

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**EGOBETS, UN SISTEMA COMPUTACIONAL DE  
ASESORÍA DE APUESTAS DE FUTBOL**

TESIS

QUE PARA OBTENER LOS TÍTULOS DE

INGENIERO EN COMPUTACIÓN Y LICENCIADO EN MATEMÁTICAS  
APLICADAS

PRESENTA

BRUNO MEDINA BOLAÑOS CACHO

ASESORES:

DR. OSVALDO CAIRÓ BATTISTUTI

DR. ADOLFO J. DE UNÁNUE TISCAREÑO

MÉXICO, D.F.

2014

“Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “**EGOBETS, UN SISTEMA COMPUTACIONAL DE ASESORÍA DE APUESTAS DE FUTBOL**”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación”.

BRUNO MEDINA BOLAÑOS CACHO

---

FECHA

---

FIRMA

*A mis padres*



# Agradecimientos

A mis amigos por todo el apoyo que me han dado para terminar este arduo proyecto. Sin ustedes, no sería quien soy hoy en día, gracias por todas sus enseñanzas y correcciones.



# Índice general

<b>Lista de figuras</b>	<b>VII</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Panorama general de las apuestas</b>	<b>5</b>
2.1. La fascinación por los juegos de azar . . . . .	5
2.2. La casa siempre gana . . . . .	8
2.3. Mercados de apuestas deportivas . . . . .	12
2.4. Bookies y momios . . . . .	15
<b>3. Las matemáticas detrás de Egobets.com</b>	<b>21</b>
3.1. Modelo de un partido de futbol . . . . .	21
3.1.1. Modelo de Poisson bivariada . . . . .	24
3.1.2. Verificando el modelo . . . . .	25
3.2. Reserva un poco para después . . . . .	26
3.3. Proceso para la selección de apuestas . . . . .	30
3.3.1. Variables necesarias . . . . .	31
3.3.2. Funciones de utilidad . . . . .	32
3.3.3. Paso 1 - Determinación de apuestas redituables .	34
3.3.4. Paso 2 - Selección de apuestas . . . . .	37
3.3.5. Paso 3 - Calcular cuanto dinero a cada apuesta .	38
3.3.6. Paso 4 - Determinar nivel de reservas . . . . .	39
3.3.7. Paso 5 - Presentando la recomendación . . . . .	42

<b>4. Sistema de Egobets.com</b>	<b>45</b>
4.1. Diseño y arquitectura . . . . .	45
4.1.1. Ventajas de correr Egobets en la nube . . . . .	46
4.1.2. Servidor LNNP . . . . .	49
4.2. Patrón de diseño MVC . . . . .	51
4.2.1. Modelos . . . . .	56
4.2.2. Vistas . . . . .	58
4.2.3. Controladores . . . . .	64
<b>5. Conclusiones</b>	<b>69</b>
<b>A. La ruina del jugador</b>	<b>73</b>
<b>B. Modelos lineales y soluciones a los problemas de optimización</b>	<b>77</b>
B.1. Decidir a favor de quien apostar . . . . .	77
B.2. Decidir la cantidad de dinero a apostar . . . . .	81
B.2.1. Ahorro precaucional . . . . .	86
B.2.2. Evolución del dinero en el tiempo . . . . .	91
<b>C. Bases de datos</b>	<b>95</b>
C.1. Dump BD Egobets.com . . . . .	95
C.1.1. Casas de apuestas . . . . .	95
C.1.2. Administradores . . . . .	95
C.1.3. Equipos . . . . .	96
C.1.4. Usuario . . . . .	99
C.1.5. Mailing . . . . .	100
C.1.6. partidos . . . . .	100
C.1.7. Resultados . . . . .	100
C.1.8. system.indexes . . . . .	101
C.1.9. usuarios . . . . .	101



# Índice de figuras

2.1. Astrágalos, los predecesores de los dados . . . . .	6
2.2. Miles de millones de dólares en apuestas . . . . .	7
3.1. La función $g(f)$ siempre encuentra un máximo y después tiende a $-\infty$ . . . . .	29
4.1. Arquitectura Cliente Servidor sobre Nube de AWS . . .	46
4.2. Patrón de diseño Modelo Vista Controlador (MVC) . . .	52
4.3. Representación UML de las colecciones de la base de da- tos de Egobets . . . . .	55
4.4. La encuesta define el perfil de riesgo del usuario . . . . .	59
4.5. La encuesta define el perfil de riesgo del usuario . . . . .	60
4.6. Partido con recomendación y estadísticas de los equipos	63
B.1. Decidir por quién apostar . . . . .	78
B.2. Decidir si apostar o no apostar . . . . .	79
B.3. Decidir si apostar en función de una utilidad . . . . .	79
B.4. Árbol de probabilidad 4 . . . . .	81



# Capítulo 1

## Introducción

Desde sus orígenes, las apuestas en los partidos de futbol han sido un controversial tema de interés [44]. Predecir los resultados de los partidos y vencer a las casas de apuestas se ha vuelto una fascinación Hollywoodense<sup>1</sup>. Muchos supuestos “oráculos” han utilizado los métodos menos ortodoxos para la predicción de los marcadores [32], e incluso se han llevado a cabo acciones fraudulentas para asegurar los marcadores finales de los partidos<sup>2</sup>. Sin embargo, en la actualidad, las matemáticas y la computación ofrecen un paradigma menos esotérico pero igual de fascinante: la predicción de resultados de partidos de futbol a través de modelos matemáticos.

En este trabajo se describe cómo funciona “**Egobets**”, una aplicación computacional de las matemáticas al estudio de las apuestas de

---

<sup>1</sup>En la película “*Moneyball*” [27] (basado en [20]) un equipo de béisbol logra resultados sorprendentes al resolver un problema de optimización con fuertes restricciones monetarias. Mientras que en el filme “*21*” [22], basado en el libro “Bringing Down the House” [28], un grupo de estudiantes del MIT (Massachusetts Institute of Technology) utiliza una estrategia de conteo infalible para ganar cientos de miles de dólares en el juego de cartas “Black Jack”.

<sup>2</sup>Por ejemplo, en 2006 se suscitó uno de los mayores escándalos en la historia del futbol: “*Calciopoli*”. Se descubrió que varios equipos de la liga italiana conspiraron para influenciar los resultados de los partidos de la temporada 2004/05 [6].

fútbol. Egobets provee asesoría de apuestas personalizadas para partidos de fútbol de las siguientes ligas europeas: alemana, española, francesa, inglesa e italiana. Su objetivo es, dado un perfil de riesgo, indicar al usuario la cantidad de dinero y las apuestas que debe realizar para buscar tener ganancias al final de la temporada. Para tal fin, se combinan un conjunto de modelos matemáticos en un sistema robusto computacional.

El sistema Egobets es interesante e innovador ya que no sólo predice el resultado de un partido de fútbol, sino que además utiliza la información de todas las ligas europeas para ofrecer una estrategia que maximice la cantidad de dinero a ganar del usuario tomando en cuenta su perfil de riesgo. Adicionalmente, el sistema le sugiere al usuario conservar un porcentaje de su dinero para apostar más agresivamente en caso de perder todas la apuestas de la jornada; garantizando así una mayor cantidad de apuestas durante la temporada y con esto, asegurar una mayor probabilidad de obtener ganancias.

El alcance de este trabajo es el de describir el sistema desarrollado para asesoría de apuestas Egobets. Se explicarán los distintos programas y sistemas que conforman el desarrollo, así como las teorías matemáticas que dan sustento al mismo. El documento cuenta con tres capítulos más las conclusiones y esta introducción; el primer capítulo que habla de las apuestas en general, el segundo describe la teoría matemática y el tercero detalla el proyecto realizado.

El primer capítulo, empieza hablando de las ilusiones que mueven a las personas a apostar, después describe como los casinos utilizan los juegos de azar para generar ganancias. En seguida, explica como funcionan los mercados de las apuestas deportivas y sus diferencias con los casinos. Y finalmente, se cierra el capítulo definiendo lo que son los “momios” y su papel como regulador en la demanda de la apuesta.

En este capítulo se presentan, sin entrar al detalle técnico, la existencia de varios modelos matemáticos que describen el comportamiento de los partidos de fútbol y como varios autores han logrado encontrar estrategias de apuestas sencillas que garantizan valores esperados positivos de un conjunto de apuestas. Finalmente se retoman estos conceptos

y se detalla el algoritmo que utiliza Egobets para la generación de las recomendaciones de apuestas.

Egobets.com proporciona al cliente los servicios de asesoría de apuestas personalizada a través de un portal Web usable, práctico y profesional. En el capítulo tercero, se presenta el sistema desarrollado con los fundamentos teóricos descritos en los capítulos anteriores, una verdadera aplicación computacional de las matemáticas. Se define el diseño y la arquitectura del sistema de Egobets en la nube. Además, se habla del conjunto de tecnologías desarrolladas y utilizadas en el sistema. Para terminar, se presentan los diagramas de base de datos y se expone el sistema a través de las tres dimensiones del patrón de diseño Modelo Vista Controlador.

En el último capítulo, se concluye que se puede llevar un apuesta simple a una estructura de portafolio de inversión. De igual manera se observa que aunque un jugador tuviera en su poder las probabilidades verdaderas de los resultados de los partidos no podría hacer nada con ellas, por lo que es necesario un enfoque de un problema de optimización. Y finalmente que teniendo un sistema metódico que decida las apuestas, remueve la emoción de la apuesta y lo convierte en un riesgo calculado. Para finalizar, se sugieren los distintos campos al que este tipo de sistemas se podría extender: otras ligas, diferentes deportes, elecciones y otros fenómenos parecidos donde se involucre la habilidad humana.



## Capítulo 2

# Panorama general de las apuestas

Este capítulo, habla de las ilusiones detrás de las apuestas y describe en general como los casinos utilizan los juegos de azar para generar ganancias. Después, explica como funcionan los mercados de las apuestas deportivas y sus diferencias con los casinos. Y finalmente, se cierra el capítulo definiendo lo que son los “momios” y su papel como regulador en la demanda de la apuesta.

### 2.1. La fascinación por los juegos de azar

Nadie conoce el origen de las apuestas, algunos dicen que todo comenzó con un anónimo paleolítico que rodó unos cuantos huesos para decidir hacia que dirección ir a cazar [40]. Más adelante, tanto los antiguos griegos como los etruscos examinarían la forma y las características del hígado de una oveja para tomar las mejores decisiones para su futuro<sup>1</sup>. Siglos después, los romanos usarían los huesos astrágalos (Ver

---

<sup>1</sup>Los adivinos llamados “Arúspices” eran los encargados de llevar la tarea de predecir el futuro en función de la examinación de las entrañas de varias bestias.

figura 2.1) de animales como precursores a los dados [40].

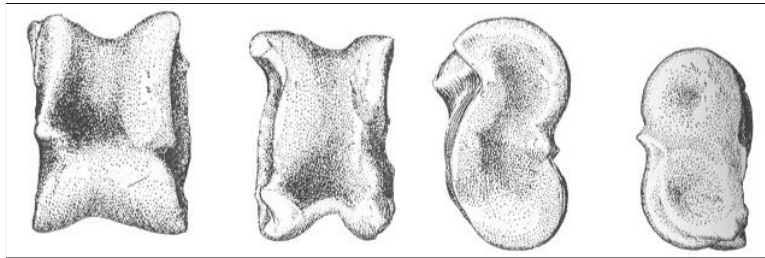


Figura 2.1: Astrágalos, los predecesores de los dados

Hoy en día, las apuestas representan uno de los negocios más rentables del mundo. En 2013, se estimó que las ganancias brutas de esta industria sumaron más de cuatrocientos cuarenta mil millones de dólares<sup>2</sup>. Como se puede observar en la figura 2.2, Estados Unidos encabeza la lista como el país que más gasta en apuestas, seguido por China. También se advierte que los residentes de Australia y Singapur apuestan mucho más agresivamente que los de cualquier otro país. Para terminar, en esta misma gráfica se estima que para el 2018 el gasto en apuestas será de más de quinientos mil millones de dólares.

Con estos datos, la relevancia de la industria de las apuestas en el mundo se vuelve evidente. Por otra parte, con respecto a las apuestas en línea, la firma KPMG [13] reporta que el mercado global de apuestas en línea creció un cuarenta y dos por ciento de Veintiún mil doscientos millones de dólares en 2008 a Treinta mil millones de dólares en 2012. Este porcentaje es notablemente superior al quince por ciento esperado para el crecimiento del total de la industria de apuestas para el mismo periodo.

Específicamente en Estados Unidos, Goldman Sachs valoró en 2009 que el mercado de apuestas en línea en caso de ser legalizado<sup>3</sup> podría

---

<sup>2</sup>Acorde a la empresa de Inteligencia de Mercado “H2 Gambling Capital” [8].

<sup>3</sup>El “Unlawful Internet Gambling Enforcement Act of 2006” (UIGEA) prohíbe a los



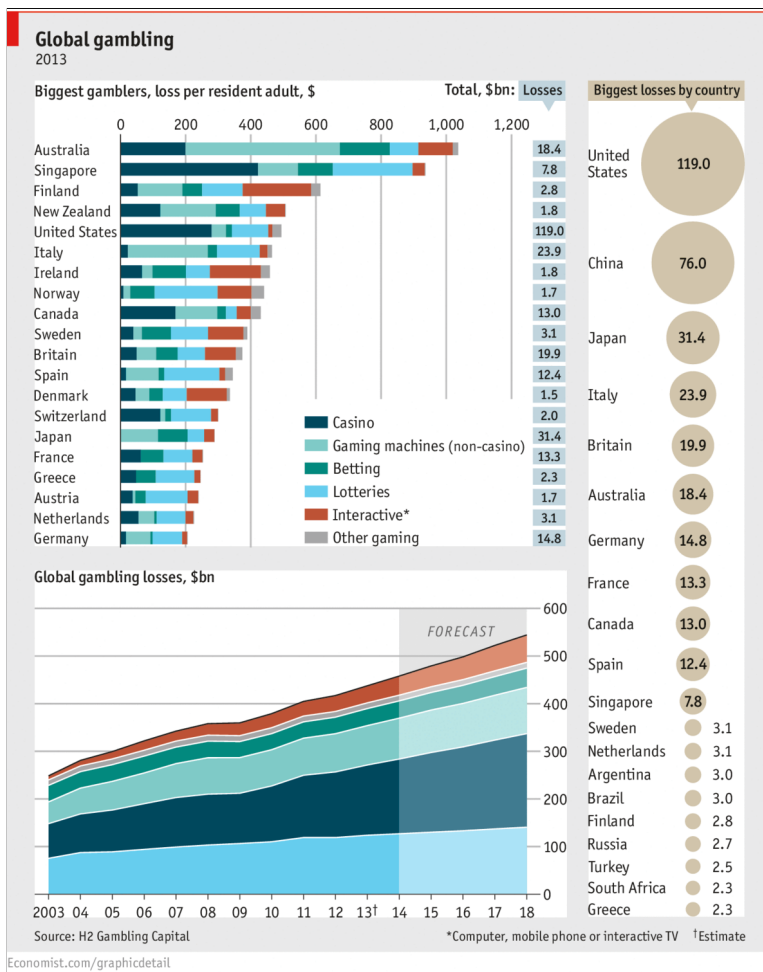


Figura 2.2: Miles de millones de dólares en apuestas

bancos y a las compañías de tarjetas de crédito procesar cargos relacionados a casinos en línea. Según Alexander, G. [1] las cuatro preocupaciones federales principales detrás de este acto son: Primero, el internet provee un acceso fácil a las apuestas, esto

valer hasta doce mil millones de dólares [38]. En este mismo documento de KPMG [13], México se propone como un mercado potencialmente lucrativo. Una de las principales razones es la legislación que permite el juego en línea<sup>4</sup>. La otra razón, el valor del mercado mexicano del juego en línea se estima en cuatro mil seiscientos millones de dólares [49].

## 2.2. La casa siempre gana

*“No hay proposición más cierta en matemáticas que la siguiente: Entre más boletos [de lotería] compre, más probabilidades tiene de ser un perdedor. Compre todos los boletos de la lotería y pierda con certeza; cuanto más boletos compre, más cerca estará de esta certeza”*

– Adam Smith, *Filósofo* [42]

El principio básico detrás de un casino es muy sencillo: *la ventaja de la casa*. Cada uno de los juegos que ofrece el casino tiene detrás un robusto sustento matemático, de manera que a pesar de aparentar ser un juego justo, le confiere a la casa una ventaja porcentual sobre el conjunto de jugadores. Al final del día, esta ventaja y la ley de los grandes números, le garantizan a los casinos que a largo plazo tendrán suficientes ganancias para subsistir, mantener su operación y gozar de utilidades sorprendentes. Sin embargo, no hay que olvidar que los fenómenos estudiados siguen siendo producto del azar, por lo que una buena racha de

---

podría exacerbar las tentaciones que enfrentan los apostadores compulsivos. Segundo, es muy complicado verificar la mayoría de edad a través de un sitio de apuestas. Tercero, los casinos en línea tienen un incentivo para defraudar a los usuarios gracias a la falta de regulación de la industria. Y cuarto, dado el volumen, la velocidad, el alcance internacional de las transacciones realizadas en internet y el alto nivel de anonimidad que tienen los operadores de casinos electrónicos; los oficiales federales creen que las apuestas en línea son particularmente susceptibles al lavado de dinero.

<sup>4</sup>En 2004, la Ley Federal de Juegos con Apuestas y Sorteos permitió y reguló los Juegos en Línea.

algunos “Grandes Apostadores” podría llegar a asustar aun a los dueños más racionales de casinos [11].

Según Hannum [11] hay dos grandes razones por las que la gente apuesta:

1. **Entretenimiento.** Un individuo podría utilizar mil pesos para ir a un casino o a un concierto. Si la ventaja de la casa es muy grande y la persona pierde su dinero rápidamente, entonces la experiencia del entretenimiento del casino no sería apreciada por el jugador. Por el otro lado, si el casino logra entretener a la persona por una tarde mientras le regala bebidas y comida, entonces puede que este individuo repita la experiencia y nunca más asista a un concierto.
2. **Cambio de Vida.** Si una persona ahorrara cien pesos semanalmente, al final del un año tendría cinco mil doscientos pesos. Pero si ese dinero lo gastara para comprar boletos de lotería, tendría la posibilidad de ganarse cuarenta millones de pesos. Claramente la probabilidad es muy cercana a cero; sin embargo, este gasto podría ser visto por esta persona como una única oportunidad para cambiar su vida.

*La ventaja de la casa*, se puede entender mejor analizando cada uno de los juegos que ofrece el casino y las probabilidades de ganar que tienen los jugadores. Tómese por ejemplo el juego de la ruleta americana<sup>5</sup>: Cuando un jugador apuesta sobre el color negro (i.e. que la pelotita caiga sobre alguna de las casillas negras) entonces se tiene que la probabilidad de que el jugador gane la apuesta es de:

$$p\{\text{La pelotita caiga en casilla negra}\} = \frac{\# \text{ casillas negras}}{\# \text{ casillas totales}} = \frac{18}{38}$$

---

<sup>5</sup>El juego de la ruleta americana consiste en 38 casillas que alternan 18 casillas rojas, 18 casillas negras y 2 casillas verdes. Cuando el crupier hace girar la ruleta, de manera aleatoria cae una pelotita en una de las casillas. Los jugadores apuestan sobre la posición final de la pelotita.

Afortunadamente para la casa, hay 2 casillas que no son color negro ni rojo, por lo que de las 38 casillas sólo 36 tienen estos colores. Por lo tanto, la probabilidad de que la pelotita caiga en una casilla verde es la siguiente:

$$p\{\text{La pelotita no caiga ni en casillas rojas ni en negras}\} = \frac{\# \text{ casillas totales} - (\# \text{ casillas rojas} + \# \text{ casillas negras})}{\# \text{ casillas totales}} = \frac{38 - 18 - 18}{38} = \frac{2}{38}$$

Estos  $\frac{2}{38}$  son la ventaja de la casa, ya que cuando un jugador apuesta al color negro en la ruleta y acierta, recibe la misma cantidad de dinero que podría perder. Sin embargo, apostó a ganar con una probabilidad de  $\frac{18}{38}$ , pero la probabilidad de perder la apuesta es igual a  $1 - \frac{18}{38} = \frac{20}{38}$ . Este detalle hace importante ver el valor esperado que tiene esta apuesta para el jugador:

$$E[\text{Apostar } k \text{ pesos al color negro}] = k \cdot p\{\text{La pelotita caiga en casilla negra}\}$$

$$-k \cdot p\{\text{La pelotita no caiga en casilla negra}\} = k \cdot \frac{18 - 20}{38} = -k \cdot \frac{2}{38}$$

Dado que siempre que se apuesta  $k > 0$ , esto implica que:

$$E[\text{Apostar } k \text{ pesos al color negro}] = -k \cdot \frac{2}{38} < 0; \forall k \in \mathbb{N}$$

Este resultado quiere decir que a la larga el jugador **siempre** va a terminar perdiendo dinero. En un principio,  $\frac{2}{38}$  de probabilidad pareciera poco, pero al multiplicarlo por la gran cantidad de jugadores y apuestas que se realizan en los casinos el monto final se vuelve exorbitante.

Este sencillo ejercicio ejemplifica como todos los juegos que se tienen en los casinos ofrecen una ventaja para la casa. Es interesante mencionar, que además de los juegos de azar como la ruleta, hay juegos que obligan al jugador a tener cierta habilidad para no perder su dinero tan rápidamente, este es el caso de juegos como el Blackjack que le dan a la casa ventajas más pequeñas al enfrentarse a jugadores expertos. La ventaja de la casa sustenta las ganancias del casino, sin embargo calcular esta ventaja puede llegar a ser complicado y requerir un análisis matemático mucho más sofisticado e incluso se puede llegar a necesitar programar el juego para correr simulaciones y estimar estas probabilidades.

Juego	Ventaja de la Casa
Ruleta (con doble cero)	5.3 %
Dados (pass/come)	1.4 %
Dados (pass/come con momios dobles)	0.6 %
Blackjack - jugador promedio	2.0 %
Blackjack - 6 barajas, estrategia básica	0.5 %
Blackjack - una baraja, estrategia básica	0.0 %
Baccarat (sin apuestas de empate)	1.2 %
Caribbean Stud	5.2 %
Let It Ride	3.5 %
Poker de tres cartas	3.4 %
Pai Gow Poker (ante/play)	2.5 %
Tragamonedas	5 % - 10 %
Video Poker	0.5 % - 3 %
Keno (promedio)	27.0 %

Tabla 2.1: Ventajas de la casa para juegos populares de casino [11]

Finalmente, aun sin la ventaja de la casa se deber recordar que existe un famoso problema y su corolario que garantizan que la casa siempre gane: *La ruina del Jugador* [36, p. 95-99]. Este problema enfrentado por

varios famosos matemáticos<sup>6</sup> deja la siguiente lección: La probabilidad de que un jugador pierda todo su dinero es igual a  $p_1 = \frac{n_2}{n_1+n_2}$  donde  $p_1$  es la probabilidad de ganar del jugador 1 y  $n_i$  es la cantidad de dinero que va a apostar el jugador  $i$ . Desafortunadamente, usualmente la casa tendrá más dinero para apostar que cualquier jugador, por lo que con esta fórmula la casa siempre gana<sup>7</sup>.

## 2.3. Mercados de apuestas deportivas

A diferencia de las máquinas tragamonedas y los juegos de mesa de los casinos, las apuestas deportivas tienen una gran ventaja para los apostadores: los resultados de los encuentros deportivos **no son completamente aleatorios**, ya que dependen en cierta medida del nivel de juego que tienen los equipos participantes. Hannum menciona en su libro [11] : *“La ventaja de la casa existe para casi todas las apuestas en un casino (ignorando las salas de Poker y las apuestas deportivas donde algunos pocos profesionales pueden vivir de de las apuestas)”*

Existen dos grandes mercados en esta rama de las apuestas [3]:

- *Bookies*.<sup>8</sup> El corredor de apuestas analiza los diferentes resultados de un encuentro deportivo en función de los participantes. Con base en este análisis, el bookie publica un número (llamado “*momio*”) para cada uno de los posibles desenlaces del partido. Este “momio” representa la cantidad potencial de retorno de esa apuesta (incluida la cantidad de dinero arriesgada)<sup>9</sup>. El bookie recibe

---

<sup>6</sup>Se dice que Blaise Pascal se lo planteó a Pierre Fermat en 1656 [9]. Después Fermat se lo replanteó a Christian Huygens en 1657 y finalmente James Bernoulli lo resolvió en su forma general como el problema de la “Duración de Juego”. Fue publicado ocho años después de la muerte de Bernoulli en 1713 [36, p. 98].

<sup>7</sup>Ver demostración en el apéndice A.

<sup>8</sup>“Bookie” proviene de la palabra en inglés “bookmakers” que en español se utiliza como: “Corredor de apuestas”.

<sup>9</sup>Por ejemplo, supóngase que se apuestan cien dólares a favor del empate de un partido con un momio igual a 3.640. En caso de que se ganara la apuesta, el jugador

apuestas sobre estos momios y cobra una comisión por cada operación recibida.

- *Sistema parimutual.* En este mercado no existen corredores, ni momios. El pago que reciben los jugadores depende de la cantidad de dinero recopilada por todas las apuestas recibidas. Por lo tanto, las ganancias de los jugadores no están determinadas hasta que todas las apuestas se reciben<sup>10</sup>.

Aunque la eficiencia<sup>11</sup> de estos mercados no es un tema que se aborde a profundidad en esta tesis, es interesante mencionar que ha sido muy cuestionada. Sauer [39] y Williams [48] critican varias de sus anomalías; por ejemplo hay un problema interesante en la eficiencia del mercado de las apuestas de los bookies conocido como: Sesgo del “*favorite - long shot*”. Estos autores explican que las apuestas a los equipos favoritos generan un mayor retorno de dinero en comparación con el retorno generado por las apuestas a equipos cuyas probabilidades de ganar son mucho menores (“*longshot bets*”). Incluso se han realizado varios estudios proponiendo un nuevo mercado de tipo “doble subasta” que consiste en tener compradores y vendedores proponiendo los precios de la apuesta, cuando dos de ellos coinciden, se lleva acabo la transacción. Referirse a Ozgit[30].

El mercado pertinente para este trabajo es el de los “bookies” o corredores de apuestas. Sin embargo, bajo las condiciones adecuadas la asesoría de apuestas podría ayudar a los jugadores en el mercado parimutual. Por ejemplo, en una apuesta entre compañeros del trabajo el usuario del sistema podría indagar la apuesta realizada por cada uno de los participantes y conocer la cantidad de retorno que tiene de cada uno de los resultados. Si el sistema de Egobets le recomendara realizar

---

recibiría la cantidad de  $(100)(3.640) = 364$  dólares. Sustrayendo los cien dólares que apostó, su ganancia neta sería de 264 dólares.

<sup>10</sup>Las afamadas “quinielas” son un tipo de apuesta de mercado parimutual.

<sup>11</sup>Fama, E [10] sugiere que la eficiencia de mercado es cuando los precios reflejan completamente toda la información disponible de una acción en particular.

esta apuesta en particular, bastaría verificar que el retorno del pago sea mejor que el pago que realizan las casas de apuestas en general. Realizando esta acción sistemáticamente, se podrían conseguir mejores ganancias que las ofrecidas por el mercado.

Sin importar el mercado, es importante tomar en cuenta el hecho de que las probabilidades reales que tienen los distintos resultados de cualquier partido **son desconocidas**. Por lo que los “bookies” contratan a empresas consultoras, o designan un área de la compañía para calcular los momios que se publican en el mercado. Este punto es básico en la creación de sistemas como Egobets, ya que el hecho de que las probabilidades sean desconocidas permite que Egobets busque estimar probabilidades mucho más cercanas a las reales. Al minimizar el error de estimación que tienen las probabilidades sugeridas por las casas de apuestas, se pueden encontrar apuestas cuyo retorno sería mayor al que realmente debería ser. Y como se verá en la siguiente sección, este error de estimación siempre existirá, debido a que los momios publicados por las casas de apuestas no buscan reflejar las probabilidades reales de los resultados de los partidos.

Una ventaja con la que cuentan los apostadores de eventos deportivos: si los bookies establecen mal sus momios<sup>12</sup>, entonces ciertos jugadores pueden llegar a tener valores esperados positivos en sus apuestas y con esto la casa podría llegar a perder mucho dinero, incluso a largo plazo (como se verá en la siguiente sección). Procédase a la siguiente sección para conocer los orígenes de los momios y las condiciones que permiten a los bookies ganar dinero con ellos.

---

<sup>12</sup>La cantidad de dinero que pagan las apuestas a eventos completamente aleatorios, como los juegos dentro de los casinos, se puede calcular explícitamente. Por lo que la única manera de que una de estas apuestas tuviera valor esperado positivo sería si se cometieron errores en los cálculos.



## 2.4. Bookies y momios

Ya que el objeto de estudio central de esta tesis toma lugar en el deporte del futbol, analícese la pregunta más recurrente de un partido de futbol: *¿Qué equipo va a ganar este partido?*

La respuesta tiene 3 posibilidades:

1. El partido lo gana el equipo de casa.
2. Es un empate.
3. El partido lo gana el equipo visitante.

En un principio se pudiera pensar que los tres eventos son equiprobables, pero como se mencionó en la sección anterior esto resulta imposible por muchísimas razones, como por ejemplo: ser el equipo visitante conlleva una desventaja importante en el desempeño del partido (Ver [35]), o la desventaja de tener a los jugadores estrella lesionados; incluso las condiciones climáticas (altitud, tipo de pasto, lluvia) durante el partido pueden ser determinantes para el resultado final. En un caso extremo, imagínese el escenario donde ambos equipos fueran igual de buenos en todos los aspectos. En este escenario, la probabilidad de empate sería mayor que las otras dos. Sin embargo, estas razones no son (tampoco) suficientes para calcular los resultados de los partidos de manera determinística. Gracias a estas particularidades es que el estudio de este mercado de apuestas se vuelve tan interesante.

Anteriormente se definió un momio como el número que indica la cantidad de dinero que obtiene un jugador al ganar una apuesta. Sin embargo, para generar los momios, las casas de apuesta realizan estudios minuciosos de mercado y de los deportes en sí mismos. Las probabilidades de cada resultado de un partido no son, de lejos, los únicos factores a considerar por el bookie a la hora de generar los momios.

Levitt [19] menciona tres escenarios que permitirían a los bookies generar ganancias:

1. **Encontrar el precio de equilibrio.** Los bookies buscan predecir los momios<sup>13</sup> que igualen el precio de equilibrio del mercado de apuestas. Es decir, se buscan los momios que equilibran la cantidad de dinero en cada lado de la apuesta. Esto implica que la diferencia de dinero entre todas las apuestas es igual a cero. Por lo que sin importar quien gane el partido, el bookie cobrará su comisión de las apuestas intacta y no tendrá deudas en ninguna de las apuestas.
2. **Predecir los resultados del partido.** Si el bookie fuera sistemáticamente mejor en predecir los resultados de los partidos, entonces podría publicar el momio “correcto” del partido (i.e. el momio que equilibra la probabilidad de que una apuesta en cualquiera de los resultados gane). Aunque la cantidad de dinero no estaría equilibrada, en promedio el bookie ganaría la comisión cobrada a los jugadores. Nótese que a diferencia del primer escenario, en este esquema existe un riesgo enorme al que se exponen las casas de apuestas. Si llegaran a fijar un momio muy alejado del precio de equilibrio del mercado, entonces podría darse el caso de que la cantidad de dinero a pagar a los ganadores de la apuesta sea mayor a la cantidad de dinero recaudada por las apuestas complementarias, lo que podría significar pérdidas millonarias para el bookie<sup>14</sup>.

Ahora, tómese de ejemplo el siguiente partido con los siguientes momios:

■ **Cagliari Vs Juventus.**

- **9.290** Gana local (Cagliari)
- **1.423** Empate.
- **4.760** Gana visitante (Juventus)

---

<sup>13</sup> Antes del partido estos momios pueden llegar a tener ajustes típicamente pequeños y relativamente infrecuentes.

<sup>14</sup> Levitt [19] cuenta el ejemplo del Epton Derby de 1946, donde los momios fueron errados y la mitad de todas las casas de apuestas británicas se fueron a bancarrota.

Sea  $m_i$  el momio que propone el bookie para que los jugadores apuesten a que se cumpla el evento  $i$ . Entonces, se define el momio de la siguiente manera:

$$m_i = \frac{1}{\hat{p}_i}$$

Donde  $\hat{p}_i$  es la probabilidad estimada que tiene el bookie de que suceda el evento  $i$ <sup>15</sup>

Con estos momios se pueden calcular las “probabilidades estimadas por el bookie”:

$$\hat{p}_L = \frac{1}{9.290} \approx 0.107643...$$

$$\hat{p}_E = \frac{1}{1.423} \approx 0.702741...$$

$$\hat{p}_V = \frac{1}{4.760} \approx 0.210084...$$

Donde  $\hat{p}_L$  es la probabilidad estimada por la casa de apuestas de que gane el equipo local,  $\hat{p}_E$  es la probabilidad de que el partido termine en empate y  $\hat{p}_V$  es la probabilidad de que gane el equipo visitante.

Estas probabilidades estimadas proponen que el escenario más probable es un empate, después la victoria del Juventus y finalmente la victoria del Cagliari. Curiosamente en este ejemplo, se puede ver que aunque el equipo Cagliari es local, se enfrenta a un adversario que puede contra la desventaja de ser visitante.

Uno de los teoremas básicos de la probabilidad, dice que la suma de las probabilidades de todos los resultados del evento debe sumar uno [36]. Sin embargo, en el caso de los bookies, las probabilidades estimadas exceden la unidad. Este excedente, es la vieja conocida: “Ventaja de la Casa”.

---

<sup>15</sup>Otros deportes tienen diferentes tipos de momios y su definición varía dependiendo del tipo de apuesta, se puede leer más de ellos en [12].

Sea  $\mathcal{O}$  el conjunto de todas las opciones sobre las que un jugador puede apostar para cierto evento. Para el caso de las apuestas sobre los resultados de los partidos, se tiene que:

$\mathcal{O} = \{\text{Apostar a que gane local,}$   
 Apostar al empate,  
 Apostar a que gane visitante}\}.

Por lo tanto:

$$\text{Ventaja de la casa} = \left( \sum_{i \in \mathcal{O}} \frac{1}{\text{Momio de la opción } i} \right) - 1$$

Siguiendo con el ejemplo, se tiene que la ventaja de la casa de este bookie para esta apuesta en particular es:

$$\text{Ventaja de la casa} = \hat{p}_L + \hat{p}_E + \hat{p}_V - 1 \approx 0.020467$$

Este 2.046 % es el punto de partida que la casa quiere obtener en ganancias.

### 3. **Predecir los resultados y encontrar el precio de equilibrio.**

Si el bookie es mejor que los jugadores prediciendo el resultado de los partidos y también puede encontrar el precio de equilibrio, entonces podría con esta información mejorar sus ganancias esperadas al publicar un momio “equivocado” de tal manera que el momio (precio) de equilibrio quede posicionado donde sus ganancias se incrementen. Ahora, hay ciertas restricciones en cuanto a la distancia a la que puede quedar posicionado el momio “equivocado”, ya que pueden existir jugadores que conozcan el momio “correcto” por lo que entre mayor sea esta distancia podría generar mayores ganancias para estos jugadores. Por ejemplo, supóngase el siguiente caso extremo: Se disputa un partido entre los equipos  $A$  y  $B$ , el equipo  $B$  es el favorito para ganar. Es por esto que la casa de apuestas arregló que el partido lo gane el equipo  $A$ . El bookie

conoce que la probabilidad de que el equipo  $A$  gane el partido es 1, es por esto que fija los momios más competitivos del mercado a favor del equipo  $B$  y a favor del empate, mientras que el momio a favor del equipo  $A$  lo presenta mucho menos competitivo que el resto del mercado. Con estas acciones, la casa de apuestas maximiza las apuestas recibidas en contra del equipo  $A$  y minimiza las apuestas a favor del equipo  $A$  dado que las otras casas de apuestas pagan mucho mejor este resultado. Ahora, si los momios fueran absurdamente buenos en contra del equipo  $A$  ciertos apostadores podrían intuir que la casa de apuestas sabe algo que los demás no saben y podrían apostar cantidades exorbitantes a favor del equipo  $A$ .

De estos tres escenarios, Levitt [19] examina de un bookie en línea, veinte mil apuestas de doscientos ochenta y cinco jugadores para los partidos de la NFL<sup>16</sup>. Esto fue lo que encontró:

- A pesar de ser un estudio para una sola casa de apuestas, sugiere que se pueden generalizar ya que casi todas las casas ofrecieron casi los mismos momios para los mismos partidos
- El bookie no parece buscar el precio de equilibrio del mercado. Acorde a los resultados en casi la mitad de todos los juegos al menos dos tercios de las operaciones caen en un solo lado de la apuesta.
- El bookie parece colocar precios estratégicamente para explotar los sesgos de los apostadores. Es decir, los jugadores parecen tener un sesgo sistemático hacia los equipos favoritos y, en menor medida, hacia los equipos visitantes. En consecuencia, el bookie logró atraer mayor atención a partidos donde estos equipos no tuvieron buenas actuaciones; estos precios lograron elevar las ganancias hasta un veinte por ciento más que con lo que hubieran obtenido en el primer escenario.

---

<sup>16</sup>La liga de fútbol americano más popular: National Football League.

- Hay poca evidencia de que existan jugadores que hayan sido capaces de vencer a los bookies sistemáticamente.

Estas conclusiones indican que, al menos en el deporte del fútbol americano, las casas de apuestas buscan siempre tener mayores ganancias. Es por esta razón que podrían estar usando el tercer escenario descrito previamente. Es interesante destacar que los momios, en realidad reflejan la probabilidad que el mercado cree que tiene cada resultado de un partido. Es por esto que sistemas como Egobets, que funcionan al aprovecharse de los momios “equivocados” propuestos por las casas de apuestas, pudieran llegar a funcionar satisfactoriamente.

## Capítulo 3

# Las matemáticas detrás de Egobets.com

En este capítulo se presentan, sin entrar al detalle técnico, los resultados comprobables de las teorías que proporcionan las probabilidades de los resultados de los partidos y las predicciones de sus resultados. De igual manera se destaca la importancia de un sistema de reservas para optimizar las tasas de crecimiento de la cantidad de dinero dedicada a las apuestas. Finalmente se retoman estos conceptos y se detalla el algoritmo que utiliza Egobets para la generación de las recomendaciones de apuestas.

### 3.1. Modelo de un partido de futbol

El tema central de esta tesis no es la predicción de resultados de futbol principal si no la generación automatizada de recomendaciones de apuestas, por este motivo y para evitar cuestiones de plagios del modelo predictivo de Egobets, en este apartado no se entrarán en los detalles de la implementación, solamente se mostrarán las teorías los modelos matemáticos prácticos que permiten predecir los resultados de

los partidos de futbol sobre los que se basa Egobets. Los resultados de los estudios analizados en esta sección son suficientes para evidenciar la existencia de un conjunto de probabilidades y resultados estimados que permitan desarrollar una estrategia de apuestas capaz de generar un valor esperado positivo.

Maher en 1982 [23] describió la intuición detrás del modelado de un partido de futbol. Los goles anotados por un equipo en un partido pudieran ser modelados por una variable Poisson, la posesión es un aspecto importante del futbol ya que cada vez que un equipo tiene la pelota tiene la oportunidad de atacar y anotar. La probabilidad de que un ataque resulte en gol es pequeña, pero son muchas las veces que el equipo tiene la posesión durante el encuentro. Si la probabilidad de anotar en cada ataque y los ataques son independientes, el número de goles sería Binomial en estas circunstancias la aproximación por una Poisson funciona correctamente.

Más aún, Maher (1982) [23] al verificar la precisión de su modelo descubrió que si bien el modelo binomial se ajustaba razonablemente bien a los datos<sup>1</sup>, pero encontró que en los valores observados había una cantidad mayor de ocasiones en las que no habían goles durante el partido, o que simplemente la cantidad de goles anotada era muchísimo mayor a la predicha; esta observación lo llevó el ajuste del modelo considerando un coeficiente de correlación entre la cantidad de goles anotados entre ambos equipos.

Dixon y Coles (1997) [7] en su contribución, mencionan algunas propiedades deseables en un modelo de futbol:

- Se deben considerar las habilidades de ambos equipos en un partido.
- Se debe dar lugar a la ventaja observada que tienen los equipos al jugar en casa.

---

<sup>1</sup>En su artículo, Maher (1982) [23] cuenta que, al comparar doce conjuntos de datos (valores observados y esperados) mediante el valor esperado de una  $\chi^2$ , diecinueve de los veinticuatro casos dan un resultado no significativo a un 5 %.



- La medida más razonable de la habilidad de un equipo debe basarse en el desempeño de sus últimos juegos.
- La habilidad de un equipo, por la naturaleza del fútbol, se comprende como la composición de la habilidad de ataque (anotar goles) y la habilidad de defensa (no recibir goles).
- Al presentar el desempeño de un equipo en sus recientes resultados, se deberá tomar en cuenta la habilidad de los equipos contra los cuales ha jugado.

En su artículo [7], consideran la doble Poisson con un parámetro de dependencia que se estima junto con los otros parámetros. Sugieren que la suposición de independencia entre los goles de los equipos, tiene cabida en los partidos con resultados: 0-0, 1-0, 0-1 y 1-1<sup>2</sup>. Además, agregan al modelo una función que pondera los partidos con el fin de “diluir” los efectos de los partidos más lejanos.

Después, Rue y Salvensen (2000) [37] retoman el análisis de Dixon y Coles [7] y lo adoptan en un modelo dinámico lineal generalizado y adoptan un procedimiento de estimación Bayesiano para estudiar las propiedades de los equipos sobre un tiempo variable<sup>3</sup>. De igual manera, Crowder, Dixon, Ledford, and Robinson (2002) [5] representan el modelo de Dixon y Coles [7] como un espacio de estados no Gaussiano con fuerzas de ataque y defensa variables en el tiempo, la estimación se realiza utilizando métodos de aproximación por su alto costo computacional. Adicionalmente, Ord, Fernandes, and Harvey (1993) [29] presentan una extensión multivariada de un modelo de datos de conteo dinámico Bayesiano para el análisis y predicción del número de goles marcados por algún equipo.

---

<sup>2</sup>Karlis y Ntzoufras (2003) [15] sugieren que la predicción de empates mejora si se considera un parámetro de dependencia (incluso aunque sea pequeño)

<sup>3</sup>En su análisis empírico truncan la cantidad de goles a no más de cinco, consideran que esos marcadores no proveen mayor información acerca del ataque y defensa de un equipo.

Finalmente, Koopman y Lit (2013) [18] retoman todas estas ideas y presentan un modelo Poisson bivariado con ataque y defensa como variables estocásticas en función del tiempo.

### 3.1.1. Modelo de Poisson bivariada

Koopman y Lit (2013) [18] dicen que el resultado de un partido entre el equipo local  $i$  y el visitante  $j$  en la semana  $t$  es tomado del par de cuentas  $(X, Y) = (X_{it}, Y_{jt})$ , para  $I \neq j = 1, \dots, J$  y  $t = 1, \dots, n$  donde  $n$  es el número de semanas disponibles en el conjunto de datos. La primera cuenta  $X_{it}$  es el numero no negativo de goles anotados por el equipo local  $i$  y la segunda cuenta  $Y_{jt}$  es el numero de goles anotados por el equipo visitante  $j$ , en la semana  $t$ . Cada par de cuentas  $(X, Y)$  se genera de una distribución Poisson bivariada con función de densidad de probabilidad:

$$p(X, Y; \lambda_x, \lambda_y, \gamma) = e^{(-\lambda_x - \lambda_y - \gamma)} \frac{\lambda_x^X \lambda_y^Y}{X! Y!} \sum_{k=0}^{\min(X, Y)} \left(\frac{X}{k}\right) \left(\frac{Y}{k}\right) k! \left(\frac{\gamma}{\lambda_x \lambda_y}\right)^k$$

para  $X = X_{it}$  y  $Y = Y_{jt}$  con  $\lambda_x$  y  $\lambda_y$  coeficientes de intensidad para  $X$  y  $Y$  respectivamente, y  $\gamma$  siendo el coeficiente que mide la dependencia entre ambas cuentas del par  $(X, Y)$ .

La notación corta se describe como<sup>4</sup>:

$$(X, Y) \sim BP(\lambda_x, \lambda_y, \gamma)$$

Su valor esperado, varianza y covarianza son los siguientes:

$$E(X) = Var(X) = \lambda_x + \gamma, E(Y) = Var(Y) = \lambda_y + \gamma, Cov(X, Y) = \gamma$$

y el coeficiente de correlación entre  $X$  y  $Y$  está dado por:

$$\rho = \frac{\lambda}{\sqrt{(\lambda_x + \gamma)(\lambda_y + \gamma)}}$$

---

<sup>4</sup>La definición que se presenta en este documento es propuesta por [18], sin embargo hay otras variantes de esta fórmula [17] y [14]

La diferencia entre las cuentas de  $X$  y  $Y$  determinan si el partido es ganado, perdido o empatado para el equipo en casa. La variable  $X - Y$  tiene una distribución de probabilidad discreta conocida como: Skellman [41]. Una propiedad curiosa de la distribución Skellman  $X - Y$  es la invariancia de  $\gamma$  cuando  $(X, Y) \sim BP(\lambda_x, \lambda_y, \gamma)$ .

### 3.1.2. Verificando el modelo

El modelo descrito es puesto a prueba en el artículo de Koopman y Lit (2013) [18], donde para la liga inglesa Premier toma una muestra de la temporada 2003/03 a 2009/10 y su pronóstico logra mejorar la precisión del modelo de Maher [23] que ya se había discutido. Esta implicación reafirma que los ajustes aplicados como la dependencia entre las cuentas obtenidas por cada equipo, así como los coeficientes variables en función del tiempo describen de mejor manera el fenómeno y presentan mejor pronósticos que los anteriores.

Koopman y Lit (2013) [18] llevan esto a un paso más allá y presentan un análisis sobre estrategia de apuestas para demostrar que se pueden obtener ganancias contra el bookie. Básicamente su estrategia de apuestas consiste en apostar si el valor esperado de la apuesta sobre el evento  $A$  es mayor a una variable  $\tau$ .

$$EV(A) = P(A) \cdot Odds(A) - 1 > \tau$$

La primera propuesta, suponiendo que la ganancia del bookie es del 7, implica  $\tau > 7\%$ . Esto trae consigo un par de observaciones, con  $0 < \tau < 0.12$ , el promedio de retorno esperado es cercano a cero. Con  $\tau > 0.12$  se tiene un valor esperado positivo de ganancias. Sin embargo, conforme  $\tau$  crece la cantidad de oportunidades de apuestas se vuelve pequeña y los intervalos de confianza de las predicciones de las apuestas empiezan a reflejar mucha mayor incertidumbre.

Ejemplificando, si se toma  $\tau = 0.40$  y se toman 50 apuestas en dos temporadas y el retorno se espera justo un poco menos de 0.5 en promedio. Cuando se juega con una unidad para cada 50 apuestas, se

espera recibir 75 unidades del bookie de regreso, es decir 25 unidades de ganancia, un 50 % de retorno en promedio. Como el retorno esperado negativo no da un intervalo de confianza del 90 % no se esperan pérdidas con una estrategia de apuestas para  $\tau = 0.4$ .

### 3.2. Reserva un poco para después

En esta sección se describe el criterio de Kelly [16] y se muestra que los estudios propuestos por Vancura [46] sugieren el uso de un sistema de reservas en la estrategia de apuestas. Al garantizar una cantidad de dinero disponible para las siguientes jornadas, se permite al usuario seguir apostando durante todas las jornadas garantizando su oportunidad de generar ganancias o recuperar pérdidas.

Browne [2] llama al criterio de Kelly [16] como la estrategia óptima de crecimiento, o como la estrategia de utilidad logarítmica. Menciona que es óptima cuando en una variedad de circunstancias la ganancia crece multiplicativamente, únicamente. Hace énfasis que la mayoría de las propiedades de optimalidad son asintóticas en naturaleza (hacia un horizonte infinito), en su trabajo analiza las propiedades de la estrategia de Kelly a corto plazo.

Supóngase que un jugador se enfrenta a un oponente infinitamente rico en apuestas sobre lanzamientos de monedas independientes. El oponente rico siempre igualará la misma cantidad de dinero apostada por el jugador en cada uno de las apuestas. Además, supóngase también que en cada lanzamiento la probabilidad de ganar es  $p > \frac{1}{2}$  y que la probabilidad de perder es  $q = 1 - p$ . El capital inicial del jugador es  $X_0$ . El objetivo de este ejercicio es el de maximizar el valor esperado  $E(X_n)$  después de  $n$  lanzamientos. Por lo que la duda que se busca responder es la siguiente: ¿Cuánto se debe apostar,  $B_k$ , en el  $k$ -ésimo lanzamiento?

Sea  $T_k = 1$ , si se gana el  $k$ -ésimo lanzamiento y  $T_k = -1$  si se pierde. Entonces  $x_k = X_{k-1} + T_k B_k$  para  $k = 1, 2, 3, \dots$ , y  $X_n = X_0 + \sum_{k=1}^n T_k B_k$ . Por lo que se puede expresar el valor esperado de la siguiente

manera:

$$E(X_n) = X_0 + \sum_{k=1}^n E(B_k T_k) = X_0 + \sum_{k=1}^n (p - p)E(B_k)$$

La intuición sugiere que como el juego tiene un valor esperado positivo ya que  $p - q > 0$ , entonces para maximizar  $E(X_n)$  se quisiera maximizar  $E(B_k)$  en cada lanzamiento, además para maximizar la ganancia esperada se debería apostar todo el dinero en cada lanzamiento. Sin embargo, la probabilidad de perderlo todo es equivalente a  $1 - p^n$  y como  $p < 1$ ,  $\lim_n [1 - p^n] = 1$  es decir, la ruina es segura. Por lo que, la intrépida decisión de apostar para maximizar la ganancia esperada es indeseable.

De la misma manera, si se juega para minimizar la probabilidad de llegar a la ruina debe de recordarse que la fórmula de la ruina del jugador A muestra que se minimiza la ruina jugando la cantidad mínima posible en cada apuesta, lo que a su vez minimiza la ganancia esperada. Al fin y al cabo ya se sabe como termina esta estrategia, por lo que la apuesta tímida no es la respuesta.

El criterio de Kelly [16] propone una estrategia de optimización asintótica para este problema. Supóngase que se apuesta la cantidad  $B_i = fX_i - 1$ , con  $0 \leq f \leq 1$ . Sean  $S$  y  $F$  la cantidad de éxitos y fallos (respectivamente), entonces se tiene que en  $n$  lanzamientos el capitán del jugador sería el siguiente:

$$X_n = X_0(1 + f)^S(1 - f)^F$$

con  $S + F = n$  y  $0 < f < 1$ ,  $p(X_n \leq 0) = 0$ . Ahora también, considérese que el jugador está arruinado cuando la cantidad de dinero después de la apuesta  $n$  del jugador es menor un entero positivo pequeño, i.e.  $X_n < \varepsilon$ .

Nótese que dado que

$$e^{n \log \frac{X_n}{X_0} 1/n} = \frac{X_n}{X_0}$$

se tiene que la cantidad

$$G_n(f) = \log \left( \frac{X_n}{X_0} \right)^{1/n} = \frac{S}{n} \log(1+f) + \frac{F}{n} \log(1-f)$$

mide la tasa de incremento exponencial por lanzamiento. Kelly sugiere maximizar el valor esperado del coeficiente de la tasa de crecimiento  $g(f)$  donde

$$\begin{aligned} g(f) &= E \left( \log \left( \frac{X_n}{X_0} \right)^{1/n} \right) = E \left( \frac{S}{n} \log(1+f) + \frac{F}{n} \log(1-f) \right) \\ &= p \log(1+f) + q \log(1-f) \end{aligned}$$

Nótese que  $g(f) = (1/n)E(\log X_n) - (1/n) \log X_0$ , así que para  $n$  fijo, maximizar  $g(f)$  es lo mismo que maximizar  $E \log(X_n)$ . Véase que:

$$g'(f) = \frac{p}{1+f} - \frac{q}{1-f} = \frac{p-q-f}{(1+f)(1-f)} = 0$$

cuando  $f = f^* = p - q$

Ahora como

$$g''(f) = -p/(1+f)^2 - q/(1-f)^2 < 0$$

entonces  $g'(f)$  es monótona estrictamente decreciente en  $[0, 1)$ . También,  $g'(0) = p - q > 0$  y  $\lim_{f \rightarrow 1} g'(f) = -\infty$ . Por lo tanto, dado que  $g'(f)$  y  $g(f)$  son continuas,  $g$  tiene un único máximo cuando  $f = f^*$ , donde  $g(f^*) = p \log(p) + q \log(q) + \log(2) > 0$ . Más aun,  $g(0) = 0$  y el  $\lim_{f \rightarrow 1} g(f) = -\infty$  por lo que existe un único número  $f_c > 0$  en  $0 < f^* < f_c < 1$  tal que  $g(f_c) = 0$ . Véase la figura 3.1 para una mejor referencia de la forma de la función  $g(f)$ .

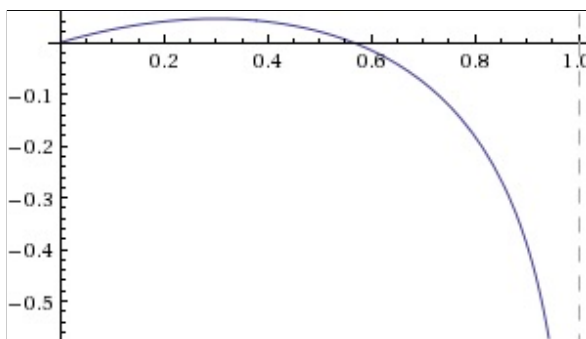


Figura 3.1: La función  $g(f)$  siempre encuentra un máximo y después tiende a  $-\infty$

## Teorema

1. Si  $g(f) > 0$ , entonces  $\lim_{n \rightarrow \infty} X_n = \infty$  casi seguro. Es decir, para cada  $M$  se tiene que  $p(\liminf_{n \rightarrow \infty} X_n > M) = 1$ .
2. Si  $g(f) < 0$ , entonces  $\lim_{n \rightarrow \infty} X_n = 0$  casi seguro. Es decir, para cada  $\varepsilon > 0$  se tiene que  $p(\limsup_{n \rightarrow \infty} X_n < \varepsilon) = 1$ .
3. Si  $g(f) = 0$ , entonces  $\limsup_{n \rightarrow \infty} X_n = \infty$  casi seguro y  $\liminf_{n \rightarrow \infty} X_n = 0$  casi seguro.
4. Dada una estrategia  $\phi^*$  que maximiza  $E(\log(X_n))$  y cualquier otra estrategia  $\phi$  “esencialmente diferente” (No necesariamente una estrategia de apuestas fraccionada fija), entonces  $\lim_{n \rightarrow \infty} X_n(\phi^*)/X_n(\phi) = \infty$  casi seguro.
5. El tiempo esperado para que la cantidad de capital actual  $X_n$  alcance alguna meta fija  $C$  es asintóticamente menor con una estrategia que maximice  $E(\log(X_n))$ .
6. Supóngase que el retorno en una unidad apostada en el  $i$ -ésimo ensayo es la variable aleatoria  $U_i$ , además supóngase que la proba-

*bilidad de éxito es  $p_i$  donde  $1/2 < p_i < 1$ . Entonces,  $E(\log(X_n))$  se maximiza escogiendo en cada ensayo la fracción  $f_i^* = p_i - q_i$  que maximiza  $E(\log(1 + f_i U_i))$ .*

De la parte 1 de este teorema se muestra que excepto para un número finito de términos, el capital del jugador  $X_n$  sobrepasará una cota fija  $M$  cuando  $f$  se escoge dentro del intervalo  $(0, f_c)$ . Pero, si  $f > f_c$ , entonces la parte 2 muestra que la ruina será casi segura. La parte 3 demuestra que si  $f = f_c$ , entonces  $X_n$  (casi seguro) oscilará aleatoriamente entre 0 y  $\infty$ . Las partes 4 y 5 establecen que la estrategia de Kelly de maximizar  $E(\log(X_n))$  es asintóticamente óptima ya que si se toma una estrategia “esencialmente diferente” como aquella tal que la diferencia,  $E(\log(X_n^*)) - E(\log(X_n))$ , entre la estrategia de Kelly y cualquier otra estrategia crece más rápido que la desviación estándar de  $\log(X_n^*) - \log(X_n)$ , asegurando  $p(\log(X_n^*) - \log(X_n) > 0) \rightarrow 1$ . Finalmente, la parte 6 establece la validez de utilizar el método de Kelly para escoger  $f_i^*$  en cada uno de los ensayos (aun cuando las probabilidades cambian de un ensayo a otro) con el fin de maximizar  $E(\log(X_n))$ .

[16] [46]

### 3.3. Proceso para la selección de apuestas

El proceso de la recomendación de apuestas se puede dividir en cinco pasos principales:

1. **Determinación de apuestas redituables.** De todas las apuestas de las cinco ligas que haya en la jornada a recomendar, se escogen únicamente aquellas que sean redituables a largo plazo. Independientemente del nivel de riesgo de cada cliente, se generarán recomendaciones únicamente de apuestas que, al analizarlas, tengan ganancias esperadas.
2. **Selección de apuestas de acuerdo al riesgo del cliente.** Las estimaciones de cada partido tienen un intervalo de confianza. Con



base en este intervalo y el perfil de riesgo del usuario se seleccionan las apuestas.

3. **Cálculo de la proporción del dinero a invertir por apuesta.**

Al conocer la selección de las apuestas, se calcula la proporción de dinero a invertir para cada apuesta, este proceso sigue siendo dependiente de la confianza de la estimación del partido y de la adversidad al riesgo del cliente.

4. **Determinación del nivel de reservas.** En este paso se cuenta con un conjunto seleccionado de apuestas y la proporción de dinero que se debe invertir para cada una de ellas. Acorde a su encuesta, se calcula el nivel óptimo de reservas que le permitan al cliente seguir apostando a pesar de que tenga una mala jornada y pierda todas las apuestas.

5. **Recomendación.** Todos estos elementos constituyen en su totalidad la recomendación de la semana que se presenta en el portal al cliente.

Para definir el escenario completo es necesario definir las variables que se utilizan en el proceso, así como las funciones de utilidad que modelan los perfiles de riesgo de los usuarios. Después de estas dos sub-secciones se retomará paso a paso la explicación del proceso de recomendación de apuestas.

### 3.3.1. Variables necesarias

En este apartado de la tesis se explica cómo se determinan los partidos que se recomiendan a los clientes. Las variables que se necesitan utilizar son las siguientes:

1. **Probabilidades.** Ya en apartados anteriores se habló de las probabilidades estimadas de que ocurran los tres diferentes resultados de cada encuentro de la jornada (Victoria del local, del visitante y

empate). Estas probabilidades son alimentadas a través del portal administrativo.

2. **Momios.** Los momios que generan las casas de apuestas para cada partido. En la mayoría de los escenarios el usuario escoge usar todos los momios de todos los bookies, en estos casos se tomarán los momios que ofrezcan mejores retornos para cada partido. Estos momios se toman de la información pública de las casas de apuestas.
3. **Variables de riesgo.** Se tienen tres elementos principales que generan la noción de riesgo en el sistema:
  - **Función de utilidad en selección.** Determina los partidos redituables de la jornada.
  - **Función de utilidad del dinero.** Dicta la proporción de dinero a apostar en cada partido.
  - **Variable de reserva.** Marca la cantidad de dinero que el cliente debería guardar para siguientes apuestas. Estas variables dependen de la encuesta que se le presenta al cliente al crear su cuenta.
4. **Variable de evolución de ingreso del cliente.** Se define como la cantidad de dinero que tiene el cliente en esta jornada entre la máxima cantidad de dinero que ha tenido. Esta variable genera una noción, en caso de que existan, de la magnitud de las pérdidas que ha tenido el cliente.

### 3.3.2. Funciones de utilidad

Para modelar el comportamiento de un individuo ante el riesgo, se utilizan funciones de utilidad. Estas funciones representan la utilidad (o beneficio) que un individuo puede obtener por una apuesta. Las funciones más arriesgadas le dan valores mayores a las apuestas que

que ofrecen las mejores ganancias y las funciones más conservadoras favorecen aquellas apuestas con mayores probabilidades. En particular, se podría decir de manera intuitiva que el valor obtenido por una función de utilidad para una apuesta en particular es la cantidad de unidades de dinero que el individuo estaría dispuesto a invertir en esa apuesta. El siguiente listado presenta las funciones que se utilizan actualmente en el sistema.

**1. Polinomial - lineal:**

$$f_1(\alpha, p, m) = \left( \frac{p}{1-p} \right)^{\frac{1}{1-\alpha}} (m-1)^{\frac{\alpha}{1-\alpha}}$$

Parámetro  $\alpha$  que va de cero a uno. Es una de las familias más volátiles en cuanto al riesgo, para valores cercanos a uno del parámetro tiende a ser muy arriesgada, casi sin diversificar, y para valores cercanos a cero tiende a ser muy conservadora y suele diversificar mucho más en las apuestas.

**2. Polinomial - cuadrática:**

$$f_2(\alpha, p, m) = \left( \frac{p}{1-p} \right)^{\frac{1}{2-\alpha}} (m-1)^{\frac{\alpha}{2-\alpha}}$$

Esta familia de funciones es muy parecida a la anterior, también tiene un parámetro  $\alpha$  entre cero y uno, sólo que en este caso no es tan extremista como en el caso anterior.

**3. Exponencial - lineal:**

$$f_3(\alpha, p, m) = \left( \frac{1}{m-1} \right) \ln \left( \frac{\alpha p(m-1)}{1-p} \right)$$

Esta familia de funciones tiene un parámetro  $\alpha$  que es mayor a uno. Es una familia de funciones conservadoras que suelen proteger muy bien adelantándose a semanas malas, a costo que en semanas buenas no logran tan buenos beneficios.

#### 4. Lineal - exponencial:

$$f_4(\alpha, p, m) = \ln \left( \frac{p(m-1)}{1-p} \right) + \ln(\alpha)$$

Esta familia de funciones tiene un parámetro  $\alpha$  que es mayor a uno. Es una familia de funciones que le gusta fuertemente diversificar, aparte tiende a favorecer un poco aquellas apuestas más riesgosas.

#### 5. Logarítmica - polinomial:

$$f_5(\alpha, p) = \left( \frac{p}{1-p} \right)^{\frac{1}{\alpha}}$$

Parámetro  $\alpha$  mayor a uno. Es una familia de funciones muy conservadora ya que toma valores sin considerar los momios del mercado, al tomar en cuenta solamente las probabilidades de los partidos tiende a ser más estable en el tiempo.

#### 6. Tangente - lineal:

$$f_6(\alpha, p, m) = \frac{1}{m-1} \sqrt{\frac{pm - (1-\alpha)p - \alpha}{1-p}}$$

Parámetro  $\alpha$  entre cero y uno. Esta familia es la más conservadora de la lista, busca tener las menores pérdidas posibles y protege las inversiones cada semana.

Estas seis funciones son las que utiliza Egobets para modelar la adversidad al riesgo de sus clientes. Cuando ellos llenan la encuesta del perfil, realmente están seleccionando una de estas funciones.

### 3.3.3. Paso 1 - Determinación de apuestas redituables

Para seguir la explicación de los pasos del proceso se tomará un ejemplo práctico y las transformaciones que sufre durante el proceso.

Al empezar las recomendaciones de una jornada se tiene un conjunto de datos muy parecido al que se da en el ejemplo. Tómesese la tabla 3.1 como base del ejemplo. En esta tabla se presentan momios de alguna casa de apuesta (usualmente los mejores del mercado) para diez partidos distintos, cada uno cuenta con los momios para los tres tipos de apuestas: victoria del Local ( $m_L$ ), empate ( $m_E$ ) y victoria del visitante ( $m_V$ ). Además se muestran las probabilidades estimadas por Egobets de los tres posibles resultados del encuentro: victoria del local ( $\hat{p}_L$ ), empate ( $\hat{p}_E$ ) y victoria del visitante ( $\hat{p}_V$ ).

Partido	$m_L$	$m_E$	$m_V$	$\hat{p}_L$	$\hat{p}_E$	$\hat{p}_V$
A	2.6	3.2	2.75	0.31	0.23	0.46
B	2.1	3.3	3.5	0.37	0.3	0.32
C	1.11	8.5	21	0.84	0.1	0.04
D	4.5	3.75	1.72	0.29	0.27	0.45
E	2	3.4	3.75	0.37	0.31	0.31
F	1.53	4	6	0.61	0.17	0.21
G	2.1	3.25	3.6	0.51	0.24	0.25
H	1.9	3.4	4	0.42	0.29	0.29
I	1.8	3.5	4.5	0.46	0.29	0.25
J	1.9	3.4	4	0.5	0.26	0.23

Tabla 3.1: Diez partidos con sus respectivos momios y las probabilidades estimadas por Egobets

Para concretar el ejemplo, se fijarán las siguientes funciones y variables.

- **Función de selección.** Tómesese la *Polinomial - lineal*  $f_1$  con pa-

rámetro  $\alpha = 0.4$ .

- **Función de utilidad del dinero.** Úsese la *Exponencial - lineal* con parámetro  $\alpha = 3$
- **Variable de reserva.** Será  $v_R = 15$ .
- **Variable de ingreso.** Será  $v_I = 0.80$ .

Ahora para proceder al primer paso, defínase el valor esperado de una apuesta como la multiplicación de la probabilidad de que suceda el resultado por el momio que se tiene para esa apuesta. Con el ejemplo bien definido se procede a filtrar las apuestas con el primer paso. De manera intuitiva, se podría enunciar que el valor esperado de una apuesta corresponde a la cantidad de unidades de dinero que se obtienen por cada unidad de dinero invertida<sup>5</sup>

Entonces, el primer paso consiste en utilizar para las recomendaciones, sólo las apuestas con ganancias a largo plazo; las otras apuestas tendrán un valor esperado no positivo, es decir, no garantizan más que pérdidas a largo plazo. Más aun, se usarán solo las apuestas que garanticen un valor esperado mayor a 1.025, esta restricción tiene dos ventajas, la primera es que las probabilidades que se usan son estimaciones por lo que ese 2.5 % funciona como protección para no caer en valores esperados negativos y la segunda es que si el rendimiento es tan pequeño toma mucho tiempo para un cliente ir acumulando ganancias.

Obsérvese en la tabla 3.2 como en negritas se han señalado las posibles apuestas a realizar. De las treinta posibles apuestas, ahora sólo quedan 9 redituables. Hay dos partidos el 'C' y el 'J' que no ameritan apuesta alguna. Y el partido 'E', por ejemplo, tiene dos posibles apuestas redituables.

---

<sup>5</sup>Recuérdese que esta cantidad incluye la cantidad de dinero invertida en la apuesta. También recuérdese que la ganancia solamente se obtiene si se gana la apuesta.

Partido	$m_L$	$m_E$	$m_V$	$\hat{p}_L$	$\hat{p}_E$	$\hat{p}_V$	$m_L \cdot \hat{p}_V$	$m_E \cdot \hat{p}_E$	$m_V \cdot \hat{p}_V$	Apuesta
A	2.6	3.2	2.75	0.31	0.23	0.46	0.81	0.74	<b>1.27</b>	Visitante
B	2.1	3.3	3.5	0.37	0.3	0.32	0.78	0.99	<b>1.12</b>	Visitante
C	1.11	8.5	21	0.84	0.1	0.04	0.93	0.85	0.84	Ninguna
D	4.5	3.75	1.72	0.29	0.27	0.45	<b>1.31</b>	1.01	0.77	Local
E	2	3.4	3.75	0.37	0.31	0.31	0.74	<b>1.05</b>	<b>1.16</b>	Empate/Visitante
F	1.53	4	6	0.61	0.17	0.21	0.93	0.68	<b>1.26</b>	Visitante
G	2.1	3.25	3.6	0.51	0.24	0.25	<b>1.07</b>	0.78	0.90	Local
H	1.9	3.4	4	0.42	0.29	0.29	0.80	0.99	<b>1.16</b>	Visitante
I	1.8	3.5	4.5	0.46	0.29	0.25	0.83	1.02	<b>1.13</b>	Visitante
J	1.9	3.4	4	0.5	0.26	0.23	0.95	0.88	0.92	Ninguna

Tabla 3.2: Escogiendo apuestas que vale la pena realizar

### 3.3.4. Paso 2 - Selección de apuestas

Para determinar el portafolio de apuestas basta con calcular el valor de la función de utilidad de selección y se deben tomar aquellas “ $n$ ” apuestas que tengan los mayores valores (Donde “ $n$ ” es el número de apuestas a recomendar). Ver tabla 3.3

Partido	Apuesta	Momio	$\hat{p}$	$f_1(\alpha, p, m)$
A	Visitante	2.75	0.46	<b>1.07</b>
B	Visitante	3.5	0.32	0.53
D	Local	4.5	0.29	0.49
E	Empate	3.4	0.31	0.48
E	Visitante	3.75	0.31	0.53
F	Visitante	6	0.21	0.31
G	Local	2.1	0.51	<b>1.12</b>
H	Visitante	4	0.29	0.48
I	Visitante	4.5	0.25	0.36

Tabla 3.3: Escogiendo apuestas que vale la pena realizar

### 3.3.5. Paso 3 - Calcular cuanto dinero a cada apuesta

Una vez determinado el portafolio de apuestas hace falta determinar la cantidad de dinero que se invertirá en éste. Esta sección y la próxima resolverán tal problema. Empezamos por determinar cuál es la proporción de dinero que se invertirá en cada apuesta. La solución es sencilla, ya determinamos en secciones anteriores que el valor definido por una función de utilidad puede ser comparado con la cantidad de unidades de dinero que una persona estaría dispuesta a invertir en una apuesta en particular. El único problema con este enfoque es que tales cantidades no se encuentran en ninguna escala (es decir, dólares, pesos, etc). Lo que hacemos es lo siguiente: para cada apuesta calculamos el valor de su función de utilidad del dinero y después lo dividimos entre la suma de todos estos valores para todas las apuestas que se van a tomar, de esta forma se tiene los números como porcentajes. Regresando al ejemplo: (En este caso la función de utilidad del dinero era Exponencial-Lineal



con parámetro 3) Ver tabla 3.4

Partido	Apuesta	Momio	$\hat{p}$	$f_3(\alpha, \mathbf{p}, \mathbf{m})$
A	Visitante	2.75	0.46	0.23
G	Local	2.1	0.51	0.35

Tabla 3.4: Calculando la proporción del dinero que se debe invertir en estas apuestas

Los valores suman a 0.59, cuando se divide cada valor por 0.59 resulta: De la cantidad a apostar, se debe de apostar 40 % en el partido A a visitante y 60 % en el partido G a local.

### 3.3.6. Paso 4 - Determinar nivel de reservas

Determinar el nivel de reservas es un algoritmo que depende del historial de ganancias del cliente así como de las circunstancias de las apuestas que se le presentan al cliente. A continuación se explica paso a paso como determinar el nivel de reservas.

#### 1. Calcular la cantidad de partidos a los que se apuesta.

Defínase  $j, k$  como la cantidad de partidos a apostar y el número de apuestas a tomar, respectivamente. En este paso bastará con calcular  $j$ . Sin embargo, es importante notar que hay veces en las que puede haber más de una apuesta sugerida para un mismo partido, i.e.  $k > j$ . En estos casos se deberá ser cuidadoso de no contar doble los partidos que ya tengan apuestas.

#### 2. Calcular el valor esperado y varianza del portafolio de apuesta.

Sean  $m_i$ ,  $p_i$  y  $prop_i$ ; el momio, la probabilidad y la proporción de dinero a apostar del partido  $i$  respectivamente. Entonces el valor

esperado de la apuesta sería:

$$\mu = \sum_{i=1}^k (prop_i)(m_i)(p_i)$$

Y el riesgo o varianza del portafolio de apuestas se calcula así:

$$\sigma = \sqrt{\sum_{i=1}^k (prop_i)^2(m_i)^2(p_i)(1-p_i)}$$

### 3. Actualizar el riesgo del portafolio.

Sea  $\rho$  el parámetro de reserva un número entero entre 10 y 25 dependiendo de la encuesta del cliente. Se actualiza el riesgo del portafolio al multiplicarlo por un escalar proveniente de la raíz cuadrada del cociente de la cantidad de partidos  $j$  entre el parámetro de reserva  $\rho$ .

$$\sigma = \left( \sqrt{\frac{j}{\rho}} \right) \sigma$$

### 4. Calcular probabilidad de riesgo.

Sea  $X$  la variable de evolución del ingreso del cliente (Esta variable es completamente dependiente de las historia de perdidas y ganancias del cliente). Y sea  $p_\sigma$  la probabilidad de riesgo y se calcula de la siguiente manera:

$$p_\sigma = \frac{1}{1.749615}(I_1 + I_2) - 1.06 + X$$

Donde  $I_1$  es

$$I_1 = 0.2925 + 1.3772\mu - 1.127\rho$$

e  $I_2$  equivale a

$$I_2 = \frac{3.11}{\mu}(1 + 0.792\sigma - \mu)((-0.567345 + 1.3772\mu - 1.17\sigma)\mu X)^{0.123}$$

### 5. Calcular variable proxy.

Sea *proxy* la variable auxiliar que tome en cuenta la probabilidad de riesgo en el cálculo.

$$proxy = 0.2925 + 1.3772\mu - 1.127\sigma - 0.9975p_\sigma$$

### 6. Calcular el nivel de reserva.

Finalmente, defínase  $c_A$  cómo el porcentaje a apostar del ingreso total. El nivel de reserva será uno menos esta proporción.

$$c_A = \begin{cases} 0.05 & \text{si } \frac{j(proxy)}{\rho} < 0.05 \\ 1 & \text{si } \frac{j(proxy)}{\rho} > 1 \\ \frac{j(proxy)}{\rho} & \text{en otro caso} \end{cases}$$

$C_A$  representa, en porcentaje, la cantidad de dinero que se le recomendará al cliente que apueste en esa semana para el portafolio de apuestas en cuestión.

### Calculando el nivel de reserva en el ejemplo.

Primero que nada es importante recordar cuanto valían las distintas variables y establecer los valores necesarios para el ejemplo.

- $k = 2$  (Cantidad apuestas a realizar)
- $\rho = 15$  (Parámetro de reserva en función de las preferencias del cliente)
- $X = 0.80$  (Evolución del ingreso del cliente)

Ahora se procederá a calcular el nivel de reservas.

- a) Se obtiene  $j = 2$  (Cantidad de partidos en los que se apuesta)
- b) Se calcula el valor esperado

$$\mu = (0.4)(2.75)(0.46) + (0.6)(2.1)(0.51) = 1.1486$$

y el riesgo del portafolio

$$\sigma = \sqrt{(0.4)^2(2.75)^2(0.46)(0.54) + (0.6)^2(2.1)^2(0.51)(0.49)} = 0.835$$

- c) Se actualiza el riesgo  $\sigma = \left(\sqrt{\frac{2}{15}}\right) 0.835 = 0.3$   
d) Se calcula la probabilidad de riesgo

$$I_1 = 0.2925 + 1.3772(1.1486) - 1.127(0.3) = 1.53$$

$$I_2 = \frac{3.11}{1.1486}(1 + 0.792(0.3) - \mu)((-0.567345 + 1.3772(1.1486) - 1.17(0.3))(1.1486)(0.8))^{0.123} = 0.2272$$

Entonces

$$p_\sigma = \frac{1}{1.749615}(1.53 + 0.2272) - 1.06 + 0.8 = 0.74$$

- e) Se calcula la variable auxiliar  $proxy = 0.2925 + 1.3772(1.1486) - 1.127(0.3) - 0.9975(0.74) = 0.81$   
f) La cantidad a apostar es entonces es igual a  $c_A = \frac{2(0.81)}{15} = 0.108$  Por lo tanto, con esta adversidad al riesgo y con base en el portafolio de apuestas de la tabla 3.4.

**El nivel de reserva para esta recomendación es del 89.2 %.**

### 3.3.7. Paso 5 - Presentando la recomendación

Ahora, con todos los pasos cumplidos lo único restante es poner las cifras juntas. Se toma la proporción calculada en 3.3.5 y se multiplica por la cantidad a apostar obtenida en 3.3.6. Más aún, si el usuario proporciona la cantidad de dinero que tiene disponible, las recomendaciones se pueden expresar en moneda.

Finalmente si para el ejemplo planteado en 3.3.3 se conoce que el usuario tiene \$10'000.00 para apostar, la recomendación de apuesta quedaría de la siguiente manera:

1. Apuesta \$432 pesos (4.32 %) en el partido A a favor del equipo visitante.
2. Apuesta \$648 pesos (8.48 %) en el partido G a favor del equipo local.



## Capítulo 4

# Sistema de Egobets.com

Egobets.com proporciona al cliente los servicios de asesoría de apuestas personalizada a través de un portal Web usable, práctico y profesional. En este capítulo se presenta el sistema desarrollado con los fundamentos teóricos descritos en los capítulos anteriores, una verdadera aplicación computacional de las matemáticas

### 4.1. Diseño y arquitectura

Egobets.com es un sistema con arquitectura cliente servidor montado sobre una máquina virtual con sistema operativo Linux en la nube de “Amazon Web Services (AWS)”. En esta sección se presentan los diagramas que describen la arquitectura del sistema, se muestran las ventajas de utilizar el computo en la nube, se presentan las tecnologías más relevantes involucradas en el sistema y también se expone el patrón de diseño Modelo Vista Controlador junto con su respectiva representación gráfica de base de datos.

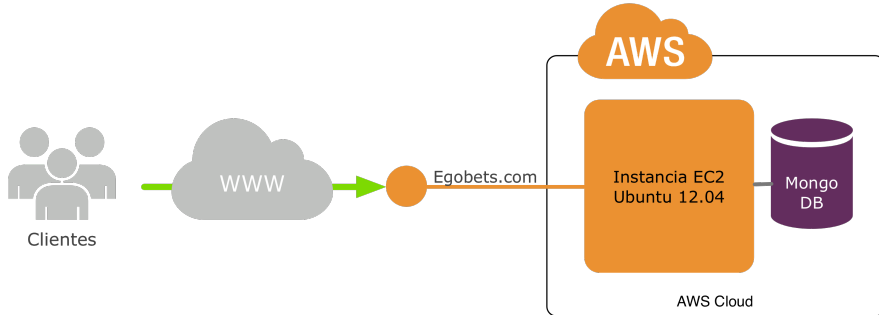


Figura 4.1: Arquitectura Cliente Servidor sobre Nube de AWS

#### 4.1.1. Ventajas de correr Egobets en la nube

El sistema usa la nube de AWS para ofrecer sus servicio a los usuarios (Véase la imagen 4.1), más aún se puede decir que el software corre en un esquema tipo “SaaS”<sup>1</sup>, esto implica que el usuario simplemente ingresa a su cuenta en un navegador de internet y puede ver las asesorías para sus apuestas. Del artículo “Cómputo en Nube: Ventajas y Desventajas” de Martínez y Gutiérrez [24] se retoman las siguientes ventajas de este paradigma:

- **Costos.** Podría ser la ventaja más atractiva que presenta el cómputo en la nube, y si no lo es, al menos es la más evidente de todas las que ofrece esta tecnología. Al dejar la responsabilidad de la implementación de la infraestructura al proveedor, el cliente no tiene que preocuparse por comprar equipos de cómputo, capacitar

---

<sup>1</sup>Software as a Service. “Es el más conocido de los niveles de cómputo en la nube. El SaaS es un modelo de distribución de software que proporciona a los clientes el acceso a éste a través de la red (generalmente Internet). De esta forma, ellos no tienen que preocuparse de la configuración, implementación o mantenimiento de las aplicaciones, ya que todas estas labores se vuelven responsabilidad del proveedor. Las aplicaciones distribuidas a través de un modelo de Software como Servicio pueden llegar a cualquier empresa sin importar su tamaño o ubicación geográfica.” [24].



personal para la configuración y mantenimiento de éstos, y en algunos casos, por el desarrollo del software. Además el usuario de estos servicios únicamente paga por los recursos que utiliza, permitiéndole diseñar un plan de pago normalmente a partir del tiempo en que éste se utiliza (memoria, procesamiento, almacenamiento). Para Egobets, esta cualidad es vital, ya que en el comienzo después de haber implantado el software, la cantidad de usuarios es mínima y los ingresos también. Conforme crece la bolsa de clientes también irá creciendo la potencia del servidor.

- **Competitividad.** Al no tener que adquirir equipos costosos, las pequeñas empresas pueden tener acceso a las más nuevas tecnologías a precios a su alcance pagando únicamente por consumo. De este modo las organizaciones de cualquier tipo podrían competir en igualdad de condiciones en áreas de TI con empresas de cualquier tamaño. La ventaja competitiva no está en aquel que tiene los recursos de cómputo sino en quien los emplea mejor. En particular a Egobets le permite utilizar este tipo de tecnología a la par de otros sistemas gigantes que tienen mucho más tiempo en el mercado y mucho mayores ingresos.
- **Disponibilidad.** El proveedor está obligado a garantizar que el servicio siempre esté disponible para el cliente. En este sentido, la virtualización juega un papel fundamental, ya que el proveedor puede hacer uso de esta tecnología para diseñar una infraestructura redundante que le permita ofrecer un servicio constante de acuerdo a las especificaciones del cliente. A manera anecdótica, en Egobets se tuvo un problema con uno de los discos duros de un servidor, bastó con crear una nueva máquina virtual de la imagen que ya se poseía y clonar el código necesario, en menos de cinco minutos el servicio estaba de vuelta en línea.
- **Abstracción de la parte técnica.** Como se mencionó al hablar de costos, el cómputo en la nube permite al cliente la posibilidad

de olvidarse de la implementación, configuración y mantenimiento de equipos; transfiriendo esta responsabilidad al proveedor del servicio. En Egobets, jamás se ha tenido que realizar ningún tipo de mantenimiento a ningún equipo de hardware de los servicios en la nube.

- **Acceso desde cualquier punto geográfico.** El uso de las aplicaciones diseñadas sobre el paradigma del cómputo en la nube puede ser accesible desde cualquier equipo de cómputo en el mundo que esté conectado a Internet. El acceso normalmente se hace desde un navegador web, lo que permite a la aplicación ser utilizada no únicamente desde una computadora de escritorio o una computadora portátil, sino que va más allá, permitiendo al usuario hacer uso de la aplicación incluso desde dispositivos móviles como smartphones. El sistema montado en Egobets.com, por ejemplo, no tiene ninguna restricción hacia ningún país del mundo.
- **Escalabilidad.** El cliente no tiene que preocuparse por actualizar el equipo de cómputo sobre el que se está corriendo la aplicación que utiliza, ni tampoco por la actualización de sistemas operativos o instalación de parches de seguridad, ya que es obligación del proveedor del servicio realizar este tipo de actualizaciones. Además, éstas son transparentes para el cliente, por lo que la aplicación debe de continuar disponible para el usuario en todo momento aún cuando se esté realizando el proceso de actualización del lado del proveedor. Las actualizaciones y nuevas funcionalidades son instaladas prácticamente de manera inmediata. Si se quisiera ampliar el poder de cómputo de los servidores de Egobets.com, bastaría con realizar un par de clicks y esperar unos minutos a que el servidor se auto-escale.

Es por estas razones que Egobets funciona tan bien en el paradigma del computo en la nube.

### 4.1.2. Servidor LNNP

Este acrónimo representa un sistema de infraestructura de internet<sup>2</sup> LNNP viene de:

- **Linux**, el sistema operativo.
- **Nginx**, el servidor Web.
- **NoSQL**, la Base de datos.
- **PHP**, el lenguaje de programación.

Una de las principales ventajas de utilizar esta configuración es que la mayoría del software utilizado corre bajo la “Licencia Pública General de GNU”, esta licencia de uso permite a los programadores que utilizan este software, ser capaces de ver el código fuente y en muchos casos les permiten modificarlo y compartirlo [21].

La máquina virtual de AWS corre Ubuntu LTS 12.04, una versión del sistema operativo creado por Linux Torvalds, **Linux** [43]. Algunas de las ventajas de usar Linux son:

- El costo de licencia es gratuito y su uso no tiene algún otro costo monetario.
- Hay miles de aplicaciones libres para hacer más robusto el servidor.
- Tener las aplicaciones en sus últimas versiones, bien configuradas y aplicar los parches de manera inteligente, garantizan que el servidor se encuentre seguro y sea funcional.

Encima del sistema operativo, se tiene **Nginx**. Un servidor Web de alto performance creado por Igor Sysoev para el sitio *www.rambler.ru*, el segundo sitio más grande de Rusia [34]. Nginx es capaz de servir más

---

<sup>2</sup>LNNP viene a retomar el acrónimo LAMP (Linux, Apache, MySQL y PHP) que será descrito en la sección ??.

peticiones por segundo y con menos recursos que sus competidores, gracias a su arquitectura. Grosso modo, consiste en un proceso maestro que delega a sus procesos *trabajadores* toda la carga de trabajo. Cada trabajador maneja varias solicitudes de manera asíncrona utilizando una funcionalidad especial del kernel de Linux<sup>3</sup>. Esto permite a Nginx, manejar un gran número de solicitudes simultáneas con muy poca sobrecarga.

A un lado, se tiene la base de datos No-SQL. A **MongoDB**, derivado de la palabra *humongous*, Membrey [26] le describe como una nueva especie de base de datos carente de conceptos de tablas, esquemas, SQL, o renglones. No tiene transacciones, joins, llaves externas, o cualquier otra de las características que suelen causar dolores de cabeza matutinos. En pocas palabras, MongoDB es una base de datos orientada a documentos, optimizada para ser veloz, escalable y fácil de integrar con cualquier lenguaje.

Finalmente el lenguaje en el que está programado el portal es **PHP**, cuyo acrónimo recursivo significa: *PHP Hypertext Preprocessor*. Según su página Web [31] es un lenguaje de scripting, el cual puede ser embebido dentro de páginas HTML. Gran parte de su sintaxis fue tomada de C, Java y Perl con un par de características específicas propias de PHP. El objetivo del lenguaje es permitir a desarrolladores Web escribir páginas generadas dinámicamente con agilidad. PHP está enfocado principalmente a la programación de scripts del lado del servidor, por lo que se puede hacer cualquier cosa como recopilar datos de formularios, generar páginas con contenidos dinámicos, o enviar y recibir cookies. Aunque PHP puede hacer mucho más. Ventajas:

- No depende del sistema operativo. Esto se debe a que corre al ser llamado por el servidor web y PHP corre sobre la mayoría de Servidores web, incluyendo Apache, IIS, lighttpd, nginx y muchos otros.

---

<sup>3</sup>Se puede leer más al respecto de como funciona internamente Nginx en Zhu [50]

- No se limita a genera HTML. Puede crear por ejemplo: imágenes, ficheros PDF e incluso películas Flash al vuelo. También puede generar archivos de texto y guardarlos en el sistema de archivo del sistema operativo.
- Permite la conexión casi cualquier Base de datos como: MySQL, SQLite, PostgreSQL, Mongo, Mssql, IBM DB2, entre muchas otras.
- PHP también cuenta con soporte para comunicarse con otros servicios usando protocolos tales como LDAP, IMAP, SNMP, NNTP, POP3, HTTP, COM (en Windows) y muchos otros.
- Existen muchas otras extensiones interesantes, las cuales están organizadas alfabéticamente y por categoría.
- Corre bajo la licencia Pública General de GNU. Por lo que se puede ser modificado y ser utilizado por cualquiera.’

Encima de PHP se utiliza una tecnología llamada CodeIgniter, esta tecnología tiene la propiedad de utilizar le patrón de diseño MVC. Esta propiedad separa la lógica del código en categorías y hace más sencillo el mantenimiento y desarrollo del mismo, en el siguiente apartado se detalla más esta arquitectura.

## 4.2. Patrón de diseño MVC

El portal público de Egobets.com está desarrollado en una de las arquitecturas más utilizadas en los sistemas de información, Modelo Vista Controlador (MVC) [47]. Esta arquitectura se basa en tres dimensiones principales: *Modelo* correspondiente a la información, *Vista* correspondiente a la presentación o interacción con el usuario y *Control* correspondiente al comportamiento. Como ya se mencionó el sistema utiliza esta arquitectura a través de *CodeIgniter*.

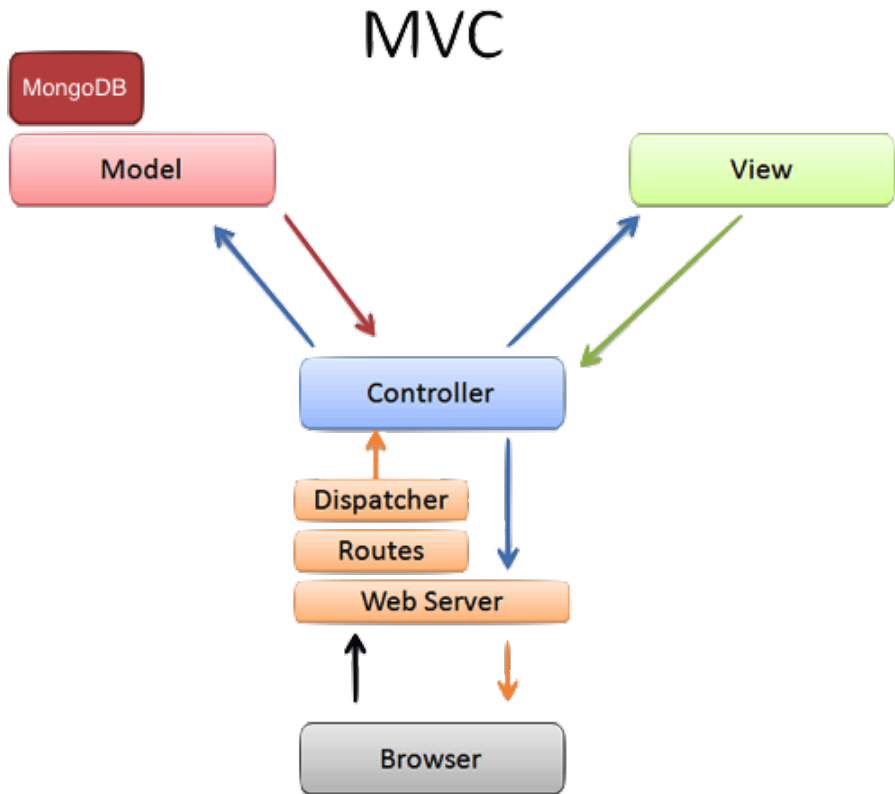


Figura 4.2: Patrón de diseño Modelo Vista Controlador (MVC)

CodeIgniter es un framework<sup>4</sup> de PHP que ahorra tiempo en le programación, robustece tu sistema y permite al programador alcanzar un

---

<sup>4</sup>Es una estructura de software compuesta de componentes personalizables e intercambiables para el desarrollo de una aplicación. En otras palabras, un framework se puede considerar como una aplicación genérica incompleta y configurable a la que se le puede añadir las últimas piezas para construir una aplicación concreta [45].

grado mayor de sofisticación en su código.

De su página Web [4] se pueden destacar las siguientes propiedades:

- Tamaño pequeño. CodeIgniter 2.2 tiene una descarga 2.2MB, incluyendo la guía del usuario.
- Documentación clara. La guía que se incluye cuenta con guía y tutoriales para empezar a trabajar de manera muy práctica.
- Compatibilidad con casi cualquier servicio de alojamiento. Sólo necesita PHP 5.1.6 y tiene soporte con las bases de datos más comunes incluido MySQL.
- Casi no necesita configuración. Todas las variables y opciones de configuración vienen predefinidas a los estándares convenidos en internet.

Por la naturaleza de la arquitectura MVC es necesario conocer la definición de la estructura de datos para después poder describir en mayor detalle los elementos del sistema. Ahora, hay que recordar que Egobets no utiliza una base de datos relacional por lo que se presenta un diagrama de clases para representar las colecciones que se tienen en la base de datos de Mongo DB. **Definición de la base de datos** La base de datos cuenta con seis colecciones principales (Véase la figura 4.3):

- Usuario. Se trata de los usuarios que se enrolan al sistema y utilizan los servicios de portal, como las asesorías de apuestas o las predicciones de los partidos.
- Administrador. Estos usuarios son los que tienen privilegios para ingresar al portal administrador y gestionar la información del sistema.
- Equipo. En esta colección se almacenan todos los equipos de las cinco ligas junto con la posición en la que se encuentra y la fortaleza con la que cuenta en ese momento.

- Partido. Se guardan los partidos de las cinco ligas junto con los marcadores estimados y después los reales. De igual manera se tienen los momios de las casas de apuestas para ese partido.
- Casa. En esta colección se almacenan los datos de las casas de apuestas de las cuales se recuperan sus momios para hacer los cálculos de las asesorías de apuestas.
- Resultado. Aquí se guarda la cantidad de dinero que obtiene o pierde un usuario en cada jornada.

Cómo se puede observar, hay conexiones entre algunas clases. Esto se debe a que las colecciones que tiene MongoDB permiten objetos complejos, de tal manera que un objeto puede contener un arreglo de otros objetos. Para ver un ejemplo de los datos almacenados en la base de datos véase el Apéndice C.

Con base en la estructura de datos definida, retómese el patrón de diseño MVC, la propiedad más interesantes de CodeIgniter. Este patrón fue descrito por el noruego Trygve Reenskaug en 1979, y facilita la descripción del sistema de Egobets de una manera más organizada [45] [47]. En los siguientes apartado se describirán los componentes de Egobets en estas tres dimensiones.



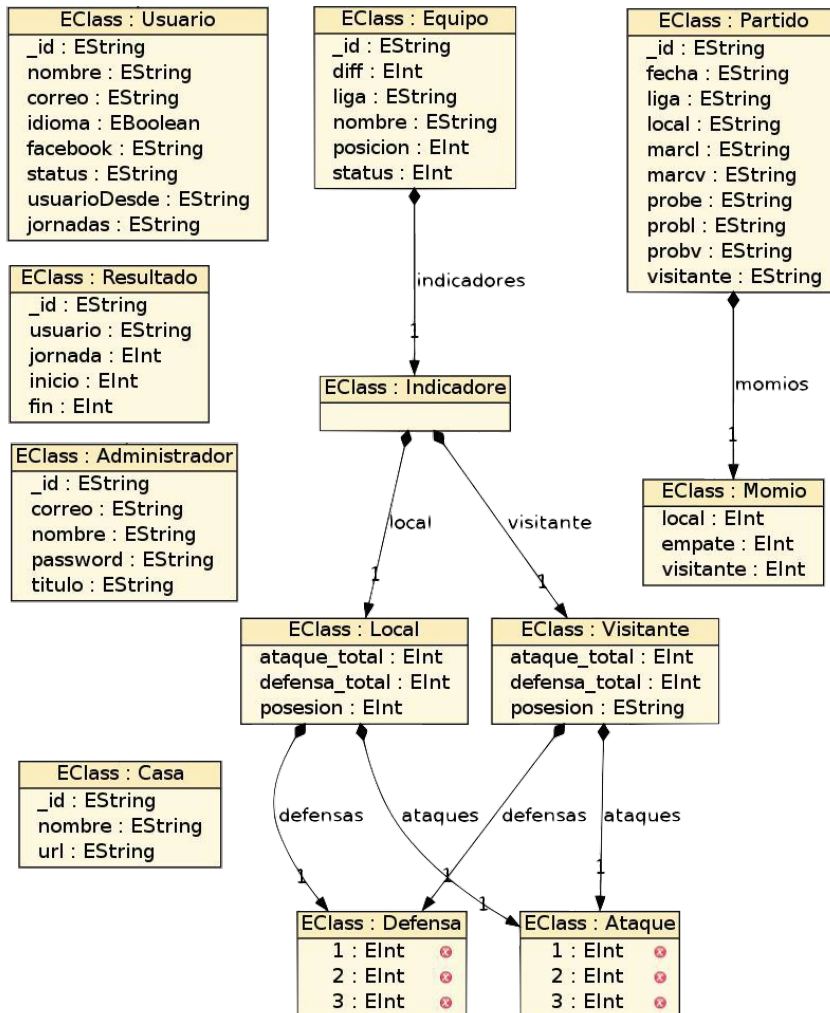


Figura 4.3: Representación UML de las colecciones de la base de datos de Egobets

### 4.2.1. Modelos

La información representa el dominio del problema y la base de datos se abstrae como un conjunto de objetos sobre los cuales recaen todas las acciones. En el modelo, los objetos básico definidos, reflejan las colecciones de la base de datos y a través del conector de base de datos, el modelo modifica los documentos en la base de datos conforme sea requerido. En particular, ya se dijo que los datos del sistema de Egobets son los definidos en la figura 4.3 y la definición de los objetos se pueden ver en el Apéndice C. Ahora, gracias a que PHP no es un lenguaje orientado a objetos<sup>5</sup>, los datos que tiene definidos en el modelo son completamente dependientes de la estructura de la base de datos, entonces los objetos utilizados por los modelos son exactamente los mismos definidos en la BD y extraídos con las consultas.

Como lo menciona Upton en su libro [45], la parte más esencial (y la más aburrida) es el escribir operaciones ABCD. Es necesario poder Agregar, Borrar, Cambiar y Desplegar la información de las colecciones de la bases de datos. Estas funciones, aunque conceptualmente fáciles de entender, son bastante complejas y escribirlas puede consumir mucho tiempo; sin embargo, son la base de un modelo bien definido, fácil de mantener y reutilizable.

En concreto, para el portal público de Egobets las siguientes son las funciones utilizadas en los modelos:

- *Usuarios*
  - Agregar - A través del modelo, se permite crear un nuevo *usuario* de la plataforma.
  - Cambiar - Se pueden cambiar los detalles del perfil de un *usuario*, así como las casas de apuestas a las que está suscri-

---

<sup>5</sup>PHP es un lenguaje que soporta objetos, y en particular PHP5 integra más este paradigma, pero no es un lenguaje orientado a objetos. Por ejemplo, miles de sus funciones principales no pertenecen a ningún objeto, ni se permite “castear” objetos de una clase a otra.

to, también es posible actualizar las jornadas ya pagadas del usuario.

- Desplegar - Se pueden obtener todos los datos de un usuario en función de su identificador o buscar algún texto en su perfil. También se pueden mostrar los amigos que tiene de Facebook en el sistema y las casas de apuestas a las que está suscrito en Egobets.

#### ■ *Resultados*

- Agregar - Se recibe la información de la cantidad de dinero ganada o perdida por cada usuario e inserta esta información en la colección de *Resultados*.
- Desplegar - Se obtienen los *resultados* de ganancias o pérdidas en función del usuario. También se genera la información base para hacer las gráficas de la evolución de dinero en el tiempo.

#### ■ *Transacciones*

- Agregar - Se permite iniciar e insertar una nueva *transacción* a la base de datos con un estatus de pendiente de pago.
- Cambiar - Se puede cambiar y asignar un estatus de pagada junto con una fecha de pago a una *transacción* a través de este modelo.
- Desplegar - Para algún usuario, se pueden obtener todas las *transacciones* realizadas.

#### ■ *Equipos*

- Desplegar - El modelo de *Equipos* despliega todos los *equipos* registrados en el sistema y estos pueden ser filtrados directamente por: el nombre equipo, el id del equipo o la liga a la que pertenecen. También se despliegan los equipos favoritos por cliente.

- *Partidos*

- Desplegar - Se muestran todos los *partidos* registrados en el sistema y se pueden filtrar por: el equipo que juega en ellos, el id del partido o la liga a la que pertenecen. También este modelo permite obtener las estrellas de cada partido, estas estrellas denotan la confiabilidad que tiene la predicción del resultado de ese partido.

#### 4.2.2. Vistas

Las vistas reflejan el estado del modelo y corresponden a las interfaces que se le presentan al usuario para el manejo de la información. Por lo general, pueden existir múltiples vistas sobre un mismo modelo, por ejemplo puede existir una vista que despliegue un listado de partidos por liga y al mismo tiempo existir una vista que muestre una gráfica con la cantidad de goles anotados por los equipos locales a través del tiempo. Estas vistas utilizan los mismos modelos, pero la información que despliega es completamente distinta y tiene fines diferentes.

También una vista puede implementar varios modelos, es decir, mostrar información compleja compuesta por varios elementos de la base de datos. Un ejemplo sería una vista que presente los datos de un partido junto con las estadísticas generales de cada equipo y una gráfica representativa de la cantidad de dinero gastada por los usuarios en las distintas apuestas del partido.

Las vistas que se tienen en el sistema se pueden enumerar de la siguiente manera:

- *Vistas relativas al perfil del usuario*

En este conjunto de vistas se le permite al usuario gestionar toda la información que tiene que ver con su perfil. Se tienen desde las pantallas para crear una cuenta, hasta las pantallas para editar la encuesta que define las recomendaciones que el usuario recibe. A continuación se listan estas vistas:

# CAMBIA TU PERFIL DE RIESGO

¿Cómo describirías tu manera de apostar?

Conservador Moderado Agresivo

¿Cuál de las siguientes respuestas describe mejor la razón por la que apuestas?

Quiero obtener ganancias a largo plazo Apuesto como pasatiempo Quiero obtener ganancias a corto plazo, aunque eso signifique perder mi dinero

¿Le gustaría que le ayudemos a llevar un sistema de "reserva"? ¿Cuál es la reserva?

Con este sistema de reserva, Egobets determinará que porcentaje de su dinero le conviene apostar, dependiendo de su nivel de riesgo

Sí No

¿Qué tanto riesgo toma al elegir tus apuestas?

Pocas ganancias, muy probable Muchas ganancias, poco probable

Figura 4.4: La encuesta define el perfil de riesgo del usuario

- Sign up. Permite al usuario crear una nueva cuenta para utilizar el sistema (Se puede utilizar Facebook para crear el perfil más rápidamente).
- Inicio de Sesión. A través de esta pantalla el usuario ingresa sus datos para entrar al sistema.
- Restablecer contraseña. Permite al usuario ingresar su correo electrónico para generar una nueva contraseña.
- Mi Perfil. En esta vista un usuario puede cambiar sus preferencias como el idioma, si requiere ser notificado cuando haya nuevas recomendaciones y seleccionar las casas de apuestas

que utiliza.

- Mis Amigos. A través de esta pantalla se pueden ingresar los correos electrónicos de otras personas que tienen cuenta en Egobets para después comparar resultados de sus ganancias.
- Encuesta. Como ya se mencionó el riesgo del usuario depende de esta pantalla. El usuario puede ajustarla cada vez que guste cambiar su estrategia de apuesta. Ver imagen 4.4

#### ■ *Vistas Estáticas*

Estas vistas no necesitan ser actualizadas casi nunca, es decir, su contenido es estático. Las páginas estáticas son:

- FAQ. Preguntas frecuentes que ayudan al usuario a comprender más a fondo el sistema (Ver imagen 4.5).

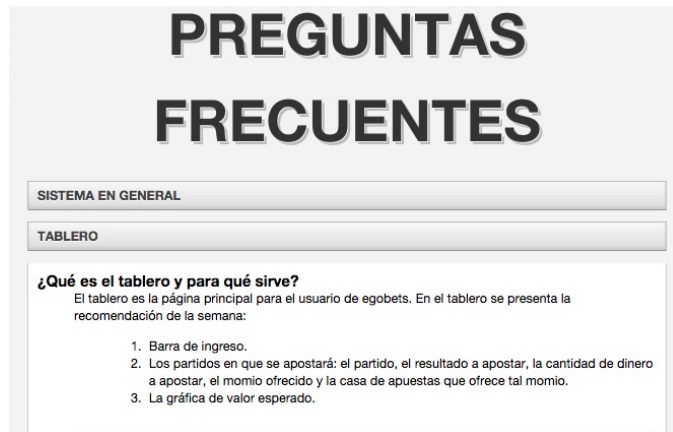


Figura 4.5: La encuesta define el perfil de riesgo del usuario

- Home Page. La página que recibe a todos los que ingresan al sitio y presenta una bienvenida a los posibles clientes junto

con una explicación de lo que es Egobets.com. Además de que contiene el aviso legal.

- Términos y condiciones. Presenta el contrato de uso al que se comprometen los clientes al utilizar el sistema.

#### ■ ***Vistas relativas a la operación del sistema***

En estas vistas se presenta la información de todas las ligas, equipos, partidos y las recomendaciones de apuestas.

- Tablero. La recomendación de apuestas de esta jornada es mostrada al usuario. Se le indica al usuario la cantidad de dinero en porcentaje a apostar en cada una de las apuestas recomendadas y los mejores momios para ese partido por casa de apuestas (Ver figura ??).
- Mis Equipos. Una tabla con los equipos preferidos del usuario junto con sus variables de ataque y de defensa.
- Tabla de posiciones de la liga. Muestra los nombres de los equipos de la liga seleccionada, variables de ataque y defensa calculadas por el sistema y el cambio en la tabla de posiciones que han tenido. Se permite al usuario agregar a sus favoritos cualquier equipo para tener un acceso directo a ellos.
- Detalle del Equipo. Se presenta el nombre del equipo, su posición en la liga y las estadísticas de ataque y defensa cuando es local o visitante. También se muestran el listado de los partidos próximos que tiene el equipo.
- Partidos de la jornada. Consiste en una tabla con los partidos próximos ordenados por fecha del juego. La tabla incluye los equipos participantes, quien es el favorito, el marcador esperado por el sistema y la fecha del juego.
- Detalle del partido. Se tienen los equipos que participan, el resultado más probable (junto con su confiabilidad en estrellas), el marcador esperado, la posesión del balón esperada

y las estadísticas de ataque y defensa. Estas estadísticas se muestran divididas en las tres categorías. Medio centro, delanteros y definición, para el ataque. Y medio centro, defensas y portero, para la defensa. Cada una de estas variables despliega al ser presionada un histograma con la magnitud esperada de esa variable durante los noventa minutos del partido en intervalos de quince minutos. (Ver imagen 4.6).

- Partidos anteriores. Finalmente esta tabla muestra un comparativo de todos los partidos predichos y su resultado real. Se remarcan los aciertos del sistema y puede ser filtrada por liga.

Todas la vistas descritas en los párrafos anteriores consisten en archivos HTML5<sup>6</sup> creados dinámicamente en función del código PHP de las vistas. Cada una de estas pantallas cuentan con un archivo CSS<sup>7</sup> que aplica el estilo y el diseño al lenguaje de marcado HTML. A su vez, se utiliza Javascript para las animaciones, validaciones y cambios en las propiedades tanto del CSS como del HTML; un ejemplo claro del uso de Javascript es la animación de los gráficos de las estadísticas de los equipos.

---

<sup>6</sup>HTML viene de HyperText Markup Language, este lenguaje de marcado llegó a la versión 5 utilizada hoy en día el 24 de Octubre del 2014 en el consorcio de la World Wide Web (Básicamente el consorcio de Internet). Para más información leer What is HTML5 de MacLauhin [25].

<sup>7</sup>Cascade Style Sheet,



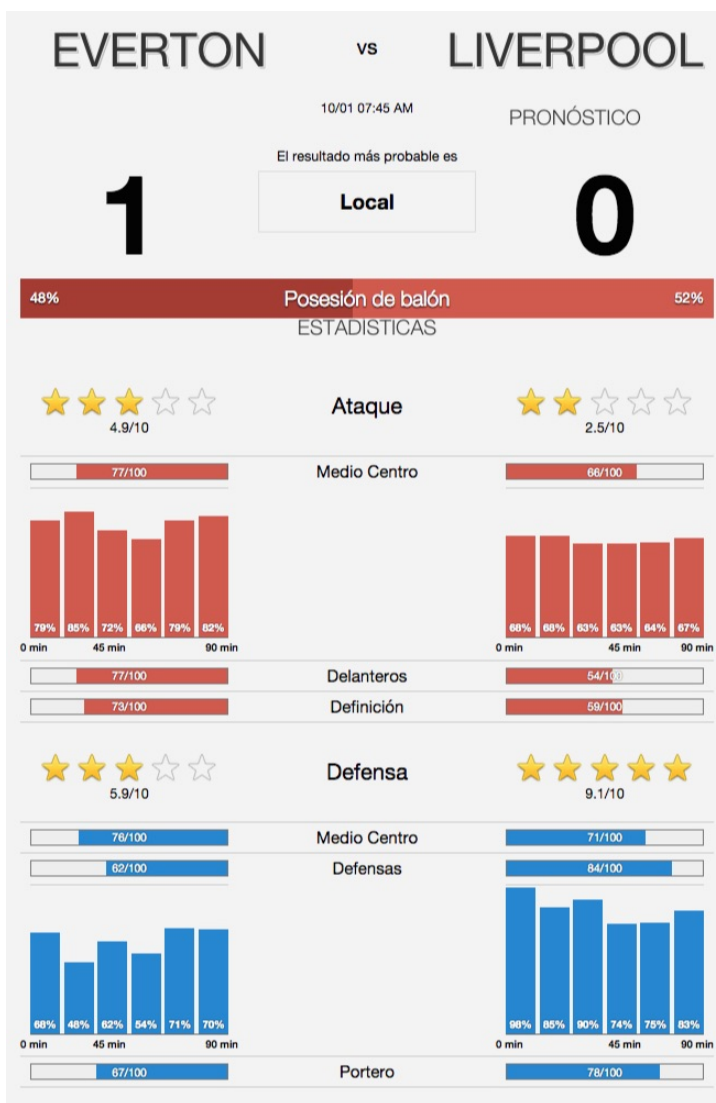


Figura 4.6: Partido con recomendación y estadísticas de los equipos

### 4.2.3. Controladores

Corresponden a la manipulación de la información a través de sus diversas presentaciones, ofrecen opciones para cambiar el estado del modelo, son los encargados de consultar los modelos y proveer a las vistas los datos dinámicos a mostrar. Básicamente, los controladores implementan el comportamiento o el control de la lógica del sistema, especificando cuándo y como el sistema cambia de estado. Este comportamiento requiere la interacción de múltiples modelos y vistas. A continuación se muestra un listado de los controladores del sistema de Egobets junto con las funciones específicas que puede realizar cada uno de ellos.:

- **Usuario.**

- Restablecer/crear/cambiar un password de una cuenta de usuario.
- Calcular el perfil de riesgo del cliente realizando y procesando los resultados de la encuesta.

- **Perfil.** Despliega la información completa del perfil de un usuario,

- Desplegar la información completa del perfil de un usuario junto con los botones necesarios para editarla.
- Agregar/quitar una cuenta de Facebook al perfil.
- Agregar/quitar casas de apuestas al usuario.
- Desplegar información del historial de pagos del usuario.
- Agregar/quitar casas de apuestas al usuario.

- **Dashboard.**

- Desplegar la información completa de la recomendación de apuestas de esta jornada al usuario en función de su perfil de riesgo, las casas de apuestas utilizadas y la reserva correspondiente a este usuario.

- Agregar la información de las ganancias o pérdidas del usuario en esta semana.
- **Amigos.**
  - Desplegar una tabla de posiciones del usuario entre sus amigos respecto a las ganancias de cada uno de ellos. También muestra el listado con la información de sus amigos.
  - Agregar/quitar amigos de Facebook o por correo.
  - Buscar amigos en el sistema con su correo electrónico.
- **Ayuda.**
  - Desplegar la información completa del perfil de un usuario junto con los botones necesarios para editarla.
  - Agregar/quitar una cuenta de Facebook al perfil.
  - Agregar/quitar casas de apuestas al usuario.
  - Desplegar información del historial de pagos del usuario.
  - Agregar/quitar casas de apuestas al usuario.
- **Sesión.**
  - Desplegar la página de inicio de sesión.
  - Iniciar la sesión del usuario con Facebook o a través de una combinación de correo y password.
  - Cerrar la sesión de un usuario.
  - Crear un nuevo usuario desde cero o con Facebook login.
- **Partido.**
  - Desplegar la información de los resultados estimados para la jornada en curso.

- Desplegar la información de los partidos de la semana pasada con sus respectivos marcadores y las predicciones dadas por el sistema.

■ **Equipo.**

- Desplegar la tabla de los equipos favoritos del usuario.
- Desplegar las ligas disponibles en el sistema.
- Desplegar la tabla de posiciones con todos los equipos que juegan en una liga.
- Desplegar la pantalla de detalle de un equipo.

■ **Pagos.**

- Desplegar las opciones para comprar jornadas de recomendaciones.
- Procesar la transferencia correcta de parte de Paypal de los pagos realizados por el usuario.

■ **Ayuda.**

- Desplegar el archivo de preguntas y respuestas de manera correcta.

■ **Favoritos.**

- Agregar/Borrar equipos de los favoritos del usuario.

Todas estas funciones realizadas por los controles, definen por completo el sistema de Egobets. Es importante mencionar una función extra que se llama `<pre>auth();</pre>` la cual corre en los constructores de cada uno de los controladores para asegurar que un usuario malintencionado no pueda realizar acciones a nombre de otro. Es decir, que si un usuario cuenta con una combinación válida de nombre de usuario y contraseña, el sistema, por ejemplo, no le permita obtener las ganancias

que ha obtenido su peor enemigo en el sistema. De no contar con estas validaciones, se podría construir una llamada HTTP de tal manera que se pudiera obtener esta información.

En la siguiente sección de este documento, se retoma la función más interesante<sup>8</sup> del sistema y se explica a mayor detalle su funcionamiento.

---

<sup>8</sup>Se considera más interesante una función que muestra recomendaciones de apuestas, que una función que permite al usuario cambiar el idioma del portal. El objetivo de esta tesis es el de mostrar una manera programática de obtener recomendaciones que permitan al usuario apostar de manera estratégica con el fin de obtener ganancias, no el de proveer documentación para la replicación de este sistema a través de ingeniería de software.



## Capítulo 5

# Conclusiones

En esta tesis se describió el sistema de recomendación de apuestas personalizadas para las ligas europeas de fútbol: Egobets, una aplicación de las matemáticas a la computación. De este trabajo se puede concluir lo siguiente:

- Existe más de un modelo que permita predecir los resultados de los partidos de las ligas europeas de fútbol. Más aún, el modelo de Maher (1982)[23] proporciona predicciones lo suficientemente confiables como para considerar su uso en una estrategia de apuestas de hoy en día.
- Conociendo las probabilidades de los partidos, aún siendo estimadas, es posible encontrar una estrategia de apuestas que genere rendimientos positivos. Por ejemplo, Koopman en su artículo [18] expone un ejemplo donde, mediante una estrategia de apuestas simple, por cada 75 unidades apostadas recibe 25 unidades de ganancia.
- La adversidad al riesgo de un apostador se puede modelar con funciones de utilidad, esto permite que las recomendaciones de apues-

tas sean personalizadas y hagan sentir al usuario cómodo mientras apuestan.

- Considerando que las temporadas del fútbol tienen varias jornadas, se integra un sistema de reservas cuya función principal consiste en maximizar la tasa de crecimiento de las ganancias del jugador. Vancura [46] muestra varias estrategias de reservas que buscan conseguir estos resultados.
- En el sistema de Egobets.com se provee de una asesoría de apuestas integral para cada semana de la temporada donde se utilizan las predicciones de los resultados de los partidos, el perfil de riesgo del usuario y el sistema de reservas.
- Los momios que publican las casas de apuestas se enfocan a predecir el mercado, por lo tanto se pueden mejorar las probabilidades propuestas.
- El porcentaje de aciertos en los pronósticos es del 70 % para la liga española, 61 % en la italiana, 52 % en la inglesa, 52 % en la alemana y 48 % en la francesa.
- En los resultados observados, Egobets reporta un rendimiento desde el 29 % hasta el 83 % dependiendo del nivel de riesgo por apuesta del usuario.
- El compute en la nube es esencial para los sistemas como Egobets, sus ventajas como escalabilidad permiten mantener costeados y funcionales los servicios.
- El enfoque de un desarrollo al diseño y la usabilidad permiten que el usuario se enfoque en realizar únicamente lo que debe realizar.
- MongoDB proporciona el beneficio de guardar información sin una estructura definida. Esto permite mucha mayor flexibilidad en los documentos que se persisten y la información que contienen[33].



**Acciones a futuro** Un sistema tan complejo como este genera muchas áreas de oportunidad, una de ellas sería implementar las nuevas investigaciones para ejorar el sistema de predicción de resultados, otro detalle sería encontrar las funciones de utilidad más prolíficas del sistema y centrar a los usuarios sobre esas. Adicionionalmente, se podría trabajar en hacer más autónomo el sistema para generar las recomendaciones. Finalmente, sobre este sistema se podría incluso podrían modificar la lógica y adaptar el modelo para abarcar más mercados, como:

- Otras ligas, por ejemplo la mexicana.
- Otros deportes, por ejemplo futbol americano.
- Incluso se varios de los autores hablan de que modelos parecidos a estos podrían servir para predecir elecciones.



# Apéndice A

## La ruina del jugador

### Planteamiento del problema

Dos apostadores Alicia y Bob participan en el siguiente juego:  
Alicia repetidamente lanza una moneda justa al aire.

- Cada que resulta *sol*, Bob le paga a Alicia un peso.
- Cada que resulta *águila*, Alicia le paga a Bob un peso.

El juego continua hasta que alguno de los dos termina sin dinero.  
Si Alicia empieza el juego con una cantidad de dinero  $A$  y Bob con  $B$ ,  
entonces:

- ¿Cuál es la probabilidad de que Alicia gane todo el dinero?
- ¿Cuál es la duración esperada del juego?

### Solución

Para esta demostración se usa el Teorema de de Paro Opcional de Doob para martingalas.

Sean  $X_1, X_2, \dots$  los incrementos de la riqueza de Alicia.

Nótese que  $X_i = 1$  depende de si resulta *águila* o *sol*.

Ahora, el cambio en el dinero de Alicia es:

$$S_n = \sum_{i=1}^n X_i$$

Se define  $\tau$  como:

$$\tau = \min\{t : S_t = B \text{ ó } S_t = -A\}$$

Claramente,  $\tau$  es un tiempo de paro relativo a filtración natural

$\mathcal{F}_n = \sigma(X_1, X_2, X_3, \dots, X_n)$ .

Y  $\tau' = \tau \wedge n$  es también un tiempo de paro.

### **La probabilidad de que Alicia gane.**

La sucesión  $S_n$ , es una martingala relativa a la filtración natural  $\mathcal{F}_n$ .

Por lo que, usando el Teorema de Paro Opcional para  $n < \infty$ ,

$$\begin{aligned} 0 &= E[S_0] = E[S_{\tau \wedge n}] = \\ &= (-A)p(\tau \leq n \text{ and } S_\tau = -A) + (B)p(\tau \leq n \text{ and } S_\tau = B) + E[S_n \chi_{\tau > n}] \end{aligned}$$

Conforme  $n \rightarrow \infty$ , la probabilidad de que  $\tau > n$  converge a cero.

El último término en la ecuación de arriba es la esperanza de una martingala acotada  $S_n$  entre  $A$  y  $B$  y converge a 0.

Por esto,

$$0 = (-A)p(S_\tau = -A) + (B)p(S_\tau = B)$$

Por lo tanto, la probabilidad de que Alicia gane todo el dinero  $S_\tau = B$  es:

$$p(S_\tau = B) = \frac{A}{A + B}.$$

### **La duración esperada del juego**

La sucesión  $(S_n^2 - n)$  es una martingala relativa a la filtración natural  $\mathcal{F}_n$ .

Entonces, usando el Teorema de Paro Opcional para  $n < \infty$ ,

$$0 = E[S_0^2] = E[S_{\tau \wedge n}^2 - (\tau \wedge n)]$$

Cuando  $n \rightarrow \infty$ , la probabilidad de que  $\tau > n$  converge a cero.  
Por lo que,

$$E[\tau] = E[S_\tau^2] = (A^2)p(S_\tau = -A) + (B^2)p(S_\tau = B) = AB$$

Por lo tanto, la duración esperada del juego es  $AB$



## Apéndice B

# Modelos lineales y soluciones a los problemas de optimización

### B.1. Decidir a favor de quien apostar

- (a) Sean  $p_L$ ,  $p_z$ ,  $p_v$  las probabilidades de que gane local, empaten o gane visitante, respectivamente. Sean  $\mu_L$ ,  $\mu_z$  y  $\mu_v$  los momios respectivos. El problema de decisión de apostar \$1 en esta situación es:

$$E_p[U(\delta_i)] = p_i \mu_i; \quad i = L, Z, V$$

Sol: Se escoge  $\rho_i \cdot \ni \cdot E_p[U(\delta_i)] = \max\{p_L \mu_L, p_z \mu_z, p_v \mu_v, 1\}$

- (b) Se quiere decidir si apostar o no en la ocurrencia de un evento: Sea  $p = p(E)$  y  $f_p$  densidad de  $p$ . Sea  $\mu$  el momio en el caso de ocurrencia. El problema de decisión asociado es el siguiente:

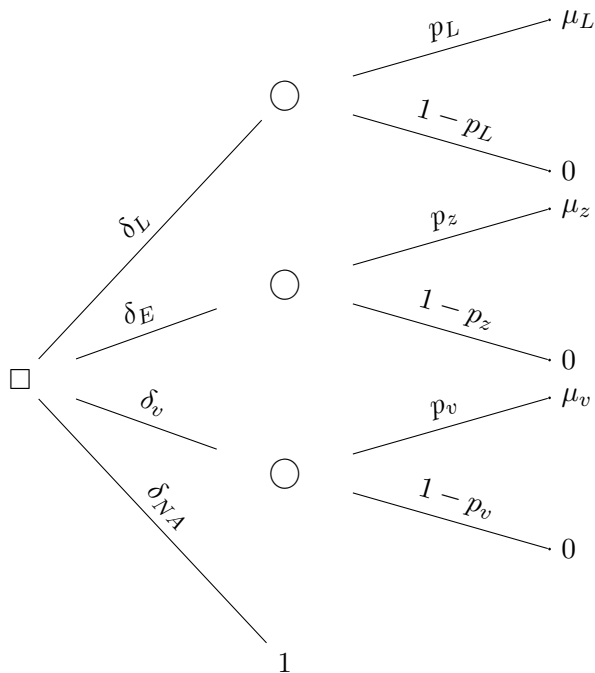


Figura B.1: Decidir por quién apostar

$$\begin{aligned}
 &\rightarrow E_p[U(\delta_A)] = E_{f_p}[p(\mu - 1) - (1 - p)] \\
 &= E_{f_p}[p(\mu) - 1] \\
 &= E_{f_p}(p)\mu - 1
 \end{aligned}$$

Apuestas si  $E_{f_p}(P) \cdot \mu \geq 1$

- (c) Mismo problema que el caso anterior, sólo que la utilidad depende de  $p$  y  $\mu$ :  $U : \Re \times [0, 1] \rightarrow \Re$   
 $(U(0, p) = 0 \quad \forall p).$

Se apuesta si:



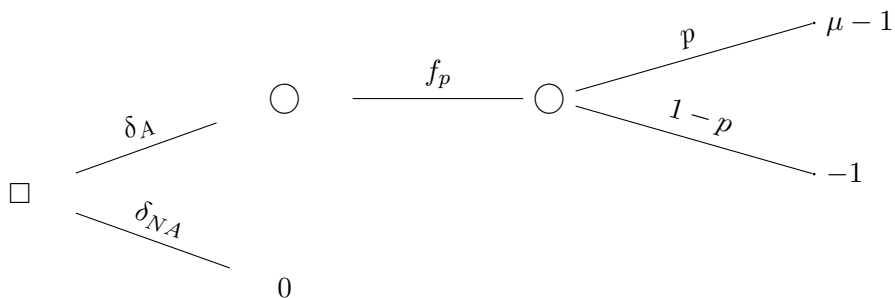


Figura B.2: Decidir si apostar o no apostar

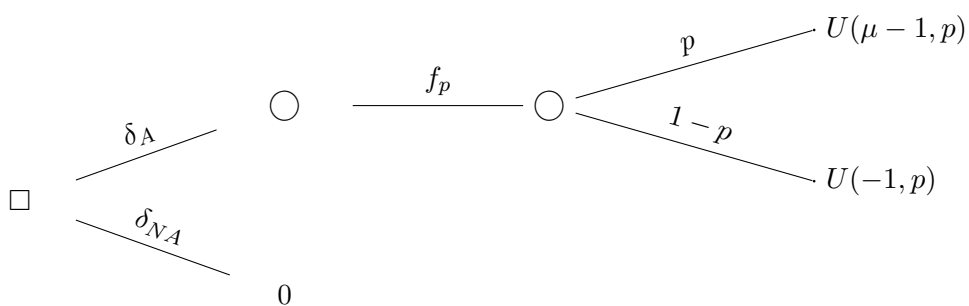


Figura B.3: Decidir si apostar en función de una utilidad

$$E_p(U(\delta_A)) = E_p[pU(\mu - 1, p) + (1 - p)U(-1, p)] \geq 0$$

Algunas funciones de utilidad posibles:

- $U_\mu(x, p) = x(\frac{1}{\mu} - p)^2$

Notese que:  $pU_\mu(\mu - 1, p) + (1 - p)U_\mu(-1, p)$

$$(\hat{p} = \frac{1}{\mu}) = (\hat{p} - p)^2(p\mu - 1)$$

Me duele más mientras más alejado esté de un trato beneficioso y me produce mayor placer mientras mayor sea el beneficio del trato.

$$\blacksquare U_{\mu,a}(x,p) = \begin{cases} ax(\hat{p} - p)^2 & \text{si } p \leq \hat{p} \\ x(\hat{p} - p)^2 & \text{si } p > \hat{p} \end{cases}$$

Notese que:

$$U_{\mu} = U_{\mu,1}$$

$$p U_{\mu,a}(\mu-1,p) + (1-p)U_{\mu}(-1,p) = \begin{cases} a(\hat{p} - p)^2(p\mu - 1) & \text{si } p \leq \hat{p} \\ (\hat{p} - p)^2(p\mu - 1) & \text{si } p > \hat{p} \end{cases}$$

Me duele “a” veces más un trato perjudicial que un trato beneficioso si me encuentro a la misma distancia que  $\hat{p}$ .

$$\blacksquare U_{\mu,a,b} = U_{\frac{\mu}{1+\mu b},a}$$

y considerar el problema de decisión con  $\mu' = \frac{\mu}{1+\mu b}$ .

$$\text{Si } \mu' = \frac{\mu}{1+\mu b} \rightarrow \hat{p}' = \hat{p} + b.$$

Los tratos empiezan a ser beneficiosos hasta que el menos sea  $b\%$  más probable que ocurra el evento de lo que sería justo.

**Nota:** En un problema de decisión sin aversión a la distribución de probabilidades (o con probabilidades fijas) si se desea apostar en apuestas con un mínimo de ganancias esperadas igual a  $b\%$  se debe comparar  $\mu_p$  con  $1 + b$  (i.e. apostar  $\leftrightarrow \mu_p \geq 1 + b$ ).

## B.2. Decidir la cantidad de dinero a apostar

Supongamos que  $\mu_p \geq 1$  y que existen 2 funciones de utilidad:

$$U_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$$

$$U_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$$

La primera es la función de utilidad del dinero para las ganancias y la segunda es la utilidad del dinero para las pérdidas monetarias.

Se harán las siguientes supuestos:

- (I)  $U_1(0) = U_2(0) = 0$ .  $U_1, U_2$  no decrecientes, una vez cont. dif.
- (II)  $U'_1(0) > U'_2(0)$  (por lo tanto convendrá apostar).
- (III)  $\forall M > 0$  fija  $\lim_{x \rightarrow \infty} \frac{U_1(\mu x)}{U_2(x)} = 0$ .  
(Perder duele muchísimo más que ganar).

El problema de decisión asociado a determinar la cantidad óptima a apostar es: (con  $0 < p < 1$  fija y  $\mu$  momio)

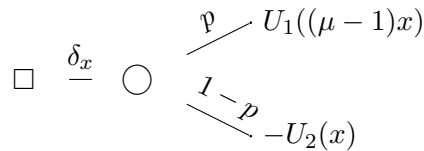


Figura B.4: Árbol de probabilidad 4

$$\rightarrow E_p[U(\delta x)] = pU_1((\mu - 1)x) - (1 - p)U_2(x)$$

Sea  $f(x) = E_p[U(\delta x)]$

Encontrar el óptimo es encontrar  $x \geq 0$  que resuelva el problema:  
 $\max_{x \geq 0} f(x)$

$$f'(x) = p(\mu - 1)U_1'((\mu - 1)x) - (1 - p)U_2'(x) = 0$$

$$\frac{p(\mu - 1)}{(1 - p)} = \frac{U_2'(x)}{U_1'((\mu - 1)x)}$$

P.d.

$$\exists x^* \cdot \exists \cdot \frac{p(\mu - 1)}{1 - p} = \frac{U_2'(x)}{U_1'(\mu x)}$$

$$(I) f'(0) = p(\mu - 1)U_1'(0) - (1 - p)U_2'(0) > p(\mu - 1)U_2'(0) - (1 - p)U_2'(0)$$

$$= U_2'(0)(p\mu - 1) \geq 0$$

Con  $U_2'(0) \geq 0$  y  $p\mu \geq 0$

Por tanto  $f'(0) > 0$

$$(II) f(0) = 0$$

$$(III) \frac{f(x)}{U_2(x)} = p \frac{U_1((\mu - 1)x)}{U_2(x)} - (1 - p)$$

$$\rightarrow \lim_{x \rightarrow \infty} \frac{f(x)}{U_2(x)} = -(1 - p)$$

$$\rightarrow \exists x \cdot \exists \cdot \frac{f(x)}{U_2(x)} = -(1 + p) + \varepsilon < 0$$

$$\rightarrow \exists x \cdot \exists \cdot f(x) < 0$$

$$\begin{aligned} \blacksquare \text{ Por T.V.M. } \exists x' \in (0, x) \cdot \exists \cdot x f'(x') &= f(x) - f(0) = \\ f(x) &< 0 \\ \rightarrow f'(x') &< 0 \end{aligned}$$

$$\blacksquare \text{ T.V.I. } \exists x^* \in (0, x') \cdot \ni \cdot f'(x^*) = 0. \text{ i.e. } \frac{p(\mu - 1)}{1 - p} = \frac{U'_2(x)}{U'_1(\mu x)}$$

Como  $f$  es primero creciente y en algún punto decreciente:  
 $\rightarrow x \cdot \ni \cdot f'(x) = 0$  es un maximizador.

Algunas funciones a considerar:

$$\blacksquare \begin{aligned} U_{1,\alpha}(x) &= x^\alpha & 0 < \alpha < 1 \\ U_2(x) &= x \end{aligned}$$

Compruébense los supuestos:

- (I)  $U_{1,\alpha}(0) = 0 = U_2(0)$ , son crecientes y una vez dif.
- (II)  $U'_{1,\alpha}(0) = +\infty$ ,  $U'_2(0) = 1 \quad \therefore U'_{1,\alpha}(0) > U'_2(0)$
- (III)  $\forall \mu > 0$

$$\lim_{x \rightarrow +\infty} \frac{U_{1,\alpha}(\mu x)}{U_2(x)} = \mu^\alpha \lim_{x \rightarrow +\infty} \frac{x^\alpha}{x} = \mu^\alpha \lim_{x \rightarrow +\infty} \frac{1}{x^{1-\alpha}} = 0$$

Para una apuesta con probabilidad  $p$  y momio  $\mu$  el óptimo se da en:

$$\frac{p(\mu - 1)}{(1 - p)} = \frac{U'_2(x)}{U'_{1,\alpha}((\mu - 1)x)} = \frac{1}{\alpha((\mu - 1)x)^{\alpha-1}} = \frac{1}{\alpha}(\mu - 1)^{1-\alpha} x^{1-\alpha}$$

$$\rightarrow \left( \frac{\alpha p}{(1 - p)} \right) (\mu - 1)^\alpha = x^{1-\alpha} \rightarrow x^* = \left( \frac{\alpha p}{1 - p} \right)^{\frac{1}{1-\alpha}} (\mu - 1)^{\alpha/1-\alpha}$$

$$\blacksquare \begin{aligned} U_{1,\alpha}(x) &= x^\alpha & 0 < \alpha < 1 \\ U_{2,\beta}(x) &= x^\beta & \beta \leq 1 \end{aligned}$$

Es fácil revisar los supuestos. Para una apuesta con probabilidad  $p$  y momio  $\mu$  el óptimo se da en:

$$\frac{p(\mu - 1)}{(1 - p)} = \frac{\beta x^{\beta-1}}{\alpha(\mu - 1)^{\alpha-1} x^{\alpha-1}} = \frac{\beta}{\alpha} (\mu - 1)^{1-\alpha} x^{\beta-\alpha}$$

$$\rightarrow \left( \frac{\alpha p}{\beta(1 - p)} \right) (\mu - 1)^{\alpha} = x^{\beta-\alpha} \rightarrow x^* = \left( \frac{\alpha p}{\beta(1 - p)} \right)^{1/\beta-\alpha} (\mu - 1)^{\alpha/\beta-\alpha}$$

- $U_1(x) = \ln(x)$   
 $U_2(x) = x$

Es fácil revisar los supuestos. Para una apuesta con probabilidad  $p$  y momio  $\mu$  el óptimo se da en:

$$\frac{p(\mu - 1)}{(1 - p)} = \frac{1}{(\frac{1}{(\mu-1)x})} = (\mu - 1)x \rightarrow x^* = \frac{p}{1 - p}$$

- $U_{1,\alpha}(x) = 1 - e^{-\alpha x} \quad \alpha \geq 1$   
 $U_2(x) = x$

Es fácil revisar los supuestos. Para una apuesta con probabilidad  $p$  y momio  $\mu$  el óptimo se da en:

$$\frac{p(\mu - 1)}{1 - p} = \frac{1}{\alpha e^{-\alpha(\mu-1)x}} \rightarrow \ln \left( \frac{\alpha p(\mu - 1)}{(1 - p)} \right) = \alpha(\mu - 1)x$$

$$\rightarrow x^* = \frac{1}{\alpha(\mu - 1)} \ln \left( \frac{\alpha p(\mu - 1)}{(1 - p)} \right)$$

Otras tres funciones de utilidad a considerar:

$$\begin{aligned} \blacksquare \quad & U_{1,\alpha}(x) = \alpha x & \alpha \geq 1 \\ & U_2(x) = e^x - 1 \end{aligned}$$

$$\rightarrow \frac{p(\mu - 1)}{1 - p} = \frac{e^x}{\alpha}$$

$$\rightarrow x^* = \ln\left(\frac{p(\mu - 1)}{1 - p}\right) + \ln(\alpha)$$

$$\begin{aligned} \blacksquare \quad & U_1(x) = \ln(x) & \alpha \geq 1 \\ & U_2(x) = x^\alpha \end{aligned}$$

$$\rightarrow \frac{p(\mu - 1)}{1 - p} = \frac{\alpha x^{\alpha-1}}{\frac{1}{(\mu-1)x}} = \alpha(\mu - 1)x^\alpha$$

$$\rightarrow x^* = \left(\frac{p}{\alpha(1 - p)}\right)^{1/\alpha}$$

$$\begin{aligned} \blacksquare \quad & U_{1,\alpha}(x) = \tan^{-1}(x) \\ & U_{2,\alpha}(x) = \alpha x & 0 < \alpha \leq 1 \end{aligned}$$

$$\rightarrow \frac{p(\mu - 1)}{1 - p} = \alpha(1 + (\mu - 1)^2 x^2)$$

$$\rightarrow \frac{p\mu - p - \alpha(1 - p)}{1 - p} = \alpha(\mu - 1)^2 x^2$$

$$\rightarrow \frac{p\mu - (1 - \alpha)p - \alpha}{1 - p} = \alpha(\mu - 1)^2 x^2$$

$$\rightarrow x^* = \frac{1}{\sqrt{\alpha}(\mu - 1)} \left(\frac{p\mu - (1 - \alpha)p - \alpha}{1 - p}\right)^{1/2}$$

$$\text{equivalentemente: } x^* = \frac{1}{\mu - 1} \left(\frac{p\mu - (1 - \alpha)p - \alpha}{1 - p}\right)^{1/2}$$

Basta probar que  $p\mu - (1 - \alpha)p - \alpha \geq 0$

$$p\mu - (1 - \alpha)p - \alpha \geq p\mu - (1 - \alpha) - \alpha = p\mu - 1 > 0$$

$x^*$  está bien definido.

### B.2.1. Ahorro precaucional

Supongamos  $F_1, \dots, F_n$  distribuciones y la siguiente sucesión de Variables aleatorias  $(x_1^t)_{t=1}^\infty, \dots, (x_n^t)_{t=n}^\infty$  independientes  $x_j^t \sim F_j \forall t \in \mathbb{N}$ .

Sean  $\alpha_1, \dots, \alpha_n \in \mathbb{R}^+ \cdot \ni \cdot \sum_{j=1}^n \alpha_j = 1$ , definimos:

$$\blacksquare z_1 = \sum_{j=1}^n \alpha_j x_j'$$

$$\blacksquare z_{t+1} = \sum_{j=1}^n \alpha_j x_j^{t+1} + z_t$$

Supongamos que  $E[x_j^t] > 1 \forall t \in \mathbb{N} \rightarrow E[z_t] = tE[z_1] = t\mu > 1$

#### Problema:

Encontrar  $y \cdot \ni \cdot (1 - ty) + yz_t \geq y \forall t \in \mathbb{N}$  con probabilidad  $(1 - \alpha) \times 100\%$ . ( $y \in [0, 1]$ ).

$$\rightarrow yz_t \geq (t+1)y - 1 \quad \rightarrow \quad z_t \geq (t+1) - \frac{1}{y}$$

$$\rightarrow z_t \geq t + k(\text{con } k = 1 - \frac{1}{y})$$



equivalentemente: Encontrar  $k \leq 0 \cdot \ni \cdot z_t \geq t + k \forall t \in \mathbb{N}$  con probabilidad  $(1 - \alpha) \times 100 \%$ .

Sol:

Sea  $\mu = E[z_1]$ ,  $\sigma^2 = Var[z_1]$

$$\rightarrow (1-\alpha) = p(z_t \geq t+k \forall t \in \mathbb{N})$$

$$= p(z_1 \geq 1+k) \cdot p(z_2 \geq 2+k | z_1 \geq 1+k) \cdots$$

$$= p(z_1 \geq 1+k) \cdot \prod_{t=1}^{\infty} p(z_{t+1} \geq (t+1) + k | z_t \geq t+k)$$

Usando el T.C.L.:  $z_t \rightarrow N(t\mu, t\sigma^2) \forall t \in \mathbb{N}$

$$(I) \quad p(z_1 \geq 1+k) = p\left(\frac{z_1 - \mu}{\sigma} \geq \frac{(1+k) - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{k - (\mu - 1)}{\sqrt{t}\sigma}\right)$$

$$(II) \quad p(z_{t+1} \geq (t+1) + k | z_t \geq t+k) = \frac{p(z_{t+1} \geq (t+1) + k, z_t \geq t+k)}{p(z_t \geq t+k)}$$

$$\begin{aligned} \blacksquare \quad p(z_t \geq t+k) &= p\left(\frac{z_t - t\mu}{\sqrt{t}\sigma} \geq \frac{k - t(\mu - 1)}{\sqrt{t}\sigma}\right) \\ &= 1 - \Phi\left(\frac{k - t(\mu - 1)}{\sqrt{t}\sigma}\right) \end{aligned}$$

$$\blacksquare \quad \begin{aligned} z_{t+1} &= y_t + z_t \text{ con } y_t \sim (\mu, \sigma^2), \quad y_t, z_t \text{ independientes} \\ z_t &\sim (t\mu, t\sigma^2) \end{aligned}$$

$$\rightarrow f(y_t, z_t) \simeq \frac{1}{2\pi\sqrt{t}\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(y_t - \mu)^2 + \frac{1}{t}(z_t - t\mu)^2]\right\}$$

Sea

$$\omega_t = y_t + z_t \qquad y_t = \omega_t - v_t$$

$$v_t = z_t \qquad z_t = v_t$$

$$\rightarrow J = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \rightarrow |det(J)| = 1$$

$$f(z_{t+1}z_t)=\frac{1}{2\pi\sqrt{t}\sigma^2}\exp\{-\frac{1}{2\sigma^2}[(z_{t+1}-z_t-\mu)^2+\frac{1}{t}(z_t-t\mu)^2]\}$$

$$\rightarrow p(z_{t+1}) \geq (t+1)+k, \, z_t \geq t+k$$

$$=\int_{t+k}^{\infty}\int_{t+1+k}^{\infty}\frac{1}{2\pi\sqrt{t}\sigma^2}\exp\{-\frac{1}{2\sigma^2}[(z_{t+1}-z_t-\mu)^2+\frac{1}{t}(z_t-t\mu)^2]\}dz_{t+1}dz_t$$

$$=\frac{1}{2\pi\sqrt{t}\sigma^2}\int_{t+k}^{\infty}\exp\{-\frac{1}{2\sigma^2t}(z_t-t\mu)^2\}\int_{t+1+k}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}[(z_{t+1}-z_t-\mu)^2]\}dz_t.$$

$$^1 = \frac{1}{2\pi\sqrt{t}\sigma^2} \int_{t+k}^{\infty} \exp\left\{-\frac{1}{2\sigma^2 t}(z_t - t\mu)^2\right\} \left[1 - \Phi\left(\frac{k+t-z_t-(\mu-1)}{\sigma}\right)\right]$$

---


$$\bar{z}_t = \frac{1}{t}z_t, \quad d\bar{z}_t = \frac{1}{t}dz_t, \quad (\bar{z}_t)_0 = 1 + \frac{k}{t}, \quad (\bar{z}_t)_1 = \infty$$


---

$$= \frac{\sqrt{t}}{2\pi\sqrt{t}\sigma^2} \int_{1+k/t}^{\infty} \exp\left\{-\frac{t}{2\sigma^2}(\bar{z}_t - \mu)^2\right\} \left[1 - \Phi\left(\frac{k+t(1-\bar{z}_t)-(\mu-1)}{\sigma}\right)\right] d\bar{z}_t$$

Por tanto:

$$p(z_{t+1}) \geq (t+1) + k, \quad z_t \geq t+k$$

$$\simeq \frac{\sqrt{t} \int_{1+k/t}^{\infty} \exp\left\{-\frac{t}{2\sigma^2}(\bar{z}_t - \mu)^2\right\} \left[1 - \Phi\left(\frac{k+t(1-\bar{z}_t)-(\mu-1)}{\sigma}\right)\right] d\bar{z}_t}{\sqrt{2\pi}\sigma \left(1 - \Phi\left(\frac{k-t(\mu-1)}{\sqrt{t}\sigma}\right)\right)}$$

Para calcular  $k$  se resuelve la siguiente ecuación:

$$\log(1-\alpha) = \log(p|z_1 \geq 1+k) + \sum_{t=1}^{\infty} \log(p(z_{t+1}) \geq (t+1) + k, \quad z_t \geq t+k)$$

$$y = \frac{1}{1-k}$$

Se realizó una muestra  $y_1, \dots, y_n$ , donde:

$$y_1 = CA(p_i, \mu_i, \sigma_i)$$

Donde:

---

<sup>1</sup>Ver Apéndice A

- $p_i$ : Un valor de probabilidad deseado.
- $M_i$ : Un valor de  $E[z_1]$  dado.
- $\delta_i$ : Un valor de  $Var(z_i)^{1/2}$ .
- $CA$ : La función que se define implícitamente de resolver las ecuaciones para calcular la cantidad de apostar.

A tales datos se les ajustó el siguiente modelo lineal:

$$y_i = \beta_0 + \beta_1 p_i + \beta_2 \mu_i + \beta_3 \sigma_i + \varepsilon_i$$

El ajuste es el siguiente:

- $\beta_0 = 0.2925$
- $\beta_1 = -0.9975$
- $\beta_2 = 1.3772$
- $\beta_3 = -1.1127$

Con  $R^2 = 0.95$ .

En adelante, se tomará como aproximación lo siguiente:

$$CA(p, \mu, \sigma) \simeq 0.2925 - 0.9975p + 1.3772\mu - 1.1127\sigma$$

### B.2.2. Evolución del dinero en el tiempo

**Problema:** Decidir  $p$  de manera óptima.

Sea  $x$  la cantidad de ingresos restantes ( $0 \leq x \leq 1$ , en porcentaje), y  $\mu, \sigma$  la media y la desviación estandar de apostar en un periodo dados.

Supongamos  $U_1, U_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  funciones de utilidad del dinero. ( $U_1$  ganancias,  $-U_2$  pérdidas)  $\cdot \ni \cdot$  son no decrecientes y una vez continuamente diferenciables. Considerese la siguiente función:

$$\begin{aligned} f(p; x, \mu, \sigma) &= [\textit{beneficio}] - [\textit{costo}] \\ f(p; x, \mu, \sigma) &= [pU_1(y(p, \mu, \sigma)\mu x)] - [(1 - p)U_2(x)] \end{aligned}$$

Suponiendo  $y(p, \mu, \sigma) = a_0 - a_1p + a_2\mu - a_3\sigma$  se obtiene:

$$a_i \geq 0, \quad i = 0, \dots, 3$$

$$f = pU_1((a_0 - a_1p + a_2\mu - a_3\sigma)\mu x) - (1 - p)U_2(x)$$

El problema es:

$$\max f$$

Sol:

$$\begin{aligned} f'(p) &= -pU_1'((a_0 - a_1p + a_2\mu - a_3\sigma)\mu x)a_1\mu x \\ &\quad + U_1((a_0 - a_1p + a_2\mu - a_3\sigma)\mu x) + U_2(x) = 0 \end{aligned}$$

Si  $U_1$  es cóncava  $\rightarrow p^*$  es un maximizador.

Forma aproximada de obtener  $p$  :

$$y = a_0 - a_1p + a_2\mu - a_3\sigma$$

$\rightarrow$  Sea  $b = a_1\mu x$ ,  $p_0$  una aproximación de  $p$ . Definimos:

$$\omega = y\mu x, \quad \omega_0 = y(p_0, \mu, \sigma)\mu x$$

$$\begin{aligned} \rightarrow U_1(\omega) &= U_1(\omega_0) + U_1'(\omega_0)(\omega - \omega_0) + O((\omega - \omega_0)) \\ &= U_1(\omega_0) + bU_1'(\omega_0)(\omega_0)(p - p_0) + O((\omega - \omega_0)) \end{aligned}$$

Se puede aproximar  $f$  por:

$$f(p) \simeq p[U_1(\omega_0) + bU_1'(\omega_0)(\omega_0)(p - p_0)] - (1 - p)U_2(x)$$

$$\Rightarrow f'(p) \simeq U_1(\omega_0) - 2bU_1'(\omega_0)p + bU_1'(\omega_0)p_0 + U_2(x) = 0$$

$$\Rightarrow p \simeq \frac{1}{2bU'_1(\omega_0)}[U_1(\omega_0) + U_2(x)] + \frac{1}{2}p_0$$

Supongamos  $U_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  dada por  $U(\omega) = \omega^\alpha$  ( $0 < \alpha \leq 1$ ) y  $U_2(x) = \beta x$

Notese que:

$$\begin{aligned} \blacksquare \quad & \frac{U_1(\omega_0)}{U'_1(\omega_0)} = \frac{\omega_0}{\alpha} \\ \blacksquare \quad & \frac{\omega_0}{b} = \frac{a_0 - a_1p + a_2\mu - a_3\sigma}{a_1} \end{aligned}$$

$$\Rightarrow p \simeq \frac{1}{2\alpha a_1}[a_0 + a_2\mu - a_3\sigma + \frac{\beta}{\mu}[(a_0 - a_1p_0 + a_2\mu - a_3\sigma)\mu x]^{1-\alpha}] + \frac{1}{2}(1 - \frac{1}{\alpha})p_0$$

Supongamos ahora que  $f$  es de la siguiente forma:

$$f(p) = pU_1(y(p, \mu, \sigma) \cdot \mu x) - (1 - p)U_2(x) - pU_3(\theta(k\sigma - \mu))$$

i.e. Hay pérdidas potenciales por el riesgo de la inversión considerar

$$U_3(\theta(k\sigma - \mu))U_2(\theta(k\sigma - \mu)x)I(\theta(k\sigma - \mu) \geq 0)$$

$\Rightarrow$  De manera análoga se obtiene:

$$p \simeq \frac{1}{2bU'(\omega_0)}[U_1(\omega_0) + U_2(x) + U_2(\theta(k\sigma - \mu)x)I_{\{m \geq 0\}}] + \frac{1}{2}p_0$$

$$\text{Si tomamos } U_1(\omega)0\omega^\alpha, \; U_2(x)=\beta x$$

$$p \simeq \frac{1}{2\alpha a_1}\{ (a_0+a_2\mu-a_3\sigma) \\ +\frac{\beta_1}{\mu}[1+(\beta_2\sigma-\beta_3\mu)I_{\{m\geq 0\}}][(a_0-a_1p+a_2\mu-a_3\sigma)\mu x]^{1-\alpha}\}+\frac{1}{2}(1-\frac{1}{\alpha})p_0$$



# Apéndice C

## Bases de datos

### C.1. Dump BD Egobets.com

En este anexo se presenta un dump BSON de la base de datos de MongoDB del sistema de Egobets. Cada apartado contiene el ejemplo de un documento de esa colección.

#### C.1.1. Casas de apuestas

---

```
{
  "_id" : ObjectId("4e4953dac67d3e1d5b000000"),
  "nombre" : "Bwin",
  "url" : "http://www.bwin.com"
}
```

---

#### C.1.2. Administradores

---

```
{
  "_id" : ObjectId("544450209273f5e31edb66b7"),
  "correo" : "bruno@egobets.com",
}
```

---

```
"nombre" : "Brus Medina",  
"password" : "9f9c1136c11828b14e6ff2c057e693d6eae9485",  
"titulo" : "Arquitecto de Software"  
}
```

---

### C.1.3. Equipos

---

```
{  
  "_id" : ObjectId("4e012952c67d3ebc5e000000"),  
  "diff" : -1,  
  "indicadores" : {  
    "local" : {  
      "ataques" : {  
        "1" : [  
          77,  
          74,  
          76,  
          77,  
          64,  
          78  
        ],  
        "2" : [  
          100,  
          100,  
          76,  
          85,  
          100,  
          91  
        ],  
        "3" : [  
          83,  
          100,  
          82,  
          78,  
          84,  
          77,  
          74,  
          76,  
          77,  
          64,  
          78  
        ]  
      }  
    }  
  }  
}
```

```

        93
    ]
},
"defensas" : {
    "1" : [
        71,
        62,
        58,
        56,
        48,
        79
    ],
    "2" : [
        67,
        68,
        68,
        59,
        64,
        60
    ],
    "3" : [
        95,
        100,
        87,
        77,
        100,
        91
    ]
},
"ataque_total" : 100,
"defensa_total" : 72,
"posesion" : 42.55401
},
"visitante" : {
    "ataques" : {
        "1" : [
            70,

```

```
        67,  
        69,  
        70,  
        58,  
        71  
    ],  
    "2" : [  
        74,  
        74,  
        56,  
        63,  
        74,  
        67  
    ],  
    "3" : [  
        65,  
        78,  
        64,  
        61,  
        66,  
        72  
    ]  
},  
"defensas" : {  
    "1" : [  
        64,  
        56,  
        52,  
        51,  
        43,  
        71  
    ],  
    "2" : [  
        84,  
        85,  
        85,  
        73,
```

```

        79,
        75
    ],
    "3" : [
        92,
        97,
        84,
        74,
        97,
        88
    ]
},
    "ataque_total" : 53,
    "defensa_total" : 79,
    "posesion" : "43.75194"
}
},
    "liga" : "alemana",
    "nombre" : "Bayern Munich",
    "posicion" : 1,
    "status" : 1
}

```

---

#### C.1.4. Usuario

```

{
    "_id" : ObjectId("4e810d6992c6e97059000002"),
    "usuario" : ObjectId("4e680ec592c6e98b2a000001"),
    "jornada" : 3,
    "inicio" : 0,
    "fin" : 0
}

```

---

### C.1.5. Mailing

---

```
{
  "_id" : ObjectId("4e0e264292c6e94939000000"),
  "correo" : "rob@surrealista.com",
  "desde" : "2011-07-01T19:55:46.648Z",
  "idioma" : "en",
  "invitado" : 1,
  "nombre" : "Roberto Hidalgo"
}
```

---

### C.1.6. partidos

---

```
{
  "_id" : ObjectId("4e23a78092c6e90b15000000"),
  "fecha" : "2011-07-24T05:00:00Z",
  "liga" : "alemana",
  "local" : "4e012952c67d3ebc5e00000f",
  "marcl" : "0.93522",
  "marcv" : "1.0974",
  "momios" : {
    "local" : 3.36516,
    "empate" : 3.20177,
    "visitante" : 2.56075
  },
  "probe" : "0.30131",
  "probl" : "0.30991",
  "probv" : "0.38878",
  "visitante" : "4e012952c67d3ebc5e00000d"
}
```

---

### C.1.7. Resultados

---

```
{
  "_id" : ObjectId("4e810d6992c6e97059000002"),
  "usuario" : ObjectId("4e680ec592c6e98b2a000001"),
  "jornada" : 3,
  "inicio" : 0,
  "fin" : 0
}
```

---

### C.1.8. system.indexes

---

```
{
  "v" : 1,
  "name" : "_id_",
  "key" : {
    "_id" : 1
  },
  "ns" : "egobets.usuarios"
}
```

---

### C.1.9. usuarios

---

```
{
  "_id" : ObjectId("5457db71c0622ba170a37d30"),
  "nombre" : "Manuel Colin",
  "correo" : "machington15@hotmail.com",
  "idioma" : true,
  "facebook" : "555255439",
  "status" : NumberLong(1),
  "usuarioDesde" : "2014-11-03T19:45:53.241Z",
  "jornadas" : NumberLong(0)
}
```

---





# Bibliografía

- [1] Gerd Alexander. “US on Tilt: Why the Unlawful Internet Gambling Enforcement Act Is a Bad Bet, The”. En: *Duke L. & Tech. Rev.* (2008), pág. 1.
- [2] Sid Browne. “Can you do better than Kelly in the short run”. En: *Finding the Edge, Mathematical Analysis of Casino Games* (2000), págs. 215-231.
- [3] Jaiho Chung y Joon Ho Hwang. “An Empirical Examination of the Parimutuel Sports Lottery Market versus the Bookmaker Market”. En: *Southern Economic Journal* 76.4 (2010), págs. 884-905.
- [4] Sitio Web de CodeIgniter. *Why CodeIgniter?* <http://www.codeigniter.com/>. [En línea; visto el 23 de enero del 2015]. 2015.
- [5] Martin Crowder y col. “Dynamic modelling and prediction of English Football League matches for betting”. En: *Journal of the Royal Statistical Society: Series D (The Statistician)* 51.2 (2002), págs. 157-168.
- [6] Walter Distaso y col. “Corruption and referee bias in football: the case of Calciopoli”. En: *20th Conference in Economy of Taxation. Tax systems, tax rates and growth*. 2008.
- [7] Mark J Dixon y Stuart G Coles. “Modelling association football scores and inefficiencies in the football betting market”. En: *Jour-*

*nal of the Royal Statistical Society: Series C (Applied Statistics)* 46.2 (1997), págs. 265-280.

- [8] The Economist. *The house wins*. <http://www.economist.com/blogs/graphicdetail/2014/02/daily-chart-0>. [eGaming Review]. 2014.
- [9] AWF Edwards. “Pascal’s problem: The ‘Gambler’s Ruin’”. En: *International Statistical Review/Revue Internationale de Statistique* (1983), págs. 73-79.
- [10] Eugene F Fama. “Market efficiency, long-term returns, and behavioral finance”. En: *Journal of financial economics* 49.3 (1998), págs. 283-306.
- [11] Robert C Hannum y Anthony N Cabot. *Practical casino math*. Institute for the study of gambling y commercial gaming, College of business administration, University of Nevada, 2005.
- [12] George Ignatin. “Sports betting”. En: *The Annals of the American Academy of Political and Social Science* (1984), págs. 168-177.
- [13] KPMG International. *Online Gaming A Gamble or a Sure Bet?* <http://www.kpmg.com/EU/en/Documents/Online-Gaming.pdf>. [En línea; visto el 12 de Diciembre del 2014]. 2010.
- [14] Norman Lloyd Johnson, Samuel Kotz y Narayanaswamy Balakrishnan. *Discrete multivariate distributions*. Vol. 165. Wiley New York, 1997.
- [15] Dimitris Karlis y Ioannis Ntzoufras. “Analysis of sports data by using bivariate Poisson models”. En: *Journal of the Royal Statistical Society: Series D (The Statistician)* 52.3 (2003), págs. 381-393.
- [16] John L Kelly. “A new interpretation of information rate”. En: *Information Theory, IRE Transactions on* 2.3 (1956), págs. 185-189.
- [17] Subrahmaniam Kocherlakota y Kathleen Kocherlakota. *Bivariate discrete distributions*. Wiley Online Library, 1992.

- [18] Siem Jan Koopman y Rutger Lit. “A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League”. En: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (2013).
- [19] Steven D Levitt. “Why are gambling markets organised so differently from financial markets?\*”. En: *The Economic Journal* 114.495 (2004), págs. 223-246.
- [20] Michael Lewis. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.
- [21] Guillermo Valdés Lozano. “Software libre”. En: (2008).
- [22] R. (Director) Luketic. *21*. Columbia Pictures. 2008.
- [23] Mike J Maher. “Modelling association football scores”. En: *Statistica Neerlandica* 36.3 (1982), págs. 109-118.
- [24] F. Martínez y B Gutiérrez. *Cómputo en la nube: Ventajas y Desventajas*. <http://revista.seguridad.unam.mx/>. [En línea; visto el 12 de Diciembre del 2014]. 2010.
- [25] Brett McLaughlin. *What is HTML5?*. O'Reilly Media, Inc.", 2011.
- [26] Peter Membrey, Eelco Plugge y DUPTim Hawkins. *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing*. Apress, 2010.
- [27] B. (Director) Miller. *Moneyball*. Columbia Pictures. 2011.
- [28] Patrick Odionikhere. *Bringing down this house*. Vol. 1. LIT Verlag Münster, 2008.
- [29] K Ord, C Fernandes y AC Harvey. “Time series models for multivariate series of count data”. En: *Developments in time series analysis*. Springer, 1993, págs. 295-309.
- [30] Alper Ozgit. “Posted-Offer vs. Double Auctions Revisited: An Investigation into Online Sports Betting”. En: *Unpublished master's thesis, Department of Economics, UCLA* (2005).

- [31] Sitio Web de PHP. *Información general*. <http://php.net/manual/es/faq.general.php>. [En línea; visto el 23 de enero del 2015]. 2015.
- [32] Peter Prevos. “The Psychic Octopus is a Fraud”. En: (2010).
- [33] Yogesh Punia y Rinkle Aggarwal. “Implementing Information System Using MongoDB and Redis”. En: ().
- [34] Will Reese. “Nginx: the high-performance web server and reverse proxy”. En: *Linux Journal* 2008.173 (2008), pág. 2.
- [35] Marcelo Roffé y col. “Las crisis durante el juego. El gol psicológico en el fútbol”. En: *Revista de psicología del deporte* 16.2 (2007), págs. 227-240.
- [36] Sheldon Ross. *A First Course in Probability 7th Edition*. Pearson, 2006.
- [37] Havard Rue y Oyvind Salvesen. “Prediction and retrospective analysis of soccer matches in a league”. En: *Journal of the Royal Statistical Society: Series D (The Statistician)* 49.3 (2000), págs. 399-418.
- [38] Goldman Sachs. *PartyGaming, 888, Playtech, bwin to Surve on Legal U.S. eGaming*. [En línea; visto el 12 de Diciembre del 2014]. 2009.
- [39] Raymond D Sauer. “The economics of wagering markets”. En: *Journal of economic Literature* (1998), págs. 2021-2064.
- [40] David G Schwartz. “Roll the bones: The history of gambling”. En: (2013).
- [41] John G SKELLAM. “The frequency distribution of the difference between two Poisson variates belonging to different populations.” En: *Journal of the Royal Statistical Society. Series A (General)* 109.Pt 3 (1946), pág. 296.
- [42] Adam Smith. *The Wealth Of Nations Vol. II*. 1963.

- [43] Linus Torvalds y David Read By-Diamond. *Just for fun: The story of an accidental revolutionary*. Harper Audio, 2001.
- [44] Ante Z Udovicic. “Special Report: Sports and Gambling a Good Mix? I Wouldn’t Bet On It.” En: *Marq. Sports LJ* 8 (1998), págs. 401-429.
- [45] David Upton. *CodeIgniter for Rapid PHP Application Development*. Packt Publishing Ltd, 2007.
- [46] Olaf Vancura, Judy A Cornelius y William R Eadington. *Finding the edge: mathematical analysis of casino games*. University of Nevada Press, 2000.
- [47] Alfredo Weitzenfeld. *Ingeniería de software orientada a objetos con UML, JAVA e INTERNET*. Thomson Editors SA, 2005.
- [48] Leighton Vaughan Williams. “Information efficiency in betting markets: A survey”. En: *Bulletin of Economic Research* 51.1 (1999), págs. 1-39.
- [49] Yogonet. *Online Gambling in Mexico: A Safe Bet*. 2007.
- [50] Joshua Zhu. *Nginx Internals*. 2010.