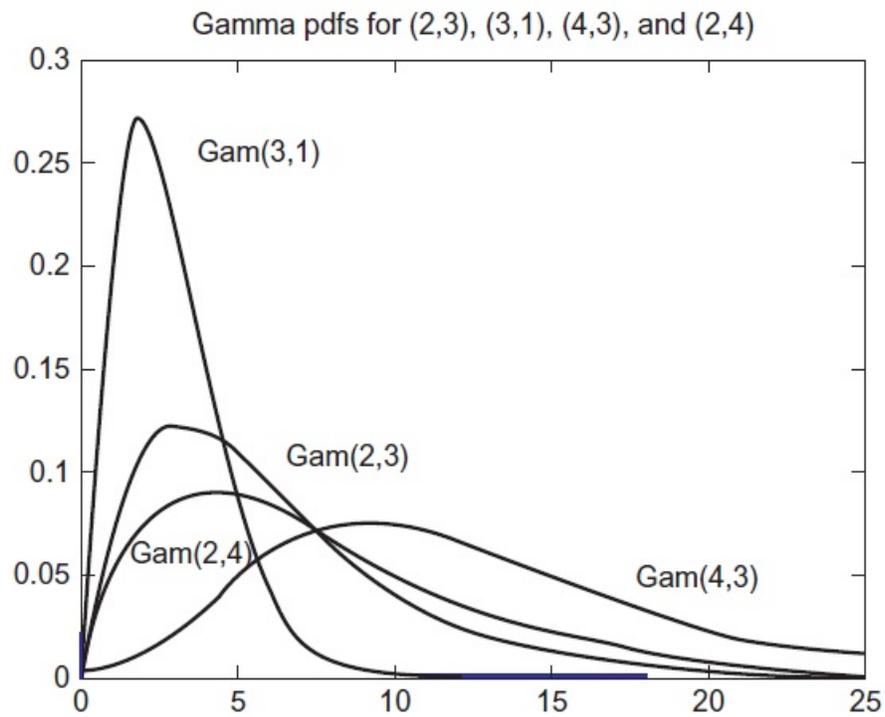


Figura 2.13: Diferentes densidades de probabilidad gamma para diferentes grados de libertad. Tomado de [18].

Capítulo 3

Marco Metodológico

En este capítulo se desarrollaran los detalles de los diversos pasos a seguir para el estudio de los resultados de fútbol y su comportamiento, partiendo desde la recolección de los datos, pasando por su completa organización hasta llegar a los análisis que se propusieron en los objetivos de esta investigación.

Donde se puntualizará todo el procedimiento para observar el mejor ajuste a las distribuciones de probabilidad, también las frecuencias de los goles de cada liga, las sumas y diferencias de goles, así como los modelos microscópicos que permitan ver el comportamiento de las variables en estudio, a fin de hacer notar la ventaja de la localía en el fútbol profesional y la retroalimentación entre los equipos de las diversas ligas de fútbol.

3.1. Recolección y organización de datos

Se realizó la selección de las ligas de fútbol a través de los siguientes parámetros:

1. Se escogieron las ligas mas importantes del fútbol europeo (España, Italia, Alemania e Inglaterra) las cuales se identifican con sus nombres comerciales La Liga (España), Serie A (Italia), Bundesliga (Alemania) y Premier League (Inglaterra).
2. Se escogió la liga de fútbol profesional venezolana (FVF).
3. Se tomó una liga joven o menos profesionalizada para realizar comparaciones (Liga Femenina de Alemania) llamada Allianz Frauen-Bundesliga.

Las ligas de fútbol europeo escogidas en conjunto con la liga femenina alemana y la liga de fútbol profesional venezolano, constan de características particulares correspondientes a la situación socio-económica y cultural de cada país a través del tiempo. Se realizará la

recolección de los datos de las 6 ligas de fútbol. Los datos consisten en los partidos realizados en cada país de donde se puede observar la siguiente cantidad de partidos [24,25]:

1. La Liga desde 1939/40 hasta 2015/16 con un total de 23106 partidos disputados.
2. Serie A desde 1945/46 hasta 2015/16 con un total de 21462 partidos disputados.
3. Premier League desde 1946/47 hasta 2015/16 con un total de 30330 partidos disputados.
4. Primera División de Venezuela desde 1957 hasta 2011 con un total de 8912 partidos disputados.
5. Bundesliga desde 1963/64 hasta 2015/16 con un total de 16160 partidos disputados.
6. Frauen-Bundesliga desde 1997/98 hasta 2015/16 con un total de 2508 partidos disputados.

La escogencia de los años se realizó para garantizar un bloque continuo de partidos ininterrumpidos en la línea temporal, lo cual no perjudica en los cálculos a realizar, por consiguiente cada fecha es particular en cada país. En España para el año 1936 hasta el año 1939 se llevó a cabo la guerra civil española donde a partir del 17 de julio de 1936 luego del fracaso parcial del golpe de estado, inició una guerra civil donde se involucró la Alemania nazi y la Italia fascista, la cual concluiría el 1ro de abril de 1939 fecha de la reanudación de la liga española de fútbol. La liga española se caracteriza por tener equipos muy técnicos y con jugadores de gran envergadura mundial, teniendo dentro de la misma a los mejores jugadores del mundo en la actualidad, no obstante tiene una gran presencia de jugadores sudamericanos en la misma lo cual enriquece a los clubes y a la competencia.

Respecto a la Italia de los años 40, se encuentra que el 25 de julio de 1943, se logra la destitución y arresto de Mussolini, lo cual inicia con el colapso de la Italia fascista, hasta que en el año de 1945 se llega al fin de la república social italiana y por consiguiente la reanudación de la Serie A. La misma es una liga caracterizada por su juego altamente defensivo y planteamientos tácticos rígidos sin dejar atrás a jugadores con excelente técnica y trabajo colectivo. Por otro lado en el mismo período se encuentra que el Reino Unido e Irlanda del Norte están involucrados desde 1939 hasta 1945 en la 2da guerra mundial, la cual al finalizar reanuda la Premier League inglesa. Inglaterra tiene una liga de fútbol altamente competitiva, donde los equipos se hacen fuerte de local, sin embargo, los equipos

visitantes tienden a buscar la victoria o alcanzar puntuar cuando se encuentran fuera de casa.

La liga venezolana inicia en el año 1957 su era profesional, año en el cual se realiza el plebiscito de Marcos Pérez Jiménez, el cual optaría por un período presidencial de más de 5 años, además ratificaría al presidente en su cargo y a todos los candidatos al congreso nacional, asambleas legislativas, estatales y consejos municipales de manera automática. El fútbol venezolano tiene raíces híbridas desde las influencias brasileras hasta las colombianas con equipos técnicamente muy buenos y una liga que crece a nivel competitivo progresivamente.

Respecto a la liga de fútbol alemán, se encuentra envuelto dentro de uno de los hechos históricos más importantes de nuestra era, ya que hacia el principio de la guerra fría se construyó entre 1958 y 1963 el muro de Berlín que separaría la República Democrática Alemana (RDA) de la República federal de Alemania (RFA), hecho que marcaba el inicio de la Bundesliga en la Alemania occidental. Es importante destacar que a finales de los años 80, termina la guerra fría y se inicia una nueva etapa en el país y en el fútbol alemán. La Bundesliga se caracteriza por equipos muy técnicos, con trabajos tácticos de gran nivel y con una competencia muy alta en el país teutón, alcanzando una selección de jugadores con muchos logros mundiales, como dijo Gary Lineker “el fútbol es un deporte en el que juegan once contra once durante 90 minutos y al final siempre gana Alemania”. La liga femenina alemana se inicia en 1990, pero no se obtienen datos hasta el año 1997 que coincide con la inclusión en 1996 del fútbol femenino por parte del Comité Olímpico Internacional (COI) para los juegos olímpicos de Atlanta 96.

3.2. Cálculo de Parámetros

Los datos serán organizados de la siguiente manera:

1. Se guardarán los resultados año por año en archivos con la etiqueta correspondiente (año+etiqueta.txt) ejemplo: 1957FVF.txt para la liga venezolana
2. Se realizará un algoritmo en c++ que agrupará los resultados de todos los años de cada liga en un solo archivo .txt ejemplo: golesFVF.txt
3. Al finalizar se obtendrán archivos para cada liga en estudio: golesBundesliga.txt, golesLaLiga.txt, golesSerieA.txt, y así para cada liga.



Figura 3.1: Ligas de fútbol

Para responder la pregunta de cuál distribución de probabilidad o modelo describe de mejor forma los datos, primero se determinarán los histogramas de frecuencia de los goles anotados por los equipos local (n_h) y visitante (n_a), donde la etiqueta (h) y la etiqueta (a) son referentes a la traducción del inglés, donde local = *home* y visitante = *away*. Se calcularán las frecuencias de los goles de local y de visita utilizando la hoja de cálculo del sistema Linux, para luego proceder a graficar dichos histogramas.

De igual forma se realizarán los histogramas de frecuencias de goles para las sumas $s = n_h + n_a$ y diferencias $d = n_h - n_a$, donde (s) es el resultado de sumar local más visitante y (d) significa restar local menos visitante, entendiendo que el resultado de un partido de fútbol es un par ordenado donde siempre los goles anotados por el equipo local está del lado izquierdo y los goles del visitante del lado derecho del par, por ejemplo: Chelsea 3 - 1 Bolton. Es importante resaltar que realizar los estudios de frecuencias de goles permitirán comprender la ventaja de la localía en el fútbol, además de observar el profesionalismo de las ligas en estudio desde el punto de vista de los goles anotados.

Por otra parte se determinarán los histogramas de las densidades de probabilidad usando el modelo de la distribución binomial negativa (DBN) tanto para local $p_h(n_h)$ como para visitante $p_a(n_a)$, de igual forma se observarán los histogramas para las sumas $p^\Sigma(s)$ y diferencias $p^\Delta(d)$ de goles basados en el mismo modelo binomial.

Se considerará el ajuste a las funciones de densidad de probabilidad de las descripciones fenomenológicas, en la distribución de probabilidad escogida, siendo ésta la distribución binomial negativa a través de algoritmos realizados en lenguaje R. Por último, se observará el comportamiento de los datos a través de los modelos microscópicos de probabilidad también realizado en lenguaje R. Donde se espera que el ajuste de los modelos que permitan discriminar sobre la competitividad de las ligas de fútbol basados en lo observado por el modelo binomial y los parámetros encontrados en el mismo.

Inicialmente basados en el trabajo de Janke et al.[2], se deberían realizar los ajustes a la distribución de Poisson, tanto para el equipo local como para el equipo visitante, donde.

$$P_{\lambda_h}^h(n_h) = \frac{\lambda_h^{n_h}}{n_h!} e^{(-\lambda_h)} \quad P_{\lambda_a}^a(n_a) = \frac{\lambda_a^{n_a}}{n_a!} e^{(-\lambda_a)} \quad (3.1)$$

Así como observar los ajustes de las sumas y diferencias de los goles para la distribución de Poisson, es decir.

$$P_{\lambda_h, \lambda_a}^{\Sigma}(s) = \sum_{n=0}^s P_{\lambda_h}^h(n) P_{\lambda_a}^a(s-n) = \frac{(\lambda_h + \lambda_a)^s}{s!} e^{[-(\lambda_h + \lambda_a)]} \quad (3.2)$$

$$P_{\lambda_h, \lambda_a}^{\Delta}(d) = \sum_{n=0}^{\infty} P_{\lambda_h}^h(n+d) P_{\lambda_a}^a(n) = e^{-(\lambda_h + \lambda_a)} \left(\frac{\lambda_h}{\lambda_a} \right)^{d/2} I_d(2\sqrt{\lambda_h \lambda_a}) \quad (3.3)$$

donde I_d es la función modificada de Bessel. Ver apéndice A. Sin embargo, dichos ajustes no se realizarán ya que la única variable involucrada en ésta distribución es el promedio λ lo cual no ofrece mucha información relevante para los objetivos de esta investigación.

Para el caso de la distribución binomial negativa (DBN), se aplicará la siguiente ecuación.

$$P_{r,p}(n) = \frac{\Gamma(n+r)}{n! \Gamma(r)} p^n (1-p)^r \quad (3.4)$$

La cual se evaluará tanto para los resultados finales de local(h) como para los visitantes(a), entonces.

Local.

$$P_{r_h, p_h}(n) = \frac{\Gamma(n + r_h)}{(n)! \Gamma(r_h)} p_h^n (1 - p_h)^{r_h}. \quad (3.5)$$

Visitante.

$$P_{r_a, p_a}(n) = \frac{\Gamma(n + r_a)}{(n)! \Gamma(r_a)} p_a^n (1 - p_a)^{r_a}. \quad (3.6)$$

Se considerarán las sumas(s) y diferencias(d) de los resultados finales para la DBN (Ver Apéndice B), donde la suma se expresa como,

$$P_{r_h, p_h, r_a, p_a}^\Sigma(s) = (1 - p_h)^{r_h} (1 - p_a)^{r_a} \frac{\Gamma(r_a + s)}{s! \Gamma(r_a)} {}_2F_1 \left(-s, r_h; 1 - s - r_a; \frac{p_h}{p_a} \right). \quad (3.7)$$

y las diferencias,

$$P_{r_h, p_h, r_a, p_a}^\Delta(d) = (1 - p_h)^{r_h} (1 - p_a)^{r_a} \frac{\Gamma(r_h + d)}{d! \Gamma(r_h)} {}_2F_1(r_h + d, r_a; 1 + d; p_h p_a). \quad (3.8)$$

donde ${}_2F_1$ es la función hipergeométrica.

Una vez definidas las funciones densidad de probabilidad a utilizar, se procederá a la realización de los algoritmos en lenguaje R para obtener los parámetros relevantes de cada una de las distribuciones, así como las gráficas que demuestren de manera visual el comportamiento de los resultados de fútbol para los equipos local(h), visitante(a), sumas(s) y diferencias(d), a fin de observar el mejor ajuste en las distribuciones escogidas.

3.2.1. Modelos microscópicos de probabilidad

Partiendo de la discusión previa en la cual las distribuciones de probabilidad sólo ofrecen información basadas en el mejor ajuste, se deben introducir modelos finos de probabilidad que permitan observar el comportamiento real de los resultados de fútbol de las diversas ligas en estudio, por lo tanto, a fin de comprender la profunda retroalimentación que existe entre los equipos, así como la probabilidad de marcar goles una vez se haya marcado con anterioridad y basados en la motivación-desmotivación de los jugadores, aunado al fortalecimiento del juego defensivo entre otros, se dice que realmente hay una componente positiva o negativa en el sistema de estudio.

Se incluirán dichas variables introduciendo un efecto de retroalimentación en el modelo de probabilidad, considerando el partido de fútbol dividido en N pasos, restringiendo

el partido a $N = 90$ pasos (minutos), con cada equipo teniendo su probabilidad de marcar durante esos pasos. La retroalimentación es introducida en el sistema, teniendo la probabilidad de marcar goles p dependiendo del número de goles marcados n hasta ahora, $p = p(n)$. Para ello como lo realizado por Janke et al.[2], se proponen los siguientes modelos de probabilidad.

$$p(n) = p(n - 1) + \kappa \quad (3.9)$$

El cual será el modelo aditivo A. Y luego.

$$p(n) = \kappa p(n - 1) \quad (3.10)$$

El cual será el modelo multiplicativo B.

La distribución binomial modificada resultante $P_N(n)$ para el total de goles anotados por un equipo puede ser calculada a través de la relación de recurrencia de Pascal,

$$P_N(n) = [1 - p(n)] P_{N-1}(n) + p(n - 1) P_{N-1}(n - 1) \quad (3.11)$$

Donde $p(n)$ será modificada como $p(n) = p_0 + \kappa n$ para el modelo A, y $p(n) = p_0 \kappa^n$ para el modelo B.

Para el caso del modelo aditivo A, se puede mostrar que para el límite continuo de $P_N(n)$, es decir $N \rightarrow \infty$ y manteniendo fijo $p_0 N$ y κN es dado por la distribución binomial negativa (DBN) con los parámetros $r = p_0 / \kappa$ y $p = 1 - e^{-\kappa N}$.

Para realizar los cálculos pertinentes a la relación de recurrencia de Pascal que permitan entender el comportamiento de las ligas de fútbol se realizará un algoritmo en lenguaje C, el cual a través de una minimización del error cuadrático se calcularán los parámetros p_0 y κ correspondientes cada liga. Una vez calculados dichos parámetros con cada modelo se realizará una comparación directa con los mismos parámetros calculados en el modelo binomial de tal manera de obtener la información necesaria que permita escoger el modelo que mejor se ajuste a los valores reales.

Pseudocódigo

El algoritmo propuesto para el cálculo de los parámetros p_0 y κ es el siguiente:

- Inicialmente se declara un arreglo con los valores reales de probabilidad para cada liga, $local(h)$ y $visitante(a)$ el cual será usado para posteriores cálculos.
- Se declara la función que calcula el error cuadrático.
- Dentro de la función se colocan las ecuaciones para la relación de recurrencia de Pascal con cada modelo microscópico.
- La función que calcula el error cuadrático depende de la resta de los valores calculados menos los valores reales.
- Dentro del programa principal se declaran las variables a utilizar, así como la semilla de los números aleatorios a usar entre 0 y 1 (basados en el cálculo de probabilidades).
- Se declaran y se llenan los arreglos con los números aleatorios y se dividen entre 100 para obtener resolución de 2 decimales.
- Se escogen de manera aleatoria dos valores dentro de los arreglos para obtener una primera pareja de p_0 y κ la cual será usada para realizar el cálculo del error cuadrático.
- Si el error cuadrático es menor que 10 (un valor muy grande puesto con la mano sabiendo que el valor nunca alcanzaría ese número), se escoge esa pareja de valores y se repite el proceso hasta obtener el error mínimo de la función.
- El proceso se repite a medida que el algoritmo descarta rangos de valores de p_0 y κ en mitades del rango, es decir si el valor del p_0 es mayor que el la diferencia del rango entre $2 h = (p_0max + p_0min)/2$ el valor del $p_0min = h$ sino $p_0max = h$. Esto se realiza de igual forma con los valores de κ , el mapeo de todos los valores se hace hasta obtener la resolución deseada.
- El paso se realiza hasta encontrar el valor del error mínimo que garantice la mejor pareja de valores de p_0 y κ .

Finalmente se puede decir que la metodología planteada permitió alcanzar de manera correcta los objetivos propuestos en esta investigación, lo cual se evidencia en el siguiente capítulo.

Capítulo 4

Resultados

En este capítulo se presentarán los resultados para los objetivos propuestos en esta investigación, donde se observarán los histogramas de frecuencias de goles para la liga inglesa (Premier League), liga femenina alemana (Allianz Frauen-Bundesliga), y la liga profesional de Venezuela (FVF). Se mostrarán además, los histogramas para los equipos local(h), visitante(a), sumas(s) y diferencias(d) de goles, así como los ajustes para la distribución binomial negativa para todas las variables ya mencionadas. Finalmente, a fin de entender el comportamiento de las ligas de fútbol en estudio, se observarán los ajustes de los modelos microscópicos propuestos en el capítulo anterior.

4.1. Frecuencia de goles

La frecuencia de goles anotados por un equipo, ya sea de local (n_h) o en condición de visitante (n_a), permite observar el comportamiento de las ligas de fútbol, a través de los resultados finales de los partidos. De igual forma sucede con las sumas (s) y diferencias (d) de los goles, las cuales nos brindan más información sobre el profesionalismo y competitividad de las ligas en estudio. Se observó la liga Premier inglesa, ya que se presenta como la mejor liga de Europa, además se observó la liga femenina alemana por su condición de fútbol amateur o emergente en la disciplina, y finalmente se dará a conocer los resultados para la liga profesional venezolana a fin de dar información relevante sobre su comportamiento, competitividad y profesionalismo.

En la figura 4.1 se muestra la frecuencia de goles de local para la liga Premier inglesa, donde se observa que los clubes locales desde la temporada 1946/47 hasta la temporada 2015/16 marcan 1 ó 2 goles al menos en más de 16.000 partidos disputados, esto significa

aproximadamente 56 % de los juegos, lo que indica que el comportamiento natural para los equipos que juegan de local es marcar al menos esa cantidad goles, es decir, ir en búsqueda de la victoria en cada partido. Dicho comportamiento se observa de igual forma en la Serie A italiana, La Liga española y en la Bundesliga alemana, (ver apéndice C).

Respecto al comportamiento de la frecuencia de goles de los partidos disputados en condición de visitante, se observa que los clubes tienden a marcar al menos 1 gol en aproximadamente 36 % de los partidos disputados, esto significa que la tendencia de los equipos visitantes es salir en la búsqueda de puntuar en dicha condición. Sin embargo comparado con los equipos locales se observa la tendencia notable de la ventaja de ser local en Inglaterra. Por otro lado, en ésta liga se han marcado 51233 goles en condición de local, que comparados con los 34959 goles de visita se demuestra la gran diferencia con la ventaja de la localía respecto de la visita. Dicho comportamiento se observa de igual forma en las ligas italiana, española y alemana, (ver apéndice C).

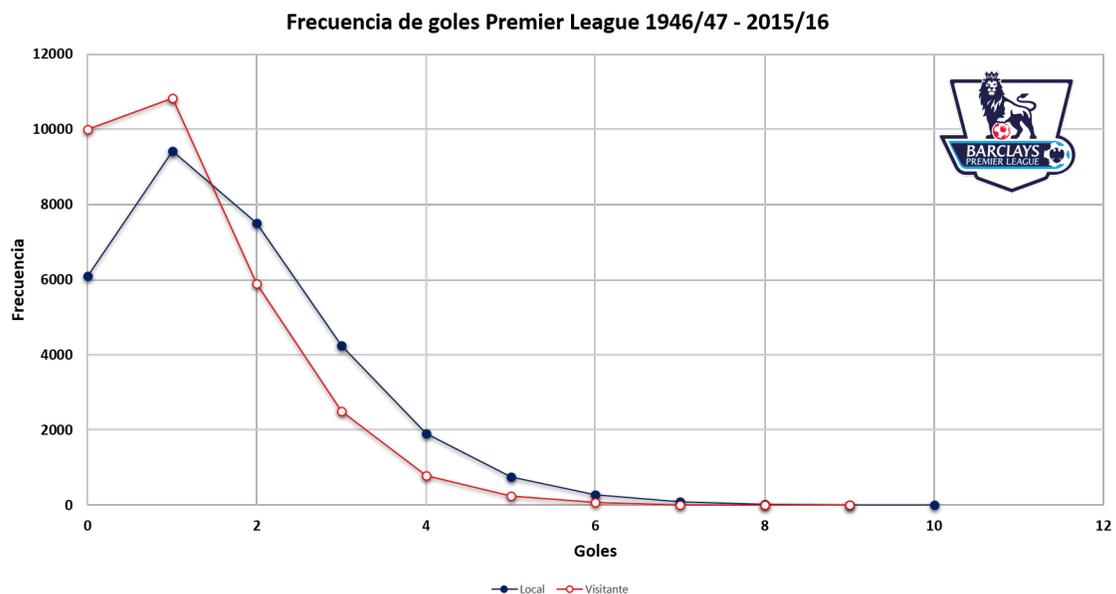


Figura 4.1: Frecuencia de goles de local y visitante de la liga Premier inglesa.

Observando la figura 4.2 para la liga femenina alemana (Allianz Frauen-Bundesliga) desde la temporada 1997/98 hasta la 2015/16, se denota que los equipos que juegan en condición de local marcan al menos 1 gol en mas de 600 partidos disputados, esto indica que en el total de 2508 partidos estudiados los locales marcan 1 gol en 24,6 % de los juegos en casa.

Por otro lado los equipos en condición visitante en la liga femenina alemana, tienen una tendencia de que los equipos suelen salir a mantener el cero en su arco, es decir, ir en búsqueda de 1 punto o el empate en 29% de los partidos. Esto denota la ventaja de los clubes locales, ya que mantener en cero al equipo visitante significa un mejor planteamiento del partido.

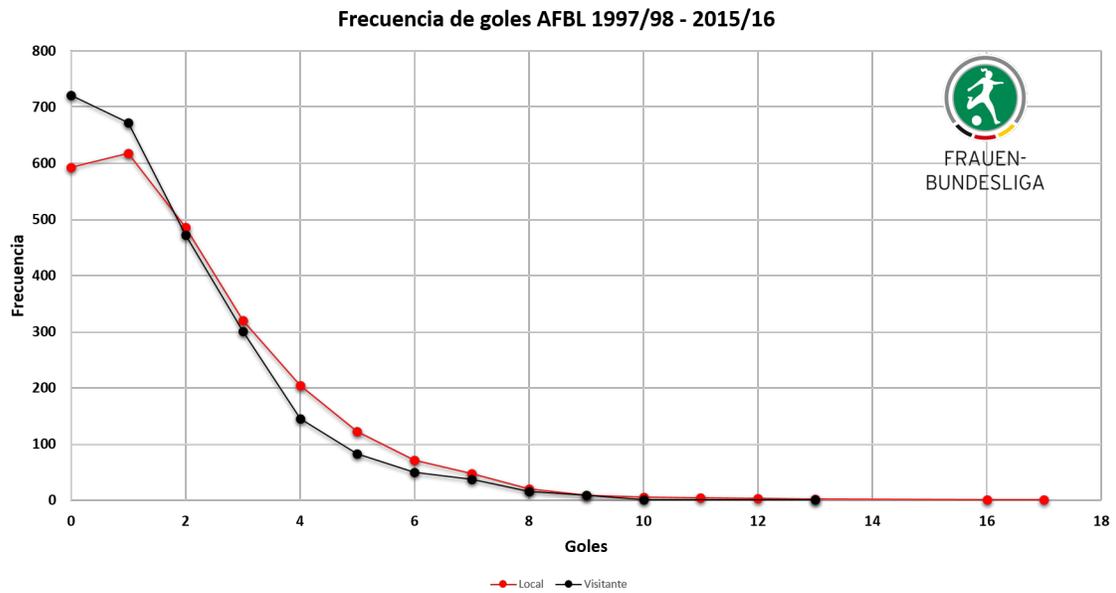


Figura 4.2: Frecuencia de goles de local y visitante de la Liga femenina alemana.

De igual forma se observa en la figura 4.3, que para la liga profesional de Venezuela desde 1957 hasta 2011 los equipos locales marcan al menos 1 gol en dicha condición en un 33 % de los partidos disputados. Dicho comportamiento indica que los clubes que juegan en casa tienden a buscar la victoria por al menos 1 gol de ventaja frente a los equipos visitantes.

Por otro lado en nuestra liga profesional venezolana, el comportamiento de los clubes visitantes tiene una tendencia parecida a la liga femenina alemana, ya que el resultado de 0 goles es el que mas se repite en dicha condición con un 38 % de los juegos disputados. Sin embargo, los partidos que finalizan con un gol marcado por el equipo visitante son alrededor de 37 % del total de juegos. En los juegos de la liga nacional se marcaron 13726 goles en condición de local y 8769 goles en condición visitante, demostrando la ventaja de ser local en el fútbol nacional.

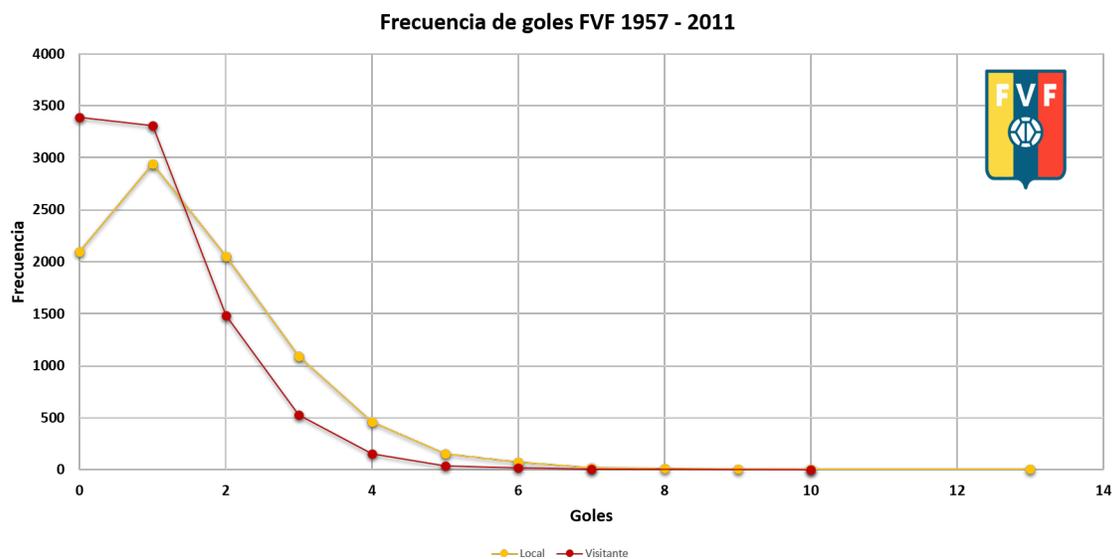


Figura 4.3: Frecuencia de goles de local y visitante de la Liga profesional de Venezuela.

4.2. Sumas y diferencias de goles

En la figura 4.4 se observa el histograma de frecuencia de sumas de goles para la liga Premier inglesa, recordando que $s = n_h + n_a$, es decir, la suma será igual a los goles de local más los goles del visitante. Es importante destacar que en ésta liga la suma más frecuente es la de 2 goles con 23 % de los partidos y 3 goles con 20,7 % de los partidos,

las posibilidades para 2 goles son 2-0, 1-1, y 0-2 y para 3 goles las posibilidades son 3-0, 2-1, 1-2, y 0-3 lo que marca una tendencia de profesionalismo en la liga inglesa, ya que son resultados esperados en ligas de alto rendimiento, además se percibe el mismo comportamiento de la curva comparado con la gráfica de los goles de local en Inglaterra.

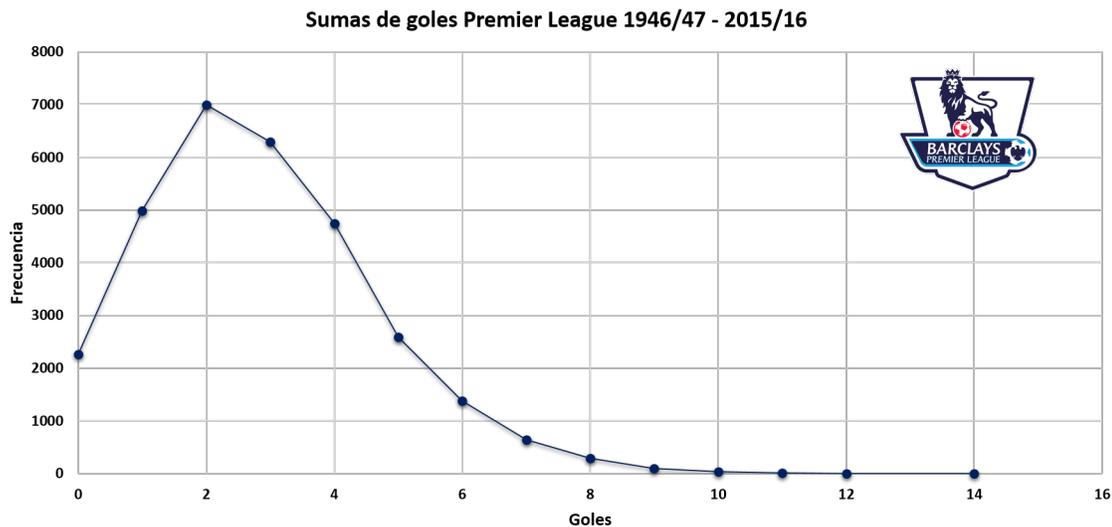


Figura 4.4: Frecuencia de las sumas de goles de la liga Premier inglesa.

En el caso de la liga femenina alemana las sumas de goles mas frecuentes son las 3 goles con 20,8 % y 4 goles que significa el 18,6 %, aumentando las posibilidades de goleadas en los partidos disputados, ya que 4 goles tiene como posibilidades 4-0, 3-1, 2-2, 1-3 y 0-4, esto implica que la liga se comporta de manera menos profesional con la aparición de goleadas, sugiriendo planteamientos más ofensivos y menor preocupación por tácticas defensivas (ver figura 4.5).

En el caso de la liga profesional venezolana, se observa que las sumas de goles mas frecuentes son las de 1 y 2 goles, con 20,1 % y 23,6 % respectivamente (figura 4.6), lo cual nos permite entender que nuestra liga se comporta de manera más competitiva en cuanto a los resultados más frecuentes, sin embargo el fútbol nacional a partir del año 2007 por recomendación de la Confederación Sudamericana de Fútbol (CONMEBOL), aumentó la cantidad de equipos a 18, esto significa que la cantidad de partidos disputados respecto a las liga europeas es mucho menor con 8912 partidos hasta el año 2011, lo cual implica que los resultados más frecuentes también tienen que ver con el progreso de nuestro fútbol a una liga más profesionalizada.

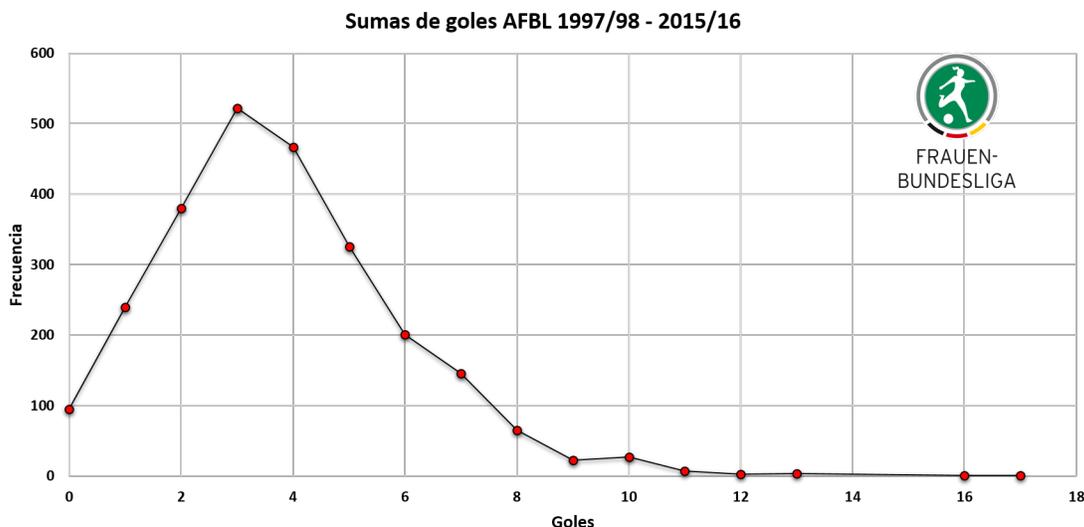


Figura 4.5: Frecuencia de las sumas de goles de la Liga femenina alemana.

Las diferencias de los goles $d = n_h - n_a$, es decir goles de local menos goles de visitante nos revelará con mas detalle la ventaja de la localía en el fútbol profesional. En la figura 4.7 se observa el comportamiento de las diferencias de goles de la liga inglesa, donde el 0 y el 1 positivo son las diferencias mas repetidas. El 0 significa todos las igualdades en el fútbol ingles con 25,9%, lo cual implica que en 7860 partidos se alcanzó el empate a goles, por lo tanto se dice que el visitante logró puntuar. No obstante el 1 positivo consta de 6713 partidos con un 22,1%, marcando una diferencia notable con el -1, el cual implica un 14,6% de los juegos disputados, marcando una ventaja notable para los equipos locales. Para completar la idea el local gana 48,73% de los partidos, se empatan 25,91% y el visitante gana el 25,35% de los juegos disputados demarcando la fuerte ventaja de la localía en Inglaterra.

Es importante destacar que ésta tendencia se repite en las ligas italiana y alemana, sin embargo en la liga española se observa una desigualdad notable, donde ser local garantiza ganar por diferencia de 1 gol al menos 24,6% de los partidos y comparando con los partidos ganados por el visitante por diferencia de -1 gol se tiene el 12,2% (ver apéndice D).

En el caso de la liga femenina alemana, se observa un comportamiento más simétrico en la figura 4.8, por un lado se repite la tendencia de que el empate es el resultado más frecuente en el fútbol con 15,1% del total de partidos, sin embargo, existe una igualdad

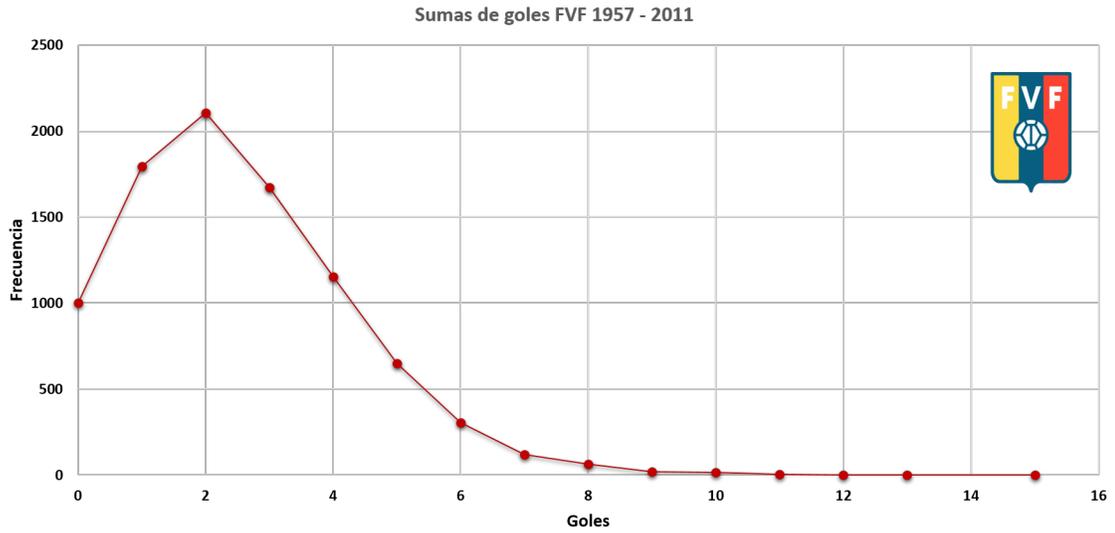


Figura 4.6: Frecuencia de las sumas de goles de la Liga profesional de Venezuela.

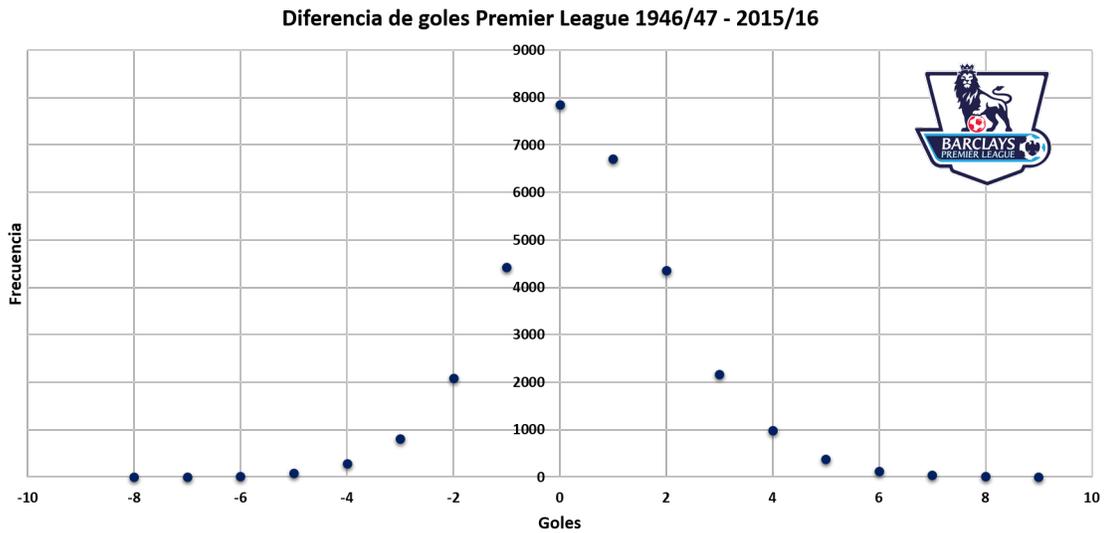


Figura 4.7: Frecuencia de las diferencias de goles de la liga Premier inglesa.

respecto a la diferencia de 1 gol (positivo y negativo), es decir, se observa que los locales ganan por diferencia de 1 gol en 13,5% de los partidos y el visitante 13,2% de los mismos. Si se observa la diferencia de dos goles el local gana 9,8% de los juegos y el visitante gana 8,9% de los partidos por diferencia de 2 goles, acentuándose cuando la diferencia es más grande. En ésta liga el local gana el 46,33% de los partidos, se empatan 15,11% de los juegos y el visitante gana 38,55% de los partidos disputados, es decir se demuestra la ventaja de ser local, pero se marca la tendencia de alta competitividad pero menor profesionalismo de la misma, lo cual es típico de un deporte emergente en los últimos años como lo es el fútbol femenino.

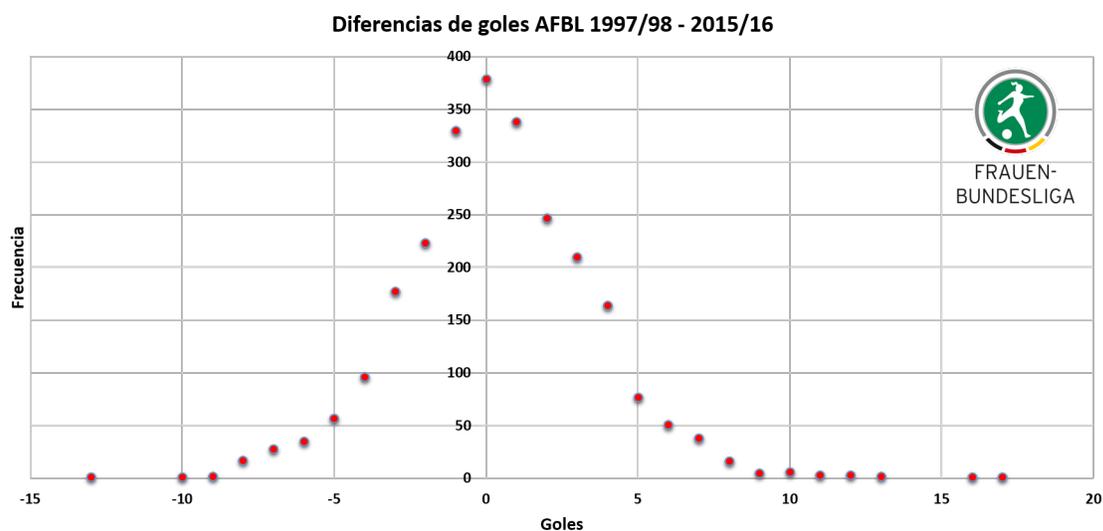


Figura 4.8: Frecuencia de las diferencias de goles de la Liga femenina alemana.

Finalmente para las diferencias de los goles en la liga profesional venezolana, de nuevo se observa la tendencia de que el empate es el resultado más frecuente en el fútbol con un 29,7% del total de los partidos disputados, además se marca la ventaja de la localía al visualizar la diferencia por 1 gol positivo (gana el local) con un 25,2% de partidos (figura 4.9). En el fútbol venezolano el local obtiene la victoria en 48,73% de los partidos, mientras que se empatan 29,67% de los juegos y el visitante gana el 21,60% de los partidos disputados, esto denota de nuevo la ventaja de ser local en el fútbol profesional y la similitud con el fútbol inglés donde es una gran ventaja ser local.

El comportamiento de las frecuencias de goles tanto de local como de visita, así como las sumas de goles en las ligas europeas, femenina alemana y la liga profesional venezolana,

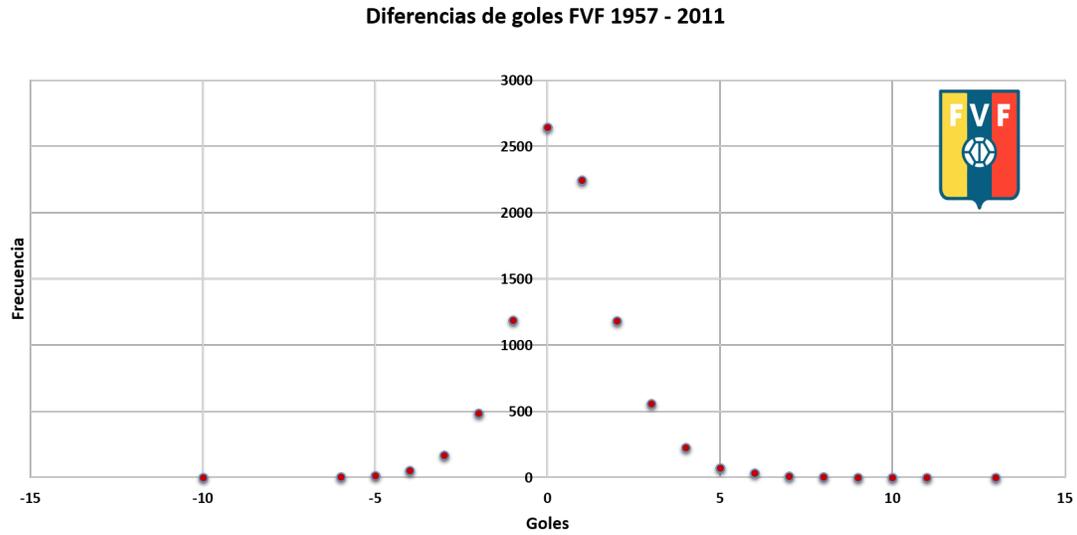


Figura 4.9: Frecuencia de las diferencias de goles de la Liga profesional de Venezuela.

sugieren con lo observado en las gráficas que pueden ajustarse a la distribución binomial negativa, como se demostrará en la siguiente sección.

4.3. Modelo de Probabilidad

Luego de observar el comportamiento de los resultados finales de los partidos de la liga Premier inglesa, la liga femenina alemana y la liga de fútbol profesional venezolano y basados en el trabajo de Janke et al.[2], se realizó el ajuste a la distribución binomial negativa (DBN), el cual brindará información en cuanto a la competitividad de las ligas, comparando el comportamiento de los datos en contraste con lo esperado por el modelo binomial.

En la tabla 4.1 se contemplan los resultados para la distribución binomial negativa en la Premier y se coloca el resultado para la distribución de Poisson (λ), a modo de comparación. Se destaca que la probabilidad de cada evento p , así como el parámetro de los casos favorables r , siempre es mayor para local que para el visitante. Altos valores del parámetro r genera bajos valores para κ , lo cual se traduce que los equipos ingleses a la hora de llevar la ventaja en el marcador tienden a reforzar tácticas defensivas, esto demuestra un comportamiento profesionalizado de los mismos.

Tabla 4.1: Parámetros de las distribuciones de probabilidad para los datos de la liga Premier inglesa 1946/47 - 2015/16

| | | Home | Away |
|---------|-----------|------------------|------------------|
| Poisson | λ | 1,68 | 1,15 |
| DBN | p | 0,11 \pm 0,01 | 0,08 \pm 0,01 |
| | r | 13,22 \pm 0,97 | 12,07 \pm 1,15 |
| | p_0 | 0,0171 | 0,0111 |
| | κ | 0,0012 | 0,0009 |

En la figura 4.10 se observa el modelo de la DBN (blanco) respecto a los datos ajustados a la distribución (azul), para los partidos de local de la liga inglesa. Se puede percibir que los resultados observados respecto al modelo propuesto se ajustan de buena manera, además no existen fuertes variaciones en la cola de la distribución, lo cual comparado con la liga española e italiana, se presentan dichas variaciones en las colas (ver apéndice E), esto indica un patrón de alta competitividad en la liga inglesa. De igual forma en la figura 4.11, se muestran los datos ajustados a la DBN respecto a los cálculos del modelo binomial en la Premier, se muestra que la curva de los valores reales se ajustan al modelo binomial, aunque mostrando un comportamiento diferente al final de la distribución en

los valores, lo cual indica la presencia de algunas goleadas en el comportamiento observado.

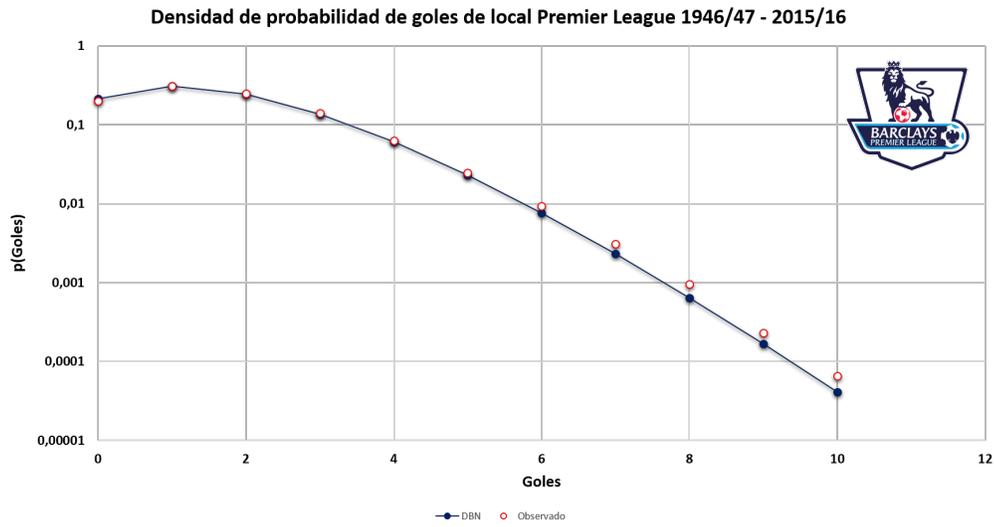


Figura 4.10: Densidad de probabilidad para los goles de local en la liga premier inglesa.

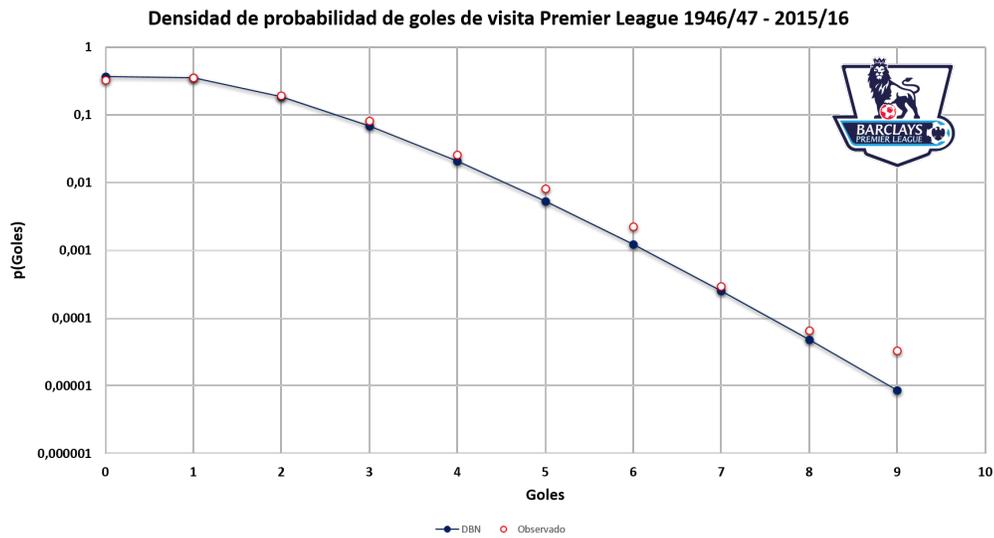


Figura 4.11: Densidad de probabilidad para los goles de visita en la liga Premier inglesa.

A continuación en la figura 4.12 se muestran los ajustes a la distribución binomial negativa para las sumas de goles, y en la figura 4.13 el ajuste para las diferencias de goles en la liga Premier inglesa, donde se observa que en el caso de las sumas de goles, los valores observados se ajustan hasta los 10 goles y se repite el cambio de comportamiento a la derecha de la distribución. Por otro lado, en cuanto a las diferencias de goles en la liga inglesa, se muestra que los datos observados se ajustan al modelo binomial sin mostrar dispersión, esto se traduce en un comportamiento competitivo respecto a las diferencias de los goles.

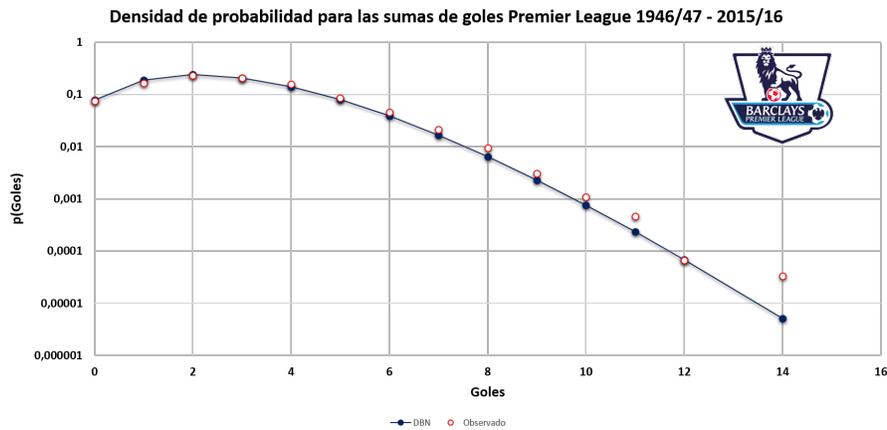


Figura 4.12: Densidad de probabilidad para las sumas de goles de la liga premier inglesa.

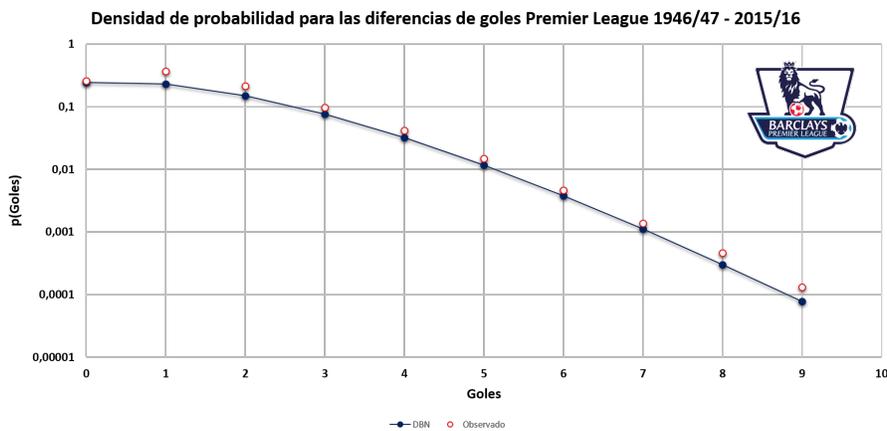


Figura 4.13: Densidad de probabilidad para las diferencias de goles de la liga Premier inglesa.

Respecto a la liga femenina alemana, se muestra como una liga menos profesional pero muy competitiva, ya que las probabilidades p son cercanas entre sí tanto para el local como para el visitante y comparados con la liga inglesa son mucho mayores, por otro lado el parámetro r es significativamente más pequeño respecto a la premier, esto se traduce en valores más grandes de κ , es decir, marcar un gol en la AFBL es más alentador que marcar un gol en la premier, sin embargo esto puede ser interpretado como la tendencia a mantener las tácticas ofensivas cuando se lleva la ventaja (ver tabla 4.2).

Tabla 4.2: Parámetros de las distribuciones de probabilidad para los datos de la liga femenina alemana (AFBL) 1997/98 - 2015/16

| | | Home | Away |
|---------|-----------|-----------------|-----------------|
| Poisson | λ | 2,06 | 1,71 |
| DBN | p | $0,49 \pm 0,03$ | $0,45 \pm 0,03$ |
| | r | $2,10 \pm 0,13$ | $2,03 \pm 0,14$ |
| | p_0 | 0,0157 | 0,0134 |
| | κ | 0,0074 | 0,0066 |

Los ajustes del modelo binomial para el fútbol femenino alemán observado en la figura 4.14 para los goles de local, y la figura 4.15 para los goles de visitante, muestran que los datos observados se ajustan de manera parcial al modelo propuesto, esto implica que los datos solo se ajustan al modelo para 8 goles o menos, denotando un cambio de comportamiento al final de la distribución. Esto se traduce en un comportamiento menos profesional y altamente competitivo, ya que se marcan una gran cantidad goles tanto de local como de visitante.

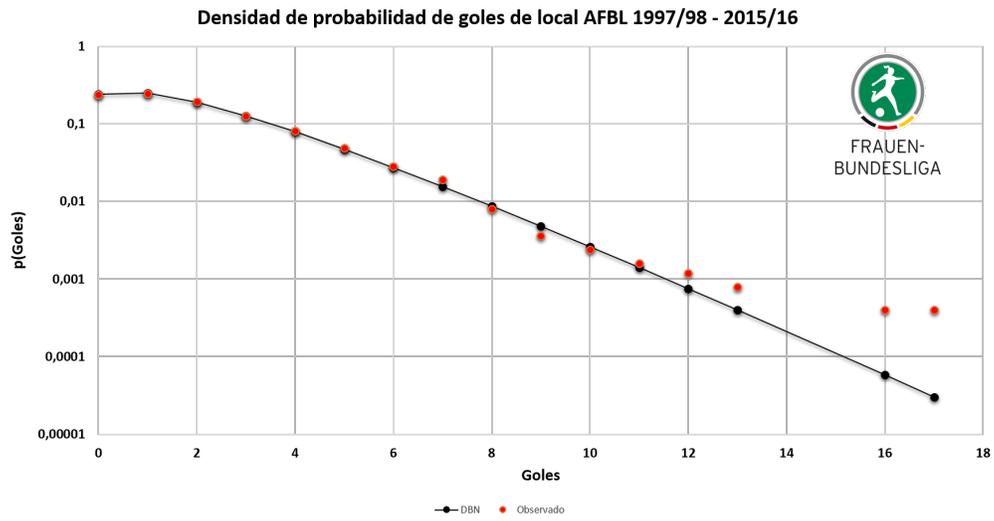


Figura 4.14: Densidad de probabilidad para los goles de local en la liga femenina alemana.

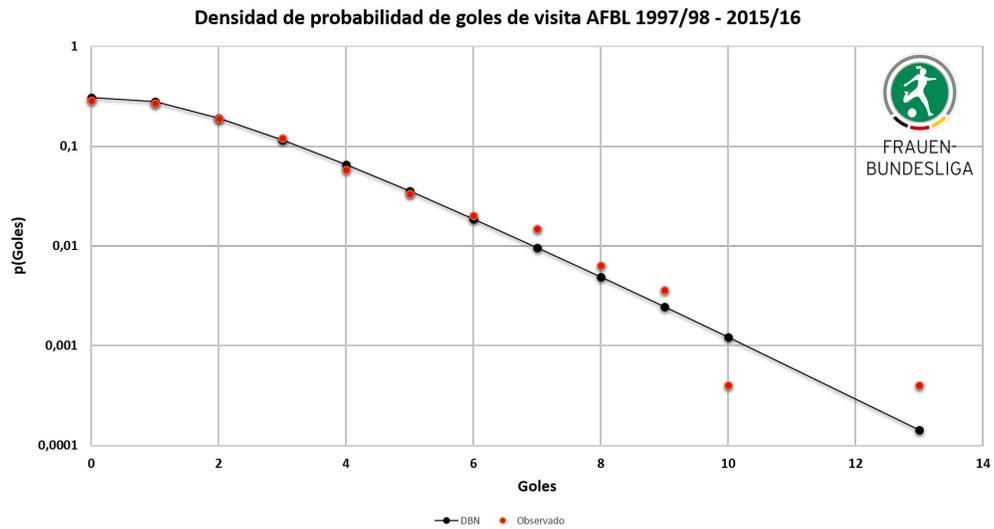


Figura 4.15: Densidad de probabilidad para los goles de visita en la liga femenina alemana.

Por otra parte observando el comportamiento para las sumas (figura 4.16) y diferencias (figura 4.17) de goles en la AFBL, se evidencia que en la sumas, el comportamiento de los valores reales se ajusta de manera parcial al modelo binomial negativo, destacando también que en las diferencias los datos observados no se ajustan al modelo. Esto significa de manera directa que la liga al ser menos profesionalizada, tiende a tener fluctuaciones bien marcadas respecto al modelo propuesto, por la aparición de diferencias muy parecidas entre los equipos locales y visitantes.

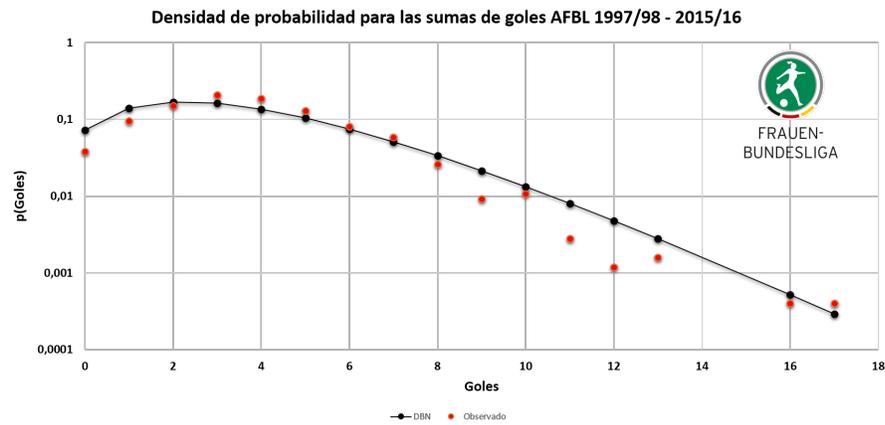


Figura 4.16: Densidad de probabilidad para las sumas de goles en la liga femenina alemana.

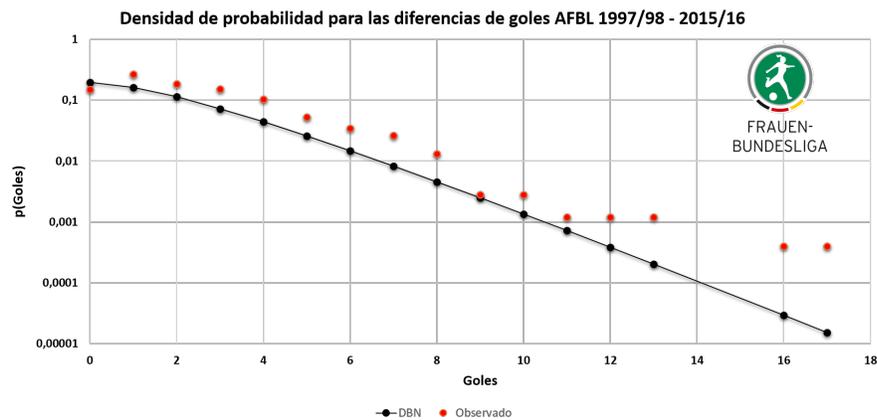


Figura 4.17: Densidad de probabilidad para las diferencias de goles en la liga femenina alemana.

Tomando en consideración que la liga inglesa es una liga profesional y muy competitiva, y por otro lado la liga femenina alemana es una liga menos profesionalizada con buenos índices de competitividad, se observa en la tabla 4.3 los resultados del ajuste al modelo binomial negativo para la liga profesional de Venezuela. Donde se percibe la ventaja de la localía en el parámetro p , además valores altos del parámetro r lo que se traduce en números más bajos en el parámetro κ . Esto marca una tendencia donde los equipos se sienten motivados a reforzar tácticas defensivas a la hora de marcar goles e ir en ventaja, sin embargo el torneo nacional muestra un comportamiento híbrido ya que suelen aparecer goleadas más frecuente de lo esperado.

Tabla 4.3: Parámetros de las distribuciones de probabilidad para los datos de la liga Profesional de fútbol de Venezuela (FVF) 1957 - 2011

| | | Home | Away |
|---------|-----------|-----------------|------------------|
| Poisson | λ | 1,54 | 0,98 |
| DBN | p | $0,16 \pm 0,01$ | $0,07 \pm 0,01$ |
| | r | $8,0 \pm 0,75$ | $12,54 \pm 2,58$ |
| | p_0 | 0,0154 | 0,0101 |
| | κ | 0,0019 | 0,0008 |

Observando el ajuste para el modelo binomial negativo, tanto para local como para visitante en el fútbol nacional, se muestra que los datos se ajustan de manera parcial al modelo propuesto, ya que se percibe un buen comportamiento hasta los 7 goles en condición de local y 5 goles de visitante perdiendo el ajuste al final de la distribución. En ambos casos se evidencian goleadas al final de la distribución denotando cierto comportamiento amateur, lo cual corrobora la conducta híbrida del torneo nacional. (Ver figuras 4.18 y 4.19).

Finalmente se analizan los ajustes para las sumas (figura 4.20) y diferencias (figura 4.21) de goles en la liga venezolana. Donde se percibe un buen ajuste al modelo binomial negativo para las sumas, excepto en las colas y de igual forma sucede con las diferencias, lo cual ratifica la idea de que el fútbol venezolano tiene un comportamiento híbrido que va en transición hasta una liga mucho más profesionalizada en el futuro.

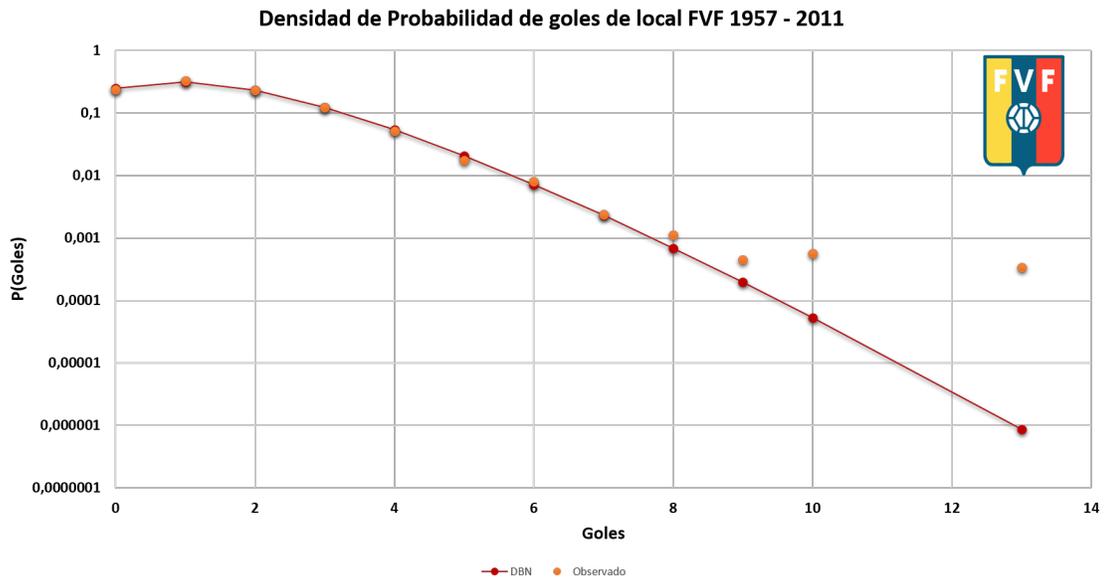


Figura 4.18: Densidad de probabilidad para los goles de local en la liga profesional de Venezuela.

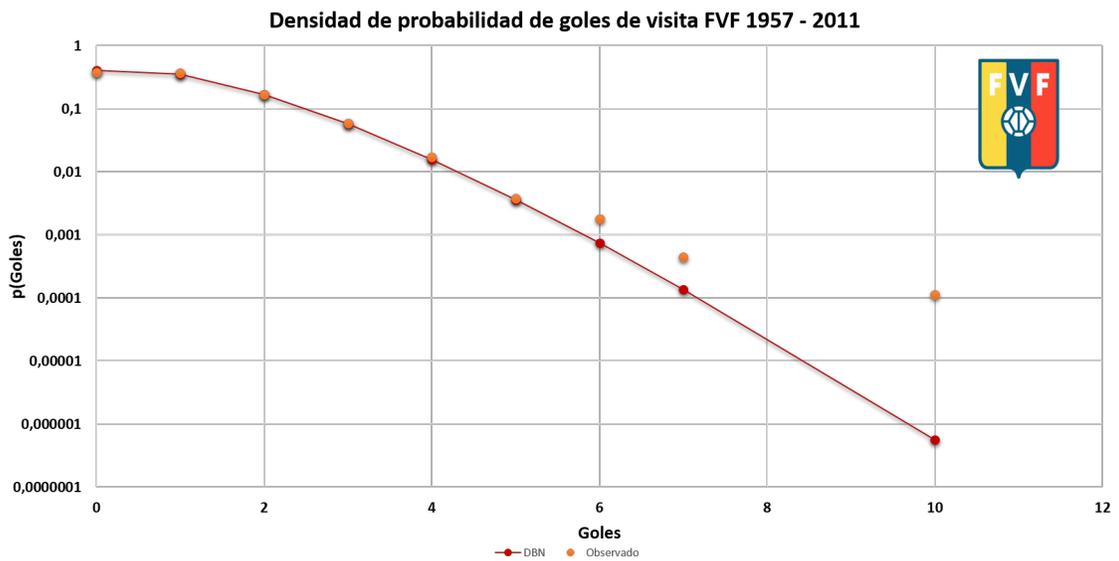


Figura 4.19: Densidad de probabilidad para los goles de visita en la liga profesional de Venezuela.

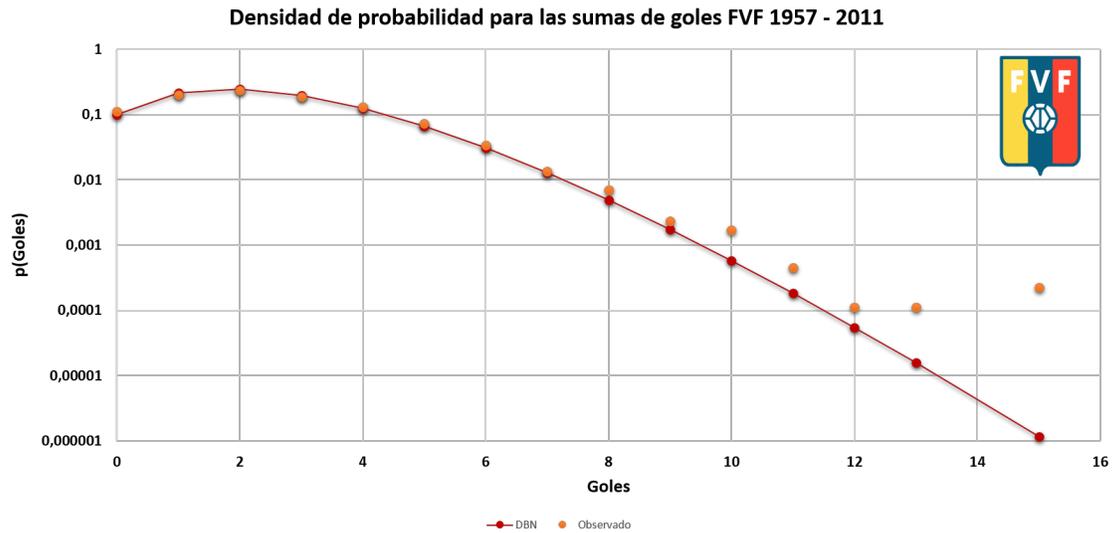


Figura 4.20: Densidad de probabilidad para las sumas de goles en la liga profesional de Venezuela.

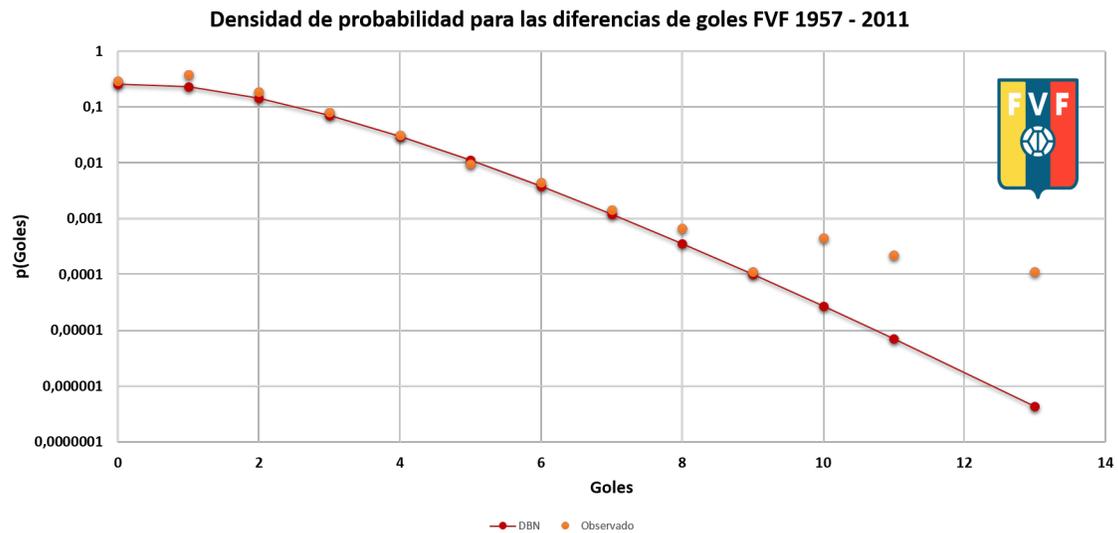


Figura 4.21: Densidad de probabilidad para las diferencias de goles en la liga profesional de Venezuela.

4.4. Modelos microscópicos

Luego de realizar la discusión previa del ajuste de los datos a la distribución binomial negativa, se propusieron dos modelos microscópicos de probabilidad, los cuales permitirán estudiar el sistema planteado, como lo es el fútbol con sus numerosas variables de manera más profunda. Por un lado entender la ventaja de ser local en las ligas estudiadas, y por el otro la fuerte retroalimentación existente entre los equipos local y visitante.

Se presentan los siguientes resultados para los modelos microscópicos en las ligas inglesa, femenina alemana y venezolana, para los modelos A y B propuestos anteriormente.

4.4.1. Liga premier inglesa

Se puede observar en la tabla 4.4 los valores calculados de los modelos microscópicos A y B para la liga Premier inglesa, donde se evidencia que el ajuste al modelo A es el mejor en comparación al modelo B ya que los valores de p_0 y κ se acercan entre sí a los calculados con el modelo binomial. Lo cual se traduce en que el modelo binomial negativo propuesto anteriormente se comporta como el límite del modelo A. De igual forma los valores de p_0 y κ del modelo A, se refieren a que la liga inglesa tiene una tendencia de comportamiento profesional ya que se evidencia el fortalecimiento de las tácticas defensivas una vez se tiene la ventaja en el partido y una gran retroalimentación entre los equipos.

Tabla 4.4: Parámetros de los modelos microscópicos para los datos de la liga Premier inglesa 1946/47 - 2015/16

| | | Home | Away |
|----------|----------|-----------------------|-----------------------|
| DBN | p_0 | 0,0171 | 0,0111 |
| | κ | 0,0012 | 0,0009 |
| Modelo A | p_0 | $0,01756 \pm 0,00003$ | $0,01218 \pm 0,00003$ |
| | κ | $0,00123 \pm 0,00003$ | $0,00087 \pm 0,00003$ |
| Modelo B | p_0 | $0,01758 \pm 0,00003$ | $0,01218 \pm 0,00003$ |
| | κ | $1,06934 \pm 0,00003$ | $1,07107 \pm 0,00003$ |

4.4.2. Liga femenina alemana (AFBL)

De igual forma que en el caso anterior, el modelo A se ajusta de mejor manera en la liga femenina alemana, aunque en este caso los valores de p_0 y κ sugieren que la liga es mucho mas competitiva y menos profesional. Por consiguiente anotar un gol en ésta liga es muy alentador y anima a los equipos a anotar mas goles y en una reafirmación de las tácticas ofensivas (ver tabla 4.5).

Tabla 4.5: Parámetros de los modelos microscópicos para los datos de la liga femenina alemana (AFBL) 1997/98 - 2015/16

| | | Home | Away |
|----------|----------|-----------------------|-----------------------|
| DBN | p_0 | 0,0157 | 0,0134 |
| | κ | 0,0074 | 0,0066 |
| Modelo A | p_0 | $0,01589 \pm 0,00003$ | $0,01375 \pm 0,00003$ |
| | κ | $0,00770 \pm 0,00003$ | $0,00679 \pm 0,00003$ |
| Modelo B | p_0 | $0,01698 \pm 0,00003$ | $0,01440 \pm 0,00003$ |
| | κ | $1,24999 \pm 0,00003$ | $1,24999 \pm 0,00003$ |

4.4.3. Liga profesional de Venezuela FVF

Finalmente para la liga profesional de Venezuela, tal como ocurrió en los dos casos anteriores, se percibe que el ajuste al modelo A, es el mejor, representando de nuevo el límite del modelo binomial negativo. Es importante destacar que los valores de p_0 y κ en nuestro fútbol nacional sugieren el comportamiento de una liga de alta competencia, donde los valores bajos de κ muestran el fortalecimiento de la táctica defensiva en pro de la obtención del mejor resultado una vez se lleva la ventaja y la fuerte retroalimentación entre los equipos (tabla 4.6).

Tabla 4.6: Parámetros de los modelos microscópicos para los datos de la liga Profesional de fútbol de Venezuela (FVF) 1957 - 2011

| | | Home | Away |
|----------|----------|-----------------------|-----------------------|
| DBN | p_0 | 0,0154 | 0,0101 |
| | κ | 0,0019 | 0,0008 |
| Modelo A | p_0 | $0,01578 \pm 0,00003$ | $0,01059 \pm 0,00003$ |
| | κ | $0,00127 \pm 0,00003$ | $0,00048 \pm 0,00003$ |
| Modelo B | p_0 | $0,01579 \pm 0,00003$ | $0,01059 \pm 0,00003$ |
| | κ | $1,08091 \pm 0,00003$ | $1,01565 \pm 0,00003$ |

Capítulo 5

Conclusiones

En éste capítulo se plantean las conclusiones a los objetivos propuestos en este trabajo de investigación provenientes de los resultados obtenidos.

Luego de realizarse la recolección de todos los resultados finales de los partidos de fútbol para las cuatro grandes ligas europeas (España, Italia, Alemania e Inglaterra), así como los datos de la liga femenina alemana, (Allianz Frauen-Bundesliga) y para la liga profesional de fútbol de Venezuela (FVF), se observaron los histogramas de frecuencias de goles de local, visitante, sumas y diferencias de goles.

Se observó basados en los histogramas de frecuencias de goles anotados por los equipos locales y visitantes que las ligas europeas representadas por la liga Premier inglesa, marcan una tendencia de ventaja importante cuando se juega de local, por otro lado, en la liga femenina alemana y en la liga profesional de Venezuela se observa la misma tendencia.

Respecto a los histogramas de frecuencias de sumas y diferencias de goles se observó para la liga Premier que las sumas mas frecuentes son las de 2 y 3 goles mostrando una alto profesionalismo y competitividad al no aparecer goleadas de mayor cantidad de goles, lo cual se asemeja al fútbol nacional indicando una alta competitividad en el torneo local, sin embargo, se puede traducir también como la baja calidad del torneo, entendiendo que el fútbol de nuestro país se encuentra en pleno desarrollo. Las diferencias de goles mostraron que el resultado más frecuente en todas las ligas de fútbol es el empate a cero goles, donde en la tendencia hacia los valores positivos marca la ventaja de ser local en las ligas profesionales exceptuando la liga femenina alemana donde el comportamiento es más competitivo pero menos profesional.

En cuanto a las distribuciones de probabilidad estudiadas en ésta investigación, se ajustaron los resultados de fútbol de las ligas antes mencionadas a la distribución binomial negativa para los valores de local, visitante, sumas y diferencias de goles. Donde se puede concluir que la liga Premier inglesa muestra un comportamiento de alta competitividad ya que los ajustes de los 4 factores estudiados fueron muy buenos, además denotando altos valores de r , que a su vez generaron valores pequeños de κ . Esto significa que las ligas de fútbol europeas estudiadas tienen un comportamiento más profesionalizado donde luego de marcar goles se ajustan las tácticas defensivas para mantener la ventaja, a diferencia de la liga femenina donde los valores más grandes de κ , sugieren que el hecho de marcar goles alientan a los equipos a marcar de nuevo.

En cuanto a la liga profesional de Venezuela, los valores obtenidos se asemejan a los de las ligas de Europa, es decir, se observó de nuevo que valores grandes de r generan valores pequeños de κ , lo cual permite concluir que se fortalecen los planteamientos defensivos a la hora de obtener la ventaja en los partidos, denotando un comportamiento profesional en nuestro fútbol nacional. Es importante resaltar que los valores bajos de κ , son un indicativo de alto profesionalismo en las ligas estudiadas, ya que los equipos al anotar goles tienen la tendencia de fortalecer las tácticas defensivas, sin embargo los equipos rivales mantienen la postura al no decaer en motivación una vez se encuentran en desventaja.

Los ajustes realizados con los modelos microscópicos de probabilidad, arrojaron que el modelo A se ajusta de mejor manera (respecto al modelo B) a los datos proporcionados por el modelo binomial negativo en las cuatro grandes ligas de Europa, la liga femenina alemana y la liga profesional de Venezuela, lo cual se traduce en que el modelo binomial se comporta como el límite de la modelo microscópico A, reafirmando la fuerte retroalimentación que existe entre los equipos a la hora de marcar goles.

Ya que el concepto sobre profesionalismo y competitividad se ajusta más al comportamiento observado en la liga Premier inglesa, en la cual todos los equipos tienen las mismas posibilidades de ganar los partidos, aunque exista la notable ventaja de la localía, se puede concluir entonces que el futuro de dichos conceptos tienden a la mejora técnico-táctica progresiva, en la cual posiblemente se marcarán menos goles y se reforzarán los planteamientos defensivos de los equipos, todo en búsqueda de la victoria con el mejor planteamiento posible.

Apéndice A

Desarrollo Numérico A

En este primer apéndice se desarrollará con detalle como se utiliza la distribución de Poisson con todas las variables de éste trabajo de investigación, donde las variables en uso son las siguientes: goles de local n_h , goles de visita n_a , promedio de goles de local λ_h y promedio de goles de visita λ_a . Por otro lado se observará con detenimiento el procedimiento a seguir para calcular las funciones de densidad de probabilidad para las sumas $P_{\lambda_h, \lambda_a}^{\Sigma}(s)$ y diferencias $P_{\lambda_h, \lambda_a}^{\Delta}(d)$ de los goles.

A.1. Distribución de Poisson

$$P_{\lambda_h}^h(n_h) = \frac{\lambda_h^{n_h}}{n_h!} e^{(-\lambda_h)}. \quad P_{\lambda_a}^a(n_a) = \frac{\lambda_a^{n_a}}{n_a!} e^{(-\lambda_a)}. \quad (\text{A.1})$$

donde n_h y n_a son los resultados finales de los partidos de local y visitante respectivamente y los parámetros λ_h y λ_a son los promedios de goles anotados por los equipos local y visitante donde $\lambda = \langle n \rangle$. Adicionalmente se puede considerar las densidades de probabilidad para las sumas y diferencias de los goles donde la suma es $s = n_h + n_a$ y la diferencia $d = n_h - n_a$ respectivamente.

Para las Sumas,

$$\begin{aligned} P_{\lambda_h, \lambda_a}^{\Sigma}(s) &= \sum_{n=0}^s P_{\lambda_h}^h(n) P_{\lambda_a}^a(s-n) \\ &= \sum_{n=0}^s \frac{\lambda_h^{s-n} \lambda_a^n}{(s-n)! n!} e^{-\lambda_h} e^{-\lambda_a} \\ &= \sum_{n=0}^s \frac{\lambda_h^{s-n} \lambda_a^n s!}{(s-n)! n! s!} e^{-\lambda_h} e^{-\lambda_a} \end{aligned}$$

$$\begin{aligned}
&= \sum_{n=0}^{n_h+n_a} \frac{\lambda_a^{n_h+n_a-n} \lambda_h^n (n_h+n_a)!}{(n_h+n_a)!(n_h+n_a)!(n_h+n_a-n)!} e^{-\lambda_h} e^{-\lambda_a} \\
&= \frac{(\lambda_h+\lambda_a)^s}{s!} e^{[-(\lambda_h+\lambda_a)]}
\end{aligned}$$

entonces para las sumas,

$$P_{\lambda_h, \lambda_a}^{\Sigma}(s) = \sum_{n=0}^s P_{\lambda_h}^h(n) P_{\lambda_a}^a(s-n) = \frac{(\lambda_h + \lambda_a)^s}{s!} e^{[-(\lambda_h+\lambda_a)]} \quad (\text{A.2})$$

Para las diferencias,

$$\begin{aligned}
P_{\lambda_h, \lambda_a}^{\Delta}(d) &= \sum_{n=0}^{\infty} P_{\lambda_h}^h(n+d) P_{\lambda_a}^a(n) \\
&= \sum_{n=0}^{\infty} \frac{\lambda_h^{n+d}}{(n+d)!} e^{-\lambda_h} \frac{\lambda_a^n}{n!} e^{-\lambda_a} \\
&= \sum_{n=0}^{\infty} = \frac{\lambda_h^{n+(n_h-n_a)} \lambda_a^n}{(n+d)! n!} e^{-(\lambda_h+\lambda_a)} \\
&= \sum_{n=0}^{\infty} = \frac{\lambda_h^n \lambda_h^{n_h-n_a} \lambda_a^n}{(n+d)! n!} e^{-(\lambda_h+\lambda_a)} = \sum_{n=0}^{\infty} = \frac{\lambda_h^n \lambda_h^d \lambda_a^n}{(n+d)! n!} e^{-(\lambda_h+\lambda_a)} \\
&= \sum_{n=0}^{\infty} = \frac{\lambda_h^n \lambda_h^d \lambda_a^n \lambda_a^{(d/2)}}{(n+d)! n! \lambda_a^{(d/2)}} e^{-(\lambda_h+\lambda_a)} \\
&= \sum_{n=0}^{\infty} = \frac{\lambda_h^n \lambda_h^{(d/2)} \lambda_h^{(d/2)} \lambda_a^n \lambda_a^{(d/2)}}{(n+d)! n! \lambda_a^{(d/2)}} e^{-(\lambda_h+\lambda_a)} \\
&= \sum_{n=0}^{\infty} = \frac{\lambda_h^n \lambda_a^n (\lambda_h \lambda_a)^{(d/2)}}{(n+d)! n!} \left(\frac{\lambda_a}{\lambda_h} \right)^{(d/2)} e^{-(\lambda_h+\lambda_a)}
\end{aligned}$$

$$\text{donde } I_d(2\sqrt{\lambda_h \lambda_a}) = \sum_{n=0}^{\infty} \frac{1}{n!(n+d)!} (\lambda_h \lambda_a)^n (\lambda_h \lambda_a)^{(d/2)}$$

$$\text{Usando } (n+d)! = \Gamma(1+n+d)$$

entonces

$$= \sum_{n=0}^{\infty} \frac{1}{n! \Gamma(1+n+d)} (\lambda_h \lambda_a)^n (\lambda_h \lambda_a)^{(d/2)}$$

finalmente para las diferencias,

$$P_{\lambda_h, \lambda_a}^{\Delta}(d) = \sum_{n=0}^{\infty} P_{\lambda_h}^h(n+d) P_{\lambda_a}^a(n) = e^{-(\lambda_h+\lambda_a)} \left(\frac{\lambda_h}{\lambda_a} \right)^{d/2} I_d(2\sqrt{\lambda_h \lambda_a}) \quad (\text{A.3})$$

donde I_d es la función modificada de Bessel.

Apéndice B

Desarrollo Numérico B

En éste apéndice se desarrollará la distribución binomial negativa (DBN) utilizada en éste trabajo de investigación, la cual es generada partiendo de la distribución gamma y la distribución de Poisson, realizando la integral del modelo de Poisson de cero a infinito en funcion del promedio λ , de donde se obtiene la distribución binomial negativa. Las variables que se ajustaran en la distribución son las siguientes: goles de local n_h , goles de visita n_a , parámetro de casos favorables de local r_h , parámetro de casos favorables de visita r_a , probabilidad de local p_h y probabilidad de visita p_a . A partir de esas variables, de igual forma se calcularán las sumas y diferencias de goles ajustados al modelo binomial negativo, donde para las sumas será $P_{r_h, p_h, r_a, p_a}^{\Sigma}(s)$ y para las diferencias de goles será $P_{r_h, p_h, r_a, p_a}^{\Delta}(d)$.

B.1. Distribución Gamma

En el caso especial donde las probabilidades de marcar λ toman la forma de la distribución gamma, se tiene

$$F(x) = \begin{cases} \frac{a^r}{\Gamma(r)} \lambda^{r-1} e^{-a\lambda}, & \lambda > 0 \\ 0, & \lambda \leq 0 \end{cases} \quad (\text{B.1})$$

B.2. Distribución Binomial Negativa (DBN)

La componente resultante de la distribución de Poisson toma la forma de la distribución binomial negativa

$$P_{r,p}(n) = \int_0^\infty d\lambda P_\lambda(n) f(\lambda)$$

Sabiendo que $\lambda = \lambda_h + \lambda_a$, $n = n_h + n_a$ y $P = \frac{1}{1+a}$.

$$\begin{aligned} P_{r,p}(n) &= \int_0^\infty d\lambda P_\lambda(n) f(\lambda) = \int_0^\infty d\lambda \frac{\lambda^n}{n!} e^{-\lambda} \frac{a^r}{\Gamma(r)} \lambda^{r-1} e^{-a\lambda} \\ &= \frac{a^r}{n! \Gamma(r)} \int_0^\infty d\lambda \lambda^{(n+r-1)} e^{-(a+1)\lambda} \\ &= \frac{a^r}{n! \Gamma(r)} \frac{\Gamma(n+r)}{(a+1)^{n+r}} \\ &= \frac{\Gamma(n+r)}{n! \Gamma(r)} \frac{a^r}{(a+1)^r} \frac{1}{(a+1)^n} \end{aligned}$$

si $P = 1/(1+a)$ entonces

$$= \frac{\Gamma(n+r)}{n! \Gamma(r)} p^n (1-p)^r$$

Finalmente

$$P_{r,p}(n) = \int_0^\infty d\lambda P_\lambda(n) f(\lambda) = \frac{\Gamma(n+r)}{n! \Gamma(r)} p^n (1-p)^r \quad (\text{B.2})$$

donde r es un parámetro de ajuste empírico, el cual corresponde a la relación inicial de la probabilidad de marcar goles.

Análogo al caso de la distribución de Poisson, desarrollado en el apéndice A, se evalúan las densidades de probabilidad para las sumas s y las diferencias d de los goles marcados por los equipos local y visitante.

Para las sumas,

$$P_{r_h, p_h}(n) = \frac{\Gamma(r_h+n)}{n!\Gamma(r_h)} p_h^n (1-p_h)^{r_h} \quad \text{y} \quad P_{r_a, p_a}(s-n) = \frac{\Gamma(r_a+s-n)}{(s-n)!\Gamma(r_a)} p_a^{s-n} (1-p_a)^{r_a}$$

$$P_{r_h, p_h, r_a, p_a}^\Sigma(s) = \sum_{n=0}^s \frac{\Gamma(r_h+n)}{n!\Gamma(r_h)} p_h^n (1-p_h)^{r_h} \frac{\Gamma(r_a+s-n)}{(s-n)!\Gamma(r_a)} p_a^{s-n} (1-p_a)^{r_a}$$

$$P_{r_h, p_h, r_a, p_a}^\Sigma(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} \sum_{n=0}^s \frac{\Gamma(r_h+n)}{n!\Gamma(r_h)} p_h^n \frac{\Gamma(r_a+s-n)}{(s-n)!\Gamma(r_a)} p_a^{s-n} p_a^{r_a}$$

$$P_{r_h, p_h, r_a, p_a}^\Sigma(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} p_a^{r_a} \sum_{n=0}^s \frac{\Gamma(r_h+n)}{n!\Gamma(r_h)} \frac{\Gamma(r_a+s-n)}{(s-n)!\Gamma(r_a)} \left(\frac{p_h}{p_a}\right)^n$$

$$P_{r_h, p_h, r_a, p_a}^\Sigma(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} p_a^{r_a} \frac{\Gamma(r_a+s)}{s!\Gamma(r_a)} \sum_{n=0}^s \frac{\Gamma(r_h+n)}{n!\Gamma(r_h)} \frac{\Gamma(r_a+s-n)}{\Gamma(r_a+s)} \frac{s!}{(s-n)!} \left(\frac{p_h}{p_a}\right)^n$$

Usando la propiedad $\Gamma(u)\Gamma(1+u) = \frac{\pi}{\text{sen}\pi u}$

$$\Gamma(n-s) = \frac{\pi}{\text{sen}\pi(s-n)\Gamma(1-n+s)}$$

Por la propiedad del seno $\text{sen}\pi(n-s) = \text{sen}\pi n \cos\pi s - \cos\pi n \text{sen}\pi s$

$$\Gamma(n-s) = \frac{-(-1)^n}{\text{sen}\pi s \Gamma(1-n+s)}$$

De igual forma

$$\Gamma(-s) = \frac{-\pi}{\text{sen}\pi s \Gamma(1+s)}$$

Entonces

$$\frac{\Gamma(n-s)}{\Gamma(-s)} = \frac{(-1)^n \Gamma(1+s)}{\Gamma(1+s-n)} = \frac{(-1)^n s!}{(s-n)!}$$

Reescribiendo

$$\frac{\Gamma(n-s)}{\Gamma(-s)(-1)^n} = \frac{s!}{(s-n)!}$$

Por otro lado y usando las mismas propiedades

$$\Gamma(r_a+s-n) = \frac{\pi}{\text{sen}\pi(r_a+s-n)\Gamma(1-r_a-s+n)} \quad \text{y} \quad \Gamma(r_a+s) = \frac{\pi}{\text{sen}\pi(r_a+s)\Gamma(1-r_a-s)}$$

Entonces

$$\frac{\Gamma(r_a+s-n)}{\Gamma(r_a+s)} = \frac{(-1)^n \Gamma(1-r_a-s)}{\Gamma(1-r_a-s+n)}$$

Reescribiendo en la ecuación principal $P_{r_h, p_h, r_a, p_a}^{\Sigma}(s)$

$$P_{r_h, p_h, r_a, p_a}^{\Sigma}(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} p_a^s \frac{\Gamma(r_a+s)}{s! \Gamma(r_a)} \sum_{n=0}^s \frac{\Gamma(r_h+n)}{n! \Gamma(r_h)} \frac{(-1)^n \Gamma(1-r_a-s)}{\Gamma(1-r_a-s+n)} \frac{\Gamma(n-s)}{\Gamma(-s)(-1)^n} \left(\frac{p_h}{p_a}\right)^n$$

$$P_{r_h, p_h, r_a, p_a}^{\Sigma}(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} p_a^s \frac{\Gamma(r_a+s)}{s! \Gamma(r_a)} \left(\frac{\Gamma(1-s-r_a)}{s! \Gamma(-s)} \sum_{n=0}^s \frac{\Gamma(-s+n)}{\Gamma(1-s-r_a+n)} \frac{\Gamma(r_h+n)}{n!} \left(\frac{p_h}{p_a}\right)^n \right)$$

Donde ${}_2F_1\left(-s, r_h; 1-s-r_a; \frac{p_h}{p_a}\right) = \left(\frac{\Gamma(1-s-r_a)}{s! \Gamma(-s)} \sum_{n=0}^s \frac{\Gamma(-s+n)}{\Gamma(1-s-r_a+n)} \frac{\Gamma(r_h+n)}{n!} \left(\frac{p_h}{p_a}\right)^n \right)$ es la función hipergeométrica.

finalmente para las sumas,

$$P_{r_h, p_h, r_a, p_a}^{\Sigma}(s) = (1-p_h)^{r_h} (1-p_a)^{r_a} p_a^s \frac{\Gamma(r_a+s)}{s! \Gamma(r_a)} {}_2F_1\left(-s, r_h; 1-s-r_a; \frac{p_h}{p_a}\right) \quad (\text{B.3})$$

Para las diferencias,

$$P_{r_h, p_h}(n+d) = \frac{\Gamma(r_h+n+d)}{(n+d)! \Gamma(r_h)} p_h^{n+d} (1-p_h)^{r_h} \quad \text{y} \quad P_{r_a, p_a}(n) = \frac{\Gamma(r_a+n)}{n! \Gamma(r_a)} p_a^n (1-p_a)^{r_a}$$

$$P_{r_h, p_h, r_a, p_a}^{\Delta}(d) = \sum_{n=0}^{\infty} \frac{\Gamma(r_h+n+d)}{(n+d)! \Gamma(r_h)} p_h^{n+d} (1-p_h)^{r_h} \frac{\Gamma(r_a+n)}{n! \Gamma(r_a)} p_a^n (1-p_a)^{r_a}$$

$$= (1-p_h)^{r_h} (1-p_a)^{r_a} \sum_{n=0}^{\infty} \frac{p_h^n p_h^d \Gamma(r_h+n+d)}{(n+d)! \Gamma(r_h)} \frac{\Gamma(r_a+n) p_a^n}{n! \Gamma(r_a)}$$

$$= (1-p_h)^{r_h} (1-p_a)^{r_a} p_h^d \sum_{n=0}^{\infty} \frac{\Gamma(r_h+n+d)}{(n+d)! \Gamma(r_h)} \frac{\Gamma(r_a+n)}{n! \Gamma(r_a)} (p_h p_a)^n$$

Usando, $(n+d)! = \Gamma(n+d+1)$, $n! = \Gamma(n+1)$ y $\Gamma(d+1) = d!$

$$= (1-p_h)^{r_h} (1-p_a)^{r_a} p_h^d \frac{\Gamma(r_h+d)}{\Gamma(r_h+d)} \frac{1}{\Gamma(r_h)} \frac{\Gamma(d+1)}{\Gamma(d+1)} \frac{1}{\Gamma(r_a)} \left(\sum_{n=0}^{\infty} \frac{\Gamma(r_h+n+d) \Gamma(r_a+n) (p_h p_a)^n}{\Gamma(n+d+1) n!} \right)$$

$$= (1-p_h)^{r_h} (1-p_a)^{r_a} p_h^d \frac{\Gamma(r_h+d)}{d! \Gamma(r_h)} \left(\frac{\Gamma(1+d)}{\Gamma(r_h+d) \Gamma(r_a)} \sum_{n=0}^{\infty} \frac{\Gamma(r_h+n+d) \Gamma(r_a+n) (p_h p_a)^n}{\Gamma(n+d+1) n!} \right)$$

donde ${}_2F_1(r_h+d, r_a; 1+d; p_h p_a) = \left(\frac{\Gamma(1+d)}{\Gamma(r_h+d) \Gamma(r_a)} \sum_{n=0}^{\infty} \frac{\Gamma(r_h+n+d) \Gamma(r_a+n) (p_h p_a)^n}{\Gamma(n+d+1) n!} \right)$ es la función hipergeométrica.

finalmente para las diferencias,

$$P_{r_h, p_h, r_a, p_a}^\Delta(d) = (1 - p_h)^{r_h} (1 - p_a)^{r_a} \frac{\Gamma(r_h + d)}{d! \Gamma(r_h)} {}_2F_1(r_h + d, r_a; 1 + d; p_h p_a) \quad (\text{B.4})$$

Apéndice C

Frecuencias de goles

El apéndice esta formado por las gráficas de frecuencias de goles de local y de visita de las ligas europeas estudiadas en este trabajo, las mismas no fueron usadas en los resultados ya que contienen en ellas el mismo comportamiento de la liga premier inglesa reportada con anterioridad. Se presentan a continuación las gráficas de frecuencias de goles para las ligas: italiana (Serie A), española (La Liga) y alemana (Bundesliga).

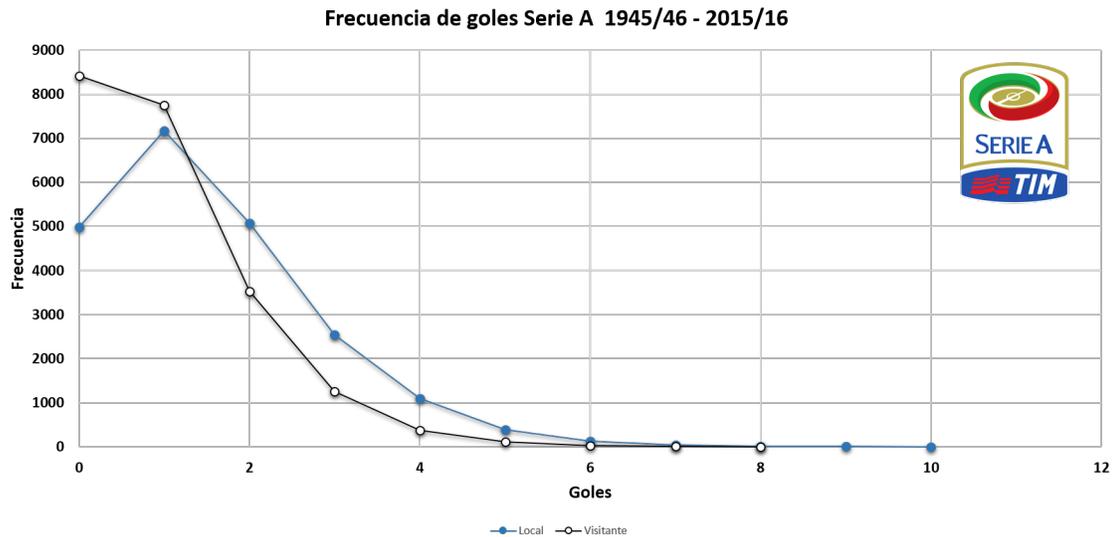


Figura C.1: Frecuencia de goles de local y visitante de la Serie A italiana.

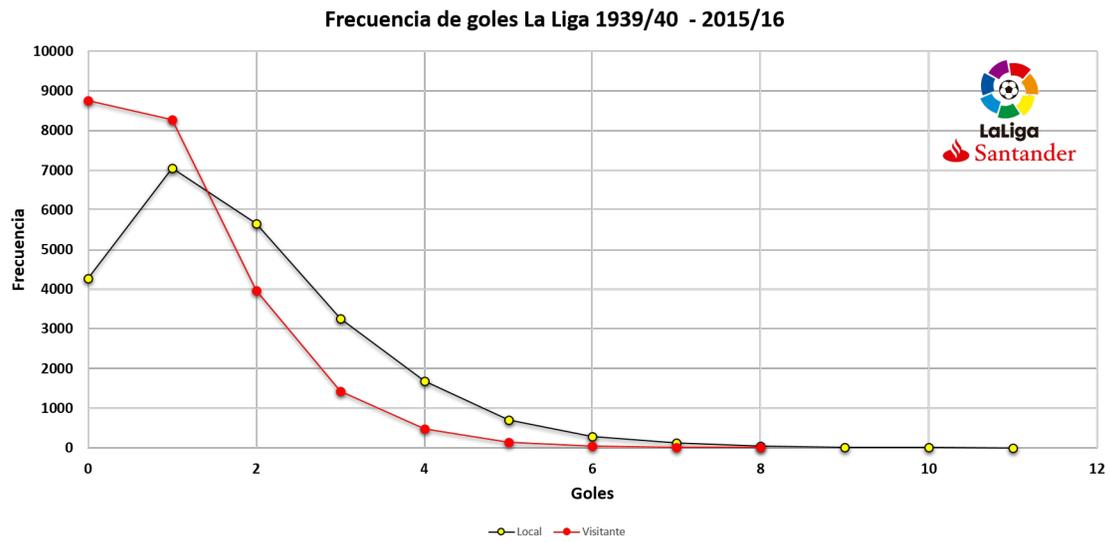


Figura C.2: Frecuencia de goles de local y visitante de la liga española.

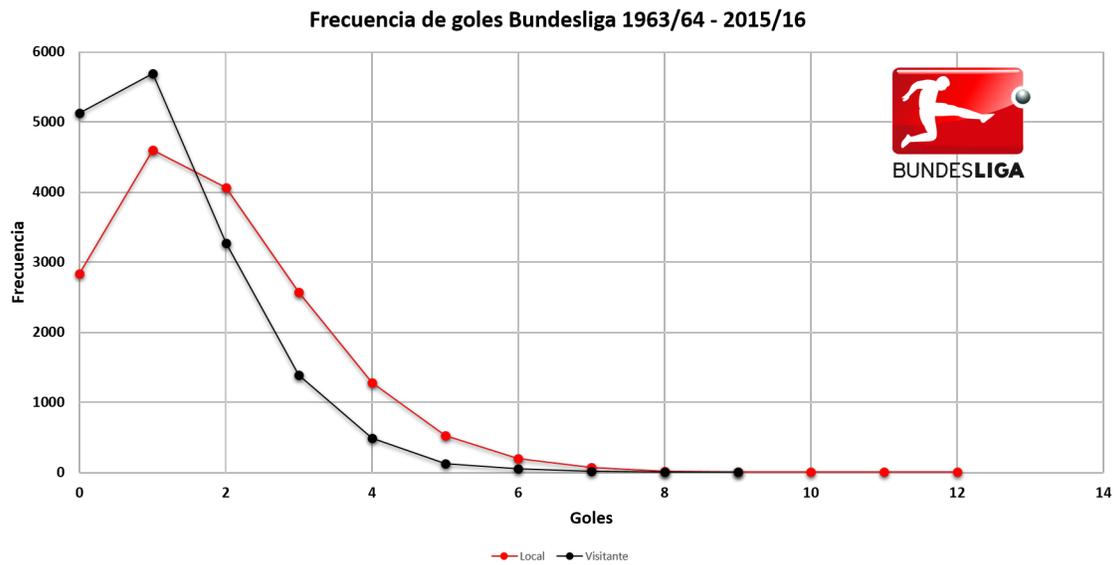


Figura C.3: Frecuencia de goles de local y visitante de la Bundesliga alemana.

Apéndice D

Sumas y diferencias de goles

A continuación se muestran las gráficas de frecuencias de sumas y diferencias de goles de las ligas europeas estudiadas en este trabajo de investigación, recordando que las sumas son $s = n_h + n_a$ y las diferencias son $d = n_h - n_a$, sin embargo las mismas no fueron usadas en los resultados ya que contienen en ellas el mismo comportamiento de la liga premier inglesa reportada con anterioridad. Se presentan a continuación las gráficas de frecuencias de sumas y diferencias de goles para las ligas: italiana (Serie A), española (La Liga) y alemana (Bundesliga).

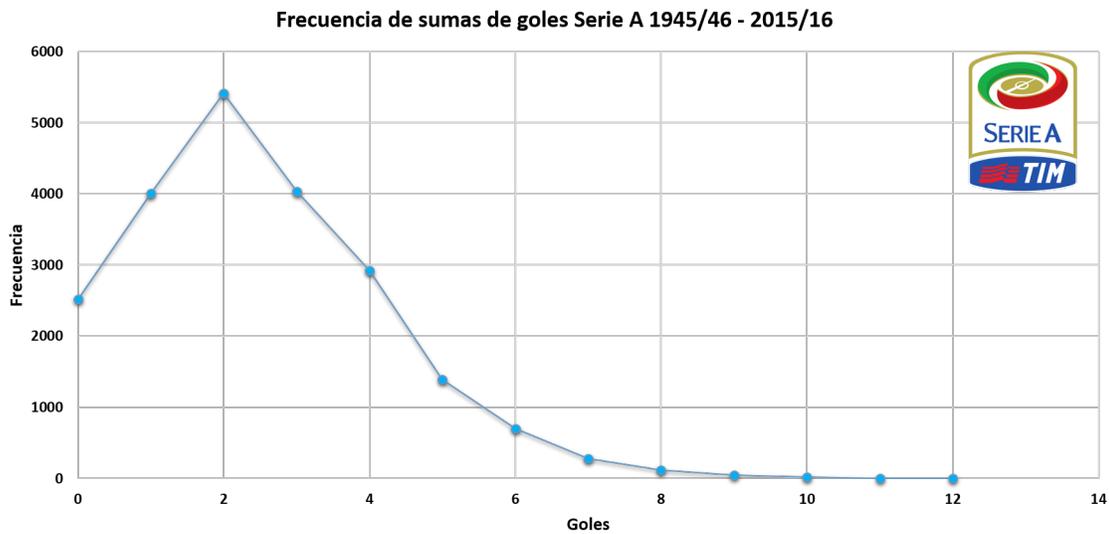


Figura D.1: Frecuencia de la suma de goles de la Serie A italiana.

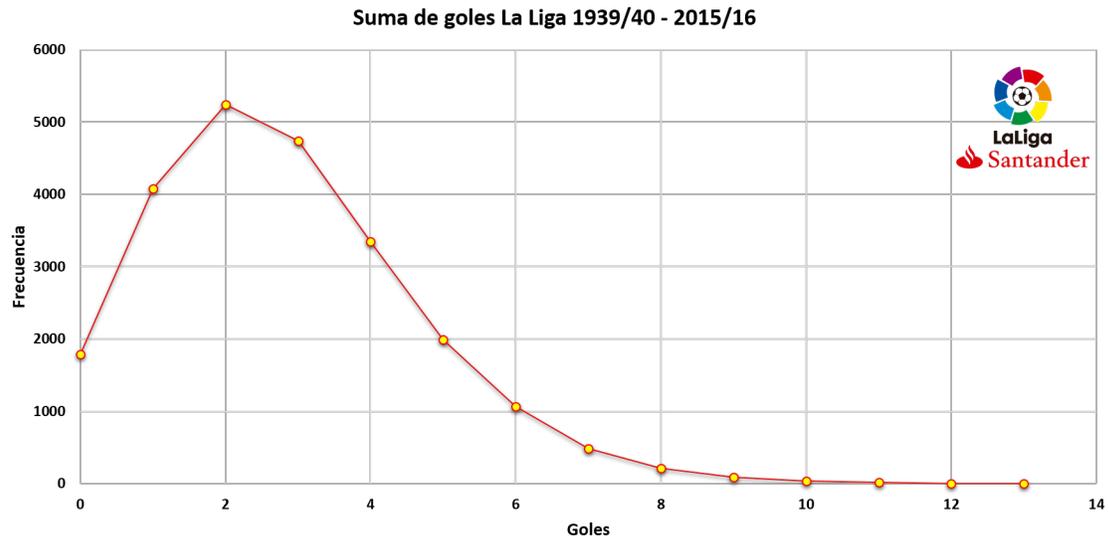


Figura D.2: Frecuencia de la suma de goles de la liga española.

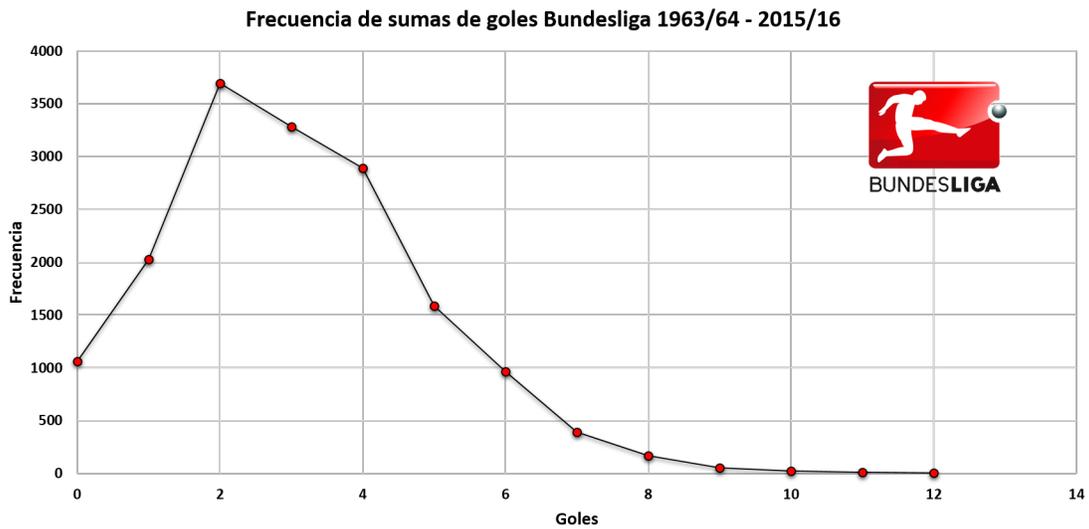


Figura D.3: Frecuencia de la suma de goles de la Bundesliga alemana.

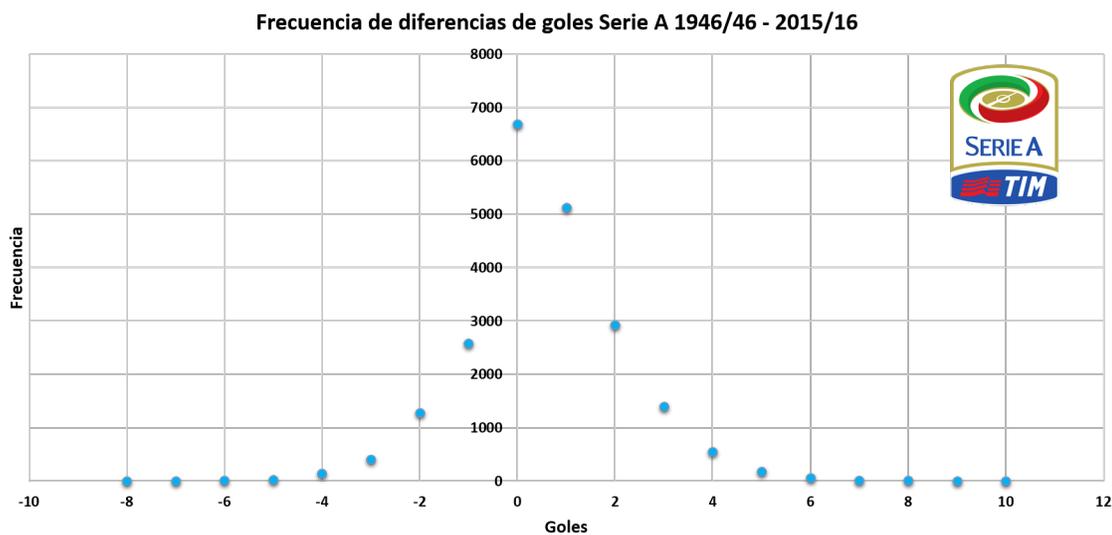


Figura D.4: Frecuencia de las diferencias de goles de la Serie A italiana.

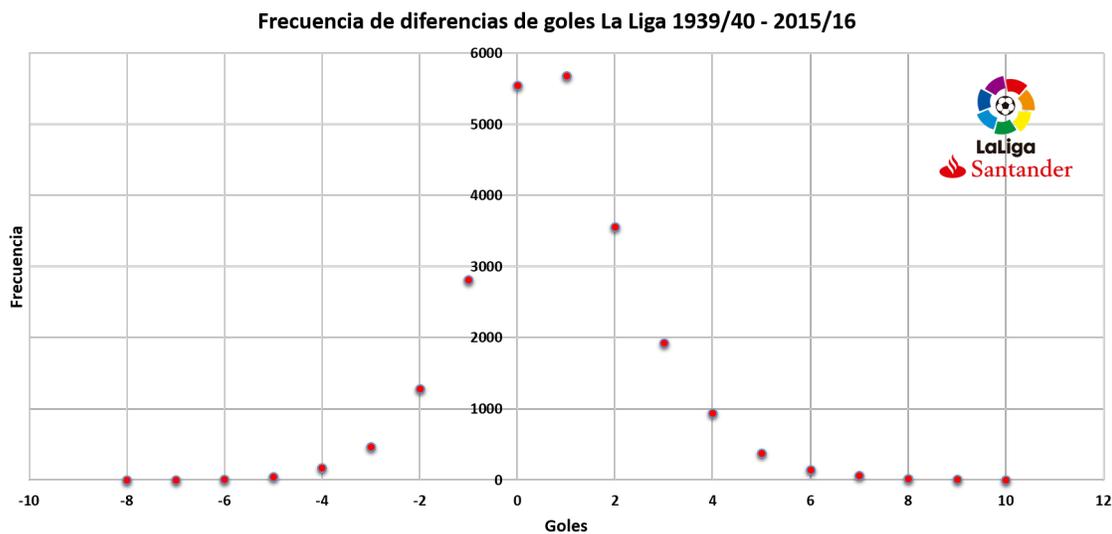


Figura D.5: Frecuencia de las diferencias de goles de la liga española.

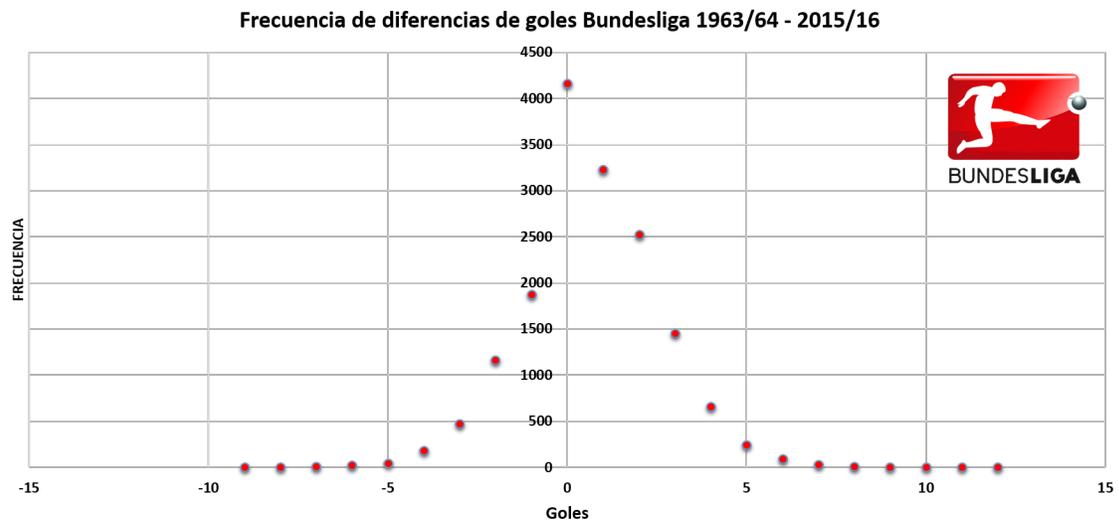


Figura D.6: Frecuencia de las diferencias de goles de la Bundesliga alemana.

Apéndice E

Modelos de Probabilidad

En éste apéndice se observan las gráficas realizadas para el ajuste a la distribución binomial negativa (DBN) para los goles de local, de visita, sumas y diferencias. Se presentan las gráficas de densidad de probabilidades para las ligas: italiana (Serie A), española (La Liga) y alemana (Bundesliga).

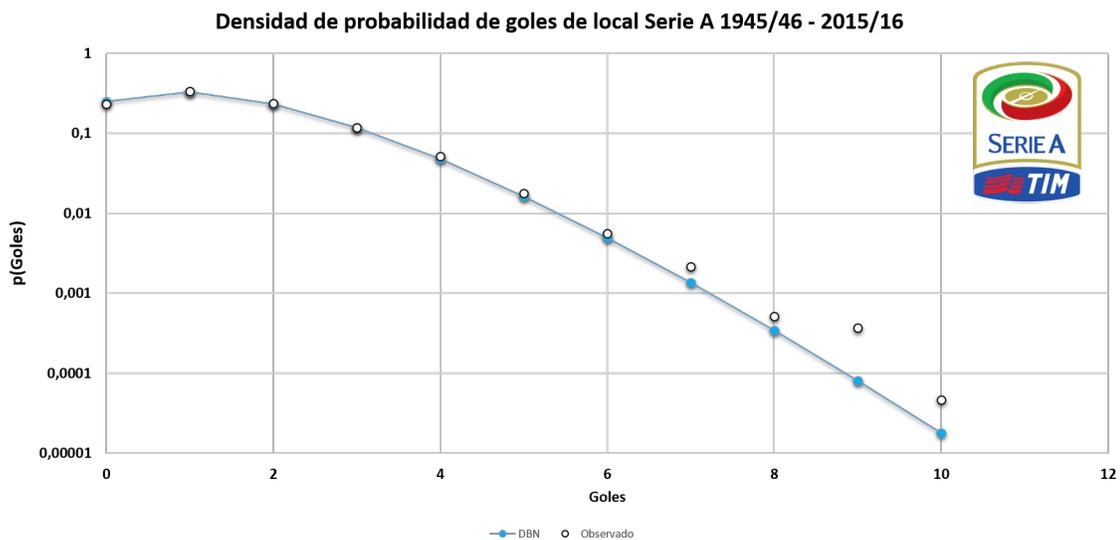


Figura E.1: Ajuste a la distribución binomial negativa para los goles de local de la Serie A italiana.

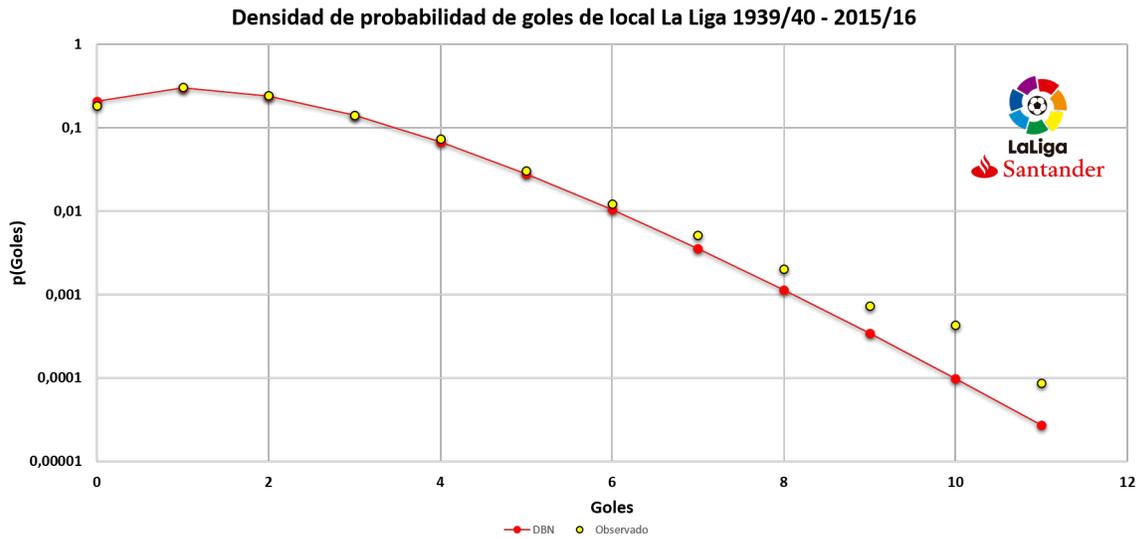


Figura E.2: Ajuste a la distribución binomial negativa para los goles de local de la liga española.

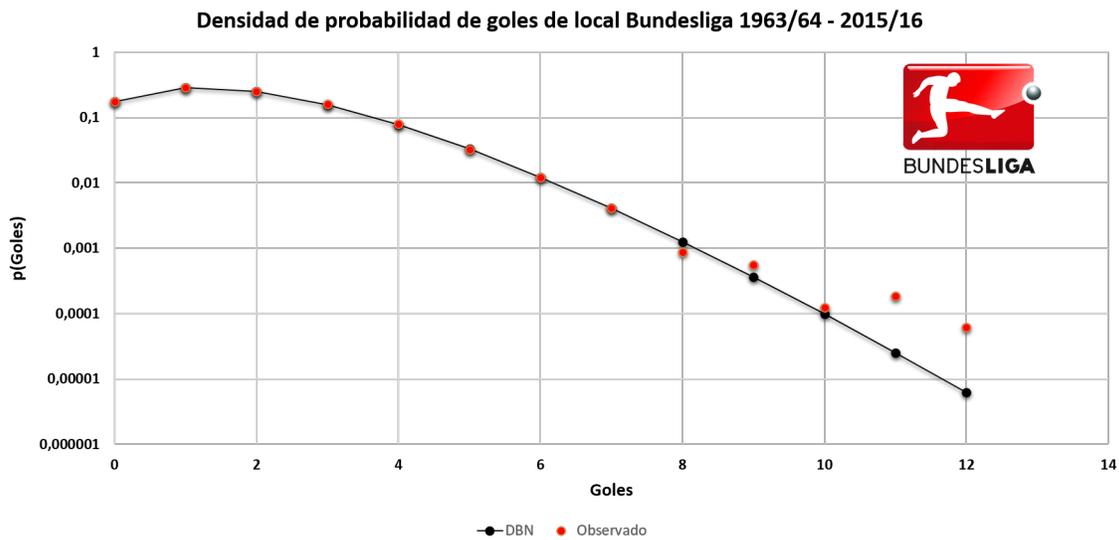


Figura E.3: Ajuste a la distribución binomial negativa para los goles de local de la Bundesliga alemana.

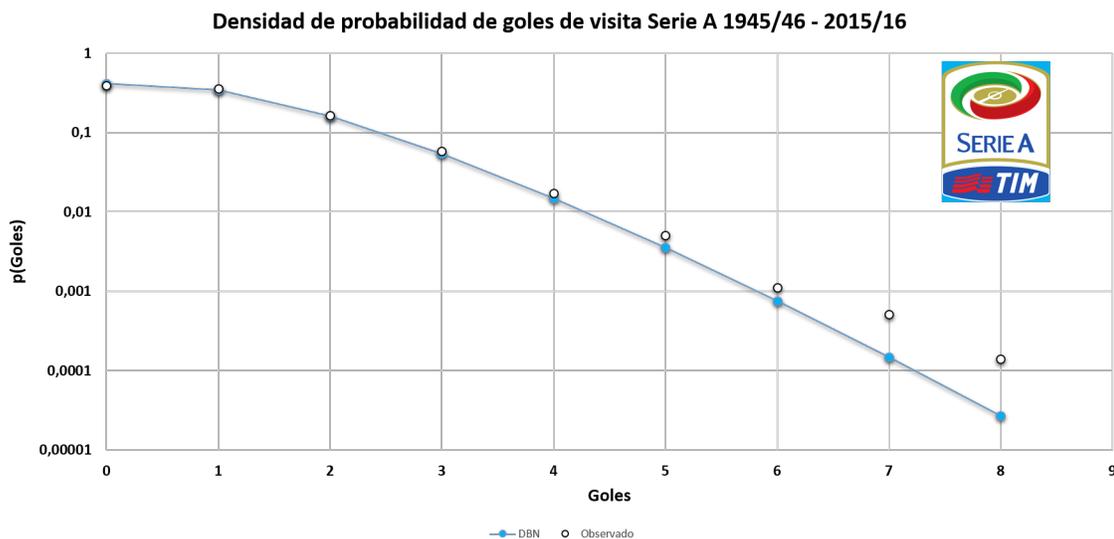


Figura E.4: Ajuste a la distribución binomial negativa para los goles de visita de la Serie A italiana.

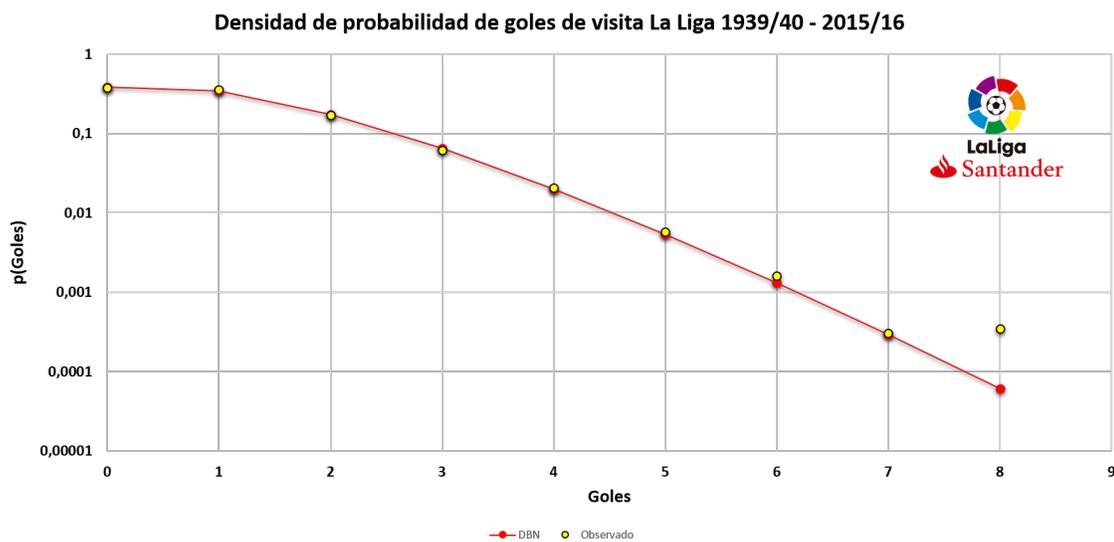


Figura E.5: Ajuste a la distribución binomial negativa para los goles de visita de la liga española.

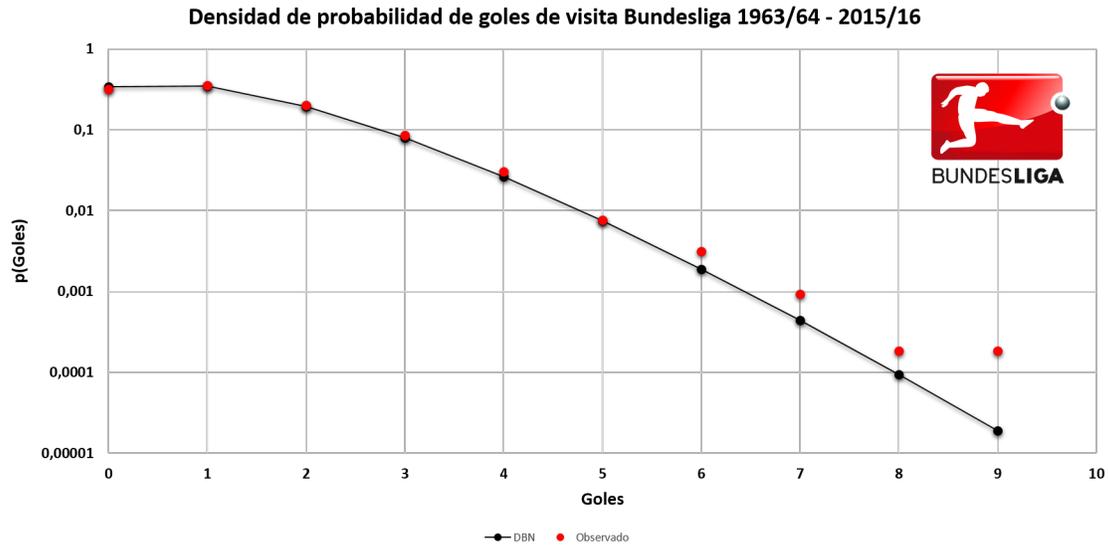


Figura E.6: Ajuste a la distribución binomial negativa para los goles de visita de la Bundesliga alemana.

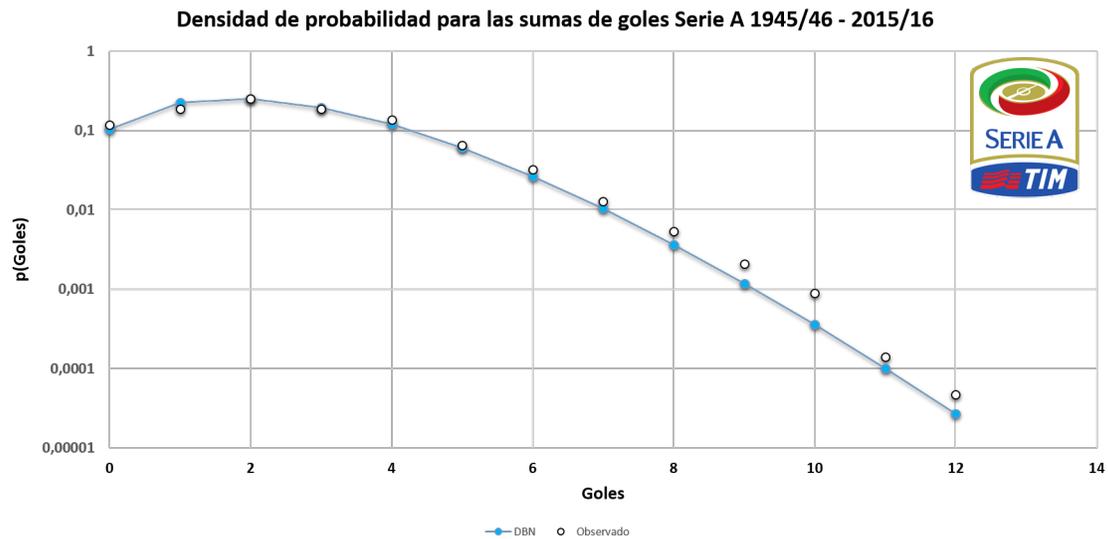


Figura E.7: Ajuste a la distribución binomial negativa para la suma de goles de la Serie A italiana.

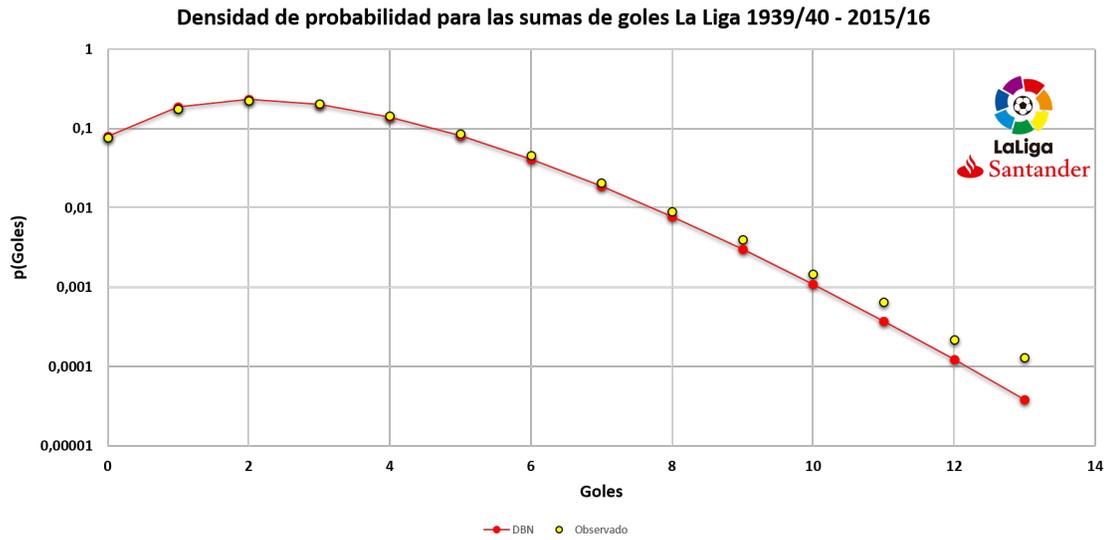


Figura E.8: Ajuste a la distribución binomial negativa para la suma de goles de la liga española.

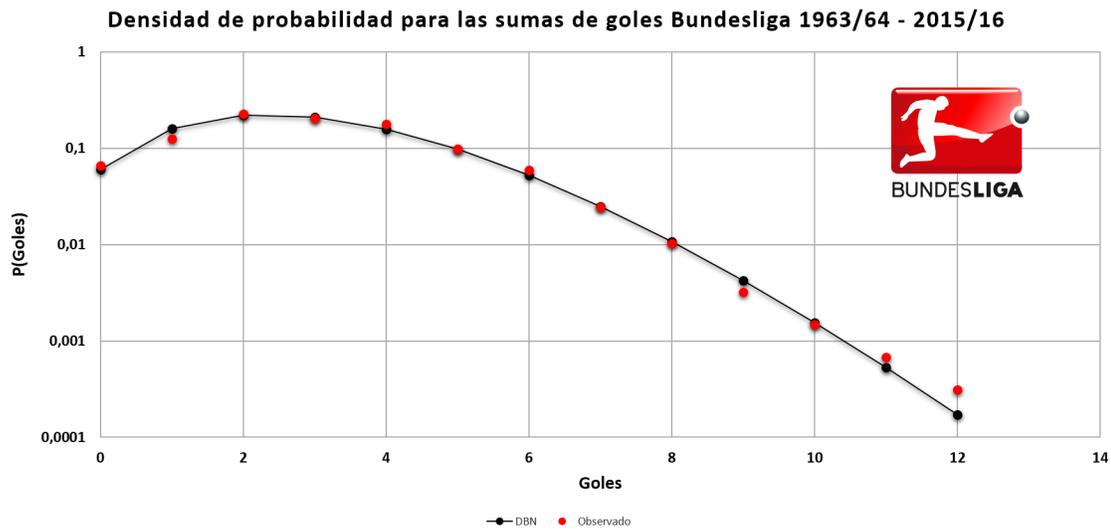


Figura E.9: Ajuste a la distribución binomial negativa para la suma de goles de la Bundesliga alemana.

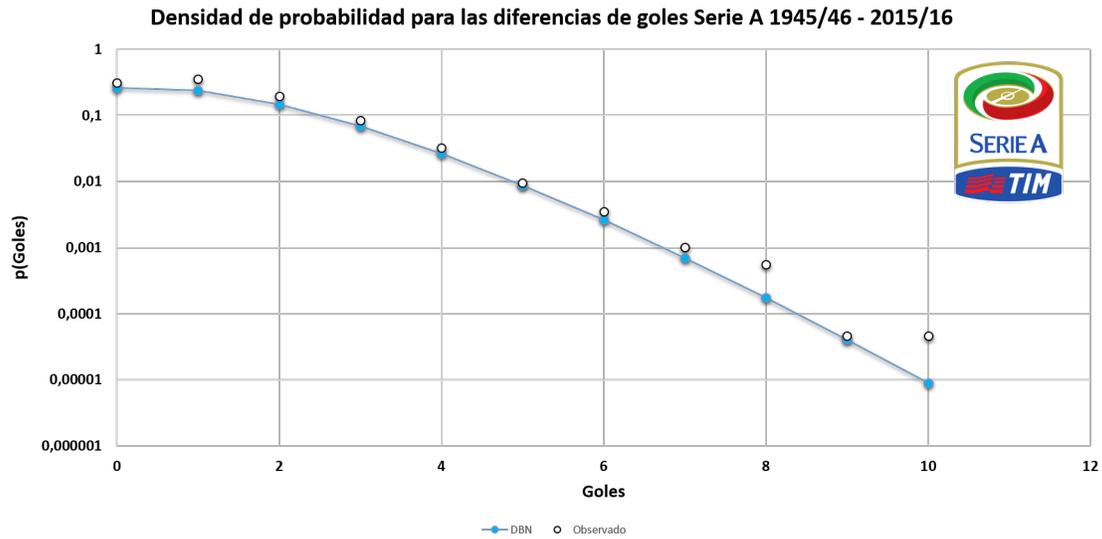


Figura E.10: Ajuste a la distribución binomial negativa para la diferencia de goles de la Serie A italiana.

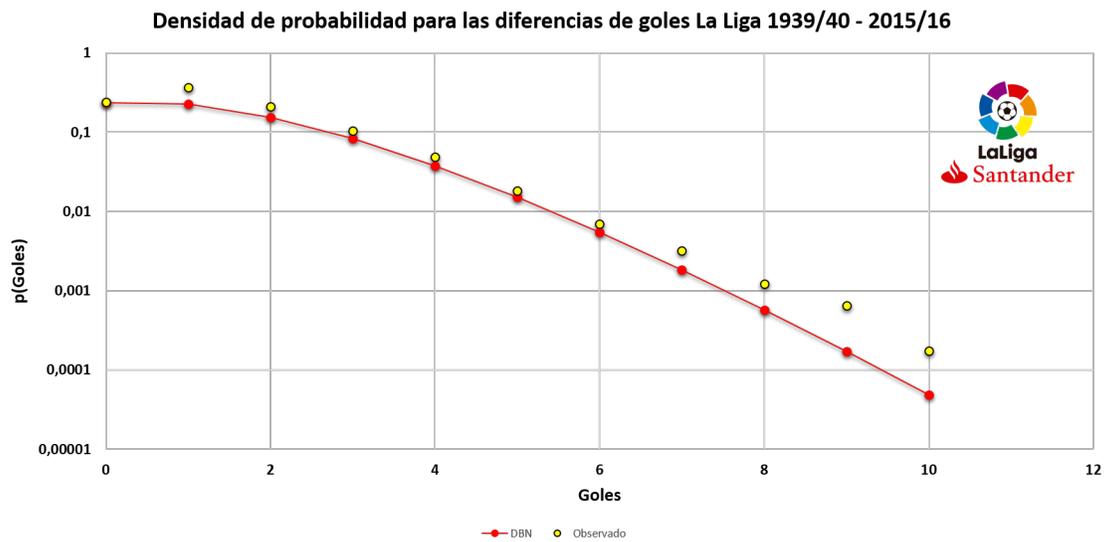


Figura E.11: Ajuste a la distribución binomial negativa para la diferencia de goles de la liga española.

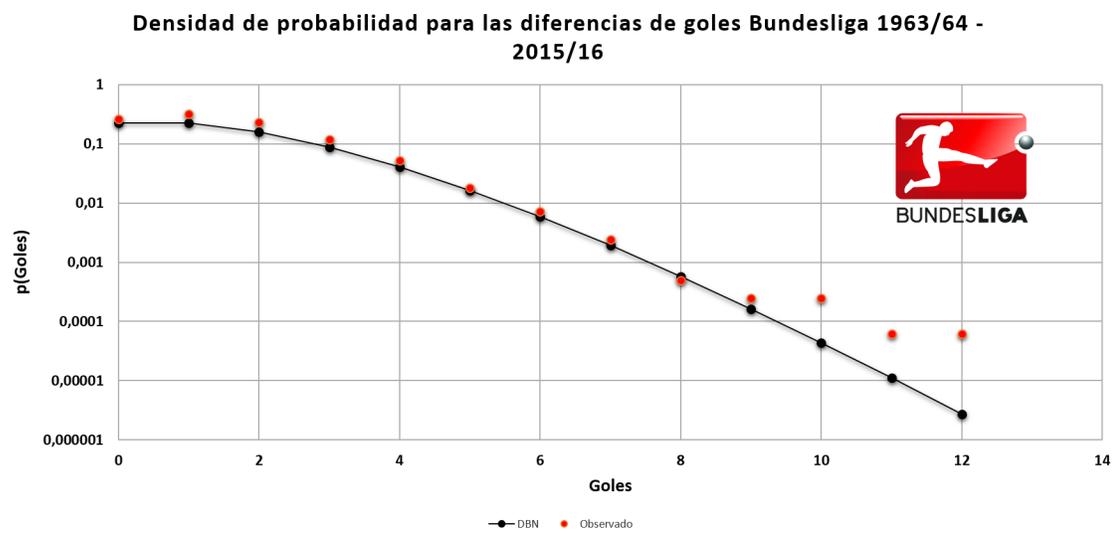


Figura E.12: Ajuste a la distribución binomial negativa para la diferencia de goles de la Bundesliga alemana.

Bibliografía

- [1] Reep C., Pollard R., Benjamin B., J. Roy. Skill and Chance in Ball Games, Journal of the Royal Statistical Society. A, Vol. 134, No 4, 1971, 134, 653.
 - [2] W. Janke, E. Bittner, A. Nußbaumer, M. Weigel. Football fever: self-affirmation model for goal distributions. Condensed Matter Physics Vol. 12, No 4, 2009, 739-752.
 - [3] <http://dle.rae.es/?id=IeptqGe> [Último acceso 11 de septiembre de 2017].
 - [4] Galeano E. Fútbol a Sol y Sombre y Otros Escritos. Editado por Ediciones La Cueva, 1995.
 - [5] Giulianotti, R., The globalization of football: a study in the glocalization of the serious life. British Journal of Sociology 55(4), 2012, 545–568.
 - [6] <http://es.fifa.com/worldcup/news/y=2015/m=12/news=mas-de-tres-mil-millones-de-telespectadores-vieron-el-mundial-de-2014-2745549.html> [Último acceso 11 de septiembre de 2017].
 - [7] Oliven, R., Damo, A., Fútbol y cultura. Editorial Norma. Enciclopedia Latinoamericana de Sociocultura y Comunicación, No. 10, 2001, 119.
 - [8] Gómez, H., Reseña de "Fútbol y cultura" de Rubén Oliven y Arlei Damo. Estudios sobre las Culturas Contemporáneas, vol. X, núm. 19, 2004, 145-149.
 - [9] Oropeza, J., El Fútbol Mexicano como Instrumento de poder Económico y Político. III Coloquio Internacional de Geocrítica 2014.
 - [10] Perez, I., El Fútbol y su situación económica en España. Análisis economico-financiero del C.D. Numancia de Soria SAD (Tesis de pregrado) 2014, Universidad de Valladolid, España.
 - [11] <http://www.superdeporte.es/futbol/2017/01/19/ranking-clubes-ricos-mundo/323132.html> [Último acceso 11 de septiembre de 2017]
-

- [12] Sumpter, D., Fútbol y Matemáticas. Editorial Ariel 1ra Edición, 2016, ISBN 978-84-344-2384-8.
- [13] Clarke, S., Home Ground Advantage of Individual Clubs in English Soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 44, No. 4 1995, 509-521.
- [14] Dyte, D., A ratings based Poisson model for World Cup soccer simulation. *Journal of the Operational Research Society* 51, 2000, 993-998.
- [15] Greenhough, J., Birch, P., Chapman, S., Rowlands, G., Football goal distributions and extremal statistics. *Physica A* 316, 2002, 615-624.
- [16] Mendenhall, W., Beaver, R., Beaver, B., *Introducción a la probabilidad y estadística*. 13ra. Edición, Cengage Learning, 2009.
- [17] Bluman, A., *Probability Demystified*. Mcgraw Hill, 2005, DOI: 10.1036/0071469990.
- [18] Ramachandran, K., Tsokos, C., *Mathematical Statistics with Applications in R*. Elsevier, Second edition, 2015, ISBN: 978-0-12-417113-8.
- [19] Devore, J., *Probabilidad y Estadística para Ingeniería y Ciencias*. Séptima edición, Cengage Learning, 2008 ISBN-13: 978-607-481-338-8.
- [20] García, J., Cardiel, N., Zamorano, J., *Estadística básica para estudiantes de Ciencias*. Universidad Complutense de Madrid, 2011, ISBN: 978-84-691-8981-8.
- [21] Hernández, J., *VARIABLES ALEATORIAS DISCRETAS*. Universidad Central de Venezuela, (Sin fecha).
- [22] <https://www.fussballdaten.de/bundesliga/1964/kreuztabelle/> [Último acceso 11 de septiembre de 2017].
- [23] <http://rssf.com/tables/venezchamp.html> [Último acceso 11 de septiembre de 2017].
-