



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

POISSON Y FUTBOL ¿COMBINACIÓN
PERFECTA?

REPORTE DE DIVULGACIÓN

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

PRESENTA:

LUIS ENRIQUE IGNACIO GÓMEZ ORDOÑEZ

TUTOR:

DR. ADRIÁN GONZÁLEZ-CASANOVA SOBERÓN



Ciudad de México, 2023

Índice general

1. Marco teórico	1
1.1. Probabilidad	1
1.2. Independencia y probabilidad condicional	3
1.3. Variables aleatorias	4
1.4. Esperanza y Varianza	6
1.5. Distribución Poisson	8
1.6. Inferencia estadística	11
1.6.1. Estimación de parámetros	11
1.6.2. Prueba de hipótesis ji-cuadrada	15
2. Aplicación de la distribución Poisson al futbol	19
2.1. Modelo de Maher	20
2.2. Modelo de Dixon-Coles	23
2.3. Un modelo con evolución dinámica	25
3. Discusión	28
A. Apéndice	29

ÍNDICE GENERAL

A.1. Proceso Poisson Puntual	29
A.1.1. Espacio de estados	30
A.1.2. Intensidad	31
A.1.3. Realizaciones	32
A.1.4. Función de verosimilitud	33
A.1.5. Esperanzas	34
A.1.5.1. Calculo de la esperanza para el PPP	35
A.1.5.2. Sumas aleatorias	37
A.1.6. Teorema de Campbell	38
A.1.6.1. Caracterización de los PPP	40
A.1.6.2. Funcional generador de probabilidad	42

1.1. Probabilidad

De acuerdo con la Real Academia Española, se define a la probabilidad como

- f. Verosimilitud o fundada apariencia de verdad.
- f. Cualidad de probable (que se verificará o sucederá).
- f. Mat. En un proceso aleatorio, razón entre el número de casos favorables y el número de casos posibles.

Debido a la extensa literatura del tema, y con el objetivo de mantener las primeras secciones accesibles a la mayoría de las y los lectores, se definirán los conceptos y enunciarán los principales resultados sin el rigor matemático que se tiene en los cursos clásicos de Probabilidad que se imparten en la Facultad de Ciencias.

Considérese un fenómeno aleatorio, es decir, un fenómeno en el cuál, al momento de iniciar no se tiene certeza del resultado que tomará, pero, en el mejor de los casos, se tiene la noción de los resultados posibles que puedan obtenerse.

Por ejemplo, cuando se lanza una moneda, se sabe que puede salir águila o sol; mientras que si se lanza un dado de Turista Mundial, se sabe que puede salir 1, 2, 3, 4, 5 o 6. Pero hasta que la moneda no caiga en una superficie plana o el dado no deje de rodar, no se sabe con exactitud cuál fue el resultado de ese experimento en particular.

Es posible entonces, poder entender a la probabilidad como una medida de, en cierta forma, determinar dentro de un posible conjunto de resultados, que tan factible es obtener un resultado dado.

El cálculo de probabilidades más básico consiste simplemente en hacer un cociente:

Definición 1.1.1. *Probabilidad frecuentista*

Considérese un experimento aleatorio del cuál se conoce el conjunto Ω de resultados posibles. La probabilidad del resultado $A \in \Omega$ (es decir, A consiste de uno o varios resultados posibles) esta dada por

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}, \quad (1.1)$$

donde $\#A$ es la notación de la cardinalidad del conjunto A (el número de elementos en A).

Es decir, se puede pensar a la probabilidad frecuentista como el número de resultados favorables entre el número de resultados totales.

Por ejemplo, la probabilidad de sacar un 6 en un dado mientras se juega Turista Mundial es $\frac{1}{6}$, mientras que la probabilidad de obtener águila en un lanzamiento de una moneda es $\frac{1}{2}$.

La medida $\mathbb{P}(A)$ además, cumple los siguientes axiomas, donde A es un evento cualquiera posible de un espacio de resultados Ω .

1. $0 \leq \mathbb{P}(A) \leq 1$

2. $\mathbb{P}(\Omega) = 1$

3. Para cualquier secuencia A_1, A_2, \dots de eventos ajenos ($A_i \cap A_j = \emptyset$ si $i \neq j$) se tiene

$$\mathbb{P}[\cup_{i=1}^n A_i] = \sum_{i=1}^n \mathbb{P}(A_i).$$

1.2. Independencia y probabilidad condicional

Si bien los axiomas de la probabilidad son suficientes para deducir propiedades útiles, e incorporan las bases para empezar a *medir* el azar, para trabajar en modelos más complejos (como el que se verá más adelante) es importante presentar los conceptos de probabilidad condicional e independencia de eventos.

Definición 1.2.1. Probabilidad condicional

Sean A y B eventos del espacio de resultados Ω . Si $\mathbb{P}(B) > 0$, entonces la probabilidad condicional de A dado B esta dada por

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.2)$$

La probabilidad condicional ayuda a incorporar la información que se tiene de un resultado obtenido anteriormente, para poder tener una estimación más precisa de un resultado que *pueda modificarse* con las consecuencias del primero.

Definición 1.2.2. Independencia de eventos

Sean A y B eventos del espacio de resultados Ω . Se dice que A y B son independientes si se cumple

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad (1.3)$$

Si no se cumple, entonces se dice que A y B son dependientes.

Ejemplo 1.2.1. Una carta se elige al azar de una baraja inglesa sin comodines (52 cartas). Si E es el evento de que la carta sea una reina y F es el evento de que sea una carta de diamantes, entonces E y F son independientes.

En primer lugar, $\mathbb{P}(E) = \frac{4}{52}$, pues cada palo de la baraja tiene una reina.

Por otro lado, $\mathbb{P}(F) = \frac{13}{52}$, pues en toda la baraja hay 13 cartas de cada palo, entre ellos los diamantes.

Al hacer el producto, se tiene que $\mathbb{P}(E)\mathbb{P}(F) = \frac{4}{52} \frac{13}{52} = \frac{52}{52^2} = \frac{1}{52} = \mathbb{P}(E \cap F)$ por lo que son eventos independientes.

1.3. Variables aleatorias

Una variable aleatoria de forma intuitiva es una función que asigna a los resultados de un experimento aleatorio, un valor numérico.

Definición 1.3.1. Variable aleatoria

Una variable aleatoria (v.a.) X es una función real asociada a un experimento aleatorio con resultados en Ω .

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned} \tag{1.4}$$

Para cualquier resultado del experimento ($\omega \in \Omega$), se le asigna un número real ($X(\omega) \in \mathbb{R}$).

Al conjunto $S_X \subset \mathbb{R}$ de valores que puede tomar X se le conoce como soporte de la variable aleatoria. Dependiendo de S_X , una v.a. puede caracterizarse como discreta, continua o mixta (cuando S_X pueda verse como unión de conjuntos discretos y continuos). Ya que se tiene una forma de asignar a cada resultado de un experimento, un número real, resta calcular

la probabilidad de este resultado, para ello se suele trabajar con funciones de probabilidad así como funciones de distribución de probabilidad.

Definición 1.3.2. *Función de probabilidad*

Para una variable aleatoria X , se define su función de probabilidad $f(x) : \mathbb{R} \rightarrow [0, 1]$ como

$$f(x) = \mathbb{P}(X = x)\mathbb{1}_{S_X}(x).$$

Es decir, ya que a los resultados de un experimento se les asigna un número, la función de probabilidad mapea dicho número a la probabilidad de que ocurra el resultado al cuál está asociado.

La función de probabilidad cuando X es discreta se conoce como función de masa de probabilidad, mientras que cuando X es continua se conoce como función de densidad de probabilidad. Para que una función real sea caracterizada como función de probabilidad, basta con que cumpla las siguientes propiedades.

Si X es discreta:

1. $f(x) \geq 0$ para toda $x \in \mathbb{R}$.
2. $\sum_{x \in S_X} f(x) = 1$.

Si X es continua:

1. $f(x) \geq 0$ para toda $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Además de la función de probabilidad, una variable aleatoria tiene asociada una función de distribución.

Definición 1.3.3. *Función de distribución*

Para una variable aleatoria X , se define su función de distribución $F(x) : \mathbb{R} \rightarrow [0, 1]$ como

$$F(x) = \mathbb{P}(X \leq x).$$

Es decir, la función de distribución evaluada en un $X(\omega)$ es la probabilidad de que la variable aleatoria X tome un valor menor o igual que $x(\omega)$, es decir, que este en el intervalo $(-\infty, x(\omega)]$. Es intuitivo pensar que de la función de probabilidad se puede encontrar la función de distribución, aunque se debe adoptar un tratamiento diferente dependiendo si X es discreta o continua.

Si X es discreta y $x \in S_X$:

$$F(x) = \sum_{u \leq x} f(u).$$

Si X es continua y $x \in S_X$:

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

Recordando los resultados de Cálculo Diferencial, si se aplica el Teorema Fundamental del Cálculo entonces, desde la función de distribución también es posible llegar a la función de probabilidad.

1.4. Esperanza y Varianza

Una vez que se tiene la función de distribución y la función de probabilidad, el arsenal para responder preguntas que impliquen el cálculo de probabilidades es enorme, sin embargo

en ocasiones es necesario poder dar información de la variable aleatoria X que ayuden a entender mejor sus características o su distribución.

Definición 1.4.1. *Esperanza*

Para una variable aleatoria X , con función de probabilidad $f(x)$ se define su valor esperado (esperanza) $\mathbb{E}(X)$ como

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in S_X} xf(x) & \text{Si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{Si } X \text{ es continua.} \end{cases} \quad (1.5)$$

Desde un punto de vista técnico, la esperanza no es más que un promedio ponderado de los valores que puede tomar la variable aleatoria, por la probabilidad de que cada resultado ocurra. Es importante notar que la esperanza no representa el valor exacto que la variable aleatoria tomará, incluso puede darse el caso que $\mathbb{E}(X) \notin S_X$.

Una de las primeras propiedades que se presentan en los cursos clásicos de probabilidad es que la esperanza es un operador lineal, es decir, “abre sumas y saca escalares”. Si $a, b \in \mathbb{R}$ y X, Y son variables aleatorias, entonces $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, la demostración del resultado anterior no implica mayor reto que usar las propiedades de la integral en el caso continuo o la suma en el caso discreto.

Se cerrará esta breve introducción con un concepto que, al usarse con la esperanza ayuda a tener una mayor comprensión de las características de una variable aleatoria.

Definición 1.4.2. *Varianza*

Sea X una variable aleatoria con función de probabilidad $f(x)$. La varianza $\mathbb{V}(X)$ se define como

$$\mathbb{V}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2. \quad (1.6)$$

Si se desarrolla el cuadrado de la definición anterior y se usa el resultado de que la esperanza es un operador lineal, se obtiene una nueva expresión para la varianza como

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

La varianza suele conocerse como una medida de dispersión, es decir, ayuda a saber en promedio que tan alejados están los datos de la media, como esta en unidades cuadradas, a menudo se suele usar su raíz cuadrada (conocida como desviación estándar) al momento de presentar resultados.

1.5. Distribución Poisson

Los modelos que se presentarán en la siguiente sección convergen en que tienen un componente aleatorio que sigue una ley de probabilidad Poisson (pronunciación: *poa-son*).

Definición 1.5.1. Distribución Poisson

Se dice que una variable aleatoria X se distribuye Poisson con parámetro $\lambda > 0$ si su soporte es $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ y su función de probabilidad es

$$f(x) = \frac{\exp(-\lambda)\lambda^x}{x!} \mathbf{1}_{\mathbb{N}_0}(x) \quad (1.7)$$

donde $\exp()$ es la función exponencial, x el número de ocurrencias y $x!$ el factorial de x .

La distribución Poisson toma su nombre de Siméon Denis Poisson, un científico francés que la introdujo a mediados del siglo XVII. Es usada principalmente para modelar el número

de veces que un evento ocurre en un intervalo de tiempo o espacio (*Spoiler*: ¡como los goles en un partido de fútbol!).

Si X es Poisson, entonces $f(x) \geq 0$ para todo $x \in \mathbb{N}_0$, pues la función exponencial tiene un rango positivo, y λ y x son no negativos. La segunda propiedad también se cumple si se recuerda el desarrollo de Taylor de la función exponencial alrededor de 0.

$$\begin{aligned} \sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} \frac{\exp(-\lambda)\lambda^x}{x!} \\ &= \exp(-\lambda) \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= \exp(-\lambda) \exp(\lambda) \\ &= 1. \end{aligned}$$

Para calcular la esperanza de una v.a. Poisson, se sigue un procedimiento análogo:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=0}^{\infty} x \frac{\exp(-\lambda)\lambda^x}{x!} \\ &= \lambda \exp(-\lambda) \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= \lambda \exp(-\lambda) \exp(\lambda) \\ &= \lambda. \end{aligned}$$

Por otro lado, para la varianza se calculará primero $\mathbb{E}(X^2)$

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{\exp(-\lambda)\lambda^x}{x!} \\
&= \lambda \sum_{x=1}^{\infty} x \frac{\exp(-\lambda)\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} \frac{(y+1) \exp(-\lambda)\lambda^y}{(y)!} \\
&= \lambda \left[\sum_{y=0}^{\infty} \frac{y \exp(-\lambda)\lambda^y}{(y)!} + \sum_{y=0}^{\infty} \frac{\exp(-\lambda)\lambda^y}{(y)!} \right] \\
&= \lambda(\lambda + 1).
\end{aligned}$$

Finalmente, para el cálculo de la varianza se tiene

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda. \quad (1.8)$$

Entonces la distribución Poisson tiene la útil propiedad de que su parámetro es igual tanto a su esperanza como a su varianza.

Para calcular la función de distribución de una variable Poisson, además de seguir la definición se tiene la siguiente fórmula recursiva.

Teorema 1.5.1. *Formula recursiva para $F(x)$ Poisson*

Si X es Poisson(λ) entonces

$$\frac{\mathbb{P}(X = x + 1)}{\mathbb{P}(X = x)} = \frac{\exp(-\lambda)\lambda^{x+1}/(x+1)!}{\exp(-\lambda)\lambda^x/x!} = \frac{\lambda}{x+1}$$

Como $\mathbb{P}(X = 0) = \exp(-\lambda)$ entonces se puede obtener $\mathbb{P}(X = 1)$ a partir de esta, y hacer tantos pasos como se requiera:

$$\mathbb{P}(X = x + 1) = \frac{\lambda}{x + 1} \mathbb{P}(X = x). \quad (1.9)$$

1.6. Inferencia estadística

La inferencia estadística, explicada coloquialmente, es la rama de las matemáticas que tiene como propósito unir datos observados en fenómenos o modelos en el mundo real con distribuciones de probabilidad conocidas. Para una presentación completa de este tópico se recomienda el libro “Inferencia estadística para estudiantes de ciencias” (4) en el cuál se fundamenta esta sección.

Si se tiene una población donde se considera una variable aleatoria X con distribución conocida pero parámetro θ desconocido es posible conseguir la mayor información posible de θ en $f_X(x, \theta)$. De la población se pueden extraer las primeras n observaciones, a esto se le conoce como muestra aleatoria: X_1, \dots, X_n y a partir de los valores muestrales se busca encontrar un estimador para θ , el cuál se denota por $\hat{\theta}$.

1.6.1. Estimación de parámetros

Al procedimiento de encontrar un estimador para un parámetro desconocido de una distribución conocida se le conoce como estimación puntual. En general, se suelen usar dos métodos para encontrar la estimación puntual: método de momentos y método de máxima verosimilitud.

Definición 1.6.1. *Momentos*

Sea X_1, \dots, X_n una muestra aleatoria de una distribución con función de densidad $f(x; \theta)$. A $\mathbb{E}(X_i^r)$ se le conoce como el r -ésimo momento poblacional y se denota por μ_r , mientras que $\frac{\sum_{i=1}^n X_i^r}{n}$ es el r -ésimo momento muestral y se denota por M_r .

El método de momentos consiste en igualar los momentos muestrales con los poblacionales y resolver para θ . Es decir, $\mu_r = M_r$ para $r = 1, \dots, k$ con k el número de parámetros a estimar.

Definición 1.6.2. *Método de estimación de momentos*

De forma general, si $\{X_i\}_{i=1}^n$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas de una población con función de densidad $f(x; \theta_1, \dots, \theta_k)$ se debe resolver el siguiente sistema de ecuaciones:

$$\begin{aligned}\mu_1 &= \mathbb{E}(X^1) = \frac{\sum_{i=1}^n X_i^1}{n} = M_1 \\ \mu_2 &= \mathbb{E}(X^2) = \frac{\sum_{i=1}^n X_i^2}{n} = M_2 \\ &\vdots \\ \mu_k &= \mathbb{E}(X^k) = \frac{\sum_{i=1}^n X_i^k}{n} = M_k\end{aligned}$$

La solución $\hat{\theta} = T(X_1, \dots, X_n)$ al sistema se conoce como el estimador por el método de momentos.

Ejemplo 1.6.1. *Los goles del Real Madrid jugando de local en la temporada 2021-2022 son:*

Fecha	Local	Goles Visitante	Goles
12/09/2021	Real Madrid	5 Celta	2
22/09/2021	Real Madrid	6 Mallorca	1
25/09/2021	Real Madrid	0 Villarreal	0
27/10/2021	Real Madrid	0 Osasuna	0
06/11/2021	Real Madrid	2 Vallecano	1
28/11/2021	Real Madrid	2 Sevilla	1
01/12/2021	Real Madrid	1 Ath Bilbao	0
12/12/2021	Real Madrid	2 Ath Madrid	0
19/12/2021	Real Madrid	0 Cadiz	0

08/01/2022	Real Madrid	4 Valencia	1
23/01/2022	Real Madrid	2 Elche	2
06/02/2022	Real Madrid	1 Granada	0
19/02/2022	Real Madrid	3 Alaves	0
05/03/2022	Real Madrid	4 Sociedad	1
20/03/2022	Real Madrid	0 Barcelona	4
09/04/2022	Real Madrid	2 Getafe	0
30/04/2022	Real Madrid	4 Espanol	0
12/05/2022	Real Madrid	6 Levante	0
20/05/2022	Real Madrid	0 Betis	0

Si se asume que ocurrieron bajo una ley de probabilidad Poisson de parámetro desconocido λ , como $\mathbb{E}(X) = \lambda$ entonces

$$\hat{\lambda} = \frac{5 + 6 + 0 + 0 + 2 + 2 + 1 + 2 + 0 + 4 + 2 + 1 + 3 + 4 + 0 + 2 + 4 + 6 + 0}{19}$$
$$\hat{\lambda} = 2.3158$$

El método de máxima verosimilitud es otra forma de estimar parámetros de poblaciones con leyes de probabilidad conocidas con parámetros desconocidos. La verosimilitud es la cualidad de verosímil, esto se define en la RAE como:

- adj. Que tiene apariencia de verdadero.
- adj. Creíble por no ofrecer carácter alguno de falsedad.

Definición 1.6.3. *Función de máxima verosimilitud*

Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$. La función de verosimilitud es la función conjunta de la muestra y se denota como y $L(\theta)$:

$$L(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_i}(x_i; \theta).$$

Definición 1.6.4. *Método de estimación por máxima verosimilitud*

Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$ y $L(\theta)$ su función de verosimilitud. A $\hat{\theta} = T(X_1, \dots, X_n)$ se le llama estimador máximo verosímil de θ , si satisface que para cualquier $\theta \in \Theta$, se tiene que $L(\hat{\theta}) \geq L(\theta)$.

Para obtener el estimador máximo verosímil se suele usar el siguiente procedimiento. Sea $f(x; \theta_1, \dots, \theta_k)$ una función de densidad con k parámetros. Si $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ satisface el sistema

$$\begin{aligned} \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_1} &= 0 \\ \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_k} &= 0 \end{aligned}$$

entonces $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ es el estimador máximo verosímil de θ . La función $\ln(L(\theta_1, \dots, \theta_k))$ se conoce como la log-verosimilitud de $f(x; \theta_1, \dots, \theta_k)$ y esta alcanza su máximo en el mismo punto que $L(x; \theta_1, \dots, \theta_k)$. Para notar lo anterior observe que para todo $i \in \{1, \dots, k\}$:

$$\frac{\partial \ln(L(\theta_1, \dots, \theta_k))}{\partial \theta_i} = \frac{1}{L(\theta_1, \dots, \theta_k)} \cdot \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_i}$$

por lo que

$$\frac{\partial \ln(L(\theta_1, \dots, \theta_k))}{\partial \theta_i} = 0 \iff \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_i} = 0$$

Por practicidad, se calcula el máximo de $\ln(\theta)$ en vez del de $L(\theta)$.

Ejemplo 1.6.2. Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $\text{Poisson}(\lambda)$. El estimador máximo verosímil de θ se calcula como sigue.

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{\exp(-\theta)\theta^{x_i}}{x_i!} \mathbb{1}_{\mathbb{N} \cup 0}(x_i) = \exp(-n\theta)\theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{\mathbb{1}_{\mathbb{N} \cup 0}(x_i)}{x_i!}.$$

Si se calcula la log-verosimilitud,

$$\begin{aligned} l(\theta) &= \ln(L(\theta)) = \ln \left(\exp(-n\theta)\theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{\mathbb{1}_{\mathbb{N} \cup 0}(x_i)}{x_i!} \right) \\ &= -n\theta + \ln(\theta) \sum_{i=1}^n x_i + \sum_{i=1}^n \ln \left(\frac{\mathbb{1}_{\mathbb{N} \cup 0}(x_i)}{x_i!} \right). \end{aligned}$$

Al calcular la derivada,

$$\frac{\partial l(\theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n x_i}{\theta}.$$

Igualando a cero,

$$\frac{\partial l(\theta)}{\partial \theta} = 0 \iff n = \frac{\sum_{i=1}^n x_i}{\theta} \iff \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

Finalmente se verifica que es un máximo,

$$\frac{\partial^2 l(\hat{\theta})}{\partial \theta^2} = -\frac{\sum_{i=1}^n x_i}{\hat{\theta}^2} < 0.$$

La última desigualdad se da pues el numerador y el denominador son cantidades positivas, pero se multiplican por un menos al calcular la segunda derivada.

1.6.2. Prueba de hipótesis ji-cuadrada

Cuando se cuenta con una serie de datos una de las preguntas que surge inmediatamente es si esa población sigue alguna ley de distribución conocida. En la inferencia estadística se cuenta con una técnica conocida como prueba de hipótesis la cuál responde dicha pregunta.

En nuestro contexto, una hipótesis estadística es una afirmación acerca de la distribución de una variable aleatoria, cuando la hipótesis especifica completamente la distribución se le conoce como hipótesis simple, mientras que si no es simple se le conoce como hipótesis compuesta.

Definición 1.6.5. *Prueba de hipótesis*

Una prueba de hipótesis es una regla de decisión mediante la cual, y con base en la muestra, se puede determinar si se acepta o se rechaza la hipótesis nula en cuestión.

Una prueba de hipótesis consta de dos hipótesis, la hipótesis nula se denota con H_0 y hipótesis alternativa se denota con H_a . La prueba se puede plantear como sigue:

H_0 : Los goles de un equipo de futbol se distribuyen Poisson vs H_a : Los goles del equipo de futbol no se distribuyen Poisson.

De esas aseveraciones el interés se centra en determinar si se puede o no confirmar una, con lo cuál la otra será falsa.

Definición 1.6.6. *Región de rechazo*

A la región C que lleva a rechazar la hipótesis nula se le llama región de rechazo o región crítica

El rol de C es fundamental pues una vez que es especificada basta con tomar una muestra y verificar si se encuentra en la región crítica o no.

Una de los aspectos que se debe tener en cuenta al realizar una prueba de hipótesis es, en la medida de lo posible, minimizar la posibilidad de cometer algún error con la decisión. En particular, existen dos tipos de errores que pueden surgir al realizar una prueba de hipótesis que causarían conclusiones erróneas:

- Error tipo I: Rechazar H_0 cuando H_0 es cierta.

- Error tipo II: No rechazar H_0 cuando H_0 es falsa.

Definición 1.6.7. *Tamaño de los errores*

El tamaño de los errores está dado por

$$\alpha = \mathbb{P}(\text{Error tipo I}) = \mathbb{P}(\text{Rechazar } H_0 | H_0 \text{ cierta})$$

y

$$\beta = \mathbb{P}(\text{Error tipo II}) = \mathbb{P}(\text{No rechazar } H_0 | H_0 \text{ falsa})$$

Además de la región de rechazo, el p -value o nivel de significancia p es una cantidad que se puede usar para tomar una decisión al realizar una prueba de hipótesis.

El p -value es la probabilidad, cuando se supone que la hipótesis nula H_0 es cierta, de obtener un resultado muestral tan extremo o más extremo que el resultado muestral observado.

Para construir una prueba de nivel α basada en el p -value se debe rechazar H_0 si y sólo si $p(x_1, \dots, x_n) \leq \alpha$.

Cuando el p -value es pequeño entonces se dice que la muestra produjo un resultado no usual si H_0 es cierta, esto quiere decir que la afirmación de H_0 es inconsistente con el resultado obtenido a partir de la muestra, por lo que se debería rechazar la hipótesis nula.

Al usar el p -value para tomar una decisión se debe considerar un valor para α , que se conoce como nivel de significancia en este contexto. Si $p \leq \alpha$, la decisión es rechazar H_0 , en caso contrario la decisión es no rechazar H_0 .

Por lo tanto, el p -value es el nivel de significancia más pequeño para el cual la hipótesis nula sería rechazada.

Finalmente, se presentará la prueba de hipótesis Ji-Cuadrada, esta es una prueba de bondad de ajuste, es decir, examina qué tan de acuerdo está una muestra de datos con una distribución dada como su población.

La prueba Ji-Cuadrada, en su forma más simple, considera k celdas, en cada una de las cuales debe caer el resultado de un experimento. Sea p_i , $i = 1, 2, \dots, k$ la probabilidad de que del experimento se obtenga un resultado que caiga en la i -ésima celda con $\sum_{i=1}^k p_i = 1$ y sea n_i , $i = 1, \dots, k$ el número de veces que el resultado cae en la i -ésima celda de un total de $n = \sum_{i=1}^k n_i$ realizaciones del experimento. Se desea probar la hipótesis

$$H_0 = p_i = \pi_i, i = 1, 2, \dots, k$$

donde p_i son las probabilidades de cada celda.

Karl Pearson propuso la estadística de prueba $T = \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i}$ donde e_i es el valor esperado de la celda i , es decir np_i y bajo H_0 , $e_i = n\pi_i$ y demostró que T tiene una distribución Ji-cuadrada con $k - 1$ grados de libertad.

Aplicación de la distribución Poisson al fútbol

En esta sección se introducirá el modelo predictor clásico de Maher y el modelo predictor modificado de Dixon-Coles. Se expondrán los resultados al aplicar el modelo a un caso particular considerando las principales ligas de la actualidad, y se dejarán las bases para aquellos lectores que quieran adentrarse más en el tema.

Los modelos, tienen el supuesto de que los goles de un equipo cualquiera, en un partido dado, siguen una distribución Poisson.

Durante los últimos años, las justificaciones para ese supuesto recaen en que los goles pueden caer en cualquier minuto del partido.

Si bien la probabilidad p de que una jugada ofensiva termine en anotación es baja, el hecho de que se tengan al menos 90 minutos para acumular ataques explica mejor esta estimación: si p es constante y los ataques son independientes, el número de goles será Binomial (“número de éxitos en intentos Bernoulli”) y la aproximación Poisson se aplicaría correctamente.

Entonces, la media de la distribución Poisson asociada a cada equipo en una competencia, tiene que considerar la *calidad* del equipo. Uno puede aventurarse más, y considerar que la media de dicha distribución deba ajustarse partido a partido, o deba tener un mecanismo de ponderación, buscando dar más importancia a los partidos más recientes.

2.1. Modelo de Maher

El modelo de Maher (2) asume que los goles de los dos equipos siguen una distribución Poisson.

Definición 2.1.1. *Modelo de Maher Si el equipo i esta jugando de local contra el equipo j , y el resultado observado es (x_{ij}, y_{ij}) , se asume que X_{ij} se distribuye Poisson con media $\alpha_i \beta_j$, que Y_{ij} se distribuye Poisson con media $\gamma_i \delta_j$, y que X_{ij} y Y_{ij} son independientes.*

Es intuitivo pensar a α_i como la representación numérica del poder ofensivo del equipo i cuando juega de local, β_j el poder defensivo del equipo j cuando juega de visitante, γ_i el poder defensivo del equipo i cuando juega de local y δ_j el poder ofensivo del equipo j jugando de visitante.

Es decir, cada equipo tiene 4 parámetros importantes a considerar. Como \mathbf{X} y \mathbf{Y} son independientes, las estimaciones de α y β se obtendrán directamente de los valores observados \mathbf{x} , mientras que la estimación de γ y δ de los valores observados de \mathbf{y} .

Usando el método de máxima verosimilitud, se tienen los siguientes estimadores:

Teorema 2.1.1. *Estimadores del modelo de Maher*

Para el modelo de Maher, considerando en general un partido entre los equipos i y j , los estimadores para los goles del equipo local (X_{ij}) están dados por

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j}$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.$$

Los estimadores para los goles del equipo visitante Y_{ij} están dados por:

$$\hat{\gamma}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\delta}_j}$$

$$\hat{\delta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\gamma}_i}.$$

Demostración. Se trabajará con la variable aleatoria X_{ij} , ya que el desarrollo para Y_{ij} es análogo.

Considerando el supuesto de que todos los partidos son independientes entre si, se tiene que la función de verosimilitud de una muestra aleatoria Poisson de parámetro λ esta dada por:

$$L(\lambda; x_1, \dots, x_n) = \prod_{j=1}^n f_X(x_j; \lambda)$$

$$L(\lambda; x_1, \dots, x_n) = \prod_{j=1}^n \exp(-\lambda) \frac{1}{x_j!} \lambda^{x_j}.$$

La función de log-verosimilitud esta dada por:

$$\begin{aligned} \ln L(\lambda; x_1, \dots, x_n) &= \ln \left(\prod_{j=1}^n f_X(x_j; \lambda) \right) \\ &= \sum_{j=1}^n \ln \left(\exp(-\lambda) \frac{1}{x_j!} \lambda^{x_j} \right) \\ &= \sum_{j=1}^n [\ln(\exp(-\lambda)) - \ln(x_j!) + \ln(\lambda^{x_j})] \\ &= \sum_{j=1}^n (-\lambda - \ln(x_j!) + x_j \ln(\lambda)). \end{aligned}$$

Volviendo al problema estudiado, la función de log-verosimilitud esta dada por:

$$\ln L(\alpha_i, \beta_j) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j - \ln(x_{ij}!) + x_{ij} \ln(\alpha_i \beta_j)).$$

Derivando respecto a α_i y β_j .

$$\begin{aligned}\frac{\partial \ln L}{\partial \alpha_i} &= \sum_{j \neq i} \left(-\beta_j + \frac{x_{ij}}{\alpha_i} \right) \\ \frac{\partial \ln L}{\partial \beta_j} &= \sum_{i \neq j} \left(-\alpha_i + \frac{x_{ij}}{\beta_j} \right).\end{aligned}$$

Igualando a 0, se obtienen los estimadores máximo verosímiles:

$$\begin{aligned}\hat{\alpha}_i &= \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \\ \hat{\beta}_j &= \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.\end{aligned}$$

Siguiendo un procedimiento análogo, se obtienen los estimadores para γ_i y δ_j . □

Maher también propone un modelo más simple, pues dada la información con la que contaba en el momento, la segmentación de poderes ofensivos y defensivos en local y visitante, no aportaba información significativa al modelo.

Definición 2.1.2. *Segundo modelo de Maher*

Si el equipo i está jugando de local contra el equipo j , y el resultado observado es (x_{ij}, y_{ij}) , se asume que X_{ij} se distribuye Poisson con media $\alpha_i \beta_j$, que Y_{ij} se distribuye Poisson con media $\gamma_i \delta_j$, y que X_{ij} y Y_{ij} son independientes. Con $\gamma_i = k \beta_i$, $\delta_j = k \alpha_j$ y $\sum \alpha_i = \sum \beta_i$.

La diferencia radica en que, si bien se sigue teniendo un factor de ajuste k para la ventaja de los equipos locales sobre los visitantes, este será el mismo para todos los equipos.

Observación. *Para el segundo modelo de Maher, los estimadores de los parámetros obtenidos por el método de máxima verosimilitud están dados por:*

$$\begin{aligned}\hat{\alpha}_i &= \frac{\sum_{j \neq i} (x_{ij} + y_{ji})}{(1 + \hat{k}^2) \sum_{j \neq i} \hat{\beta}_j} \\ \hat{\beta}_j &= \frac{\sum_{i \neq j} (x_{ij} + y_{ji})}{(1 + \hat{k}^2) \sum_{i \neq j} \hat{\alpha}_i} \\ \hat{k}^2 &= \frac{\sum_i \sum_{j \neq i} y_{ij}}{\sum_i \sum_{j \neq i} x_{ij}}.\end{aligned}$$

Es decir, Maher define el factor de ventaja de los equipos locales, como el total de goles anotados en todos los partidos por los equipos visitantes, entre el total de goles anotados en todos los partidos por los equipos locales.

2.2. Modelo de Dixon-Coles

Partiendo del modelo de Maher, Dixon y Coles (1) presentan una propuesta de modelo con el objetivo de predecir resultados de partidos de futbol, aceptando que el modelo general tiene detalles que no son tan precisos, pero conservando la estructura global pues es adecuada para modelar este fenómeno.

Definición 2.2.1. Modelo de Dixon-Coles

Si el equipo i esta jugando de local contra el equipo j , sean $X_{i,j}$ y $Y_{i,j}$ las variables aleatorias que modelan el número de goles anotados por los equipos local y visitante respectivamente; el resultado observado es $(x_{i,j}, y_{i,j})$ entonces

$$\mathbb{P}(X_{i,j} = x, Y_{i,j} = y) = \tau_{\lambda, \mu}(x, y) \frac{\lambda^x \exp(-\lambda)}{x!} \frac{\mu^y \exp(-\mu)}{y!} \quad (2.1)$$

con

$$\lambda = \alpha_i \beta_j \gamma$$

$$\mu = \alpha_j \beta_i$$

y

$$\tau_{\lambda,\mu}(x, y) = \begin{cases} 1 - \lambda\mu\rho & \text{Si } x = y = 0 \\ 1 + \lambda\rho & \text{Si } x = 0, y = 1 \\ 1 + \mu\rho & \text{Si } x = 1, y = 0 \\ 1 - \rho & \text{Si } x = 1, y = 1 \\ 1 & \text{en otro caso.} \end{cases}$$

Se asume que $X_{i,j}$ y $Y_{i,j}$ son independientes, $\alpha_i, \beta_i > 0, \forall i$, α_i mide el poder ofensivo del equipo i mientras que β_i mide el poder defensivo del equipo i y $\gamma > 0$ es el parámetro del efecto de la ventaja del equipo local.

El parámetro ρ tiene la siguiente restricción

$$\max\left(\frac{-1}{\lambda}, \frac{-1}{\mu}\right) \leq \rho \leq \min\left(\frac{1}{\lambda\mu}, 1\right)$$

y entra como un parámetro de dependencia: $\rho = 0$ corresponde a la independencia entre los equipos, mientras que cualquier otro valor causa que la independencia de la distribución esté perturbada para eventos donde $x \leq 1$ o $y \leq 1$.

Las distribuciones marginales de X y Y continúan siendo Poisson con parámetros λ y μ , respectivamente.

Se sigue de 2.1 que en una liga con n equipos hay parámetros de ataque $(\alpha_1, \dots, \alpha_n)$, parámetros de defensa $(\beta_1, \dots, \beta_n)$, el parámetro de dependencia ρ y el parámetro de ventaja del local γ a ser estimados. Basta suponer la siguiente restricción, para que el modelo no este *sobre parametrizado*:

$$\sum_{i=1}^n \frac{\alpha_i}{n} = 1.$$

Si se etiquetan los partidos de 1 hasta t , y se denotan los correspondientes marcadores con

(x_k, y_k) , la función de verosimilitud tiene la siguiente forma:

$$L(\alpha_i \beta_i \rho, \gamma; i = 1, \dots, t) = \prod_{i=1}^t \tau_{\lambda_k \mu_k}(x_k, y_k) \exp(-\lambda_k) \lambda_k^{(x_k)} \exp(-\mu_k) \mu_k^{(y_k)}$$

donde

$$\lambda_k = \alpha_{i(k)} \beta_{j(k)} \gamma$$

$$\mu_k = \alpha_{j(k)} \beta_{i(k)} \gamma$$

con $i(k)$ y $j(k)$ denotan respectivamente los índices de los equipos local y visitante que se enfrentaron en el partido $k \in \{1, \dots, t\}$. En el modelo de Dixon y Coles, a diferencia del modelo de Maher, plantear un sistema de ecuaciones para obtener los estimadores no es trivial por lo que, sin pérdida de generalidad, los autores proponen utilizar métodos numéricos para maximizar la función de verosimilitud y de este modo encontrar los estimadores de los parámetros del modelo.

2.3. Un modelo con evolución dinámica

En esta sección se propondrá el planteamiento para poder desarrollar un modelo alternativo de predicción de resultados, asumiendo una naturaleza en la que los parámetros cambiarán después de cada partido. Se conservará el supuesto de que los goles de los equipos siguen una ley de probabilidad Poisson.

Suponiendo un campeonato con n equipos, en el cuál se enfrentan todos contra todos a visita recíproca, se tendrá que para el partido entre el equipo local i y el equipo visitante j se tendrán las variables aleatorias $X_{i,j}$, $Y_{i,j}$ que modelan los goles anotados por los equipos local y visitante respectivamente; el resultado observado es $(x_{i,j}, y_{i,j})$.

La diferencia de este modelo a los anteriores, es la forma en la que se define el parámetro de las variables aleatorias Poisson usadas en cada enfrentamiento. Sean $\alpha_{i,k}$, $\beta_{i,k}$, $\alpha_{j,k}$, $\beta_{j,k}$ los

poderes ofensivos y defensivos de los equipos i y j en el enfrentamiento k . Estos parámetros se podrán determinar de forma recursiva, de modo que a medida que avance los poderes ofensivos y defensivos de los equipos irán cambiando dinámicamente a lo largo del campeonato. Para cada partido, la variable que modela los goles de los equipos local y visitante tendrán como parámetro el producto entre los poderes ofensivo y defensivo de los equipos en cuestión, incorporando un ajuste para incorporar la ventaja del equipo local. Los parámetros para el equipo local:

$$X_{i,j,k} \sim \text{Poisson}(\lambda_{i,j,k}) \text{ con } \lambda_{i,j,k} = \alpha_{i,k} \beta_{j,k} \gamma.$$

Los parámetros para el equipo visitante:

$$Y_{i,j,k} \sim \text{Poisson}(\mu_{i,j,k}) \text{ con } \mu_{i,j,k} = \alpha_{j,k} \beta_{i,k}.$$

El cálculo recursivo para determinar los parámetros ofensivo y defensivo se define como sigue, si $k \geq 2$:

$$\alpha_{i,k} = m \bar{\alpha}_{i,k} + (1 - m) \alpha_{i,k-1}.$$

$$\alpha_{i,1} = \alpha_{i,0}.$$

$$\beta_{i,k} = m \bar{\beta}_{i,k} + (1 - m) \beta_{i,k-1}.$$

$$\beta_{i,1} = \beta_{i,0}.$$

Donde $\bar{\alpha}_{i,k} = \sum_{j=1}^{k-1} \frac{(\alpha_{i,j})}{k-1}$ y $\bar{\beta}_{i,k} = \sum_{j=1}^{k-1} \frac{(\beta_{i,j})}{k-1}$. De lo anterior es importante observar que para cualquier equipo s , α_s y β_s son variables aleatorias con una distribución desconocida, pero que con base en un valor inicial se puede ir construyendo el siguiente valor.

El valor m tiene la función de ponderador, por lo que estará restringido a tomar valores en el intervalo $[0, 1]$. Conforme m tiende a 1, la historia del equipo tiene una influencia mayor

para determinar el poder ofensivo más reciente; por otro lado, conforme m tiende a 0 se le da cada vez menos peso a la historia del equipo y más peso al último resultado.

Finalmente, para determinar los valores iniciales de cada equipo se sugiere tomar en consideración no sólo los resultados de la temporada anterior, sino también factores cualitativos que influyan en el desempeño de un equipo de una temporada a otra, como pudiera ser cambio de entrenador, nuevos fichajes, cambios en las reglas de la competición, entre otras.

El objetivo de este trabajo fue el divulgar, tanto al público en general (Probabilidad, distribución Poisson) como a los estudiantes de la Facultad de Ciencias (Proceso Poisson Puntual) temas cruciales con una aplicación a un fenómeno que es conocido universalmente: el fútbol.

Aunque en ese trabajo se enfocó la distribución Poisson a dicho fenómeno, esta distribución se ha aplicado exitosamente a problemas de múltiples áreas.

La comparación de los modelos de Maher y Dixon Coles, aunque no da resultados idénticos a los observados, si genera mucha información. En particular, se ha observado que la cantidad de partidos acertados es amplia, pues la tendencia de cada liga se ha mantenido.

Esto es crucial, pues aunque en cada enfrentamiento normalmente suele haber un "favorito", si los equipos tienen a priori la información de la fuerza ofensiva y defensiva de sus contrincantes, podrían preparar cada partido con base en ello. Además, las asociaciones de fútbol podrían incorporar dicha información al proceso de realizar los calendarios de las competiciones, buscando que los partidos más atractivos queden separados en el tiempo buscando mantener la atención del público neutral. Finalmente, se pueden usar estos modelos como base para realizar una estrategia de apuestas, bajo el riesgo del lector.

A.1. Proceso Poisson Puntual

Esta sección sigue un nivel más avanzado que las dos anteriores pues cubre un tema que no se incluye en los cursos de Licenciatura en la Facultad de Ciencias, esta basado en el libro “Poisson Point Processes Imaging, Tracking, and Sensing” de R. L. Streit ([3](#)).

Un proceso puntual es una variable aleatoria cuyas realizaciones son conjuntos aleatorios, estos conjuntos contienen únicamente puntos de un espacio de estados S .

Un proceso puntual finito es una variable aleatoria cuyas realizaciones en un subconjunto acotado $R \subset S$ son conjuntos que contienen únicamente un número finito de puntos de R .

El número de puntos y sus ubicaciones pueden elegirse de muchas maneras, por ejemplo, existen los procesos puntuales finitos con puntos independientes e idénticamente distribuidos.

Un Proceso Poisson Puntual (PPP) es aquel cuyos número de puntos se distribuye Poisson en los enteros no negativos $\{0, 1, \dots\}$. El parametro de la distribución Poisson en un subconjunto R esta relacionado con una variable aleatoria espacial $X|R$ restringida a (condicionada en) R . En este caso, la función de densidad de probabilidad de $X|R$ es proporcional a una función no negativa llamada función de intensidad, y la constante de proporcionalidad es el

número de puntos esperados en la realización.

Los PPP son una clase estrecha de procesos puntuales finitos independientes e idénticamente distribuidos. La propiedad que les da una importancia única es que los PPP en subconjuntos disjuntos de S son estadísticamente independientes (esta propiedad, en el contexto de los procesos estocásticos se conoce como incrementos independientes).

A.1.1. Espacio de estados

Los puntos obtenidos en las realizaciones de los PPPs con una función de intensidad definida en un espacio continuo son distintos con probabilidad 1. Los puntos de una realización de un PPP también son conocidos como listas o conjuntos finitos aleatorios.

Los puntos de un PPP ocurren en un espacio de estados S , usualmente S es el espacio euclidiano, $S = \mathbb{R}^m$, $m \geq 1$, o algún subconjunto de este.

Las realizaciones de PPPs en un subconjunto R de S constan de el número $n \geq 0$ y las ubicaciones x_1, x_2, \dots, x_n de los puntos en R . Una realización se denota por el par ordenado

$$\xi = (n, \{x_1, \dots, x_n\}). \quad (\text{A.1})$$

Se usa una notación de conjuntos para indicar que el orden de los x_j es irrelevante, sin embargo los puntos no son necesariamente distintos, por lo que debe pensarse a $\{x_1, \dots, x_n\}$ como una lista sin orden. Otra observación importante es que si bien n se incluye explícitamente en ξ , esta determinada por el tamaño de la lista $\{x_1, \dots, x_n\}$.

Si $n = 0$ entonces ξ es el evento trivial $(0, \emptyset)$, con \emptyset el conjunto vacío.

Definición A.1.1. *Espacio de eventos*

El espacio de eventos es la colección de todos los posibles subconjuntos finitos de R :

$$E(R) = \{(0, \emptyset)\} \bigcup_{n=1}^{\infty} \left\{ (n, \{x_1, \dots, x_n\}) \quad : \quad x_j \in R, \quad j = 1, \dots, n \right\}. \quad (A.2)$$

A.1.2. Intensidad

Todo PPP esta parametrizado por una cantidad denominada intensidad.

Definición A.1.2. *Intensidad del PPP*

La función de intensidad generalizada para un PPP general en el espacio continuo S es

$$\lambda_D(s) = \lambda(s) + \sum_{j=1}^{\infty} w_j \delta(s - a_j), \quad s \in S, \quad (A.3)$$

donde $\lambda(s)$ es la función de intensidad, $\delta(\cdot)$ es la función delta de Dirac y, para toda j , los pesos w_j son no negativos y los puntos $a_j \in S$ son distintos: $a_i \neq a_j$ si $i \neq j$.

Aunque la intensidad suele ser un concepto intuitivo, puede tomar distintas formas matemáticas dependiendo en el espacio de estados S .

De acuerdo a su función de intensidad, un PPP puede caracterizarse como ordenado, homogéneo o no homogéneo:

Definición A.1.3. *PPP ordenado*

Un PPP en un espacio continuo $S \subset \mathbb{R}^m$ es ordenado si la intensidad es una función $\lambda(s) \geq 0$ para toda $s \in S$.

Definición A.1.4. *PPP homogéneo*

Si $\lambda(s) \equiv \alpha$ para una constante $\alpha \geq 0$, entonces el PPP se denomina homogéneo.

Definición A.1.5. *PPP no homogéneo*

Si

$$0 \leq \int_{\mathbf{R}} \lambda(s) ds < \infty$$

para todos los subconjuntos acotados \mathbf{R} de \mathbf{S} .

A.1.3. Realizaciones

Se considerará un PPP ordenado, es decir, un PPP con una función de intensidad bien definida en un espacio continuo $\mathbf{S} \subset \mathbb{R}^m$.

Las realizaciones son conceptualmente de simular para subconjuntos acotados de espacios continuos $\mathbf{S} \subset \mathbb{R}^m$. Se puede pensar a los subconjuntos acotados como “ventanas” en las cuales las realizaciones del PPP son observadas.

Cada realización de un PPP en un conjunto acotado \mathbf{R} es un elemento del espacio de eventos $\mathcal{E}(\mathbf{R})$. La realización ξ por lo tanto se compone del número $n \geq 0$ y las ubicaciones $\{x_1, \dots, x_n\}$ de los puntos en \mathbf{R} .

Un proceso de dos pasos, uno discreto y otro continuo, permite simular una realización $\xi \in \mathcal{E}(\mathbf{R})$ de un PPP no homogéneo con intensidad $\lambda(s)$ en un subconjunto acotado \mathbf{R} de \mathbf{S} . El procedimiento también sirve para ilustrar la estructura estadística básica del PPP. Si $\int_{\mathbf{R}} \lambda(s) ds = 0$, entonces ξ es el evento trivial. Si $\int_{\mathbf{R}} \lambda(s) ds > 0$, la realización se obtiene como sigue:

1. El número $n \geq 0$ de puntos es determinado mediante el muestreo de una variable aleatoria discreta Poisson, denotada por N , con función de masa de probabilidad dada por

$$p_N(n) = \frac{(\int_{\mathbf{R}} \lambda(s) ds)^n}{n!} \exp \left(- \int_{\mathbf{R}} \lambda(s) ds \right) \quad (\text{A.4})$$

Si $n = 0$, la realización es $\xi = (0, \emptyset)$ y el paso 2 no se ejecuta.

2. Los n puntos $x_j \in \mathbb{R}$, $j = 1, \dots, n$, son obtenidos como realizaciones independientes e idénticamente distribuidas de una variable aleatoria X en \mathbb{R} con función de densidad de probabilidad dada por

$$p_X(s) = \frac{\lambda(s)}{\int_{\mathbb{R}} \lambda(s) ds} \text{ para } s \in \mathbb{R}. \quad (\text{A.5})$$

El resultado es el par ordenado $\xi_0 = (n, (x_1, \dots, x_n))$. Si se sustituye la lista ordenada (x_1, \dots, x_n) con la lista $\{x_1, \dots, x_n\}$ da la realización del PPP $\xi = (n, \{x_1, \dots, x_n\})$.

La distinción entre ξ_0 y ξ se hace para evitar confusiones cuando se trabajen problemas donde el orden sea importante.

Para espacios continuos $S \subset \mathbb{R}^m$ y como consecuencia inmediata del paso 2, se tiene que los puntos $\{x_1, \dots, x_n\}$ son distintos con probabilidad uno. En teoría elementos repetidos están permitidos, pero en la práctica esto nunca ocurre (con probabilidad uno). Otra forma de decir lo anterior es que la lista $\{x_1, \dots, x_n\}$ es un conjunto con probabilidad uno. Esta afirmación falla cuando el PPP no está ordenado, esto es, cuando la intensidad generalizada tiene uno o más componentes de la función Delta de Dirac.

A.1.4. Función de verosimilitud

La variable aleatoria Ξ con realizaciones en $\mathcal{E}(\mathbb{R})$ para todo subconjunto acotado \mathbb{R} de S es un PPP si sus realizaciones son generadas mediante el procedimiento de dos pasos [A.1.3](#). Sea $p_{\Xi}(\xi)$ la función de densidad de probabilidad de Ξ evaluada en $\Xi = \xi$. Sea $\Xi \equiv (N, \mathcal{X})$, donde N es el número de puntos y $\mathcal{X} = \{x_1, \dots, x_N\}$ es el conjunto de puntos. Sean $\xi = (n, \{x_1, \dots, x_n\})$ las realizaciones. Se tiene entonces condicionando,

$$p_{\Xi}(\xi) = p_N(n) p_{\mathcal{X}|N}(\{x_1, \dots, x_n\} | n), \quad (\text{A.6})$$

donde $p_N(n)$ es la función de masa de probabilidad de N . La función de densidad condicional de $\mathcal{X}|N$ es

$$p_{\mathcal{X}|N}(\{x_1, \dots, x_n\}|n) = n! \prod_{j=1}^n p_X(x_j), \quad (\text{A.7})$$

donde X es la variable aleatoria correspondiente a un punto único que sigue una distribución $p_X(s)$. El $n!$ resulta del hecho de que hay $n!$ ensayos igualmente ordenados, independientes e idénticamente distribuidos que generan el conjunto no ordenado \mathcal{X} . Por lo tanto, haciendo sustituciones se tiene que la función de densidad de Ξ evaluada en $\xi = n, \{x_1, \dots, x_n\} \in \mathcal{E}(\mathbf{R})$:

$$\begin{aligned} p_{\Xi}(\xi) &= p_N(n) p_{\mathcal{X}|N}(\{x_1, \dots, x_n\}|n) \\ &= \exp\left(-\int_{\mathbf{R}} \lambda(s) ds\right) \frac{(\int_{\mathbf{R}} \lambda(s) ds)^n}{n!} n! \prod_{j=1}^n \frac{\lambda(x_j)}{\int_{\mathbf{R}} \lambda(s) ds} \\ &= \exp\left(-\int_{\mathbf{R}} \lambda(s) ds\right) \prod_{j=1}^n \lambda(x_j) \end{aligned} \quad (\text{A.8})$$

A.1.5. Esperanzas

Las esperanzas se obtienen de funciones reales \mathcal{F} definidas en el espacio de eventos $\mathcal{E}(\mathbf{R})$, donde \mathbf{R} es un subconjunto acotado de \mathbf{S} . Por lo tanto, $\mathcal{F}(\xi)$ mapea a un número real a cada $\xi \in \mathcal{E}(\mathbf{R})$. La esperanza de $\mathcal{F}(\xi)$ en su forma más general esta dada por

$$\mathbb{E}(\mathcal{F}) \equiv \mathbb{E}_{\Xi}(\mathcal{F}) = \sum_{\xi \in \mathcal{E}(\mathbf{R})} \mathcal{F}(\xi) p_{\Xi}(\xi), \quad (\text{A.9})$$

donde la suma está bien definida y relacionada con la función de verosimilitud del proceso puntual. En el caso de un PPP, la verosimilitud es conocida. Dicha suma es conocida como un “promedio conjunto” sobre todas las realizaciones del proceso puntual.

Sin embargo, la suma es desalentadora debido al gran tamaño del conjunto $\mathcal{E}(\mathbf{R})$.

A.1.5.1. Cálculo de la esperanza para el PPP

Sea $\xi = (n, \{x_1, \dots, x_n\})$, para usos analíticos es conveniente reescribir la función $\mathcal{F}(\xi) = \mathcal{F}(n, \{x_1, \dots, x_n\})$ en una función que use un argumento más fácil de manejar, es decir, sea

$$\mathcal{F}(n, \{x_1, \dots, x_n\}) \equiv F(n, x_1, \dots, x_n). \quad (\text{A.10})$$

La función F hereda una propiedad importante de simetría de \mathcal{F} . Sea $S(n)$ el conjunto de todas las permutaciones en los primeros n enteros. Para todas las permutaciones $\sigma \in S(n)$ se tiene

$$\begin{aligned} F(n, x_{\sigma(1)}, \dots, x_{\sigma(n)}) &= \mathcal{F}(n, \{x_{\sigma(1)}, \dots, x_{\sigma(n)}\}) \\ &= \mathcal{F}(n, \{x_1, \dots, x_n\}) \\ &= F(n, x_1, \dots, x_n). \end{aligned} \quad (\text{A.11})$$

Es decir, $F(n, x_1, \dots, x_n)$ es simétrica, o invariante bajo permutaciones de sus parámetros de ubicación.

Si los parámetros están en forma ordenada, entonces la esperanza de F esta dada por

$$\mathbb{E}(F) = \sum_{n, x_1, \dots, x_n} F(n, x_1, \dots, x_n) p_{\Xi}(n, x_1, \dots, x_n). \quad (\text{A.12})$$

La suma anterior es una suma discreta-continua y su interpretación es la siguiente. La factorización condicional

$$p_{\Xi}(\xi) = p_N(n) p_{|N}(x_1, \dots, x_n | n) \quad (\text{A.13})$$

de la realización ordenada $\xi = (n, x_1, \dots, x_n)$ es la clave para la suma sobre n , al hacerla la suma más extrema, y para la parte continua se hace de forma natural como integrales sobre conjuntos de la forma $\mathbb{R} \times \dots \times \mathbb{R}$. Así, se tiene que la esperanza de F son un par de esperanzas anidadas. La primera, \mathbb{E}_N es sobre N , y la segunda $\mathbb{E}_{\mathcal{X}|N}$ es sobre $\mathcal{X}|N$. Por lo

tanto, la esperanza con respecto al proceso puntual esta dada por

$$\begin{aligned}\mathbb{E}(F) &= \mathbb{E}_N(\mathbb{E}_{X|N}(F)) \\ &= \sum_{n=0}^{\infty} p_N(n) \int_{\mathbf{R}} \cdots \int_{\mathbf{R}} F(n, x_1, \dots, x_n) p_{X|N}(x_1, \dots, x_n | n) dx_1 \cdots dx_n\end{aligned}\tag{A.14}$$

La esperanza de funciones no simétricas no esta definida, la definición se extiende formalmente a funciones más generales, llámense $G(n, x_1, \dots, x_n)$ usando la versión “simetrizada”

$$G_{Sim}(n, x_1, \dots, x_n) = \frac{1}{n!} \sum_{\sigma \in Sim(n)} G(n, x_{\sigma_1}, \dots, x_{\sigma_n}).\tag{A.15}$$

La esperanza de G esta definida por $\mathbb{E}(G) = \mathbb{E}(G_{Sim})$. Esta definición funciona porque G_{Sim} es una función simétrica de sus argumentos. La definición es compatible con la definición para funciones simétricas, ya que $G_{Sim}(n, x_1, \dots, x_n) \equiv G(n, x_1, \dots, x_n)$ si G es simétrica. La esperanza definida anteriormente para cualquier proceso puntual finito con eventos en $\mathbf{E}(R)$, no solo PPPs. Para los PPP y otros procesos puntuales finitos independientes e idénticamente distribuidos se tiene que

$$p_{\xi|N}(x_1, \dots, x_n | n) = \prod_{j=1}^n p_X(x_j)\tag{A.16}$$

por lo que la esperanza es

$$\mathbb{E}(F) \equiv \sum_{n=0}^{\infty} p_N(n) \int_{\mathbf{R}} \cdots \int_{\mathbf{R}} F(n, x_1, \dots, x_n) \prod_{j=1}^n p_X(x_j) dx_1 \cdots dx_n.\tag{A.17}$$

Para el caso de los PPP, se conoce la distribución de probabilidad discreta de $p_N(n)$ y la función de probabilidad de $p_X(x)$.

El número esperado de puntos en \mathbf{R} es $\mathbb{E}(N(\mathbf{R}))$. Cuando el contexto claramente identifica el conjunto \mathbf{R} , esta esperanza sólo se escribe como $\mathbb{E}(N)$. Al sustituir $F(n, x_1, \dots, x_n) \equiv n$ y observando que todas las integrales integran 1, se sigue inmediatamente que:

$$\mathbb{E}(N) = \sum_{n=0}^{\infty} np_N(n) = \int_{\mathbb{R}} \lambda(s) ds. \quad (\text{A.18})$$

De forma similar, la varianza:

$$\mathbb{V}(N) = \sum_{n=0}^{\infty} (n)(n - \mathbb{E}(N))^2 p_N(n) = \int_{\mathbb{R}} \lambda(s) ds. \quad (\text{A.19})$$

A.1.5.2. Sumas aleatorias

Calcular esperanzas suele presentar dificultades significativas para muchas elecciones de la función F . Estas son, afortunadamente, dos clases importantes de funciones que se simplifican considerablemente. La primera clase comprende funciones llamadas sumas aleatorias. Estas son especialmente útiles en física y procesamiento de señales. Las esperanzas de las sumas aleatorias se reducen a una integral ordinaria sobre \mathbb{R} .

Sea $f(x)$ una función dada real. La variable aleatoria

$$F(\Xi) = \sum_{j=1}^N f(X_j) \quad (\text{A.20})$$

es llamada una suma aleatoria. Dada una realización $\Xi = \xi$, una realización de la suma aleatoria está dada por

$$F(n, x_1, \dots, x_n) = \sum_{j=1}^n f(x_j), \text{ para } n \geq 1, \quad (\text{A.21})$$

y, para $n = 0$, por $F(0, \emptyset) = 0$. El caso especial de la ecuación anterior para el cual $f(x) = 1$ se reduce a $F(n, x_1, \dots, x_n) = n$, el número de puntos en \mathbb{R} . La media de F está dada por

$$\mathbb{E}(F) = \mathbb{E} \left(\sum_{j=1}^N f(X_j) \right) = \int_{\mathbb{R}} f(x) \lambda(x) (dx). \quad (\text{A.22})$$

La esperanza de la ecuación anterior se obtiene tras sustituir A.21 en A.17 y al intercambiar la suma sobre j y las integrales sobre \mathbf{R} da

$$\mathbb{E}(F) = \sum_{n=0}^{\infty} p_N(n) \sum_{j=1}^n \int_{\mathbf{R}} \dots \int_{\mathbf{R}} f(x_j) \prod_{j=1}^n p_x(x_j) dx_1 \dots dx_n.$$

De las integrales, todas excepto una valen 1, por lo que

$$\mathbb{E}(F) = \sum_{n=0}^{\infty} p_N(n) n \int_{\mathbf{R}} f(x) p_x(x) dx.$$

Al sustituir A.4 y A.5 y simplificando se tiene A.22.

A.1.6. Teorema de Campbell

El teorema de Campbell es el resultado más famoso para sumas aleatorias, ya que da la función característica para las sumas aleatorias de la forma A.21.

Bajo ciertos supuestos, el teorema de Campbell dice que cuando θ es puramente imaginario,

$$\mathbb{E}(\exp(\theta F)) = \exp\left(\int_{\mathbf{R}} ([\exp(\theta f(x))] - 1) \lambda(x) dx\right), \quad (\text{A.23})$$

donde $f(x)$ es una función real. La esperanza existe para cualquier complejo θ para el cual la integral converja. Esto se hace con manipulación algebraica. Se sustituye la forma explícita de ?? en la definición de esperanza y se tiene:

$$\begin{aligned}
\mathbb{E}(\exp(\theta F)) &= \sum_{n=0}^{\infty} p_N(n) \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \exp\left(\theta \sum_{j=1}^n f(x_j)\right) p_{\mathcal{X}|N}(x_1, \dots, x_n|n) dx_1 \dots dx_n \\
&= \exp\left(-\int_{\mathbb{R}} \lambda(s) ds\right) \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \prod_{j=1}^n \exp(\theta f(x_j)) \lambda(x_j) dx_1 \dots dx_n \\
&= \exp\left(-\int_{\mathbb{R}} \lambda(s) ds\right) \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_{\mathbb{R}} \exp(\theta f(s)) \lambda(s) ds\right)^n \\
&= \exp\left(-\int_{\mathbb{R}} \lambda(s) ds\right) \exp\left(\int_{\mathbb{R}} \exp(\theta f(s)) \lambda(s) ds\right).
\end{aligned} \tag{A.24}$$

La última expresión es equivalente a [A.23](#).

La función característica de F esta dada por [A.23](#) cuando $\theta = i\omega$, donde $\omega \in \mathbb{R}$, $i = \sqrt{-1}$ y $\mathbb{R} = \mathbb{R}$. La convergencia de la integra requiere que la transformada de Fourier de f exista. De los cursos introductorios de probabilidad se sabe que la función generadora de momentos está estrechamente relacionada a la función característica. Al expandir la exponencial se tiene

$$\begin{aligned}
\mathbb{E}(\exp(i\omega F)) &= \mathbb{E}\left(1 + i\omega F + (i)^2 \frac{(\omega)^2 F^2}{2!} + \dots\right) \\
&= 1 + i\omega \mathbb{E}(F) + (i)^2 \frac{(\omega)^2}{2!} \mathbb{E}(F^2) + \dots,
\end{aligned}$$

asumiendo que la integración termino a termino se preserva. Por lo tanto, por diferenciación, el momento de orden $n \geq 1$ es

$$\mathbb{E}(F^n) = (-1)^n \frac{d^n}{d\omega^n} \mathbb{E}(\exp(i\omega F))_{\omega=0} \tag{A.25}$$

La forma ([A.23](#)) de la función característica también caracteriza a los PPP, esto es, un proceso puntual finito cuyas esperanzas de sumas aleatorias satisfacen [A.23](#) es necesariamente un PPP.

A.1.6.1. Caracterización de los PPP

Un proceso puntual finito es necesariamente un PPP si su esperanza de sumas aleatorias iguala la forma dada por el Teorema de Campbell. Sea Ξ un proceso puntual finito cuyas realizaciones $\xi = (n\{x_1, \dots, x_n\})$ están en el espacio de eventos $R(S)$. La función de probabilidad de Ξ es $p_\Xi(\xi)$, y la esperanza está definida como en [A.9](#).

La esperanza de la suma aleatoria

$$F(\Xi) = \sum_{j=1}^N f(X_j), \quad n \geq 1 \quad (\text{A.26})$$

se asume que satisface el Teorema de Campbell (con $\theta = -1$) para una clase de funciones f suficientemente grande. Esta clase de funciones se definirá en breve. Por lo tanto, para toda f en esta clase, se asume que

$$\mathbb{E}(\exp(-F)) = \exp\left(\int_{\mathbb{R}} [\exp(-f(x)) - 1] \lambda(x) dx\right) \quad (\text{A.27})$$

para alguna función no negativa $\lambda(x)$. El objetivo es mostrar que [A.27](#) implica que el proceso puntual finito Ξ es necesariamente un PPP con función de intensidad $\lambda(x)$. Esto se hace mostrando que Ξ satisface la propiedad de dispersiones independientes para cualquier número $k < \infty$ de conjuntos A_j tales que $S = \cup_{j=1}^k A_j$ y $A_i \cap A_j = \emptyset$ para $i \neq j$.

Considérese una función f no negativa con valores f_1, f_2, \dots, f_k en los conjuntos especificados A_1, A_2, \dots, A_k respectivamente, entonces

$$A_j = \{x : f(x) = f_j\}.$$

Sea

$$m_j = \int_{A_j} \lambda(x) dx.$$

El lado derecho de A.27 es entonces

$$\mathbb{E}(\exp(-F)) = \exp\left(\sum_{j=1}^k (\exp(-f_j) - 1)m_j\right). \quad (\text{A.28})$$

Se tiene también que

$$\sum_{j=1}^N f(X_j) = \sum_{j=1}^k f_j N(A_j), \quad (\text{A.29})$$

donde $N(A_j)$ es el número de puntos en A_j . Para la función dada f , la identidad asumida en A.27 es equivalente a

$$\mathbb{E}(\exp[-\sum_{j=1}^k (f_j N(A_j))]) = \exp[\sum_{j=1}^k (\exp(-f_j) - 1)m_j]. \quad (\text{A.30})$$

Sea $z_j = \exp(-f_j)$. El último resultado es

$$\mathbb{E}[\prod_{j=1}^k z_j^{N(A_j)}] = \prod_{j=1}^k \exp(m_j(z_j - 1)). \quad (\text{A.31})$$

Al variar la elección de los valores de la función $f_j \geq 0$, el resultado de A.31 es válido para toda $z \in (0, 1)$.

La función conjunta característica de varias variables aleatorias es el producto de las funciones características individuales si y solo si las variables aleatorias son independientes, y la función característica de una distribución Poisson con media m_j es (en esta notación) $\exp(m_j(z_j - 1))$. Por lo tanto, los conteos $N(A_j)$ son independientes y se distribuyen Poisson con media m_j . Como los conjuntos A_j son arbitrarios, el proceso puntual finito Ξ es un Proceso Poisson Puntual.

La clase de funciones para las cuales la identidad A.27 se mantiene debe incluir la clase de todas las funciones no negativas que son constantes a pedazos, con valores arbitrariamente especificados f_j , en un número finito arbitrario de conjuntos A_j disjuntos.

A.1.6.2. Funcional generador de probabilidad

Un funcional es un operador que mapea una función a un número real. Con esto en mente, el operador esperanza es un funcional pues $\mathbb{E}(f) = \int_{\mathcal{S}} f(x) dx \in \mathbb{R}$.

El funcional de Laplace evaluado para la función f esta definido por el proceso puntual finito Ξ por

$$L_{\Xi}(f) = \mathbb{E}(\exp(-F(\Xi))) = \mathbb{E}\left(\exp\left(-\sum_{j=1}^N f(X_j)\right)\right). \quad (\text{A.32})$$

La función característica de la suma aleatorá F es $L_{\Xi}(-i\omega f)$. Como en el Teorema de Campbell, f es una función no negativa para la cual la esperanza existe.

Para funciones f tales que $0 < f(x) < 1$, se define el funcional generador de probabilidad como el funcional de Laplace de $-\log(f)$:

$$G_{\Xi}(f) = L_{\Xi}(-\log(f)) = \mathbb{E}\left(\exp\left(\sum_{j=1}^N \log(f(X_j))\right)\right) = \mathbb{E}\left(\prod_{j=1}^N f(X_j)\right). \quad (\text{A.33})$$

El funcional generador de probabilidad es el análogo para los procesos finitos puntuales de la función generadora de probabilidad para variables aleatorias.

Los funcionales de Laplace y los funcionales generadores de probabilidad están definidos para procesos puntuales Ξ en general, no solo para PPP. Si Ξ es un PPP con función de intensidad $\lambda(x)$ entonces

$$G_{\Xi}(f) = \exp\left(\int_{\mathcal{R}} (f(x) - 1)\lambda(x)dx\right). \quad (\text{A.34})$$

Bibliografía

1. M. J. Dixon y S. G. Coles, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **46**, 265-280, ISSN: 00359254, 14679876, (<http://www.jstor.org/stable/2986290>) (1997).
2. M. J. Maher, *Statistica Neerlandica* **36**, 109-118, (<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9574.1982.tb00782.x>) (1982).
3. R. L. Streit, *Poisson Point Processes Imaging, Tracking, and Sensing* (Springer US, 2010).
4. J. Vázquez, L. Naranjo, R. Fuentes y M. Chávez, *Inferencia estadística para estudiantes de ciencias* (Prensas de Ciencias, 2019).