Unbetting Football, Capítulo I  
La Falsa Irregularidad del Futbol Mexicano (v2.0)

*En el 2013 publiqué un artículo en el que demostré que la percepción de La Liga MX como una liga irregular, era un error. A pesar de que creo que ese texto fue un trabajo aceptable, siempre sentí que pude haberlo hecho mejor. Ocho años, una cuarentena y muchos tutoriales de YouTube después, este es mi segundo intento.*

Cuando era niño tuve la suerte de que en una navidad me regalaran una PlayStation y uno de los primeros juegos que tuve fue el famosísimo *Winning Eleven*, mejor conocido como “El Winin”. Aunque antes del Winin ya había tenido unos cuantos videojuegos de futbol, desde hacía unos años ya venía intentando explotar algo que me despertaba mucho interés: hacer mis propios torneos. Sin embargo, no fue sino hasta la llegada del famoso Winin que, por fin, me pareció que lo que ocurría en los partidos era realista (o lo que entendía por ello). Empecé primero con los mundiales incluyendo hasta sorteos de grupos con cabezas de serie y pools. Después configuraba cada partido en modo “CPU” a ambos equipos (es decir que eran controlados por “la máquina”), y así observaba cada uno de los 64 partidos mientras anotaba los resultados, estadísticas y alguna que otra cosa que me llamaba la atención. Recuerdo que me sorprendía mucho ver cómo los equipos hacían cambios de jugadores o hasta de formación, según lo que ocurría en el partido, sin ninguna intervención de mi parte.

Luego pensé que podría hacer lo mismo con la entonces llamada “Primera División del Futbol Mexicano” y me puse manos a la obra. Tomé a las selecciones nacionales y las adapté, cambiando desde el uniforme hasta cada uno de los atributos de los jugadores, incluida su apariencia, para que se parecieran a lo que yo entendía como lo que era el futbol mexicano. Recuerdo que la idea se me vino a la mente cuando vi que el entonces equipo de los Tecos tenía prácticamente el mismo uniforme que la Croacia de Suker, Boban y Jarni. Pero había un problema.

Una de las razones por las que decidí simular mi propia liga era porque en mis mundiales, casi sin excepción, prácticamente siempre había algo que a mi entender echaba a perder mis queridos torneos. Sucedía que un equipo favorito perdía en la fase de grupos, que un equipo malo calificaba a segunda ronda, que los dos favoritos se enfrentaban en octavos de final o que, de repente, en la final aparecía uno o hasta dos equipos que nunca en la historia habían llegado a esas instancias. Recuerdo que aquello me molestaba y mucho, y cuando empezaron a suceder el mismo tipo de sucesos raros en mis simulaciones del futbol mexicano, creo que aquello me superó ¿por qué pasaba eso? ¿estaba yo haciendo algo mal? Me dejaba una sensación difícil de explicar el no entender qué estaba pasando y confieso que lo abandoné, hasta que algunos años después ocurrió algo que me ayudó a empezar el camino para entender qué pasaba. De hecho, no fue una cosa, sino dos: leí al matemático Nassim Nicholas Taleb, autor del libro El Cisne Negro, y el Leicester City, mi equipo de toda la vida (ay sí), ganó la Premier League. Veinte años después, mi intención en estas entregas es responderle a mi yo del pasado por qué ni sus mundiales, ni sus ligas, ni nada de lo que simulaba, tenían algo de malo. Para ello, primero, necesitamos entender de una manera muy sencilla cómo funcionan las apuestas y lo que implican.

# Unbetting Football

Lo que haremos a continuación es aplicar Ingeniería Inversa al sistema de apuestas. Para elaborar el presente estudio y varios de los siguientes que publicaré pedí permiso a Joseph Buchdahl de Football-data.co.uk para emplear sus datos sobre resultados deportivos y apuestas vinculadas a ellos. Se trata de una base de datos de alrededor de 150,000 partidos de 28 ligas europeas, a lo que añadí resultados y datos de la Liga MX. Todos los datos presentados en este capítulo involucran solo resultados de partidos de liga o temporada regular.

Un ejemplo clásico de una apuesta en el futbol (en el formato de Europa Continental) para un partido es el que se muestra en la tabla de abajo. Lo expresado en la primera fila de la tabla representa el resultado final del partido y, por su parte, los números de la segunda fila son los Momios o *multiplicadores* que establecen el pago de una apuesta exitosa.

|  |  |  |
| --- | --- | --- |
| Victoria del Local | Empate | Victoria del Visitante |
| 2 | **3.3** | **5** |

Si una persona apuesta $100.00 a que, al final del partido el equipo local ganará, de suceder, su pago será:

En caso de que el resultado del partido fuese distinto a la victoria del local el apostador perdería su dinero. Sin embargo, un momio no es solo un multiplicador, pues también representa una probabilidad implícita. Para expresar un momio en términos de probabilidad se requiere realizar la siguiente operación:

Que, aplicado al ejemplo anterior, quedaría así:

|  |  |  |
| --- | --- | --- |
| Victoria del Local | Empate | Victoria del Visitante |
| 2 | **3.3** | **5** |

|  |  |  |
| --- | --- | --- |
| Victoria del Local | Empate | Victoria del Visitante |
| 0.5 (50%) | **0.3 (30%)** | **0.20 (20%)** |

Nota: el ejemplo previo está simplificado, pues no incluye el margen de las casas de apuestas que funciona, básicamente, como una especie de comisión. Esto será explicado en el siguiente artículo: “La Caja de Pandora” .

Ahora traigamos a la mesa tres partidos que ocurrieron hace poco: el primero, Liverpool vs Chelsea, se jugó el 4 de marzo de 2021 en Inglaterra (EPL); el segundo, Cruz Azul vs Monterrey, fue el 19 de agosto de 2021 en México (Liga MX) y el tercero, Boavista vs Moreirense, el 19 de febrero de 2021 en Portugal (Primeira Liga). Los tres partidos, expresados en probabilidades son:

|  |  |  |
| --- | --- | --- |
| Liverpool | Empate | Chelsea |
| 42% | **26%** | **32%** |

|  |  |  |
| --- | --- | --- |
| Cruz Azul | Empate | Monterrey |
| 42% | **26%** | **32%** |

|  |  |  |
| --- | --- | --- |
| Boavista | Empate | Moreirense |
| 42% | **26%** | **32%** |

Uno de los partidos anteriores terminó con victoria para el local por 1 a 0, otro lo ganó el visitante por 0 a 1 y otro terminó en empate 1 a 1. Por supuesto, esto no es un reto de memoria ni una adivinanza, lo relevante aquí es que, incluso si no conociéramos a alguno de los equipos que jugaron en estos partidos, cualquiera de los tres resultados es perfectamente aplicable a cualquiera de los partidos. Entonces ¿qué diferencia hay entre esos partidos? Uno es de los mejores partidos posibles a nivel mundial, mientras que los otros dos son claramente de perfil más modesto, sin embargo, esta diferencia se trata solo de una diferencia en *escala*, la cual solo es posible apreciarla cuando los equipos compiten en torneos internacionales. Pero, en cuanto a lo que ocurre dentro de cada partido, es decir, los eventos como faltas, disparos, tarjetas, goles o incluso los resultados, son sumamente estables *en el largo plazo*.

Imaginemos que los 3 partidos listados anteriormente ocurriesen al mismo tiempo, ¿cambiaría algo? Suponiendo que los equipos no se viesen afectados por cuestiones de husos horarios y que los resultados de un partido no afectasen a los otros dos (algo cierto, puesto que el resultado de un torneo es independiente de los otros), resulta que la concurrencia en el tiempo no afecta a los partidos; ¿o si los partidos hubiesen ocurrido con 5 años de diferencia entre uno y otro? Siempre y cuando las reglas del juego se mantuviesen más o menos intactas, los partidos seguirían ocurriendo de la misma forma. Considerando lo anterior, resulta que tenemos una poderosa carta a nuestro favor: ¿cuántos partidos hubo, entre 2012 y 2021, en los que los equipos locales tuvieron 42% de probabilidad de ganar? Para ser exacto, 2484.

Ahora ha llegado el momento más importante: tomemos ese conjunto de partidos y hagámosles preguntas. Por ejemplo, ¿cuál fue el promedio de goles en ese *escenario*? Habiendo anotado un total de 6185 goles en esos 2484 partidos, el promedio fue: **2.49 goles por partido**.

Pero ¿es eso poco o mucho? O planteemos la pregunta desde otro ángulo: ¿cómo hacemos para saber si alguien es más alto que otra persona? Lo que hacemos es que tomamos a ambos individuos y los comparamos con lo que hemos definido que es una medida (previamente convenida y conocida por ambas partes), en este caso el *metro*. Si uno de nuestros hipotéticos amigos midiese 2 veces lo que mide un metro y el otro 1.5 veces, habremos resuelto el dilema y sabremos quién es el más alto.

Retomando el promedio de 2.49 goles en el escenario de 42% de probabilidad de victoria del equipo local, ¿cómo podemos saber si esto es mucho o poco? Comparando el promedio de ese escenario con los otros 99. Observemos, por favor, con mucha calma la siguiente gráfica:

Chart, bar chart, histogram

Description automatically generated

La gráfica anterior tiene unas características de lo más interesantes. En primer lugar, vemos que la gráfica describe una curvatura en forma de U, es decir que el promedio de goles por escenario (la altura de la barra) se comprime hacia el intervalo 42% a 38% llegando a lo que parece un valor mínimo. Para descubrir por qué ocurre esto (nuestra hipótesis) deberás esperar un poco, a un capítulo posterior de *Unbetting Football*, y que se llamará “Los tiempos de todas las jugadas, en La Jugada de Todos Los Tiempos”. Pero, por ahora, traigamos en este contexto a nuestra querida Liga MX. Si recordamos uno de los 3 partidos que mencionamos con anterioridad:

|  |  |  |
| --- | --- | --- |
| Cruz Azul | Empate | Monterrey |
| 42% | **26%** | **32%** |

El partido anterior es un ejemplo perfecto de lo que es competir en la Liga MX. Para la siguiente medición se tomaron todos los partidos de la Liga MX entre agosto de 2012 y diciembre de 2021 y se promediaron las probabilidades de victoria de cada equipo. ¿Dónde se ubican los equipos de la Liga MX en el abanico de escenarios presentado en la gráfica anterior? Aquí:

A picture containing timeline

Description automatically generated

¿Es esto poco o mucho? ¿bueno o malo? Veamos la siguiente gráfica para comparar a la Liga MX con las 5 principales ligas del mundo:

Timeline

Description automatically generated

A primera vista pareciera como si los detractores de la Liga MX hubieran encontrado su Tierra Prometida, pero ¿será así? Hagamos más preguntas, pero hagámoslas al conjunto de los partidos. Consideremos los partidos que caen en el siguiente escenario: **Victoria de Local 42%**.

Sabemos que los partidos en los que los locales tienen 42% de probabilidad de victoria el promedio de goles por partido es de 2.49, pero este promedio tiene que, necesariamente, componerse de la suma de los promedios de los equipos locales más el promedio de los visitantes:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Promedio de goles en el escenario 42%** |  | **Promedio de goles del  Local** |  | **Promedio de goles del Visitante** |  |  |
|  |  |

Hablemos un poco de lo que nos dicen los números, de lo que significan. Por supuesto que, según las reglas del juego, es imposible anotar 1.4 o 1.09 goles en un partido, pero ¿y si fueran 100 partidos? ¿cuántos goles *esperaríamos* que anoten ambos equipos si jugaran un total de 100 partidos en las mismas circunstancias?

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Promedio de goles en el escenario** |  | **Promedio de goles del Local** |  | **Número de partidos** |  | **Promedio de goles del Visitante** |  | **Número de partidos** |  | **goles** |
|  |  |  |  |

Lo que estamos diciendo en este caso es que, por cada 100 partidos jugados en esta circunstancia, esperaríamos que se anotasen un total de 249 goles. Pero observemos otro detalle curioso en la operación que acabamos de hacer, ¿qué porcentaje de los goles totales fueron anotados por los equipos locales y los equipos visitantes? Observemos:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Promedio de goles en el escenario** |  | **Promedio de goles del Local** |  | **Número de partidos** |  | **Promedio de goles del Visitante** |  | **Número de partidos** |  | **goles** |
|  |  |  |  |
|  |  | **140  goles del local**  **56%** | | |  | **109 goles del visitante**  **44%** | | |  |  |
|  |  |  |  |

¿Esta diferencia es poca o mucha? ¿cómo podemos saber si es poca o mucha? Así es, tenemos que comparar. Reunamos 3 tipos de partidos a partir de sus probabilidades:

|  |  |  |
| --- | --- | --- |
| **Probabilidad de victoria del Local** | **Probabilidad de victoria del Local** | **Probabilidad de victoria del Local** |
| **80%** | **50%** | **20%** |

A screenshot of a computer

Description automatically generated with medium confidence Ahora observemos la siguiente gráfica que muestra cómo se distribuye la cantidad de goles según cada escenario:

Chart, bar chart

Description automatically generated Por supuesto, hay varias observaciones interesantes que podemos hacer sobre la gráfica anterior. Pero, para entender el peso de los escenarios en los partidos debemos mirar con mayor precisión. Acerquémonos a los partidos y hagamos otra pregunta. Si consideramos solo los partidos en los que se anotaron 4 goles tendríamos 5 posibles resultados. Observemos cómo se distribuyen según el escenario:

A partir de la gráfica anterior podemos hacer algunas observaciones. Primero, es claro que las barras azules están cargadas hacia los resultados que favorecen al local (hacia la izquierda). Por su parte, las barras verdes están cargadas hacia los resultados que favorecen a los visitantes (hacia la derecha). Las barras rojas, en cambio, parecen estar más balanceadas hacia el empate (hacia el centro). ¿Cuál de estos tres tipos de escenario se parecerá más a lo que sucede con mayor frecuencia en la Liga MX? Sin lugar a duda, el que tiene la tendencia hacia el centro. Ahora recordemos la siguiente gráfica:

Timeline

Description automatically generated

Entre más alejado esté un escenario del intervalo entre 46% y 38%, más tenderán a aumentar los goles por partido, sin embargo (y esto es mi opinión), en lo personal prefiero ver partidos que enfrenten a rivales más parejos donde tiendan a resultados más equilibrados o, cuando menos, que no fuesen humillantes. Nuevamente, en mi opinión, golear al grado de humillar a un rival no es una muestra de competitividad: es la prueba de su ausencia.

Si lo comparamos con el Box, ver un partido como Bayern Múnich contra, prácticamente, cualquier otro equipo de la Bundesliga es algo así como el equivalente deportivo de ver a Tyson Fury (2.06 metros, 126 kilos) enfrentarse a El Canelo Álvarez (1.75 metros, 76 kilos) o Gervonta Davies (1.66 metros, 61 kilos).

Entre los torneos Apertura 2012 y Apertura 2021, en la Liga MX participaron 25 equipos que anotaron 7425 goles en fase regular, siendo León y América los equipos que más goles anotaron, 502, lo que representa casi 7% de los goles totales; por su parte, en el mismo periodo y con la misma cantidad de partidos en la Bundesliga se anotaron 8028 goles, siendo Bayern Múnich el equipo que más goles anotó, 698, representando casi 9%. La diferencia entre el porcentaje de León y América comparado con Bayern Múnich parece muy pequeña, pero ¿qué implicaciones tendrá? Para apreciarlo, ha llegado la hora de irnos de vacaciones.

# Lo llevamos a Monte Carlo

Cuando estuve en la Universidad, en el Campus Monterrey, llevé una materia que se llamaba “Análisis de Señales y Sistemas de Comunicación” con el gran Doc Olvera. El Doc, que no se andaba escaso de carisma, fue brillante en muchos sentidos, pero, en particular, cuando llegaba la hora de resolver algo en clase tenía una generosa baraja de “viejos trucos” que hoy en día no sé si eran ciencia o magia. Uno de los viejos trucos que más usaba El Doc (o, seguramente, el que más recuerdo) era lo que se conoce como “El Teorema de L’Hôpital” y, cada vez que lo aplicaba para resolver algo, nos decía “¿y ahora como le hacemos? ¡Pues lo llevamos al Hospital!”.

Así que, para fines de este capítulo y con el permiso de El Doc, vayamos a un destino de lo más encantador. Para llegar necesitamos una máquina muy simple, pero al mismo tiempo muy especial, se trata de una caja que nos entrega *al azar* una pelotita que tiene un número impreso entre el 1 y el 100. No necesitamos más.

Ahora recordemos el partido que mencionamos hace algunos párrafos:

|  |  |  |
| --- | --- | --- |
| Victoria Cruz Azul | Empate | Victoria  Monterrey |
| 42% | **26%** | **32%** |

Lo que haremos es distribuir esos porcentajes entre los números del 1 al 100, asignando 42 números a Cruz Azul, 26 al Empate y 32 a Monterrey:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Victoria  Cruz Azul** | **Empate** | **Victoria Monterrey** |  | **Victoria Cruz Azul** | **Empate** | **Victoria Monterrey** |
| **42%** | **26%** | **32%** |  | **1 a 42** | **43 a 68** | **69 a 100** |

A picture containing text, electronics, keyboard, calculator

Description automatically generatedA picture containing text, electronics, calculator, keyboard

Description automatically generated

De tal forma que los porcentajes ahora se traducirían en tres rangos de valores. Con esto listo, ahora le pedimos a nuestra caja que nos dé una pelotita, la cual resulta ser la siguiente:

**27**

Como el número de la pelotita quedó dentro del rango de números que asignamos a la Victoria de Cruz Azul, entonces, diremos que el equipo Celeste ganó el partido. Así:



A picture containing text, electronics, calculator, keyboard

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Pero ese fue solo un partido y en la Liga MX hay, cada temporada, 17 jornadas de 9 partidos que da un total de 153 encuentros. Los resultados determinan las posiciones en la tabla al final de la temporada regular otorgando 3 puntos por partido ganado y, en caso de empate, se reparte 1 punto a cada equipo. Para simular un torneo posible, gracias a la Base de Datos de partidos de y apuestas de https://www.football-data.co.uk/ que, amablemente, me permitieron usar, solo tendríamos que repetir la operación de la pelotita para cada uno de los 153 partidos. Por último, con los resultados obtenidos, acumulamos los puntos y, al final, tendríamos nuestra tabla general:

Table

Description automatically generated

Y si ya que estamos aquí, ¿qué tal que simulamos otro torneo? O mejor de una vez que sean 4:

Table

Description automatically generated

Table

Description automatically generated

A picture containing table

Description automatically generated

Table

Description automatically generated

Los cinco torneos simulados hasta ahora son torneos posibles que terminan formando un conjunto al que podemos hacer preguntas, pero ¿por qué quedarnos cortos? ¿Y si le seguimos? Un torneo de la Liga MX se compone de 153 partidos y cada uno de ellos tiene 3 posibles resultados (victoria del equipo A, empate y victoria del equipo B). Si lo comparamos, por ejemplo, con la fase de grupos de Copa Mundial de la FIFA, en un grupo de 4 equipos hay un total de 6 partidos (3 jornadas con 2 partidos cada una) y cada uno tiene 3 posibles resultados. De ahí que, el total de posibles resultados se calcula como que es igual a 729 posibles combinaciones de resultados en 6 partidos, entonces ¿cuántos torneos posibles hay en nuestra querida Liga MX si su temporada tiene 153 partidos? Para ser exactos: lo cual es… una cifra de más de 70 números y para la cual ni siquiera existe una palabra para nombrarla en ningún idioma del mundo.

Así de grande es el futbol.

Sin embargo, dado que enumerar todos esos posibles torneos y calcular sus probabilidades tomaría un largo rato, y como sería preferible obtener resultados en un lapso menor al tiempo de vida de un ser humano y, como aquí nadie quiere dejar sin chamba al Dr. Strange, cerremos la cantidad de simulaciones en 5000.

La siguiente tabla muestra el resultado de lo que empezó con una PlayStation y unos mundialitos en El Winin. Para entender la tabla, considera los siguientes puntos. Te recomiendo que, después de leer cada punto, intentes entenderlo en la tabla:

* En la tabla **los 18 equipos aparecen ordenados** según la posición que ocuparon en el torneo Apertura 2021.
* Las **columnas** Líder, Sublíder, 3, 4, 5, 6, etcétera, representan las 18 posiciones posibles en la Liga MX al final de la temporada. Cada equipo, en principio, podría quedar en cualquiera de las 18 posiciones.
* Los porcentajes en cada una de las casillas representan el porcentaje de veces que, de los 5000 torneos simulados, quedaron en esa posición.
  + Por ejemplo, de las 5000 simulaciones León quedó como Tercer Lugar, aproximadamente, 500 veces, es decir, 10%.
* **Los números en color rojo representan el porcentaje de veces que cada equipo quedó en la posición que ocupó en el Apertura 2021**.
  + Por ejemplo, de las 5000 simulaciones, en 10% de los casos Tigres quedó en la posición 4.
* El color en las celdas es una escala de tonos que, **entre más verde, mayor es el porcentaje de que el equipo ocupase esa posición en la tabla**. En contra parte, entre más blanca sea la casilla, menor es el porcentaje de que el equipo ocupase dicha posición.
  + Por ejemplo, Mazatlán tuvo 22% de quedar en último lugar general y, al ser su porcentaje más alto, aparece en el tono de verde más oscuro.
* Considera que los porcentajes están redondeados.
  + Redondeando a 1 decimal, un 0.05% redondea a 0.1%, mientras que 0.04% redondea a 0%.

Table

Description automatically generated

Con temor a sonar aún más repetitivo, recuerda que no basta con solo obtener resultados. Para poner la tabla anterior en perspectiva, observa la siguiente tabla que corresponde al mismo tipo de simulación, pero aplicada a la Bundesliga en su temporada 2019-2020.

Table, Excel

Description automatically generated

Miles de torneos después, lo que hemos construido es una *Aproximación de Probabilidades* utilizando lo que se conoce como *Simulación Monte Carlo* y, con estas tablas en mano, es necesario hacerles preguntas:

¿En cuál de las dos tablas aparecen más casillas con 0% de probabilidad? ¿Qué interpretación darías a ese detalle? U otra forma de apreciar lo anterior: ¿en cuál de las dos tablas aparecen más casillas coloreadas con algún tono de verde? ¿Qué podría indicarnos esto? Otra manera de apreciar la idea de este capítulo: ¿en cuál de las dos tablas hay más equipos que tienen 0% de probabilidad de quedar en último lugar? O, al revés, ¿en cuál de las dos tablas hay más equipos que tienen 0% de probabilidad de alcanzar el primer lugar?

¿Cuál de las dos tablas muestra un torneo más competitivo?

La Falsa Irregularidad del Futbol Mexicano

Ing. Jaime G. Meléndez

Twitter: [@JamesGMelendez](https://twitter.com/JamesGMelendez)  
  
Web: <https://academiathi.com/>   
Email: [james@academiathi.com](mailto:james@academiathi.com)   
LinkedIn: [jaimemelendezthi](https://www.linkedin.com/in/jaimemelendezthi/)   
GitHub: [@castorx9000](https://github.com/castorx9000)

Síguenos en el podcast: [Enrachados Deportes](https://www.facebook.com/EnrachadosDeportes) todos los jueves a las 21:00 horas   
Twitter e Instagram: [@Enrachados\_deportes](https://www.instagram.com/enrachados_deportes/?hl=en)

P.D.: en el largo plazo todas las ligas del mundo son regulares.