

ELEMENTOS DEL PROCESO DE POISSON.

ABRAHAM PLAZAS OSPINA.



**UNIVERSIDAD DEL VALLE.
FACULTAD DE CIENCIAS NATURALES Y
EXACTAS.
PROGRAMA ACADÉMICO DE MATEMÁTICAS.
SANTIAGO DE CALI.**

ELEMENTOS DEL PROCESO DE POISSON

ABRAHAM PLAZAS OSPINA.

abraham.plazas@correounivalle.edu.co

Trabajo de grado presentado como requisito parcial
para optar al título de Matemático.

Director: Dr. MIGUEL ÁNGEL MARMOLEJO.

UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y
EXACTAS.
PROGRAMA ACADÉMICO DE MATEMÁTICAS.
SANTIAGO DE CALI.

UNIVERSIDAD DEL VALLE.
FACULTAD DE CIENCIAS NATURALES Y
EXACTAS.
PROGRAMA ACADÉMICO DE MATEMÁTICAS

Abraham Plazas Ospina.

ELEMENTOS DEL PROCESO DE POISSON.

Palabras claves: Distribución de Poisson, Distribución Exponencial, Proceso de conteo, Proceso Estocástico, Tiempo de Interrarribo, Proceso de Poisson.

SANTIAGO DE CALI.



UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS
ACTA DE EVALUACIÓN DE TRABAJO DE GRADO
PROGRAMA ACADÉMICO DE MATEMÁTICAS

JURADO CONFORMADO POR:

DR. JAIME ALFONSO ARANGO

El día 19 de noviembre de 2020 a las 2:00 pm. se llevó a cabo la SUSTENTACIÓN del TRABAJO DE GRADO titulado “**ELEMENTOS DEL PROCESO DE POISSON**”, presentado por el estudiante PLAZAS OSPINA ABRAHAM, código: 200932085, bajo la dirección del profesor MIGUEL ANGEL MARMOLEJO.

RESULTADO DE LA EVALUACIÓN:

- (X) **APROBADO.**
- () **RECOMENDACIÓN DE DISTINCIÓN DE MERITORIO.**
- () **RECOMENDACIÓN DE DISTINCIÓN DE LAUREADO.**
- () **REPROBADO.**

Se recomienda modificaciones: SI () NO ().

OBSERVACIONES

La sustentación se realiza de manera virtual debido a la actual pandemia ocasionada por el COVID-19.

MIGUEL ANGEL MARMOLEJO
DIRECTOR TRABAJO DE GRADO

JAIME ALFONSO ARANGO
JURADO

LILIANA POSADA VERA
DIRECTORA PROGRAMA ACADÉMICO

Agradecimientos.

Como dice el refrán “Sentir gratitud y no expresarla es como envolver un regalo y no darlo”. Por lo tanto, me gustaría agradecer a todas aquellas personas que de una u otra forma, me han ayudado a realizar este trabajo.

Sé que es costumbre agradecer al director de tesis por su inestimable ayuda. Desconozco cuánto de esos agradecimientos sean de verdad o hacen parte de las frases de cajón de protocolo. Lo que sí tengo claro es que, en mi caso, este trabajo nunca habría salido adelante sin la ayuda de mi director. Así que desde aquí, profesor Miguel Marmolejo, quiero agradecerle no sólo por la cantidad de su ayuda, sino por la paciencia y por siempre empujarme a lograr este cometido. También, destaco mi agradecimiento al profesor Juan Miguel Velásquez quien además de ser un gran pilar en mi formación académica, logró incentivar la confianza en mi mismo como matemático y me brindó otra de las muchas oportunidades para poder continuar, permitiendo que este trabajo haya visto la luz. A ustedes dos, de verdad, gracias.

En segundo lugar, agradezco a mis compañeros, Lina, Alejandro, Jhon, John quienes me brindaron un ambiente más agradable en mi paso por la academia, y a mis grandes amigos Daniel, Esteban, Oscar, que desde siempre me han apoyado, me han acompañado, me han escuchado y han estado para brindarme su hombro en necesidad. Pero sobre todo agradezco a la persona que tanto en la cercanía como en la distancia siempre ha estado guiándome, alentándome y “metiéndome presión” para acabar este trabajo, siempre con buen humor y mucho cariño: Carlos, simplemente no tengo palabras... ¡eres un Pro!.

En tercer lugar, agradezco a Diana por ser parte de mi vida, en esta y otras aventuras. En pocas palabras por aguantar mis defectos, manías, peticiones, enfados, etc. y darme su apoyo incondicional en éste y todos mis trabajos.

Finalmente, agradezco a mi madre y a mi abuela por su apoyo, comprensión, amor, paciencia, dedicación y demás detalles que me brindaron y me siguen brindando, detalles que son mi bastón y horizonte en mi vida. Gracias por dedicar todos sus esfuerzos en mi futuro, en mi formación y en mi todo; las amo muchísimo y éste trabajo es dedicado a ustedes porque es el fruto de su perseverancia y tenacidad.

A todos ustedes, ¡mil y mil gracias!

Abraham Plazas Ospina.

UNIVERSIDAD DEL VALLE

Índice general

Resumen	II
Introducción	1
1. Preliminares	1
1.1. Espacios de probabilidad.	1
1.2. Variables aleatorias y sus distribuciones.	3
1.3. Vectores Aleatorios Multivariados ($n > 2$) y sus distribuciones.	17
2. La Distribución de Poisson	24
3. La Distribución Exponencial	29
4. El Proceso de Poisson	35
5. Algunas Aplicaciones del Proceso de Poisson	54
5.1. Desintegración radioactiva	54
5.2. Clientes que compran un producto	56
5.3. El problema de la colección de cupones	57
Bibliografía	63

Resumen

El proceso de Poisson, es un proceso de conteo estocástico que surge naturalmente en diversas situaciones cotidianas. El presente trabajo abarca diferentes definiciones del proceso de Poisson y discute varias de sus propiedades. Para ello, se estructuraron los temas de la siguiente manera:

En el Capítulo 1, de los preliminares, se dan a conocer todas las herramientas probabilísticas implementadas para su desarrollo. El Capítulo 2, introduce todo lo pertinente a la distribución de Poisson (definiciones, propiedades, etc.), haciendo énfasis en las relaciones que tiene con las distribuciones Binomial y Multinomial. A partir del Capítulo 3, se explica qué es la distribución Exponencial, su caracterización por medio de la propiedad de falta de memoria y, se deducen las distribuciones de varias variables aleatorias relacionadas con sumas finitas de variables aleatorias independientes e idénticamente distribuidas con distribución Exponencial.

El contenido central se encuentra en el Capítulo 4, se desarrolla con detalle la construcción del proceso de Poisson dando algunas definiciones y mostrando las equivalencias entre estas. Además, se observan las propiedades y las relaciones a distribuciones de probabilidad bien conocidas y se menciona brevemente el proceso compuesto de Poisson. Por último, en el Capítulo 5, se evidencian algunas aplicaciones de procesos de Poisson que nos muestran la gran versatilidad del mismo.

Introducción

En la presente investigación se pretende realizar una recopilación de varios aspectos del proceso de Poisson, partiendo de varias definiciones que se tienen de este. Las definiciones que se desarrollarán son las siguientes:

(1.) Un proceso estocástico de tiempo continuo $\{N_t : t \geq 0\}$ es un proceso de Poisson con parámetro $\lambda > 0$, si y solo si,

- (i) $N(0) = 0$.
- (ii) $N(t + s) - N(s) \sim Pois(\lambda t)$.
- (iii) $N(t)$ tiene incrementos independientes.

(2.) Se dice que un proceso estocástico $\{N(t) : t \geq 0\}$ es un proceso de conteo si representa el número de eventos ocurridos hasta el tiempo t , es decir, satisface:

- (i) $N(0) \geq 0$.
- (ii) Si $s < t$, entonces $N(s) \leq N(t)$.
- (iii) Para $s < t$, $N(s) - N(t)$ es igual al número de eventos que ocurrieron en el intervalo $(s, t]$.

(3.) Una colección de variables aleatorias $\{N(t) : t \geq 0\}$ es un Proceso de Poisson de parámetro $\lambda > 0$ si:

- (i) $N(0) = 0$.
- (ii) $N(t)$ tiene incrementos estacionarios e incrementos independientes.
- (iii) Para h pequeña:
 - (a) $P(N(t + h) - N(t) = 1) = \lambda h + o(h)$.
 - (b) $P(N(t + h) - N(t) > 1) = o(h)$.

Se demostrará la equivalencia entre todas ellas, con el fin de realizar una construcción de manera concreta del proceso mismo. Para esto seguiremos los conceptos dado por [6, 7, 8, 10].

Este proceso de investigación se dividirá en cinco temas, entendiendo que el primero de ellos representa los preliminares que se encuentran en el Capítulo 1. En estos preliminares vamos a dar todas las herramientas base que representan los conceptos fundamentales y el lenguaje que se usará a lo largo del desarrollo para llegar a la construcción del proceso de Poisson, aquí se definirán conceptos como espacios de medida, variables aleatorias, esperanza, vectores aleatorios y distribuciones de probabilidad, siguiendo como guía a [2, 5, 7, 8].

En las siguientes partes del proyecto mostraremos dos de las distribuciones que actúan de manera más directa con el proceso de Poisson. Así, en el Capítulo 2 haremos énfasis en la distribución de Poisson, en este se da una caracterización de la distribución por medio de la función generadora de momentos, además, se muestra la convergencia que tiene con la distribución binomial y multinomial, y se calcula la distribución que tiene la suma de variables aleatorias independientes con distribución de Poisson, para todo esto seguiremos como guía a [2, 9].

En el Capítulo 3 comenzaremos dando una caracterización de la distribución exponencial mediante la aplicación de una de las propiedades más importantes de la distribución y una herramienta indispensable para las demostraciones del proceso de Poisson, la propiedad de falta de memoria. Hablaremos de la función generadora de momentos y se demostrarán propiedades importantes, tales como, la distribución del mínimo de variables aleatorias idénticamente distribuidas y la suma de variables aleatorias idénticamente distribuidas con distribución exponencial, además se mostrará la relación de esta distribución con la distribución de Erlang, en este capítulo se tiene como guía a [2, 9].

El proceso de Poisson puede definirse de tres maneras distintas, como se mencionó inicialmente, cada una tiene un punto de partida distinto, siendo la primera de ellas partiendo de los incrementos estacionarios e incrementos independientes, la otra se centra en los intervalos de tiempo no sobrepuestos y las longitudes de estos intervalos y por último es la definición a través de los tiempos entre llegadas. A pesar de parecer que dichas definiciones son distintas entre si, en el Capítulo 4 se desarrolla en su totalidad uno de los objetivos, que es demostrar la equivalencia entre cada una de ellas, además, a partir de ensayos independientes de Bernoulli realizamos una construcción del proceso de Poisson, logrando así culminar el objetivo principal del trabajo. Por último se define a grosso modo, el proceso compuesto de Poisson mediante el uso de la función generadora de momentos.

Para culminar con el trabajo, la última parte compuesta del Capítulo 5, se desarrollan varias aplicaciones en las que se puede emplear el proceso de Poisson, entre estas miramos de cerca uno de los problemas más conocidos en probabilidad, el problema de la colección de cupones, aquí mostraremos una solución al problema empleando la relación que tiene la distribución multinomial con el proceso de Poisson.

Capítulo 1

Preliminares

El propósito de esta sección es suministrar un breve contexto acerca de los principales temas en la probabilidad clásica: Espacios de probabilidad, Variables aleatorias, Funciones de distribución, entre otros.

El siguiente apartado lo desarrollaremos sin mayores pruebas y seguiremos [2, 5, 7, 8, 9] como referencia para esta sección.

1.1. Espacios de probabilidad.

Definición 1.1.1 Sea Ω un conjunto no vacío llamado espacio muestral. Sea $\mathcal{F}$ una colección de conjuntos de Ω . $\mathcal{F}$ se dice una σ -álgebra de conjuntos sobre Ω , ó un σ -álgebra sobre Ω , si y solo si $\Omega \in \mathcal{F}$ y $\mathcal{F}$ es cerrado bajo complementos y uniones numerables, es decir:

$$(\sigma_1) \quad \Omega \in \mathcal{F}.$$

$$(\sigma_2) \quad \text{Si } A \in \mathcal{F} \text{ entonces } A^c \in \mathcal{F}.$$

$$(\sigma_3) \quad \text{Si } A_n \in \mathcal{F}, n \in \mathbb{N} \text{ entonces } \cup_{n=1}^{\infty} A_n \in \mathcal{F},$$

Sea C una colección no vacía de subconjuntos de Ω , tal que $C \subseteq \mathcal{P}(\Omega)$, entonces existe una única σ -álgebra, denotada por $\sigma(C)$, llamada la σ -álgebra generada por C , que verifica:

$$(\sigma_{g1}) \quad \sigma(C) \text{ es } \sigma\text{-álgebra.}$$

$$(\sigma_{g2}) \quad C \subseteq \sigma(C).$$

$$(\sigma_{g3}) \quad \text{Si } \mathcal{A} \text{ es una } \sigma\text{-álgebra y } C \subseteq \mathcal{A}, \text{ entonces } \sigma(C) \subseteq \mathcal{A}.$$

Además, se tiene que $\sigma(C)$ es la σ -álgebra más pequeña que contiene a C , explícitamente es $\sigma(C) = \bigcap \{ \alpha : \alpha \text{ es } \sigma\text{-álgebra sobre } \Omega \text{ y } \alpha \supseteq C \}$.

Definición 1.1.2 Sea $(\Omega, \mathcal{F})$ un espacio medible. Una función $\mu : \mathcal{F} \rightarrow [0, \infty]$ se llama una medida en $(\Omega, \mathcal{F})$ si satisface que:

$$(\mu_1) \quad \mu(\emptyset) = 0$$

(μ_2) Para $\{A_n\}_{n \in \mathbb{N}}$ una familia disjunta y contable de conjuntos de $\mathcal{F}$, es decir, $A_i \cap A_j = \emptyset$ para $i \neq j$, se satisface

$$\mu \left(\biguplus_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Un espacio de medida es una tripla $(\Omega, \mathcal{F}, \mu)$ donde Ω es un conjunto, $\mathcal{F}$ es una σ -álgebra de subconjuntos de Ω , y μ es una medida definida en $\mathcal{F}$. μ es una medida de probabilidad, la cual denotaremos por P , si P satisface:

$$(P_1) \quad 0 \leq P(A) \leq 1, \quad A \in \mathcal{F}.$$

$$(P_2) \quad P(\Omega) = 1.$$

$$(P_3) \quad P \left(\biguplus_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} P(A_n), \text{ donde } \{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F} \text{ tal que } A_i \cap A_j = \emptyset, \text{ para } i \neq j.$$

La tripla $(\Omega, \mathcal{F}, P)$ se llama espacio de probabilidad.

Ejemplo 1.1.1 El ejemplo clásico de un espacio de probabilidad es:

$$(\Omega, \mathcal{F}, P) \equiv ([0, 1], \mathcal{B}([0, 1]), \lambda)$$

donde $\mathcal{B}([0, 1])$ es la σ -álgebra de los conjuntos de Borel de $[0, 1]$ y λ es la medida de Lebesgue.

Así pues, llamaremos espacio medible a la dupla $(\Omega, \mathcal{F})$ donde Ω es un conjunto no vacío y $\mathcal{F}$ es una σ -álgebra de subconjuntos de Ω . Los elementos de $\mathcal{F}$ se llaman conjuntos medibles.

Sea un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Se dice que dos σ -álgebras $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{F}$, son independientes sí y sólo sí para todo $A_1 \in \mathcal{A}_1$, y para todo $A_2 \in \mathcal{A}_2$ se tiene que $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$, adicionalmente se tiene que $A, B \in \mathcal{F}$ son independientes sí y sólo sí $\sigma(A)$ y $\sigma(B)$ son independientes, donde $\sigma(A)$ y $\sigma(B)$ son las σ -álgebras generadas por A y B respectivamente.

1.2. Variables aleatorias y sus distribuciones.

Una forma que se tiene para trasladar la información de un espacio a otro es mediante el uso de una aplicación. En nuestro caso, dicha aplicación habrá de trasladar el concepto de evento, lo que exige una mínima infraestructura en el espacio receptor de la información semejante a la σ -álgebra que contiene a los eventos. Para tratar el caso unidimensional, una sola característica numérica del espacio imagen $\mathbb{R}$ será asociada a los puntos del espacio muestral. Dado que en $\mathbb{R}$, los intervalos son el lugar habitual de trabajo, se exige a la infraestructura que los contenga, y puesto que en $\mathbb{R}$ la σ -álgebra de Borel, $\mathcal{B}(\mathbb{R})$, es la menor de las σ -álgebras que contienen a los intervalos, esta permite convertir a $\mathbb{R}$ en espacio probabilizable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definición 1.2.1 Sea $(\Omega, \mathcal{F})$ y $(\mathcal{S}, \mathcal{A})$ espacios medibles. Sea $f: \Omega \rightarrow \mathcal{S}$ una función que satisface $f^{-1}(A) \in \mathcal{F}$ para cada $A \in \mathcal{A}$. Entonces, se dice que f es $\mathcal{F}/\mathcal{A}$ -medible.

Definición 1.2.2 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ el espacio medible de los reales con la σ -álgebra de Borel. Una función $X: \Omega \rightarrow \mathbb{R}$ es una variable aleatoria, que denotaremos v.a., si X es una función $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -medible.

Teorema 1.2.1 Sea un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Dadas dos v.a. X e Y y $k \in \mathbb{R}$, se tiene que $X+Y$, XY y kX , son v.a. con las operaciones usuales de funciones.

Un ejemplo sencillo de v.a. es la función la indicatriz $X(\omega) = I_A(\omega)$ de un conjunto arbitrario $A \in \mathcal{F}$ definida de la siguiente manera:

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

Una v.a. X que tiene representación

$$X(\omega) = \sum_{n=1}^{\infty} x_n I_{A_n}(\omega), \quad x_n \in \mathbb{R}$$

donde $\bigcup_{i=1}^{\infty} A_i = \Omega$, $\{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F}$ se llama discreta. Si la suma anterior es finita, la v.a. se llama simple.

La distribución de probabilidad de una v.a. $X: \Omega \rightarrow \mathbb{R}$ es la medida de probabilidad $P_X: \mathcal{F} \rightarrow [0, 1]$ definida por:

$$P_X(B) := P(X^{-1}(B)), \quad B \in \mathcal{F}.$$

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, la función $F: \mathbb{R} \rightarrow [0, 1]$ definida por,

$$F(x) := P((-\infty, x]),$$

se llama función de distribución asociada a P , la cual satisface:

(F₁) $0 \leq F(x) \leq 1$, para todo $x \in \mathbb{R}$.

(F₂) Para todo $a, b \in \mathbb{R}$, si $a \leq b$ entonces $F(a) \leq F(b)$.

(F₃) Para todo $a \in \mathbb{R}$, $F(a^+) = F(a)$, $F(a^+) \equiv \lim_{x \rightarrow a^+} F(x)$.

(F₄) $F(+\infty) = 1$, $F(-\infty) = 0$, donde $F(+\infty) = \lim_{x \rightarrow \infty} F(x)$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$.

Recíprocamente, asociada a tal función hay una única medida de probabilidad $P : \mathcal{F} \rightarrow [0, 1]$, tal que para todo $a, b \in \mathbb{R}$ con $a < b$ se tiene:

$$P(a \leq X \leq b) = F(b) - F(a).$$

Debido a la anterior correspondencia entre medidas de probabilidad y funciones de distribución (ver Teorema 1 pag. 152 [9]), podemos observar varios tipos de medida de probabilidad con base en cada función de distribución, así pues, se clasifican en: medidas discretas, absolutamente continuas y singulares. Las medidas discretas, son medidas P a las que corresponden funciones de distribución F escalonadas con saltos en los conjuntos contables de puntos $x_1, x_2, \dots$. La medida en los casos discretos está concentrada en los puntos $x_1, x_2, \dots$ y satisface:

$$p_i = P_X(x_i) := P(X = x_i) = \Delta F(x_i) = F(x_i) - F(x_{i-1}) > 0, \quad \sum_i p_i = 1,$$

donde p_i la llamamos función de masa, el conjunto de números $\{p_1, p_2, \dots\}$ es llamado distribución de probabilidades discreta y la función de distribución es llamada discreta. Ejemplos característicos de este tipo de medidas son: Uniforme discreta, Binomial, Poisson, Geométrica y Binomial Negativa.

Las medidas absolutamente continuas, son medidas a las que corresponden funciones de distribución de la forma,

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

donde $f(x)$ es una función no negativa, tal que, $\int_{-\infty}^{\infty} f(t)dt = 1$, denominada función de densidad de la función de distribución $F(x)$. Algunos ejemplos de este tipo son: las distribuciones Uniforme, Normal, Gamma, Beta, Exponencial y Cauchy (mencionadas en [4]). Por último, las medidas singulares P son aquellas donde su función de distribución F es continua, más no es posible generar una expresión en forma de una integral. La escalera de Cantor es el ejemplo más común de este tipo (en [9], pag. 157).

Para una v.a discreta la función de distribución está dada por:

$$F(x) = \sum_{x_i \leq x} P(X = x_i)$$

en donde la suma se realiza sobre aquellos valores de i donde se tiene que $x_i \leq x$.

Una función asociada a una v.a discreta o a su respectiva distribución, se llamará función de masa de probabilidad, denotada por f_x , si se tiene que:

$$p_{x_i} = \begin{cases} P(X = x_i), & \text{si } x = x_i \\ 0 & , \text{ si } x \neq x_i, \text{ para todo } i=1, 2, \dots \end{cases}$$

Como un resultado inmediato, tenemos lo siguiente:

$$(1) \quad 0 \leq f(x) \leq 1.$$

$$(2) \quad \sum_{i=1}^{\infty} f(x_i) = 1.$$

$$(3) \quad F(x) = \sum_{x_i \leq x} f(x_i).$$

Es importante tener en cuenta que la clasificación de las v.a. en discretas y continuas, no implica que toda distribución de probabilidad encaje en esta categorización.

Si f es la función de densidad de una v.a. absolutamente continua X y F es su función de distribución,

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

entonces:

$$(1) \quad f(x) \geq 0 \text{ para todo } x \in \mathbb{R}.$$

$$(1) \quad F \text{ es continua.}$$

$$(3) \quad P(X = a) = 0 \text{ para todo } a \in \mathbb{R}.$$

$$(4) \quad F'(a) = f(a), \text{ si } f \text{ es continua en } a \in \mathbb{R}.$$

$$(5) \quad P(a < X \leq b) = \int_a^b f(x)dx.$$

algunos ejemplos de las distribuciones de probabilidad más conocidos se encuentran resumidos en la siguiente tabla:

Distribución	Notación	Función de densidad	Función de distribución
Binomial	$B(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	$\sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$
Poisson	$Pois(\lambda)$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!} e^{-\lambda}$
Uniforme discreta	$U(x_1, \dots, x_n)$	$\frac{1}{n}$	$\frac{1}{n} \sum_i 1_{(-\infty, x_i]}(x_i)$
Uniforme continua	$U(a, b)$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$
Exponencial	$Exp(\lambda)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$
Normal	$N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt$
Gamma	$Gamma(\alpha, \beta)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$	$\int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt$

Tabla 1.1. Distribuciones de probabilidad

Definición 1.2.3 Sea X una v.a. sobre un espacio de probabilidades $(\Omega, \mathcal{F}, P)$ y sea g una función de variable real x . Entonces, $Y = g(X)$ es una nueva función real definida sobre Ω tal que,

$$Y(\omega) = g(X(\omega)).$$

Se impone sobre Y la siguiente condición para que satisfaga que Y sea v.a.:

$$\{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F}.$$

Teorema 1.2.2 Sea X una v.a. discreta y Y una v.a. definida por, $Y = g(X)$. Entonces la función de distribución de Y viene dada por:

$$F_Y(y) = \sum_{g(x) \leq y} P(X = x).$$

Teorema 1.2.3 Sea X una v.a. absolutamente continua con función de densidad f_X y sea Y una v.a. definida por $Y = g(X)$. Entonces la función de distribución de Y esta dada por:

$$F_Y(y) = \int_D f_X(x) dx,$$

donde D es el subconjunto de la recta real definido por $g(x) \leq y$.

Teorema 1.2.4 Sea X una v.a. absolutamente continua con función de densidad f_X continua. Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función monótona y derivable, entonces $Y = g(X)$ es una v.a. absolutamente continua con función de densidad:

$$f_Y(y) = f_X(x) \cdot \frac{1}{|g'(x)|},$$

para todo $y \in \mathbb{R}$, donde $y = g(x)$. Además, si g es creciente, la función de distribución de Y es:

$$F_Y(y) = F_X(x).$$

Si g es decreciente, entonces la función de distribución de Y es:

$$F_Y(y) = 1 - F_X(x).$$

Definición 1.2.4 Sea un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Se define una v.a. bidimensional, que denotaremos v.a.b., $\mathbf{X} = (X, Y)$, tal que, X e Y son v.a. definidas sobre el mismo espacio de probabilidad.

Definición 1.2.5 Dada una v.a.b. $\mathbf{X} = (X, Y)$ sobre $(\Omega, \mathcal{F}, P)$, llamamos función de distribución conjunta a la función real de dos variables definida por

$$F_{\mathbf{X}}(x, y) = P(X \leq x, Y \leq y),$$

donde,

$$[X \leq x, Y \leq y] = [X \leq x] \cap [Y \leq y].$$

Teorema 1.2.5 La función de distribución conjunta $F_{\mathbf{X}}$ satisface lo siguiente:

(1) $0 \leq F_{\mathbf{X}}(x, y) \leq 1$ para todo $(x, y) \in \mathbb{R}^2$.

(2) $F_{\mathbf{X}}$ es creciente en cada variable,

$$x_1 < x_2 \Rightarrow F_{\mathbf{X}}(x_1, y) \leq F_{\mathbf{X}}(x_2, y),$$

$$y_1 < y_2 \Rightarrow F_{\mathbf{X}}(x, y_1) \leq F_{\mathbf{X}}(x, y_2).$$

(3) $F_{\mathbf{X}}(+\infty, +\infty) = 1$ y $F_{\mathbf{X}}(-\infty, y) = F_{\mathbf{X}}(x, -\infty) = 0$.

(4) $P(a < X \leq b, c < Y \leq d) = F_{\mathbf{X}}(b, d) - F_{\mathbf{X}}(a, d) - F_{\mathbf{X}}(b, c) + F_{\mathbf{X}}(a, c)$

(5) $F_{\mathbf{X}}$ es continua por la derecha en cada variable.

(6) Si $F_{\mathbf{X}}$ es la función de distribución conjunta de una v.a.b $\mathbf{X} = (X, Y)$, se tiene que:

$$\lim_{y \rightarrow +\infty} F_{\mathbf{X}}(x, y) = F_X(x) \quad y \quad \lim_{x \rightarrow +\infty} F_{\mathbf{X}}(x, y) = F_Y(y),$$

Observación 1.2.1 De (6) se tiene que, $F_X(x)$ y $F_Y(y)$ son las funciones de distribución de las v.a. X e Y por separado, respectivamente y estas se denominan distribuciones marginales de $\mathbf{X} = (X, Y)$.

Definición 1.2.6 Sea un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y una v.a.b. discreta $\mathbf{X} = (X, Y)$, con X e Y v.a. discretas. Si estas v.a toman valores x_i e y_j con $P(X = x_i)$ y $P(Y = y_j)$ tal que $i, j = 1, 2, \dots$ respectivamente. Su función de masa de probabilidad conjunta está dada por:

$$f(x, y) = \begin{cases} P(X = x_i, Y = y_j), & \text{si } x = x_i, y = y_j \\ 0 & , \quad \text{si } x \neq x_i, y \neq y_j, \text{ para } i, j = 1, 2, \dots \end{cases}$$

en donde

$$P(X = x_i, Y = y_j) = P([X = x_i] \cap [Y = y_j]) \quad (i, j = 1, 2, \dots),$$

así, la función de distribución conjunta viene dada por:

$$F(x, y) = P(X \leq x_i, Y \leq y_j) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j).$$

La función de masa de probabilidad conjunta satisface,

$$(1) \quad 0 \leq f(x_i, y_j) \leq 1 \text{ para todo } i, j = 1, 2, \dots$$

$$(2) \quad \sum_i \sum_j f(x_i, y_j) = 1.$$

$$(3) \quad F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j).$$

Se puede ver que,

$$\lim_{y \rightarrow +\infty} F_{\mathbf{X}}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j) = \sum_{x_i \leq x} \sum_{y_j} P(X = x_i, Y = y_j).$$

Además, se sabe que la distribución marginal de X está dada por:

$$\lim_{y \rightarrow +\infty} F_{\mathbf{X}}(x, y) = F_X(x) = \sum_{x_i \leq x} P(X = x_i, Y = y_j).$$

Así, se puede obtener lo siguiente:

$$F_X(x) = P(X = x_i) = \sum_{y_j} P(X = x_i, Y = y_j),$$

De manera similar, obtenemos este otro resultado:

$$F_Y(y) = P(Y = y_j) = \sum_{x_i} P(X = x_i, Y = y_j).$$

Definición 1.2.7 Si f es la función de densidad conjunta de una v.a.b. absolutamente continua $\mathbf{X} = (X, Y)$. Se define la distribución conjunta de $\mathbf{X}$, $F(x, y)$ por:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$$

se satisface lo siguiente,

- (1) $\frac{\partial^2 F(x, y)}{\partial y \partial x} = f(x, y)$, en aquellos puntos (x, y) donde f es continua.
- (2) $f(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}^2$.
- (3) Sea $D \in \mathcal{B}(\mathbb{R}^2)$, entonces,

$$P((X, Y) \in D) = \int_D f(x, y) dx dy,$$

$$\text{En particular, } P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y) dx dy.$$

Sea $\mathbf{X} = (X, Y)$ una v.a.b. absolutamente continua con distribución conjunta f . Sean f_X y f_Y las funciones de densidad de X y Y respectivamente, luego, sus funciones de distribución están dadas por

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{y} \quad F_Y(y) = \int_{-\infty}^y f_Y(t) dt.$$

Por otro lado, por el teorema anterior se tiene que las marginales de $\mathbf{X} = (X, Y)$ están dadas por:

$$\lim_{y \rightarrow +\infty} F_{\mathbf{X}}(x, y) = F_X(x) \quad \text{y} \quad \lim_{x \rightarrow +\infty} F_{\mathbf{X}}(x, y) = F_Y(y).$$

En particular,

$$F_X(x) = F_{\mathbf{X}}(x, +\infty) = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f(u, v) dv \right) du.$$

Comparando las anteriores ecuaciones se puede concluir que las funciones de densidad para X e Y , dadas por:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f(x, y) dy, \\ F_Y(y) &= \int_{-\infty}^y f(x, y) dx. \end{aligned}$$

Observación 1.2.2 Sea un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y $\mathbf{X} = (X_1, X_2)$ una v.a.b.. Entonces, $\mathbf{Y} = (Y_1, Y_2)$ es una v.a.b. definida por,

$$Y_1 = g_1(X_1, X_2) \quad y \quad Y_2 = g_2(X_1, X_2),$$

en donde g_1, g_2 son funciones de $\mathbb{R}^2$ en $\mathbb{R}$ que satisfacen:

$$\{\omega \in \Omega : Y_1(\omega) = g_1(X_1(\omega), X_2(\omega)) \leq y_1\} \cap \{\omega \in \Omega : Y_2(\omega) = g_2(X_1(\omega), X_2(\omega)) \leq y_2\} \in \mathcal{F}.$$

Teorema 1.2.6 Sea $\mathbf{X} = (X_1, X_2)$ una v.a.b. discreta y $\mathbf{Y} = (Y_1, Y_2)$ definida por,

$$\begin{aligned} Y_1 &= g_1(X_1, X_2), \\ Y_2 &= g_2(X_1, X_2). \end{aligned}$$

La función de distribución de $\mathbf{Y}$ está dada por:

$$F_{\mathbf{Y}}(y_1, y_2) = \sum_{\substack{g_1(x_1, x_2) \leq y_1 \\ g_2(x_1, x_2) \leq y_2}} P(X = x_1, X = x_2).$$

Para las v.a. absolutamente continuas se tienen los siguientes resultados:

Teorema 1.2.7 Sea $\mathbf{X} = (X_1, X_2)$ una v.a.b. absolutamente continua con función de densidad conjunta $f_{\mathbf{X}}$ y sean Y_1, Y_2 dos v.a. definidas por $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$. Entonces, la función de distribución conjunta de $\mathbf{Y} = (Y_1, Y_2)$ está dada por:

$$F_{\mathbf{X}}(y_1, y_2) = \int_D f_{\mathbf{X}}(x_1, x_2) dx_1 dx_2,$$

donde D es la región de integración definida por $g_i(x_1, x_2) \leq y_i$, $i = 1, 2$.

Teorema 1.2.8 Sea $\mathbf{X} = (X_1, X_2)$ una v.a.b. absolutamente continua con función de densidad conjunta $f_{\mathbf{X}}$ y sean $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$, de manera que $g = (g_1, g_2)$ una aplicación de $\mathbb{R}^2$ en sí misma de clase $\mathcal{C}^1$ e inyectiva en el conjunto A en que $f_{\mathbf{X}}$ no es nula. Entonces, se define la transformación inversa $g^{-1} = (g_1^{-1}, g_2^{-1})$ por $X_1 = g_1^{-1}(Y_1, Y_2)$, $X_2 = g_2^{-1}(Y_1, Y_2)$, tal que es continua y su jacobiano $J_{g^{-1}}$ no se anula en $g^{-1}(A)$. Con las condiciones anteriores, $\mathbf{Y} = (Y_1, Y_2)$ es una v.a.b. absolutamente continua y,

$$f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{X}}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) \cdot |J_{g^{-1}}|.$$

Definición 1.2.8 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Dos v.a. X e Y son independientes si su función de distribución conjunta se factoriza como producto de funciones de distribución individuales

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y).$$

La condición de independencia para dos v.a. X, Y discretas es,

$$P(X = x, Y = y) = P_X(X = x) \cdot P_Y(Y = y),$$

y para v.a. absolutamente continuas se tiene,

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y).$$

Teorema 1.2.9 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Si X e Y son v.a. independientes, entonces las v.a. $U = g(X)$ y $V = h(Y)$ también son independientes, para g y h funciones reales.

Definición 1.2.9 Sea X una v.a. definida sobre el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Sea $B \in \mathcal{F}$ con $P(B) > 0$. Se llama función de distribución condicionada de la variable X dado el evento B , denotada por $F(\cdot|B)$, a la función definida por,

$$F(x|B) = P(X \leq x|B) = \frac{P(X \leq x, B)}{P(B)}.$$

La función de distribución condicionada satisface lo siguiente:

1. $0 \leq F(x|B) \leq 1$ para todo $x \in \mathbb{R}$.
2. $F(\cdot|B)$ es monótona no decreciente:

$$x_1 < x_2 \rightarrow F(x_1|B) \leq F(x_2|B),$$

para todo $x_1, x_2 \in \mathbb{R}$.

3. $F(-\infty|B) = 0$ y $F(+\infty|B) = 1$.
4. $P(a < X \leq b|B) = F(b|B) - F(a|B)$ para todo $a, b \in \mathbb{R}$ con $a \leq b$.
5. $F(\cdot|B)$ es continua por derecha en cada punto de $\mathbb{R}$.

En el caso de que X sea una v.a. absolutamente continua, la función de densidad condicionada, denotada por $f(\cdot|B)$, está dada por,

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

Esta función cumple las siguientes propiedades análogas:

1. $\int_{-\infty}^{+\infty} f(x|B)dx = 1$.
2. $f(x|B) \geq 0$ para todo $x \in \mathbb{R}$.
3. $f(x|B) = F'(x|B)$ en los puntos x donde $F(\cdot|B)$ es continua.

$$4. P(a < X \leq b) = \int_a^b f(x)dx.$$

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Sean X e Y v.a. absolutamente continuas y $B = [Y \leq y]$, con $P(B) > 0$. Entonces se tiene que:

$$\begin{aligned} F(x|Y \leq y) &= \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} \\ &= \frac{F(x, y)}{F_Y(y)}, \quad F_Y(y) > 0. \end{aligned}$$

Además, si f es continua, se satisface:

$$\frac{\partial}{\partial x} (F(x|Y \leq y)) = \frac{\partial}{\partial x} \left(\frac{F(x, y)}{F_Y(y)} \right) = \frac{1}{F_Y(y)} \frac{\partial F(x, y)}{\partial x}.$$

Ahora, sean X e Y v.a. absolutamente continua y $B = [Y = y]$, con $P(B) > 0$. Entonces las funciones de densidad condicionada están dadas por:

$$f_X(x|Y = y) = \begin{cases} \frac{f(x, y)}{f_Y(y)}, & \text{si } f_Y(y) > 0 \\ 0, & \text{en otro caso} \end{cases},$$

y

$$f_Y(y|X = x) = \begin{cases} \frac{f(x, y)}{f_X(x)}, & \text{si } f_X(x) > 0 \\ 0, & \text{en otro caso} \end{cases}.$$

Finalmente, si X e Y son independientes, entonces,

$$f(x, y) = f_X(x) \cdot f_Y(y),$$

por lo tanto,

$$f(x|Y = y) = f_X(x) \quad \text{y} \quad f(y|X = x) = f_Y(y).$$

En el caso discreto,

$$P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)},$$

en donde,

$$P(Y = y_j) = \sum_{x_i} P(X = x_i, Y = y_j) \neq 0.$$

Teorema 1.2.10 Sean X e Y dos v.a. absolutamente continuas sobre el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Sean f_X y f_Y las distribuciones de densidad y sean $f_X(\cdot|y)$, $f_Y(\cdot|x)$ las funciones de densidad condicionadas, respectivamente. Así, las formulas de la probabilidad total y la fórmula de Bayes se satisfacen y están dadas por:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x|y) \cdot f_Y(y) dy,$$

y

$$f(y|x) = \frac{f(x|y) \cdot f_Y(y)}{f_X(x)},$$

donde $f(x|y) = f(x|Y = y)$.

Definición 1.2.10 Dada una v.a. X definida sobre un espacio muestral Ω numerable, su esperanza o valor esperado, es el número $E(X)$ dado por

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega),$$

siempre que,

$$\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\omega),$$

sea convergente.

Teorema 1.2.11 Sea X una v.a. discreta que toma los valores x_i , $i = 1, 2, \dots$. Sea $f(x_i) = P(X = x_i)$, $i = 1, 2, \dots$ su función de densidad y sea g una función de variable real tal que $g(X)$ es v.a.. Entonces,

$$E(g(X)) = \sum_n g(x_n) \cdot f(x_n),$$

si la serie es absolutamente convergente.

Sean X, Y v.a discretas. Sean $E(X), E(Y)$ sus valores esperados respectivamente, entonces se cumplen las siguientes propiedades:

1. $E(1) = 1$, donde 1 es la v.a discreta constante igual a 1.
2. $E(\alpha X) = \alpha E(X)$, para todo $\alpha \in \mathbb{R}$.
3. $E(X + Y) = E(X) + E(Y)$.
4. $|E(X)| \leq E(|X|)$.
5. Si $Y \leq X$, entonces $E(Y) \leq E(X)$. En particular, si $X \geq 0$, entonces $E(X) \geq 0$.
6. Si X, Y son v.a. independientes, entonces $E(X \cdot Y) = E(X) \cdot E(Y)$.

Definición 1.2.11 Si X es una v.a. absolutamente continua y sea $f(x_i) = P(X = x_i)$ su función de densidad, entonces, se define la esperanza o valor esperado de X por:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx,$$

si,

$$\int_{-\infty}^{+\infty} |x| \cdot f(x) dx,$$

sea convergente.

Teorema 1.2.12 Sea X una v.a. absolutamente continua y sea g una función derivable, entonces,

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$$

es absolutamente convergente.

Sean X, Y v.a. absolutamente continuas. Sean $E(X), E(Y)$ sus valores esperados respectivamente, entonces tenemos propiedades análogas a las discretas, tales que:

1. $E(1) = 1$, donde 1 es la v.a. discreta constante igual a 1.
2. $E(\alpha g(X) + h(X)) = \alpha E(g(X)) + E(h(X))$, por tanto $\alpha \in \mathbb{R}$ y g, h funciones reales de variable real.
3. Si X tiene una función de densidad en un punto c , entonces $E(X) = c$.
4. $|E(g(X))| \leq E(|g(X)|)$.
5. Si $0 \leq g(x) \leq h(x)$ para todo $x \in \mathbb{R}$, entonces $0 \leq E(g(X)) \leq E(h(X))$.
6. Si $m \leq g(x) \leq M$ para todo $x \in \mathbb{R}$, entonces $m \leq E(g(X)) \leq M$.
7. Si X, Y son v.a. independientes, entonces $E(X \cdot Y) = E(X) \cdot E(Y)$.

Definición 1.2.12 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Sean X, Y v.a. y sea la distribución condicionada $Y|X = x$. Entonces se define la esperanza condicionada de $Y|X = x$ de la siguiente manera:

$$E(Y|X = x) = \sum_i y_i \cdot f_Y(y_i|X = x),$$

se tiene para el caso discreto. Y

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y \cdot f_Y(y|X = x) dy.$$

para el caso absolutamente continuo.

Análogamente se define la esperanza de X condicionada por $Y = y$.

En particular, las esperanzas de ciertas funciones de X son de especial interés.

Definición 1.2.13 Sea X una v.a. en $(\Omega, \mathcal{F}, P)$. Si $k > 0$, el número $E(X^k)$ es llamado el k -ésimo momento de X . $E[|X|^k]$ se llama el k -ésimo momento absoluto de X . $E[(X - E(X))^k]$ se denomina el k -ésimo momento central de X y el número $E[|X - E(X)|^k]$ lo llamaremos el k -ésimo momento absoluto central de X . Los momentos centrales sólo pueden ser definidos si $E(X)$ es finito.

Observación 1.2.3 De los momentos hay dos que merecen especial atención. El primer momento central ($k = 1$), si existe, está dado por $E[(X - E(X))^{k=1}]$ y gracias a su linealidad, vale siempre cero, y el segundo momento central, $\sigma^2 = E[(X - E(X))^2]$, que se denomina la varianza de X y lo denotaremos como $\text{Var}(X)$ y la raíz positiva de la varianza, σ , la llamaremos desviación estandar.

Teorema 1.2.13 (a) Si X es una v.a. no negativa, $0 < p < \infty$ y $0 < \varepsilon$,

$$P\{X \geq \varepsilon\} \leq \frac{E(X^p)}{\varepsilon^p}.$$

(b) Si X es una v.a. con media finita m y varianza σ^2 , y $0 < k < \infty$,

$$P\{|X - m| \geq k\sigma\} \leq \frac{1}{k^2}.$$

Este teorema es conocido como la Desigualdad de Chebyshev.

Teorema 1.2.14 Sean $X_1, X_2, \dots$ v.a. independientes no necesariamente idénticamente distribuidas, cada una con media y varianza finita. Asumimos que las varianzas están uniformemente acotadas por $M < \infty$. Sea $S_n = X_1 + \dots + X_n$. Entonces $[S_n - E(S_n)]/n$ converge en probabilidad a 0, es decir, dado un $\varepsilon > 0$,

$$P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \rightarrow 0 \quad \text{siempre que } n \rightarrow \infty.$$

A este teorema lo denominaremos, la Ley Débil de los Grandes Números.

Hay dos casos de interés particular.

- (1) Si $E(X_i) = m$ para todo i , entonces, $[S_n - E(S_n)]/n = (S_n/n) - m$; por lo tanto $S_n/n \rightarrow m$ en probabilidad. Por lo que para valores de n muy grandes, la media aritmetica de n v.a. independientes, cada una con esperanza finita m y con la varianza uniformemente acotada, es probable que esté muy cerca a m .

- (2) Si $X_1, X_2, \dots$ son independientes, y para cada i , $P\{X_i = 1\} = p$ y $P\{X_i = 0\} = q = 1 - p$, i.e. tenemos una sucesión infinita de ensayos de Bernoulli, entonces, $X_1, X_2, \dots$ es el número de éxitos en n ensayos y por tanto S_n/n es la frecuencia relativa de éxitos. Ya que $E(X_i) = p$, se tiene que $S_n/n \rightarrow p$ en probabilidad. Por lo tanto para grandes valores de n , la frecuencia relativa de éxitos esta muy cerca a p .

1.3. Vectores Aleatorios Multivariados ($n > 2$) y sus distribuciones.

Cuando estudiamos simultáneamente n características numéricas ligadas al resultado de un experimento, por ejemplo la altura y el peso de las personas, nos movemos en $\mathbb{R}^n$ como espacio imagen de nuestra aplicación. La σ -álgebra de sucesos con la que dotamos a $\mathbb{R}^n$ para hacerlo probabilizable es la correspondiente σ -álgebra de Borel $\mathcal{B}(\mathbb{R}^n)$, que tiene la propiedad de ser la menor que contiene a los rectángulos $(a, b] = \prod_{i=1}^n (a_i, b_i]$, con $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$ y $-\infty \leq a_i \leq b_i < +\infty$. De entre ellos merecen especial mención aquellos que tiene el extremo inferior en $-\infty$, $S_x = \prod_{i=1}^n (-\infty, x_i]$.

Definición 1.3.1 *Un vector aleatorio n -dimensional en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, es un mapeo Borel medible de Ω a $\mathbb{R}^n$.*

Si $X : \Omega \rightarrow \mathbb{R}^n$ y $X_i = \rho_i(X)$, donde ρ_i es la proyección de $\mathbb{R}^n$ en el i -ésimo espacio de coordenadas, entonces X es Borel medible, si y solo si, cada X_i es Borel medible. Así, un vector aleatorio puede ser considerado como una n -tupla $(X_1, \dots, X_n)$ de variables aleatorias.

Gran parte del desarrollo de la sección anterior se “traslada”, y del mismo modo vamos a empezar definiendo la medida de probabilidad inducida por un vector aleatorio X , la cual está definida por:

$$P_X(B) := P\{\omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Definición 1.3.2 *La función de distribución de X es la función $F = F_X$ de $\mathbb{R}^n$ a $[0, 1]$ definida por,*

$$F(x) = P_X(-\infty, x] = P\{\omega : X_i(\omega) \leq x_i, i = 1, \dots, n\};$$

también se denomina a F como la función de distribución conjunta de $X_1, \dots, X_n$. F es creciente y continua por derecha en $\mathbb{R}^n$, y P_X es la medida de Lebesgue-Stieltjes determinada por F . Entonces se tiene:

$$F(x_1, \dots, x_n) \rightarrow \begin{cases} 1 & \text{si } x_i \uparrow +\infty \text{ para todo } i \\ 0 & \text{si } x_i \downarrow -\infty \text{ para un } i \text{ particular con las otras coordenadas fijas} \end{cases}.$$

Si F es una función de distribución en $\mathbb{R}^n$ que satisface lo anterior, entonces F es la función de distribución de algún vector aleatorio X , y el espacio de probabilidad subyacente puede ser construido de la siguiente forma: sea $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, con P la medida Lebesgue-Stieltjes determinada por F , y X la función identidad en Ω . Ya que, $P_X(B) = P\{\omega : X(\omega) \in B\} = P(B)$, X ha inducido la medida de probabilidad P . Además, la función de distribución de X está dada por,

$$F_X(x) = P_X(-\infty, x] = P(-\infty, x] = F(X).$$

Existe una propiedad que nos permite obtener P_X a partir de F_X , para esto consideremos primero lo siguiente:

Supongamos $n = 2$ y consideremos el rectángulo $(a, b] = (a_1, b_1] \times (a_2, b_2]$. Tenemos que $P_X((a, b]) \geq 0$, además, dado lo anterior se puede escribir,

$$P_X((a, b]) = F_X(b_1, b_2) - F_X(b_1, a_2) - F_X(a_1, b_2) + F_X(a_1, a_2) \geq 0.$$

Este resultado es cierto, para cualquier k . Se introducirá el operador diferencia, a fin de obtener una representación sencilla del resultado para el caso general. Sea $a_i \leq b_i$, Δ_{a_i, b_i} representa el operador diferencia definido de la siguiente forma,

$$\Delta_{a_i, b_i} F_X(x_1, \dots, x_n) = F_X((x_1, \dots, x_{i-1}, b_i, \dots, x_n) - F_X((x_1, \dots, x_{i-1}, a_i, \dots, x_n).$$

Aplicaciones sucesivamente lo anterior se obtiene,

$$\Delta_{a, b} F_X(x) = \Delta_{a_1, b_1}, \dots, \Delta_{a_n, b_n} F_X(x_1, \dots, x_n) = P_X((a, b]),$$

lo que permite generalizar el anterior resultado mediante la expresión,

$$\Delta_{a, b} F_X(x) \geq 0, \text{ para todo } a, b \in \mathbb{R}^n, \text{ con } a_i \leq b_i.$$

Por lo tanto se tiene que,

$$P_X((a, b]) = \Delta_{a, b} F_X(x),$$

obteniendo así $\frac{1}{2}$ la equivalencia entre P_X y F_X .

Definición 1.3.3 Sea X un vector aleatorio. Se dice que X es discreto si y solo si las variables aleatorias $X_1, \dots, X_n$ que la componen, son todas discretas. Así pues, la función de probabilidad p , para este tipo de vectores aleatorios estaría dada por $p(x) = p\{X = x\}$, más explícitamente,

$$P\{X \in B\} = \sum_{x \in B} p(x).$$

Definición 1.3.4 Sea X un vector aleatorio. Diremos que X es absolutamente continua si y solo si, hay una función Borel medible y no negativa, f en $\mathbb{R}^n$, llamada función de densidad de X , tal que,

$$F(x) = \int_{(-\infty, x]} f(t) dt, \quad x \in \mathbb{R}^n.$$

De esto se tiene lo siguiente,

$$P_X(B) = \int_B f(x)dx, \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}^n).$$

Sea μ una medida definida por $\mu(B) = \int_B f(x)dx$, $B \in \mathcal{B}(\mathbb{R}^n)$. Como esta satisface que $\mu[a, b] = \int_{[a, b]} f(x)dx = F(a, b]$, entonces, μ es la medida de Lebesgue-Stieltjes determinada por F , es decir, $\mu = P_x$.

Definición 1.3.5 Sea $(\Omega, \mathcal{F}, P)$ el espacio de probabilidad y sean $X_1, \dots, X_n$ v.a. definidas sobre este. Entonces, se dice que $X_1, \dots, X_n$ son independientes, si y solo si, para todos los conjuntos $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, se tiene:

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\}.$$

También se puede expresar la independiencia de las v.a. $X_1, \dots, X_n$, si se dice que $X = (X_1, \dots, X_n)$, entonces P_X se puede ver como el producto de P_{x_i} , $i = 1, \dots, n$.

Las siguientes implicaciones surgen de la definición de independiencia.

Si $X_1, \dots, X_n$ son independientes, también lo son $X_1, \dots, X_k$ para $k < n$. Esto se comprueba tomando $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$, luego se sigue que,

$$\begin{aligned} P\{X_1 \in B_1, \dots, X_k \in B_k\} &= P\{X_1 \in B_1, \dots, X_k \in B_k, X_{k+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}\} \\ &= P\{X_1 \in B_1\} \cdots P\{X_k \in B_k\}. \end{aligned}$$

Y por lo tanto, no es necesario verificar todas las subfamilias de la colección de eventos $\{X_i \in B_i\}$, $i = 1, \dots, n$.

La independiencia para las v.a. extendidas se define de la misma forma que para las v.a. tomando a $\mathcal{B}(\mathbb{R})$ en lugar de $\mathcal{B}(\mathbb{R})$. El siguiente teorema nos muestra como la independiencia de las v.a. puede ser caracterizada en terminos de las funciones de distribución.

Teorema 1.3.1 Sean $X_1, \dots, X_n$ v.a. en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Sea F_i la función de distribución de X_i , $i = 1, \dots, n$, y F la función de distribución de $X = (X_1, \dots, X_n)$. Entonces $X_1, \dots, X_n$ son independientes si y solo si:

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n) \quad \text{para todo real } x_i, i = 1, \dots, n.$$

Demostración Si $X_1, \dots, X_n$ son independientes, entonces,

$$F(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \prod_{i=1}^n P\{X_i \leq x_i\} = \prod_{i=1}^n F_i(x_i).$$

Demostremos el recíproco. Para esto asumimos que $F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i)$ para todo $x_i, i = 1, \dots, n$. De esto se sigue:

$$P_X(a, b] = F(a, b] = \prod_{i=1}^n [F_i(b_i) - F_i(a_i)] = \prod_{i=1}^n P_{X_i}(a_i, b_i].$$

Así,

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdots P\{X_n \in B_n\},$$

cuando los B_i son intervalos de los reales semicerrados por derecha. Ahora, fijando los intervalos $B_2, \dots, B_n$. La colección $\mathcal{C}$ de $B_1 \in \mathcal{B}(\mathbb{R})$, para la cual la anterior igualdad sostiene que es una clase monótona que incluye las uniones finitas disjuntas de intervalos semicerrados por derecha, se tiene que $\mathcal{C} = \mathcal{B}(\mathbb{R})$ por el teorema de la clase monótona. Aplicando el mismo razonamiento a cada coordenada a su vez, obtenemos la independencia de los X_i . □

También se puede caracterizar la independencia en terminos de las densidades.

Teorema 1.3.2 Si $X = (X_1, \dots, X_n)$ es un vector aleatorio con densidad f , entonces cada X_i es una v.a. con densidad f_i . En este caso, $X_1, \dots, X_n$ son independientes si y solo si $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ para todo $(x_1, \dots, x_n) \in \mathcal{B}(\mathbb{R}^n)$ casi en todas partes.

Demostración Sea,

$$\begin{aligned} F_1(x_1) &= P\{X_1 \leq x_1\} \\ &= P\{X_1 \leq x_1, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}\} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \cdots dt_n. \end{aligned}$$

Por la definición de continuidad absoluta, X_1 tiene una densidad dada por,

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \cdots dx_n,$$

Por el teorema de Fubini se tiene que f_1 es Borel medible. Similarmente, cada X_i tiene una densidad de f_i , obtenida por la integración de todas las variables excepto x_i .

Ahora, si $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$, entonces,

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n = F_1(x_1) \cdots F_n(x_n),$$

entonces los X_i son independientes.

Recíprocamente, si los X_i son independientes, entonces,

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(t_1) \cdots f_n(t_n) dt_1 \cdots dt_n.$$

Entonces, si $g(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$, se tiene que,

$$P_X(B) = \int_B g(x) dx, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Pero, $P_X(B) = \int_B f(x) dx$, entonces se concluye que $f = g$, por fuera de un conjunto de medida cero.

□

Observación 1.3.1 Si $X_1, \dots, X_n$ son v.a. independientes y cada X_i es una v.a. con densidad f_i , $i = 1, \dots, n$, entonces $X = (X_1, \dots, X_n)$ es un vector aleatorio con densidad f dada por $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$.

Teorema 1.3.3 Sea X un vector aleatorio definido en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$ con función de densidad f_X . Sea $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ una función medible tal que:

1. $g = (g_1, \dots, g_n)$ admite inversa $g^{-1}(y) = (g_1^*, \dots, g_n^*)$.

2. Para todo $i, j = 1, \dots, n$, existe $\frac{\partial g_i^*(y_1, \dots, y_n)}{\partial y_j}$.

3. $J_{g^{-1}} = \left| \left(\left(\frac{\partial g_i^*(y_1, \dots, y_n)}{\partial y_j} \right) \right)_{i,j} \right| \neq 0$.

Con estas condiciones, el vector aleatorio n -dimensional $Y = g(X)$ es de tipo continuo y su función de densidad es

$$f_Y(y) = f_X(g^{-1}(y)) |J_{g^{-1}(y)}|, \quad \text{para todo } y \in \mathbb{R}^n.$$

Demostración Vamos a calcular la función de distribución del vector Y , ya que X es continua, la probabilidad de que tome valores en un determinado conjunto se obtiene integrando f_X en dicho conjunto,

$$F_Y(y) = P(Y \leq y) = P(X \in g_{1-}((-\infty, y])) = \int_{g^{-1}((-\infty, y])} f_X(x) dx, \quad \text{para todo } y \in \mathbb{R}^n,$$

se aplica un cambio de variable en la integral y, teniendo en cuenta las propiedades de la función g se obtiene,

$$F_Y(y) = \int_{g^{-1}((-\infty, y])} f_X(x) dx = \int_{(-\infty, y]} f_X(g^{-1}(t)) |J(t)| dt, \quad \text{para todo } y \in \mathbb{R}^n,$$

La expresión de F_Y implica que el vector aleatorio Y es continuo, y su función de densidad es la indicada en el teorema.

□

Observación 1.3.2 *El teorema de cambio de variable exige que el vector aleatorio transformado tenga la misma dimensión que el original. En caso de que se quiera calcular la función de densidad de un vector que tenga dimensión menor que el original, se consideran variables transformadas auxiliares hasta igualar la dimensión de este; es decir, si $X = (X_1, \dots, X_n)$ e $Y = g(X) = (Y_1, \dots, Y_m)$, con $m < n$, se considera un vector auxiliar $Y' = g'(X) = (Y_1, \dots, Y_m, Y_{m+1}, \dots, Y_n)$, tal que, la transformación de X a Y' satisfaga las condiciones del teorema. Se aplica la fórmula de cambio de variable para obtener la función de densidad de Y' , y a partir de ella se calcula la función de densidad marginal de Y .*

Teorema 1.3.4 *Sea X un vector aleatorio definido en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$, con función de distribución F . Sea g una función Borel medible de $\mathbb{R}^n$ a $\mathbb{R}$.*

Se define $Y = g \circ X$, se tiene que,

$$E(Y) = \int_{\mathbb{R}} g(x) dF_X = \left(\int_{\mathbb{R}} g dP_X \right)$$

Demostración *Sea g un indicador $I_B, B \in \mathcal{B}(\mathbb{R})$. Entonces,*

$$E(Y) = E(I_B \circ X) = E(I_{X \in B}) = P_X(B) = \int_{\mathbb{R}} g dP_X,$$

luego, $E(Y)$ y $\int_{\mathbb{R}} g dP_X$ existen y son iguales.

Ahora, sea g una función simple no negativa, tal que, $g(x) = \sum_{j=1}^n x_j I_{B_j}(x)$, donde los B_j son conjuntos disjuntos en $\mathcal{B}(\mathbb{R})$. Se sigue que,

$$\begin{aligned} E(Y) &= \sum_{j=1}^n x_j E(I_{B_j} \circ X) \\ &= \sum_{j=1}^n x_j \int_{\mathbb{R}} I_{B_j} dP_X \\ &= \int_{\mathbb{R}} \left(\sum_{j=1}^n x_j I_{B_j} \right) dP_X \\ &= \int_{\mathbb{R}} g dP_X. \end{aligned}$$

Nuevamente se tiene que ambas integrales existen y son iguales.

Si g es una función Borel medible no negativa y sean $g_1, \dots, g_n$ funciones simples no negativas con $g_n \uparrow g$. Por lo anterior,

$$E(g_n \circ X) = \int_{\mathbb{R}} g_n dP_X;$$

por lo tanto por el toerema de la convergencia monotona,

$$E(g \circ X) = \int_{\mathbb{R}} g dP_X,$$

y se llega a la existencia e igualdad entre ambas integrales.

Finalmente, si $g = g^+ - g^-$ es una función arbitraria Borel medible y $Y = g \circ X$, se tiene,

$$\begin{aligned} E(Y) &= E(Y^+) - E(Y^-) \\ &= E(g^+ \circ X) - E(g^- \circ X) \\ &= \int_{\mathbb{R}} g^+ dP_X - \int_{\mathbb{R}} g^- dP_X \\ &= \int_{\mathbb{R}} g dP_X. \end{aligned}$$

Por lo anterior, si $E(Y)$ existe y, $E(Y^-)$ es finita, entonces $\int_{\mathbb{R}} g^- dP_X$ es finita, y se tiene que $\int_{\mathbb{R}} g dP_X$ existe. Usando el mismo razonamiento, la existenccia de $\int_{\mathbb{R}} g dP_X$ implica la de $E(Y)$.

□

Definición 1.3.6 La función generadora de momentos (f.g.m.) de una variable aleatoria X es una función $m_X(t)$ definida como

$$m_X(t) = E(e^{tX}).$$

Decimos que existe la f.g.m de X , si existe una constante positiva a tal que $m_X(t)$ es finita para todo $t \in [-a, a]$.

En particular, si X es una v.a. discreta, entonces

$$m_X(t) = \sum_{n=0}^{\infty} e^{tn} p_X(n).$$

Observación 1.3.3 La f.g.m. de X genera todos los momentos de X . Es por eso que se llama función generadora de momentos. Los momentos de las v.a. están determinados de la siguiente forma $m'_X(0) = E(X)$, $m''_X(0) = E(X^2)$ y en general $m_X^{(n)}(0) = E(X^n)$.

Teorema 1.3.5 La f.g.m (si existe) determina de forma unívoca la distribución. Es decir, si dos variables aleatorias tienen la misma f.g.m., entonces deben tener la misma distribución. Por lo tanto, si encuentra la f.g.m de una variable aleatoria, se puede determinar su distribución.

Capítulo 2

La Distribución de Poisson

En este capítulo presentamos la distribución de Poisson, una de las distribuciones de probabilidad más utilizadas para analizar eventos independientes que ocurren a una velocidad constante dentro de un intervalo de tiempo dado, también permite analizar recuentos de eventos que se supone ocurren aleatoriamente en el tiempo o pueden surgir como una “ley de números pequeños”, es decir, como un límite de la distribución Binomial con gran tamaño de muestra y parámetro pequeño.

La distribución de Poisson también tiene propiedades convenientes, tales como; la suma de variables aleatorias independientes, cada una con distribución de Poisson, tiene una distribución de Poisson y la convergencia con otras distribuciones conocidas como la Binomial y Multinomial.

Definición 2.0.1 *Una variable aleatoria X tiene distribución de Poisson de parámetro λ , en símbolos: $X \sim \text{Pois}(\lambda)$, si X es discreta y tiene una función de densidad de probabilidad (o función de masa):*

$$p_X(n) = e^{-\lambda} \frac{\lambda^n}{n!} I_{\mathbf{N}_0}(n).$$

La función de distribución de tal variable aleatoria X es $F_X(x) = \sum_{n=0}^{\lceil x \rceil} e^{-\lambda} \frac{\lambda^n}{n!} I_{[0, \infty)}(x)$, donde $\lceil x \rceil$ indica el mayor entero contenido en x .

Una caracterización de la distribución de Poisson

Una variable aleatoria discreta X que toma valores en $\mathbf{N}_0$ tiene distribución de Poisson(λ) si y sólo si $E[Xf(X)] = \lambda E[f(X+1)]$ para toda función $f : \mathbf{N}_0 \rightarrow [0, \infty)$. En efecto, si $X \sim \text{Pois}(\lambda)$, entonces

$$\begin{aligned}
E[Xf(X)] &= \sum_{n=0}^{\infty} nf(n)e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \sum_{n=1}^{\infty} nf(n)e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \sum_{j=0}^{\infty} (j+1)f(j+1)e^{-\lambda} \frac{\lambda^{j+1}}{(j+1)!} \\
&= \lambda \sum_{j=0}^{\infty} f(j+1)e^{-\lambda} \frac{\lambda^j}{j!} \\
&= \lambda E[f(X+1)].
\end{aligned}$$

Recíprocamente, tomando $f(x) = I_n(x)$; $n = 1, 2, \dots$, de una parte, $E[Xf(X)] = nP(X = n)$ y, de otra parte, $\lambda E[f(X+1)] = \lambda P(X = n-1)$. Por tanto, para $n = 1, 2, \dots$ se cumple $P(X = n) = \frac{\lambda}{n} P(X = n-1)$. De esto se sigue que $X \sim \text{Pois}(\lambda)$.

Funciones generadoras y momentos de la distribución de Poisson

Sea $X \sim \text{Pois}(\lambda)$. La f.g.m. de X está dada por

$$m_X(t) = \sum_{n=0}^{\infty} e^{tn} p_X(n) = \sum_{n=0}^{\infty} e^{tn} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda(1-e^t)}, \quad t \in \mathbf{R}.$$

Por tanto, $E[X] = m'_X(0) = \lambda$, $E[X^2] = m''_X(0) = \lambda^2 + \lambda$ y $\text{Var}[X] = \lambda$.

Suma de variables aleatorias independientes con distribución de Poisson

Si $X_1, X_2, \dots, X_k$ son variables aleatorias independientes tales que $X_i \sim \text{Pois}(\lambda_i)$; $i = 1, 2, \dots, k$, entonces la f.g.m. de la variable aleatoria $S = X_1 + X_2 + \dots + X_k$ es:

$$m_S(t) = \prod_{i=1}^k e^{-\lambda_i(1-e^t)} = e^{-\Lambda(1-e^t)},$$

donde $\Lambda = \sum_{i=1}^k \lambda_i$. Esto significa que $S \sim \text{Pois}(\Lambda)$.

De manera general, se tiene el siguiente resultado.

Proposición 2.0.1 Si $\{X_n\}_{n \in \mathbf{N}}$ es una sucesión de variables aleatorias independientes tales que $X_n \sim \text{Pois}(\lambda_n)$; $n = 1, 2, 3, \dots$, $\Lambda = \sum_{n=1}^{\infty} \lambda_n$ y $S = \sum_{n=1}^{\infty} X_n$

- Si $\Lambda < \infty$, entonces S converge y $S \sim \text{Pois}(\Lambda)$.
- Si $\Lambda = +\infty$, entonces S diverge.

Demostración ■ Sabemos que $S_k = \sum_{i=1}^k X_i \sim \text{Pois} \left(\Lambda_k = \sum_{i=1}^k \lambda_i \right)$.

Para $j \in \{0, 1, 2, 3, 4, \dots\} \equiv \mathbb{N}_0$ fijo. se tiene que:

$$P(S_k \leq j) = \sum_{m=0}^j e^{-\Lambda_k} \frac{\Lambda_k^m}{m!}.$$

Ahora, la sucesión de eventos $A_k := \{S_k \leq j\}$; $k \in \mathbb{N}$ es decreciente, pues $x_1 + x_2 + \dots + x_{i-1} + x_i \leq j$ implica $x_1 + x_2 + \dots + x_{i-1} \leq j$. Por tanto

$$\begin{aligned} P \left(\bigcap_{k=1}^{\infty} A_k \right) &= \lim_{k \rightarrow \infty} P(A_k), \\ P \left(\sum_{k=1}^{\infty} X_k \leq j \right) &= \lim_{k \rightarrow \infty} \left[e^{-\Lambda_k} \sum_{m=0}^j \frac{\Lambda_k^m}{m!} \right] \\ &= e^{-\lim \Lambda_k} \lim \left[\sum_{m=0}^j \frac{\Lambda_k^m}{m!} \right] \equiv e^{-\Lambda} \sum_{m=0}^j \frac{\Lambda^m}{m!} \\ &\equiv P(\text{Pois}(\Lambda) \leq j). \end{aligned}$$

Esto implica que $P(S = j) = e^{-\Lambda} \frac{\Lambda^j}{j!}$; $j = 0, 1, 2, \dots$

■ Si $\Lambda = +\infty$, para $j \in \mathbb{N}_0$ fijo,

$$P(S_k \leq j) = e^{-\Lambda_k} \sum_{m=0}^j \frac{\Lambda_k^m}{m!},$$

y

$$P(S \leq j) = \lim_{k \rightarrow \infty} \frac{\sum_{m=0}^j \frac{\Lambda_k^m}{m!}}{e^{\Lambda_k}} = 0.$$

Esto implica que para cada $j \in \mathbb{N}_0$

$$P(S_k > j) = 1.$$

Si $B_j = \{S > j\}_{j \in \mathbb{N}}$, entonces $\{B_j\} \downarrow$,

$$P(S = +\infty) = 1.$$

□

Relación con la distribución Binomial

Una propiedad importante de la distribución de Poisson es que puede usarse para aproximar una v.a. con distribución Binomial cuando el parámetro binomial n es grande y p es pequeño. Para ver esto, sea una variable aleatoria X con distribución Binomial de parámetros n, p y sea $\lambda = np$, entonces,

$$\begin{aligned} \binom{n}{x} p^x (1-p)^{n-x} &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \frac{n!}{(n-x)!} \frac{1}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \frac{n(n-1) \cdots (n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^x}, \end{aligned}$$

tomando el límite de $n \rightarrow \infty$,

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1) \cdots (n-x+1)}{n^x} \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^x \approx 1,$$

luego,

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!},$$

obteniendo la función de masa de probabilidad para la distribución de Poisson con parámetro λ .

Relación con la distribución Multinomial

Si $X_1, X_2, \dots, X_k, X_{k+1}$ son variables aleatorias independientes tales que $X_i \sim \text{Pois}(\lambda_i)$; $i = 1, 2, \dots, k, k+1$, entonces la distribución condicional de $X_1, X_2, \dots, X_k$ dado que $S \equiv X_1 + X_2 + \dots + X_k + X_{k+1} = n$ es multinomial de parámetros $n, p_1, p_2, \dots, p_k$, donde $p_i = \frac{\lambda_i}{\Lambda}$, $i = 1, 2, \dots, k$ y $\Lambda = \sum_{i=1}^{k+1} \lambda_i$. En efecto; $S \sim \text{Pois}(\Lambda)$ y, por la hipótesis de independencia,

$$\begin{aligned} P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k \mid S = n) &= \frac{P(X_1 = n_1, \dots, X_k = n_k, X_{k+1} = n_{k+1})}{P(S = n)} \\ &= \frac{\prod_{i=1}^{k+1} P(X_i = n_i)}{P(S = n)} \\ &= \frac{\prod_{i=1}^{k+1} e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}}{e^{-\Lambda} \frac{\Lambda^n}{n!}} \\ &= \binom{n}{n_1, n_2, \dots, n_k, n_{k+1}} p_1^{n_1} \cdots p_k^{n_k} p_{k+1}^{n_{k+1}}, \end{aligned}$$

donde $n_{k+1} = n - \sum_{i=1}^k n_i$ y $p_{k+1} = 1 - \sum_{i=1}^k p_i$.

También, si $S \sim Pois(\Lambda)$ y $(X_1, X_2, \dots, X_k)$ dado $S = n$ tiene distribución multinomial de parámetros $n, p_1, p_2, \dots, p_k$, entonces $X_1, X_2, \dots, X_k$ son variables aleatorias independientes tales que $X_i \sim Pois(\lambda_i = \Lambda p_i)$; $i = 1, 2, \dots, k$. De hecho, si hacemos $n_{k+1} = n - \sum_{i=1}^k n_i$ y $p_{k+1} = 1 - \sum_{i=1}^k p_i$, entonces para $n_1, n_2, \dots, n_k \in N_0$ se verifica

$$\begin{aligned}
P(X_1 = n_1, \dots, X_k = n_k) &= \sum_n P(X_1 = n_1, \dots, X_k = n_k \mid S = n) P(S = n) \\
&= \sum_n \frac{n!}{n_1! \dots n_k! n_{k+1}!} p_1^{n_1} \dots p_k^{n_k} p_{k+1}^{n_{k+1}} \frac{e^{-\Lambda} \Lambda^n}{n!} \\
&= \frac{(\Lambda p_1)^{n_1} (\Lambda p_2)^{n_2} \dots (\Lambda p_k)^{n_k}}{n_1! n_2! \dots n_k!} \sum_{n_{k+1}=0}^{\infty} \frac{e^{-\Lambda} (\Lambda p_{k+1})^{n_{k+1}}}{n_{k+1}!} \\
&= \frac{(\Lambda p_1)^{n_1} (\Lambda p_2)^{n_2} \dots (\Lambda p_k)^{n_k}}{n_1! n_2! \dots n_k!} e^{-\Lambda(1-p_{k+1})} \\
&= \prod_{i=1}^k e^{-\Lambda p_i} \frac{(\Lambda p_i)^{n_i}}{n_i!}.
\end{aligned}$$

Capítulo 3

La Distribución Exponencial

La distribución exponencial es una de las distribuciones más ampliamente utilizadas en la práctica estadística, probablemente debido a su naturaleza simple, sin embargo posee varias propiedades estadísticas importantes y exhibe una gran capacidad de modelación matemática.

La distribución exponencial se utiliza para modelar datos de tiempo de vida y varios problemas relacionados con eventos de tiempo de espera, algunos ejemplos que se tienen son, la duración de las llamadas telefónicas, la vida útil de una bombilla y el tiempo entre impulsos sucesivos en las médulas espinales de varios mamíferos. Este capítulo está dedicado al estudio de la distribución exponencial, sus propiedades y caracterizaciones, y los modelos que se inducen de ella e ilustran sus aplicaciones.

Definición 3.0.1 *Se dice que una variable aleatoria continua X tiene distribución exponencial de parámetro $\lambda > 0$, en símbolos: $X \sim \text{Exp}(\lambda)$, si su función de distribución es $F_X(x) = (1 - e^{-\lambda x})I_{[0,\infty)}(x)$, o lo que es lo mismo, si X tiene función de densidad*

$$f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x).$$

Una caracterización de la distribución Exponencial

Si $X \sim \text{Exp}(\lambda)$, entonces para $s, t \in [0, \infty)$ se verifica

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-(s+t)}}{e^{-s}} = P(X > t).$$

Si X modela el tiempo de vida de una bombilla, la propiedad anterior, llamada falta de memoria de la distribución exponencial, establece que si la bombilla está funcionando en el tiempo s , entonces el restante tiempo de vida de esta bombilla es el mismo tiempo de vida de una bombilla que recién se instala en el tiempo s .

La distribución exponencial, es la única distribución continua que toma valores no negativos y también satisface la propiedad de falta de memoria. De forma precisa, si X

es una variable aleatoria continua tal que $F_X(0) = 0$, $F_X(x) < 1$ para $x > 0$ y verifica la propiedad anterior, entonces $X \sim \text{Exp}(\lambda)$. En efecto; si

$$g(t) := P(X > t) = 1 - F_X(t), \quad t \in [0, \infty),$$

entonces la propiedad de falta de memoria dice que para $s, t \in [0, \infty)$, se cumple $g(t+s) = g(t)g(s)$. Ahora, tomando $s = t = 1$ se obtiene $g(2) = g(1)^2$ y, usando inducción sobre n , se llega a $g(n) = g(1)^n$; $n = 1, 2, \dots$. Puesto que $1 = \frac{m}{m} = \frac{1}{m} + \frac{1}{m} + \dots + \frac{1}{m}$ (m -veces); $m = 1, 2, 3, \dots$, entonces $g(1) = g(\frac{1}{m})^m$, i.e., $g(\frac{1}{m}) = g(1)^{\frac{1}{m}}$. Similarmente; para un número racional positivo $r = \frac{n}{m} = \frac{1}{m} + \frac{1}{m} + \dots + \frac{1}{m}$ (n -veces) se llega a $g(r) = g(\frac{1}{m})^n = g(1)^{\frac{n}{m}} = g(1)^r$. Por último, si t es un número irracional positivo y $\{r_n\}_{n \in \mathbf{N}}$ es una sucesión de números racionales con $r_1 > r_2 > \dots > t$ y $\lim_{n \rightarrow \infty} r_n = t$, puesto que $g(\cdot)$ es una función continua por la derecha:

$$g(t) = \lim_{n \rightarrow \infty} g(r_n) = g(1)^t.$$

Ahora bien; como $g(t) = 1 - F_X(t) \rightarrow 0$ si $t \rightarrow \infty$, entonces $g(1) < 1$. De otra parte, $F_X(x) < 1$ para todo x implica $0 < g(1)$. En resumen

$$g(t) = g(1)^t = e^{(\ln g(1))t} = e^{\lambda t},$$

donde $\lambda = \ln g(1)$.

Función generadora de momentos y momentos de la distribución Exponencial

Si $X \sim \text{Exp}(\lambda)$, su f.g.m. es

$$m_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{-(\lambda-t)x} dx.$$

Esta integral converge sólo cuando $\lambda - t > 0$, en cuyo caso se obtiene

$$m_X(t) = \frac{\lambda}{\lambda - t}; \quad t < \lambda.$$

Ahora, para $-\lambda < t < \lambda$ se puede escribir [Ver 21.22 de [2]]

$$m_X(t) = \frac{1}{1 - (t/\lambda)} = \sum_{n=0}^{\infty} (t/\lambda)^n = \sum_{n=0}^{\infty} E[X^n] \frac{t^n}{n!}.$$

De aquí que $E[X^n] = \frac{n!}{\lambda^n}$, $n = 1, 2, \dots$. En particular, $E[X] = \frac{1}{\lambda}$ y $\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{\lambda^2}$.

Distribución del mínimo de las variables aleatorias i.i.d. con distribución Exponencial.

Sean $X_1, X_2, \dots, X_n$ variables aleatorias i.i.d. con distribución exponencial con sus respectivos parámetros $\lambda_1, \dots, \lambda_n$. Sea $M = \min\{X_1, X_2, \dots, X_n\}$, entonces su distribución y función de probabilidad están dadas por:

1. Para $t > 0$,

$$P(M > t) = e^{-t(\lambda_1 + \dots + \lambda_n)}.$$

Es decir, M tiene distribución exponencial con parámetro $\lambda_1 + \dots + \lambda_n$.

2. Para $k = 1, \dots, n$,

$$P(M = X_k) = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}.$$

En efecto,

1. Para $t > 0$,

$$\begin{aligned} P(M > t) &= P(X_1 > t, \dots, X_n > t) = P(X_1 > t) \cdot P(X_n > t) \\ &= e^{-\lambda_1 t} \dots e^{-\lambda_n t} = e^{-t(\lambda_1 + \dots + \lambda_n)}. \end{aligned}$$

2. Lo anterior muestra que la f.g.m de M es la de una distribución exponencial de parámetro $\lambda_1 + \dots + \lambda_n$.

La función de probabilidad de M se puede obtener mediante la siguiente condición:

Para $1 \leq k \leq n$, fijando X_k tenemos,

$$\begin{aligned} P(M = X_k) &= P(\min(X_1, \dots, X_n) = X_k) \\ &= P(X_1 \geq X_k, \dots, X_n \geq X_k) \\ &= \int_0^\infty P(X_1 \geq t, \dots, X_n \geq t | X_k = t) \lambda_k e^{-\lambda_k t} dt \\ &= \int_0^\infty P(X_1 \geq t, \dots, X_{k-1} \geq t, X_{k+1} \geq t, \dots, X_n \geq t) \lambda_k e^{-\lambda_k t} dt \\ &= \int_0^\infty P(X_1 \geq t) \dots P(X_{k-1} \geq t) P(X_{k+1} \geq t) \dots P(X_n \geq t) \lambda_k e^{-\lambda_k t} dt \\ &= \lambda_k \int_0^\infty e^{-t(\lambda_1 + \dots + \lambda_n)} dt = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}. \end{aligned}$$

Suma de variables aleatorias i.i.d. con distribución Exponencial. Relación con la distribución de Erlang

Sean $X_1, X_2, \dots, X_n, X_{n+1}$ variables aleatorias i.i.d. con distribución exponencial de parámetro λ . Sean ahora, $S_1 = X_1$, $S_2 = X_1 + X_2, \dots$, $S_{n+1} = X_1 + X_2 + \dots + X_{n+1}$. En lo que sigue se determinan las distribuciones de:

1. La variable aleatoria S_k ; $k = 2, 3, \dots, n + 1$.
2. El vector aleatorio $(S_1, S_2, \dots, S_k)$; $k = 2, 3, \dots, n + 1$.
3. El vector aleatorio (S_n, S_{n+1}) .

Distribución de la variable aleatoria S_k

Por la propiedad reproductiva, la f.g.m. de S_k es

$$m_{S_k}(t) = \prod_{i=1}^k m_{X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^k, \quad t < \lambda,$$

que es la f.g.m. de una variable aleatoria con distribución $Erlang(k, \lambda)$. Esto significa que S_k es una variable aleatoria continua con función de densidad

$$f_{S_k}(s) = \frac{\lambda^k s^{k-1}}{(k-1)!} e^{-\lambda s} I_{[0, \infty)}(s),$$

y función de distribución

$$F_{S_k}(s) = \int_0^s \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x} dx, \quad s > 0.$$

Integrando por partes se obtiene la relación

$$F_{S_k}(s) = 1 - e^{-\lambda s} \sum_{j=0}^{k-1} \frac{(\lambda s)^j}{j!}; \quad s \geq 0,$$

que relaciona las distribuciones de Erlang y de Poisson. Dicho de otra forma:

$$P[S_k \leq s] = 1 - P[Pois(\lambda s) \leq k - 1].$$

Distribución del vector aleatorio $(S_1, S_2, \dots, S_k)$

Se usa el método de las transformaciones para determinar la función de densidad del vector aleatorio $(S_1, S_2, \dots, S_k)$. Si $\mathcal{X} = [0, \infty)^k$ y $\mathcal{D} = \{(s_1, s_2, \dots, s_k) : 0 < s_1 < s_2 < \dots < s_k\}$, la transformación $g : \mathcal{X} \mapsto \mathcal{D}$ dada por

$$g(x_1, x_2, \dots, x_k) = (x_1, x_1 + x_2, \dots, x_1 + x_2 + \dots + x_k) = (s_1, s_2, \dots, s_k)$$

tiene inversa $g^{-1} : \mathcal{D} \mapsto \mathcal{X}$ definida por

$$g^{-1}(s_1, s_2, \dots, s_k) = (s_1, s_2 - s_1, \dots, s_k - s_{(k-1)}) = (x_1, x_2, \dots, x_k)$$

y su jacobiano es igual a 1. De aquí que el vector aleatorio $(S_1, S_2, \dots, S_k)$ es continuo con función de densidad

$$\begin{aligned} f_{(S_1, S_2, \dots, S_k)}(s_1, s_2, \dots, s_k) &= f_{(X_1, X_2, \dots, X_k)}(s_1, s_2 - s_1, \dots, s_k - s_{(k-1)}) I_{\mathcal{D}}(s_1, s_2, \dots, s_k) \\ &= \lambda e^{-\lambda s_1} e^{-\lambda(s_2 - s_1)} \dots e^{-\lambda(s_k - s_{(k-1)})} I_{\mathcal{D}}(s_1, s_2, \dots, s_k) \\ &= e^{-\lambda s_k} I_{\mathcal{D}}(s_1, s_2, \dots, s_k). \end{aligned}$$

Distribución del vector aleatorio (S_n, S_{n+1})

La función de densidad del vector aleatorio (S_n, S_{n+1}) se obtiene de la del vector aleatorio $(S_1, S_2, \dots, S_n, S_{n+1})$ de la siguiente forma; para $0 < s_n < s_{n+1}$:

$$\begin{aligned} f_{(S_n, S_{n+1})}(s_n, s_{n+1}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{(S_1, \dots, S_n, S_{n+1})}(s_1, \dots, s_n, s_{n+1}) ds_1 \dots ds_{n-1} \\ &= \int_0^{s_n} \int_{s_2}^{s_n} \dots \int_{s_{n-1}}^{s_n} \lambda^{n+1} e^{-\lambda s_{n+1}} ds_1 \dots ds_{n-1} \\ &= \lambda^{n+1} e^{-\lambda s_{n+1}} \int_0^{s_n} \int_{s_2}^{s_n} \dots \int_{s_{n-1}}^{s_n} ds_1 \dots ds_{n-1} \\ &= \lambda^{n+1} e^{-\lambda s_{n+1}} \frac{s_n^{n-1}}{(n-1)!}. \end{aligned}$$

Note que para $t > 0$:

$$\begin{aligned} P(S_n \leq t < S_{n+1}) &= \int_0^t \int_t^{\infty} \lambda^{n+1} e^{-\lambda s_{n+1}} \frac{s_n^{n-1}}{(n-1)!} ds_n ds_{(n-1)} \\ &= \int_0^t \lambda^n e^{-\lambda t} \frac{s_n^{n-1}}{(n-1)!} ds_n \\ &= \lambda^n e^{-\lambda t} \frac{t^n}{n!} \\ &= (\lambda t)^n \frac{e^{-\lambda t}}{n!} = P(Pois(\lambda t) = n). \end{aligned}$$

Otra forma de obtener el resultado anterior es observar que $\{S_n < t\} \subseteq \{S_{(n+1)} > t\}$, y por tanto:

$$\begin{aligned}
P(S_n \leq t < S_{n+1}) &= P(S_{n+1} > t) - P(S_n > t) \\
&= e^{-\lambda t} \sum_{j=0}^n \frac{(\lambda t)^j}{j!} - e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!} \\
&= (\lambda t)^n \frac{e^{-\lambda t}}{n!}.
\end{aligned}$$

Ahora obtenemos una expresión para la función de distribución del vector aleatorio $(S_n, S_{(n+1)})$. Para $0 < s_n < s_{(n+1)}$

$$\begin{aligned}
F_{(S_n, S_{n+1})}(s_n^*, s_{n+1}^*) &= \int_0^{s_n^*} \left[\int_{s_n}^{s_{n+1}^*} f_{(S_n, S_{n+1})}(s_n, s_{n+1}) ds_{n+1} \right] ds_n \\
&= \int_0^{s_n^*} \frac{s_n^{n-1}}{(n-1)!} \lambda^n \left[-e^{-\lambda s_{n+1}} \Big|_{s_n}^{s_{n+1}^*} \right] ds_n \\
&= \int_0^{s_n^*} \frac{s_n^{n-1}}{(n-1)!} \lambda^n \left[e^{-\lambda s_n} - e^{-\lambda s_{n+1}^*} \right] ds_n \\
&= \int_0^{s_n^*} \frac{s_n^{n-1}}{(n-1)!} \lambda^n e^{-\lambda s_n} ds_n - \frac{\lambda^n e^{-\lambda s_{n+1}^*}}{(n-1)!} \int_0^{s_n^*} s_n^{n-1} ds_n \\
&= P(S_n \leq s_n^*) - \frac{\lambda^n e^{-\lambda s_{n+1}^*}}{n!} (s_n^*)^n \\
&= 1 - e^{-\lambda s_n^*} \sum_{j=0}^{n-1} \frac{(\lambda s_n^*)^j}{j!} - e^{-\lambda s_{n+1}^*} \frac{(\lambda s_n^*)^n}{n!}.
\end{aligned}$$

Así, se tiene que

$$\begin{aligned}
P(S_n \leq s_n^*, s_{(n+1)}^* < S_{(n+1)}) &= \int_0^{s_n^*} \int_{s_{n+1}^*}^{\infty} \frac{s_n^{n-1} \lambda^{n+1} - e^{-\lambda s_{n+1}}}{(n-1)!} ds_{n+1} ds_n \\
&= \int_0^{s_n^*} \frac{s_n^{n-1} \lambda^{n+1}}{(n-1)!} \left[e^{-\lambda s_{(n+1)}^*} \right] ds_n \\
&= e^{-\lambda s_{n+1}^*} \frac{(\lambda s_n^*)^n}{n!}.
\end{aligned}$$

Capítulo 4

El Proceso de Poisson

En el presente capítulo estudiaremos el proceso de Poisson. Un modelo relevante tanto en las aplicaciones como en la teoría general de los procesos estocásticos. Se definirá este proceso de varias maneras y estudiaremos algunas de sus propiedades, sus generalizaciones y algunas de sus aplicaciones. Seguiremos [1, 2, 8, 9, 10] como referencias para esta sección.

Antes de dar una primera definición del proceso de Poisson, definiremos las nociones de Proceso estocástico, incrementos estacionarios e incrementos independientes.

Un proceso estocástico $\mathbf{X} = \{X(t) : t \in T\}$ es una colección de variables aleatorias para cada t en el conjunto de índices T , es decir, $X(t)$ es una variable aleatoria para cada $t \in T$. Denominamos t como tiempo y llamamos $X(t)$ el estado del proceso en el tiempo t . Si el conjunto de índices t es un conjunto contable, llamamos a X un proceso estocástico de tiempo discreto, y si T es un intervalo de $\mathbb{R}$, lo llamamos un proceso estocástico de tiempo continuo.

El otro componente fundamental, además del conjunto de índices I para definir la estructura de un proceso estocástico, es el espacio de estado S . Este es el conjunto de todos los valores que puede tomar el proceso, es decir, es el espacio muestral de cada una de las variables aleatorias de la colección $\mathbf{X}$ en la sucesión que conforma el proceso.

Para un proceso estocástico de tiempo continuo $\{X(t) : t \geq 0\}$, un incremento es la diferencia en el proceso en dos tiempos s y t . Para $s < t$, el incremento de tiempo s al tiempo t es la diferencia $X(t) - X(s)$.

Se dice que un proceso tiene incrementos estacionarios si la distribución del incremento $X(t) - X(s)$ depende de s y t solo a través de la diferencia $t - s$, para todos los $s < t$. Entonces la distribución de $X(t_1) - X(s_1)$ es la misma que la distribución de $X(t_2) - X(s_2)$ si $t_1 - s_1 = t_2 - s_2$. Hay que tener en cuenta que los intervalos $[s_1, t_1]$ y $[s_2, t_2]$ pueden superponerse.

Se dice que un proceso tiene incrementos independientes si dos incrementos que implican intervalos disjuntos son independientes. Es decir, si $s_1 < t_1 < s_2 < t_2$, entonces los dos incrementos $X(t_1) - X(s_1)$ y $X(t_2) - X(s_2)$ son independientes.

Ahora procedamos con una definición de un proceso de Poisson.

Definición 4.0.1 Sean $T_1, T_2, \dots$ variables aleatorias independientes idénticamente distribuidas, que denotaremos como v.a.i.i.d., con distribución exponencial de parámetro λ . Sea $\tau_0 = 0$ y $\tau_n = T_1 + \dots + T_n$ para $n \geq 1$. Definimos el proceso de Poisson de parámetro (intensidad) λ por,

$$N(s) = \text{máx}\{n : \tau_n \leq s\}, \quad s \geq 0.$$

Las variables T_n representan los intervalos de tiempo entre eventos sucesivos, $\tau_n = T_1 + \dots + T_n$ es el instante en el ocurre el n -ésimo evento y $N(s)$ es el número de eventos que han ocurrido hasta el instante s .

Una caracterización fundamental del proceso de Poisson es la siguiente:

Teorema 4.0.1 Un proceso estocástico de tiempo continuo $\{N(t) : t \geq 0\}$ es un proceso de Poisson con parámetro $\lambda > 0$ si y solo si,

- (i) $N(0) = 0$.
- (ii) $N(t + s) - N(s) \sim \text{Pois}(\lambda t)$.
- (iii) $N(t)$ tiene incrementos independientes.

Demostración Para demostrar la primera afirmación calculemos la distribución de $N(s)$.

$N(s) = n$ si y solo si $\tau_n \leq s < \tau_{n+1}$, es decir, el n -ésimo evento ocurre antes del instante s pero el $(n + 1)$ -ésimo ocurre después de s . Usando la ley de la probabilidad total, condicionando respecto al instante en el cual ocurre τ_n , obtenemos,

$$P(N(s) = n) = \int_0^s P(\tau_{n+1} > s | \tau_n = t) f_{\tau_n}(t) dt = \int_0^s P(T_{n+1} > s - t) f_{\tau_n}(t) dt.$$

y por referencia anterior,

$$f_{\tau_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \quad \text{para } t \geq 0,$$

entonces,

$$\begin{aligned}
P(N(s) = n) &= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dt \\
&= \frac{(\lambda)^n}{(n-1)!} e^{-\lambda s} \int_0^s t^{n-1} dt = \frac{(\lambda s)^n}{n!} e^{-\lambda s}.
\end{aligned}$$

Sea τ_n el instante en el cual ocurre el n -ésimo evento. El primer evento ocurre después de t si y solo si no ocurre ningún evento en $[0, t]$. Por la fórmula para la distribución de Poisson,

$$P(\tau_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

lo cual muestra que $\tau_1 = T_1 \sim \text{Exp}(\lambda)$. Para $T_2 = \tau_2 - \tau_1$ observamos que,

$$\begin{aligned}
P(T_2 > t | T_1 = s) &= P(\text{no ocurre ningún evento en } (s, s+t] | T_1 = s) \\
&= P(N(t+s) - N(t) = 0 | N(r) = 0 \text{ para } r < s, N(s) = 1) \\
&= P(N(t+s) - N(t) = 0) = e^{-\lambda t}.
\end{aligned}$$

por la propiedad de incrementos independientes, de modo que $T_2 \sim \text{Exp}(\lambda)$ y es independiente de T_1 . Repitiendo este argumento vemos que $T_1, T_2, \dots$ son v.a.i.i.d. con distribución exponencial de parámetro λ .

Por lo tanto hemos demostrado que $N(s)$ tiene distribución de Poisson de parámetro λs .

Demostremos primero que $N(0) = 0$. Como $N(s)$ tiene distribución de Poisson de parámetro λs , esto implica que se tiene $P(N(s) = 0) = e^{-\lambda s}$, entonces:

$$P(N(s) = 0) = P(N(0) = 0) = e^0 = 1.$$

Ahora, supongamos que $N(s) = n$ y que el n -ésimo evento ocurrió en el instante τ_n . Sabemos que el intervalo de tiempo para el siguiente evento debe satisfacer $T_{n+1} > s - \tau_n$, pero por la propiedad de falta de memoria de la distribución exponencial,

$$P(T_{n+1} > s - \tau_n + t | T_{n+1} > s - \tau_n) = P(T_{n+1} > t) = e^{-\lambda t}.$$

Lo anterior indica que la distribución del tiempo de espera hasta el primer evento después de s es exponencial de parámetro λ y es independiente de T_i , $1 \leq i \leq n$. Por otro lado $T_{n+1}, T_{n+2}, \dots$ son independientes de T_i , $1 \leq i \leq n$ y por lo tanto también de τ_i , $1 \leq i \leq n$. Esto muestra que los intervalos entre eventos que ocurren después de s son v.a.i.i.d., con distribución exponencial de parámetro λ , y por lo tanto $N(t+s) - N(s)$ es un proceso de Poisson.

Ahora, $N(s)$ tiene distribución de Poisson de parámetro λs lo que implica que $N(t_n) - N(t_{n+1})$ es independiente de $N(r)$, $r \geq t_{n-1}$ y en consecuencia también de $N(t_{n-1}) - N(t_{n-2}), \dots, N(t_1) - N(t_0)$. Procediendo por inducción tenemos que si $t_0 < t_1 < \dots <$

t_n , entonces $N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$ son independientes, es decir, $N(t)$ tiene incrementos independientes.

Con lo cual se concluye la demostración. $\square$

Observación 4.0.1 Cabe resaltar, que el resultado anterior da una caracterización de los procesos de Poisson de una manera que dependen únicamente de las distribuciones en el tiempo t .

De esta definición se sigue que el espacio de estado del proceso es $S = \{0, 1, 2, \dots\}$, también que, para incrementos estacionarios, la distribución de $N(t) - N(s)$ para $s < t$, tiene la misma distribución que $N(t - s) - N(0) = N(t - s)$, la cual es una distribución de Poisson con parámetro $(t - s)$. Por último, el proceso no es decreciente, es decir, $N(t) - N(s) \geq 0$ con probabilidad 1 para cualquier $s < t$, ya que $N(t) - N(s)$ tiene una distribución de Poisson.

El proceso de Poisson también es visto como un conteo de eventos a medida que se avanza en el tiempo, es decir, sea $N(t)$ el número de eventos que han ocurrido hasta el momento t y al cabo del tiempo $t + s$, se habrán contado más eventos ocurridos de la forma $N(t + s) - N(t)$, con $N(t + s) - N(t)$ teniendo distribución de Poisson con parámetro λs .

Por tal motivo se puede denotar el proceso de Poisson como un proceso de conteo. Una forma muy común en que se emplea el proceso de Poisson consiste en interpretar a $N(t)$ como el número de llegadas de tareas/trabajos/clientes a un sistema por tiempo t .

Se define entonces un proceso de conteo de la siguiente forma:

Definición 4.0.2 Se dice que un proceso estocástico $\{N(t) : t \geq 0\}$ es un proceso de conteo si representa el número de eventos ocurridos hasta el tiempo t , es decir, satisface:

- (i) $N(0) \geq 0$.
- (ii) Si $s < t$, entonces $N(s) \leq N(t)$.
- (iii) Para $s < t$, $N(s) - N(t)$ es igual al número de eventos que ocurrieron en el intervalo $(s, t]$.

Para poder observar más de la estructura de un proceso de Poisson, consideremos un análogo de este en el tiempo discreto, el proceso de Bernoulli, descrito como sigue:

Tomemos el intervalo $[0, 1)$. Ahora, dividamos en subintervalos disjuntos, cada uno de longitud h , con h muy pequeño, tal que se tienen los siguientes intervalos $[0, h), [h, 2h), [2h, 3h), \dots$. Supongamos que cada intervalo corresponde a un ensayo independiente

de Bernoulli, de modo que en cada intervalo, independientemente de cualquier otro intervalo, hay un evento exitoso con probabilidad λh . Definimos entonces el proceso de Bernoulli $\{B(t) : t = 0, h, 2h, \dots\}$, donde $B(t)$ es el número de ensayos exitosos hasta el momento t .

Por como se definió anteriormente, se puede ver que el proceso de Bernoulli $B(t)$ tiene incrementos estacionarios e independientes, además, $B(0) = 0$. Así, el proceso de Bernoulli es una aproximación de tiempo discreto al proceso de Poisson con parámetro λ si $B(t) \sim \text{Pois}(\lambda t)$.

Para un t de la forma nh , conocemos la distribución exacta de $B(t)$. Ahora bien, como hasta el momento t hay n ensayos independientes y cada uno con probabilidad λh de éxito, se dice que $B(t)$ posee una distribución binomial con parámetros n y λh . Por lo tanto, el número medio de éxitos hasta el momento t es $n\lambda h = \lambda t$, lo cual coincide con $E[B(t)]$. El hecho de que la distribución de $B(t)$ es aproximadamente $\text{Pois}(\lambda t)$ se deduce de la aproximación de Poisson a la distribución binomial, como se observó en el Capítulo 2.

Sea k un entero no negativo y $t > 0$. Teniendo en cuenta que $t = nh$ para algún número entero positivo n ,

$$\begin{aligned} P(B(t) = k) &= \binom{n}{k} (\lambda h)^k (1 - \lambda h)^{n-k} \\ &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &= \frac{n!}{(n-k)!n^k} \left(1 - \frac{\lambda t}{n}\right)^{-k} \frac{(\lambda t)^k}{k!} \left(1 - \frac{\lambda t}{n}\right)^n \\ &\approx \frac{n!}{(n-k)!n^k} \left(1 - \frac{\lambda t}{n}\right)^{-k} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \end{aligned}$$

para n muy grande (o h muy pequeño). También se tiene para n grande que,

$$\left(1 - \frac{\lambda t}{n}\right)^{-k} \approx 1,$$

y

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \approx 1.$$

Por lo tanto, $P(B(t) = k) \approx (\lambda t)^k / k! e^{-\lambda t}$

En el proceso de Bernoulli la probabilidad de éxito en cualquier intervalo dado es λh y la probabilidad de dos o más éxitos es 0, es decir, $P(B(h) = 1) = \lambda h$ y $P(B(h) \geq 2) = 0$. Por lo tanto, en el proceso de Poisson tenemos la aproximación de que $P(N(h) = 1) \approx \lambda h$ y $P(N(h) \geq 2) \approx 0$.

Una manera más precisa de escribir lo anterior es,

$$P(N(h) = 1) = \lambda h + o(h) \text{ y } P(N(h) = 2) = o(h),$$

tal que, $o(h)$ se denomina notación $o(h)$ de Landau, y significa cualquier función de h que sea de menor orden que h . Esto implica que, si $f(h)$ es $o(h)$ entonces $f(h)/h \rightarrow 0$ como $h \rightarrow 0$, i.e., $f(h)$ va a 0 más rápido que h va a 0.

Con lo anterior definamos el proceso de Poisson de la siguiente forma:

Definición 4.0.3 *Una colección de variables aleatorias $\{N(t) : t \geq 0\}$ es un Proceso de Poisson de parámetro $\lambda > 0$ si:*

- (i) $N(0) = 0$.
- (ii) $N(t)$ tiene incrementos estacionarios e incrementos independientes.
- (iii) Para h pequeña:
 - (a) $P(N(t+h) - N(t) = 1) = \lambda h + o(h)$.
 - (b) $P(N(t+h) - N(t) > 1) = o(h)$.

La condición *iii.a* nos dice que la probabilidad de que ocurra exactamente un evento en un intervalo de tiempo pequeño, es proporcional a la longitud del intervalo, y la condición *iii.b* nos dice que la probabilidad de que ocurran dos o más eventos en un intervalo de tiempo pequeño es despreciable.

Construcción de un proceso de Poisson a partir de ensayos independientes de Bernoulli

Un proceso de Poisson de parámetro λ puede verse como el resultado de realizar un ensayo independiente de Bernoulli con probabilidad de éxito $p = \lambda dt$ en cada intervalo de tiempo “infinitesimal” de longitud dt , y colocar un punto si la prueba correspondiente es un éxito (caso contrario no se coloca ningún punto). Intuitivamente, esto produciría un proceso puntual con ambos tipos de incrementos, estacionarios e independientes, es decir, un proceso de Poisson. El número de ensayos de Bernoulli que pueden ajustarse en cualquier intervalo sólo depende de la duración del intervalo y, por lo tanto, la distribución del número de éxitos en ese intervalo también dependerá únicamente de la duración. Para dos intervalos no superpuestos, los ensayos de Bernoulli en cada uno serían independientes entre sí, ya que todos los ensayos son i.i.d., por lo tanto, el número de éxitos en un intervalo sería independiente del número de éxitos en el otro intervalo.

La distribución Exponencial se puede obtener como un límite de la distribución Geométrica: Se fija un n grande, y se realiza, utilizando la probabilidad de éxito $p_n = \lambda/n$, un ensayo independiente de Bernoulli en cada punto de tiempo i/n , $i \geq 1$. Sea Y_n el tiempo en el cual ocurrió el primer éxito. Entonces $Y_n = K_n/n$ donde K_n denota el

número de ensayos hasta el primer éxito, y tiene distribución Geométrica con probabilidad de éxito p_n . Como $n \rightarrow \infty$, Y_n converge a un v.a. Y con distribución Exponencial de parámetro λ . Por lo tanto, Y puede servir como el primer tiempo de llegada t_1 para un proceso de Poisson de parámetro λ . Los pequeños intervalos de longitud $1/n$ se convierten (en el límite) en los intervalos dt infinitesimales. Una vez que se tiene el primer éxito, en el tiempo t_1 , continuamos adelante en el tiempo (en el intervalo (t_1, ∞)) con nuevas pruebas de Bernoulli hasta que obtengamos el segundo éxito en el tiempo t_2 y así sucesivamente hasta que obtengamos todos los tiempos de llegada $\{t_n : n \geq 1\}$. Por construcción, cada tiempo entre llegadas, $X_n = t_n - t_{n-1}$, $n \geq 1$, es una v.a.i distribuida exponencialmente con parámetro λ ; Por lo tanto, se construyó proceso de Poisson de parámetro λ .

Otra forma de comprender cómo se puede construir el proceso de Poisson a partir de los ensayos de Bernoulli es el hecho de que la distribución de Poisson se puede usar para aproximar la distribución Binomial, como se observa en el Capítulo 2. Con dicho resultado se construye el proceso de conteo $N(t)$ en el tiempo t , para un proceso de Poisson de la siguiente manera: Se fija $t > 0$. Se divide el intervalo $(0, t]$ en n subintervalos, $((i-1)t/n, it/n]$, $1 \leq i \leq n$, de igual longitud t/n . En el punto final derecho it/n de cada subintervalo, se realiza una prueba de Bernoulli con probabilidad de éxito $p_n = \lambda t/n$, y se coloca un punto si tiene éxito (no se coloca el punto en caso contrario). Sea $N_n(t)$ el número de tales puntos colocados (éxitos). Luego $N_n(t) \sim \text{Binom}(n, p_n)$ y converge en distribución a $N(t) \sim \text{Pois}(\lambda t)$, cuando $n \rightarrow \infty$. Además, los puntos colocados en $(0, t]$ de los ensayos de Bernoulli convergen (cuando $n \rightarrow \infty$) a los puntos $t_1, \dots, t_{N(t)}$ del proceso de Poisson durante $(0, t]$. Por lo que en realidad se obtuvo el proceso de Poisson hasta el tiempo t .

Teorema 4.0.2 *El Teorema 4.0.1 y la Definición 4.0.3 son equivalentes.*

Demostración *Veamos primero que Teorema 4.0.1 $\Rightarrow$ Definición 4.0.3.*

Sabemos que un proceso de Poisson tiene incrementos estacionarios, esto es Teorema 4.0.1 $\Rightarrow$ Definición 4.0.3.ii.

Falta ver que Teorema 4.0.1 $\Rightarrow$ Definición 4.0.3.iv. Para esto tenemos que:

$$\begin{aligned} P(N(t+h) - N(t) = 1) &= P(N(h) = 1) \\ &= (\lambda h)e^{-\lambda h} \\ &= \lambda h + \lambda h(e^{-\lambda h} - 1). \end{aligned}$$

Luego, como

$$\lim_{h \rightarrow 0} \frac{\lambda h(e^{-\lambda h} - 1)}{h} = h \lim_{h \rightarrow 0} (e^{-\lambda h} - 1) = 0,$$

tenemos que,

$$P(N(t+h) - N(t) = 1) = \lambda h + o(h).$$

Ahora, para la condición (b) tenemos,

$$\begin{aligned} P(N(t+h) - N(t) > 1) &= P(N(h) > 1) \\ &= \sum_{k=2}^{\infty} (\lambda h)^k \frac{e^{-\lambda h}}{k!}. \end{aligned}$$

y como,

$$\lim_{h \rightarrow 0} \frac{(\lambda h)^k \frac{e^{-\lambda h}}{k!}}{h} = \frac{h}{k!} \lim_{h \rightarrow 0} h^{k-1} e^{-\lambda h} = 0, \text{ para } k \geq 2,$$

tenemos que,

$$\lim_{h \rightarrow 0} \sum_{k=2}^{\infty} (\lambda h)^k \frac{e^{-\lambda h}}{k!} = 0,$$

esto es,

$$P(N(t+h) - N(t) > 1) = o(h).$$

Por lo tanto Teorema 4.0.1 $\Rightarrow$ Definición 4.0.3.

Veamos ahora que Definición 4.0.3 $\Rightarrow$ Teorema 4.0.1. Falta ver que Definición 4.0.3 $\Rightarrow$ Teorema 4.0.1.ii. Para esto definamos una función $f(t, n) := P(N(t) = n)$ y mostremos que $f(t, 0) = e^{-\lambda t}$, entonces,

$$\begin{aligned} f(t+h, 0) &= P(N(t+h) = 0) \\ &= P(N(t) = 0, N(t+h) - N(t) = 0) \\ &= P(N(t) = 0)P(N(t+h) - N(t) = 0) \\ &= f(t, 0)(1 - P(N(t+h) - N(t) \geq 1)) \\ &= f(t, 0)(1 - (P(N(t+h) - N(t) = 1) + P(N(t+h) - N(t) > 1))) \\ &= f(t, 0)(1 - ((\lambda h + o(h)) + o(h))) \\ &= f(t, 0)(1 - (\lambda h + o(h))) \\ &= f(t, 0)(1 - \lambda h + o(h)). \end{aligned}$$

Tenemos entonces,

$$\lim_{h \rightarrow 0} \frac{f(t+h, 0) - f(t, 0)}{h} = \lim_{h \rightarrow 0} \frac{f(t, 0)(-\lambda h + o(h))}{h}$$

$$f'(t, 0) = -\lambda f(t, 0),$$

de donde,

$$f(t, 0) = e^{-\lambda t}.$$

Notemos que para todo $n > 1$,

$$\frac{d}{dt}f(t, n) = \lambda(f(t, n-1) - f(t, n)).$$

Ahora, veamos que,

$$\begin{aligned}
f(t+h, n) &= P(N(t+h) = n) \\
&= \sum_{k=0}^n P(N(t) = n-k, N(t+h) - N(t) = k) \\
&= \sum_{k=0}^n P(N(t) = n-k)P(N(t+h) - N(t) = k) \\
&= P(N(t) = n)P(N(t+h) - N(t) = 0) \\
&\quad + P(N(t) = n-1)P(N(t+h) - N(t) = 1) \\
&\quad + \sum_{k=2}^n P(N(t) = n-k)P(N(t+h) - N(t) = k) \\
&= f(t, n)(1 - P(N(t+h) - N(t) \geq 1)) + f(t, n-1)(\lambda h + o(h)) \\
&\quad + \sum_{k=2}^{\infty} (\lambda t)^{n-k} \frac{e^{-\lambda t}}{(n-k)!} o(h) \\
&= f(t, n)(1 - (P(N(t+h) - N(t) = 1) + P(N(t+h) - N(t) > 1))) \\
&\quad + \lambda h f(t, n-1) + o(h) \\
&= f(t, n)(1 - ((\lambda h + o(h)) + o(h))) + \lambda h f(t, n-1) + o(h) \\
&= f(t, n) + \lambda h(f(t, n-1) - f(t, n)) + o(h).
\end{aligned}$$

Entonces,

$$\lim_{h \rightarrow 0} \frac{f(t+h, n) - f(t, n)}{h} = \lim_{h \rightarrow 0} \frac{(-\lambda h(f(t, n-1) - f(t, n)) + o(h))}{h}$$

$$\frac{d}{dt}f(t, n) = \lambda(f'(t, n-1) - f(t, n)).$$

Luego, demostremos que para todo $n \geq 0$, $f(t, n) = (\lambda t)^n \frac{e^{-\lambda t}}{n!}$.

$$\begin{aligned}
\frac{d}{dt} \left((\lambda t)^{n+1} \frac{e^{-\lambda t}}{(n+1)!} \right) &= \frac{\lambda^{n+1}}{(n+1)!} \frac{d}{dt} (t^{n+1} e^{-\lambda t}) \\
&= \frac{\lambda^{n+1}}{(n+1)!} ((n+1)t^n e^{-\lambda t} - \lambda t^{n+1} e^{-\lambda t}) \\
&= \lambda \left((\lambda t)^n \frac{e^{-\lambda t}}{n!} - (\lambda t)^{n+1} \frac{e^{-\lambda t}}{(n+1)!} \right).
\end{aligned}$$

Así, por el teorema de existencia y unicidad de soluciones de ecuaciones diferenciales, tenemos el resultado y por lo tanto $\{N(t) : t \geq 0\}$ es un proceso de Poisson de parámetro λ .

Por consiguiente se tiene que el Teorema 4.0.1 y la Definición 4.0.3 son equivalentes. $\square$

Definición 4.0.4 *Un proceso de llegada es una sucesión creciente de variables aleatorias, $0 < X_1 < X_2 < \dots$, donde $X_i < X_{i+1}$ significa que $X_{i+1} - X_i$ es una variable aleatoria positiva, es decir, una v.a X tal que $F_X(0) = 0$.*

Sea $N(t)$ un proceso de Poisson con parámetro λ . Sea X_1 el momento de la primera llegada. Entonces,

$$P(X_1 > t) = P(\text{no hay llegadas en } (0, t]) = e^{-\lambda t}.$$

Por lo tanto, $X_1 \sim \text{Exp}(\lambda)$. Sea X_2 el tiempo transcurrido entre la primera y la segunda llegada. Sean $s > 0$ y $t > 0$. Como los intervalos $(0, s]$ y $(s, s+t]$ son disjuntos, se tiene,

$$\begin{aligned} P(X_2 > t | X_1 = s) &= P(\text{no hay llegadas en } (s, s+t] | X_1 = s) \\ &= P(\text{no hay llegadas en } (s, s+t]) \\ &= e^{-\lambda t}. \end{aligned}$$

Luego $X_2 \sim \text{Exp}(\lambda)$, y X_1 y X_2 son independientes. Del mismo modo podemos argumentar que para las v.a. $X_1, X_2, \dots$ todos los X_i son independientes y $X_2 \sim \text{Exp}(\lambda)$ para $i = 1, 2, \dots$. Las variables aleatorias $X_1, X_2, \dots$ se denominan tiempos entre llegadas del proceso de conteo $N(t)$.

Teorema 4.0.3 *Si $N(t)$ un proceso de Poisson con parámetro λ , los tiempos entre llegadas $X_1, X_2, \dots$ son i.i.d con $X_i \sim \text{Exp}(\lambda)$, para $i = 1, 2, 3, \dots$.*

Demostración *Procedamos por inducción. Por lo anterior tenemos que $X_1 \sim \text{Exp}(\lambda)$. Ahora veamos el caso general. Sea X_n el tiempo entre el evento $(n-1)$ y el n entonces $X_1, \dots, X_n$ son i.i.d. con distribución exponencial de parámetro λ . Mostremos que se satisface X_{n+1} . Para hacer esto, fijamos $t, x_1, \dots, x_n > 0$. El teorema será verdadero si mostramos que la distribución de X_{n+1} condicionada a $X_1 = x_1, \dots, X_n = x_n$ no depende de $x_1, \dots, x_n$, que muestra que X_{n+1} es independiente de $X_1, \dots, X_n$ y $P(X_n > t) = e^{-\lambda t}$. Entonces consideremos la probabilidad condicional,*

$$P(X_{n+1} > t | X_1 = x_1, \dots, X_n = x_n)$$

Primero, reexpresaremos el evento $X_n = x_n, \dots, X_1 = x_1$. Así pues, sea $S_k = X_1 + \dots + X_k$ el k -ésimo tiempo de llegada y sea $s_k = x_1 + \dots + x_k$, para $k = 1, \dots, n$. Entonces,

$$\{X_n = x_n, \dots, X_1 = x_1\} = \{S_n = s_n, \dots, S_1 = s_1\},$$

reescribiendo la probabilidad condicional tenemos,

$$P(X_{n+1} > t | X_n = x_n, \dots, X_1 = x_1) = P(X_{n+1} > t | S_n = s_n, \dots, S_1 = s_1)$$

El hecho de que el evento $\{X_{n+1} > t\}$ es independiente del evento $\{S_n = s_n, \dots, S_1 = s_1\}$ se debe a que son incrementos independientes, para esto veamos que, si el evento $\{S_n = s_n, \dots, S_1 = s_1\}$ ocurre entonces el evento $\{X_{n+1} > t\}$ ocurre si y solo si no hay llegadas en el intervalo $(s_n, s_n + t]$, así se tiene que,

$$P(X_{n+1} > t | S_n = s_n, \dots, S_1 = s_1) = P(N(s_n + t) - N(s_n) = 0 | S_n = s_n, \dots, S_1 = s_1).$$

Por lo tanto, deseamos expresar el evento $\{S_n = s_n, \dots, S_1 = s_1\}$ en términos de incrementos disjuntos del incremento $N(s_n + t) - N(s_n)$. Definimos los incrementos,

$$\begin{aligned} I_1^{(k)} &= N(s_1 - 1/k) - N(0) \\ I_i^{(k)} &= N(s_i - 1/k) - N(s_{i-1} + 1/k) \text{ para } i = 2, \dots, n, \end{aligned}$$

para $k > M$, donde se elige M para que $1/k$ sea más pequeño que el más pequeño tiempo entre llegadas, y también definir los incrementos,

$$\begin{aligned} B_i^{(k)} &= N(s_i + 1/k) - N(s_i - 1/k) \text{ para } i = 1, \dots, n-1 \\ B_n^{(k)} &= N(s_n) - N(s_n - 1/k), \end{aligned}$$

para $k > M$. Los incrementos $I_1^{(k)}, B_1^{(k)}, \dots, I_n^{(k)}, B_n^{(k)}$ son todos disjuntos y representan todo el intervalo $[0, s_n]$. Ahora definimos el evento,

$$A_k = \{I_1 = 0\} \cap \dots \cap \{I_n = 0\} \cap \{B_1 = 1\} \cap \dots \cap \{B_n = 1\}.$$

Entonces A_k está contenido en A_{k-1} , luego la sucesión $\{A_k\}_{k=M}^{\infty}$ es una sucesión decreciente de conjuntos tal que,

$$\{S_n = s_n, \dots, S_1 = s_1\} = \bigcap_{k=M}^{\infty} A_k,$$

Sin embargo (y es por eso que construimos los eventos A_k), para cualquier k el evento A_k es independiente del evento $\{N(s_n + t) - N(s_n) = 0\}$ porque el incremento $N(s_n + t) - N(s_n)$ es independiente de todos los incrementos $I_1^{(k)}, B_1^{(k)}, \dots, I_n^{(k)}, B_n^{(k)}$, ya que todos son incrementos disjuntos. Pero si el evento $\{N(s_n + t) - N(s_n) = 0\}$ es independiente de A_k para cada k , este es independiente de la intersección de A_k . Por lo tanto,

$$\begin{aligned} P(X_{n+1} > t | S_n = s_n, \dots, S_1 = s_1) &= P(N(s_n + t) - N(s_n) = 0 | S_n = s_n, \dots, S_1 = s_1) \\ &= P\left(N(s_n + t) - N(s_n) = 0 \middle| \bigcap_{k=M}^{\infty} A_k\right) \\ &= P(N(s_n + t) - N(s_n) = 0) \\ &= P(N(t) = 0) = e^{-\lambda t}, \end{aligned}$$

Así, se tiene que $X_{n+1} \sim \text{Exp}(\lambda)$ y es independiente de $X_1, \dots, X_n$ y por lo tanto queda demostrado que la sucesión de tiempos entre llegadas $X_1, X_2, \dots$ son v.a.i.i.d con distribución exponencial de parámetro λ .

□

Nótese que, si $X \sim \text{Exp}(\lambda)$, entonces X es una v.a. sin memoria como se muestra en el Capítulo 3. Pensando en el proceso de Poisson, la propiedad de pérdida de memoria de los tiempos entre llegadas es consistente con la propiedad de los incrementos independientes de la distribución de Poisson. En cierto sentido, ambos implican que el número de llegadas en intervalos no superpuestos son independientes.

Como ya tenemos la distribución de los tiempos entre llegadas, encontraremos la distribución de los tiempos de llegada. Para eso tenemos el siguiente teorema.

Teorema 4.0.4 *Sea S_n el tiempo de la n -ésima llegada en un proceso de Poisson, donde $S_n = X_1 + \dots + X_n$ es la suma de los primeros n tiempos entre llegadas, entonces $S_n \sim \text{Gamma}(n, \lambda)$.*

Demostración *Mostraremos que la suma de n v.a.i.i.d con distribución exponencial de parámetro λ tienen una distribución Gamma de parámetros n y λ . Para $n = 1$, el resultado es inmediato. Para $n > 1$, observemos que el tiempo de la n -ésima llegada es menor o igual a t si y solo si el número de llegadas en el intervalo $[0, t]$ es mayor o igual que n , es decir, los dos eventos,*

$$\{S_n \leq t\} \text{ y } \{N(t) \geq n\}$$

son equivalentes. Sin embargo, como la probabilidad del primer evento $\{S_n \leq t\}$ da la función de distribución acumulada de S_n , entonces calculemos la función de distribución acumulada de S_n :

$$F_{S_n}(t) \equiv P(S_n \leq t) = P(N(t) \geq n) = \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t}.$$

Para obtener la densidad de S_n , derivamos con respecto a t , entonces,

$$\begin{aligned} f_{S_n}(t) &= - \sum_{j=n}^{\infty} \lambda \frac{(\lambda t)^j}{j!} e^{-\lambda t} + \sum_{j=n}^{\infty} \lambda \frac{(\lambda t)^{j-1}}{(j-1)!} e^{-\lambda t} \\ &= \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}. \end{aligned}$$

Por lo tanto, se tiene que $S_n \sim \text{Gamma}(n, \lambda)$.

□

Por lo tanto, hemos demostrado en el Teorema 4.0.1 una caracterización de la Definición 4.0.3. Ahora demostraremos la equivalencia entre la Definición 4.0.1 y la Definición 4.0.2.

Teorema 4.0.5 *La Definición 4.0.1 y la Definición 4.0.2 son equivalentes.*

Demostración *Partamos de la Definición 4.0.1. Es decir, sean $X_1, X_2, \dots$ una sucesión de v.a.i.i.d. con parámetro λ . Para cada n , sea $S_n = X_1 + X_2 + \dots + X_n$, con*

$S_0 = 0$, y sea $N(t) = \max\{n : S_n \leq t\}$, con $N(t) \sim \text{Exp}(\lambda t)$ como se mostró anteriormente.

Para $k \geq 0$, $N(t) = k$ si y solo si la k -ésima llegada ocurre en el tiempo t y la llegada $(k+1)$ ocurre después de t . Es decir, $S_k \leq t < S_k + X_{k+1}$. Ya que S_k es una función de $X_1, \dots, X_k$, S_k y X_{k+1} son v.a. independientes, y su densidad conjunta es el producto de sus densidades marginales, por tanto,

$$f_{S_k, X_{k+1}}(s, x) = f_{S_k}(s)f_{X_{k+1}}(x) = \left(\frac{\lambda^k s^{k-1} e^{-\lambda s}}{(k-1)!} \lambda e^{-\lambda x} \right), \quad s, x > 0.$$

Para $k \leq 0$,

$$\begin{aligned} P(N(t) = k) &= P(S_k \leq t \leq S_k + X_{k+1}) \\ &= P(S_k \leq t, X_{k+1} \geq t - S_k) \\ &= \int_0^t \int_{t-s}^{\infty} f_{S_k, X_{k+1}}(s, x) dx ds \\ &= \int_0^t \int_{t-s}^{\infty} \left(\frac{\lambda^k s^{k-1} e^{-\lambda s}}{(k-1)!} \right) \lambda e^{-\lambda x} dx ds \\ &= \frac{\lambda^k}{(k-1)!} \int_0^t (s^{k-1} e^{-\lambda s}) e^{-\lambda(t-s)} ds \\ &= \frac{e^{-\lambda t} \lambda^k}{(k-1)!} \int_0^t s^{k-1} ds = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \end{aligned}$$

por lo tanto tiene distribución de Poisson.

El hecho de que los tiempos entre llegadas de un proceso de Poisson no tengan memoria nos indica que el patrón de llegadas desde un tiempo arbitrario en adelante se comporta igual que el patrón de llegadas desde el momento 0 en adelante. Por tanto se deduce que el número de llegadas en el intervalo $(s, s+t]$ tiene la misma distribución que el número de llegadas en $(0, t]$, por lo que los incrementos estacionarios, como independientes son consecuencias directas de lo anterior.

Ahora, supongamos que se tiene Definición 4.0.2. Sea X_1 el momento de la primera llegada. Entonces $X_1 > t$ si y solo si no hay llegadas en $[0, t]$, entonces,

$$P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t}, \quad \text{para } t > 0.$$

Por lo tanto, $X_1 \sim \text{Exp}(\lambda)$. Sea X_2 el tiempo transcurrido entre la primera y la segunda llegada. Sea $s > 0$ y $t > 0$, como los intervalos $(0, s]$ y $(s, s+t]$ son disjuntos, se tiene,

$$\begin{aligned} P(X_2 > t | X_1 = s) &= P(N_{s+t} - N_s | X_1 = s) \\ &= P(N_{s+t} - N_s = 0) \\ &= P(N(t) = 0) = e^{-\lambda t} \quad \text{para } t > 0. \end{aligned}$$

Luego $X_2 \sim \text{Exp}(\lambda)$, y X_1 y X_2 son independientes. Para el caso general, consideremos $P(X_{k+1} > t | X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$. Si la k -ésima llegada ocurre entre el tiempo $x_1 + \dots + x_k$ y $X_{k+1} > t$, entonces no hay llegadas en el intervalo $(x_1 + \dots + x_k + t]$ y se tiene que,

$$\begin{aligned} P(X_{k+1} > t | X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) \\ &= P(N_{x_1 + \dots + x_{k+1}} - N_{x_1 + \dots + x_k} = 0 | X_1 = x_1, \dots, X_k = x_k) \\ &= P(N_{x_1 + \dots + x_{k+1}} - N_{x_1 + \dots + x_k} = 0) \\ &= P(N(t) = 0) = e^{-\lambda t}, \quad \text{para } t > 0. \end{aligned}$$

Se sigue que, X_{k+1} es independiente de $X_1, \dots, X_k$ y $X_{k+1} \sim \text{Exp}(\lambda)$, se deduce entonces que $X_1, X_2, \dots$ es una sucesión de v.a.i.i.d. con parámetro λ , por lo que se tiene la Definición 4.0.1

Por lo tanto queda demostrado que la Definición 4.0.1 y la Definición 4.0.2 son equivalentes. □

El siguiente teorema confirma el hecho intuitivamente natural de que al superponer varios procesos de Poisson mutuamente independientes obtenemos un proceso de Poisson. En la suma debajo del conjunto de índices puede ser finita o infinitamente contable. En lo ultimo caso debemos suponer que $\sum_j \lambda_j < \infty$.

Teorema 4.0.6 Si $N_1(t), N_2(t), \dots$ son procesos independientes de Poisson con intensidades λ_j , entonces $N(t) = \sum_j N_j(t)$ es un proceso de Poisson con intensidad $\lambda = \sum_j \lambda_j$.

El siguiente resultado auxiliar se utiliza para probar el teorema.

Lema 4.0.1 Si $N_j(t) \sim \text{Pois}(\lambda_j)$ son independientes, entonces $\sum_j N_j(t) \sim \text{Pois}(\sum_j \lambda_j)$.

Demostración (Teorema) Ahora demostraremos el Teorema 4.0.6, para esto verificaremos las tres condiciones en el Teorema 4.0.1.

(i) Como cada N_j es tal que $N_j(0) = 0$, entonces $N(0) = \sum_j N_j(0) = 0$.

(ii) Con la ayuda de Lema 4.0.1, se tiene que

$$N(t) - N(s) = \sum_j (N_j(t) - N_j(s)) \sim \text{Pois}(\lambda(t - s)),$$

donde $\lambda = \sum_j \lambda_j$.

(iii) Si los intervalos de tiempo $(s_1, t_1]$ y $(s_2, t_2]$ son disjuntos, entonces $N_j(s_1, t_1]$ es localmente independiente sobre $N_j(s_2, t_2]$, o dicho en símbolos,

$$N_j(s_1, t_1] \perp\!\!\!\perp N_j(s_2, t_2] \text{ para todo } j.$$

Dado que los N_j son mutuamente independientes, esto permite concluir que

$$\sum_j N_j(s_1, t_1] \perp\!\!\!\perp \sum_j N_j(s_2, t_2].$$

Por lo tanto, los enteros aleatorios $N(s_1, t_1]$ y $N(s_2, t_2]$ son independientes. Lo anterior funciona de la misma manera para múltiples intervalos de tiempo disjuntos. Por lo tanto N tiene incrementos independientes.

Como se satisfacen las tres condiciones del Teorema 4.0.1, se tiene que $N(t) = \sum_j N_j(t)$ es un proceso de Poisson.

□

Por lo anterior, se concluye que las Definiciones 4.0.1, 4.0.2 y 4.0.3 son equivalentes.

El siguiente teorema demuestra una especie de convergencia. Si los eventos llegan acorde a un proceso de Poisson y estos eventos son de diferentes tipos, entonces si el tipo de evento es independiente del proceso de llegada, cada tipo específico de proceso de llegada es un proceso de Poisson.

Teorema 4.0.7 Sea $N(t)$ un proceso de Poisson de parámetro λ . Supongamos que cada llegada es de tipo 1, con probabilidad p_1 , o de tipo 2, con probabilidad $p_2 = 1 - p_1$, independiente de las demás llegadas. Sea $N_i(t)$ el número de llegadas del tipo i hasta el tiempo t . Entonces $N_1(t)$ y $N_2(t)$ son procesos de Poisson independientes con parámetros λp_1 y λp_2 respectivamente.

Demostración Sea $\{\xi_k\}_{k \geq 0}$ una sucesión de v.a. Bernoulli de parámetro p_1 . Entonces

$$N_1(t) = \sum_{k=1}^{N(t)} \xi_k$$

y

$$N_2(t) = \sum_{k=1}^{N(t)} (1 - \xi_k) = N(t) - N_1(t).$$

Es claro que $N_1(t)$ y $N_2(t)$ tienen incrementos independientes y estacionarios, pues $N(t)$ los tiene. Además, se tiene que $P(N_i(t+h) - N_i(t) > 1) = o(h)$, puesto que $P(N(t+h) - N(t) > 1) = o(h)$ y $N_i(h) = s \Rightarrow N(t) \geq s$.

Para h pequeña se tiene que, $P(N_1(t+h) - N(t) = 1) = p_1 \lambda h + o(h)$.

$$\begin{aligned}
P(N_1(t+h) - N(t) = 1) &= P\left(\sum_{k=1}^{N_{t+h}} \xi_k - \sum_{k=1}^{N_t} \xi_k = 1\right) \\
&= P\left(\sum_{N_t+1}^{N_{t+h}} \xi_k = 1\right) \\
&= P(N(t+h) - N(t) = 1, \xi_k = 1) \\
&= P(N(t+h) - N(t) = 1)P(\xi_k = 1) \\
&= p_1(\lambda h + o(h)) \\
&= \lambda p_1 h + o(h).
\end{aligned}$$

Análogamente se tiene que $P(N_2(t+h) - N_2(t) = 1) = \lambda p_2 h + o(h)$. Por lo tanto $N_1(t)$ y $N_2(t)$ son procesos de Poisson con parámetros λp_1 Y λp_2 .

Veamos ahora que son independientes. Sea $t \geq 0$ y $s_1, s_2 \in \mathbb{R}$.

$$\begin{aligned}
P(N_1 \leq s_1, N_2(t) \leq s_2) &= \sum_{k_1=0}^{s_1} P(N_1(t) = k_1, N_2(t) \leq s_2) \\
&= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} P(N_1(t) = k_1, N_2(t) = k_2),
\end{aligned}$$

dado que $N_1 + N_2(t) = N(t)$,

$$P(N_1(t) = k_1, N_2(t) = k_2) = P(N_1(t) = k_1, N_2(t) = k_2 | N(t) = k_1 + k_2) P(N(t) = k_1 + k_2)$$

y

$$P(N_1(t) = k_1, N_2(t) = k_2 | N(t) = k_1 + k_2) = \binom{k_1 + k_2}{k_1} p_1^{k_1} p_2^{k_2},$$

entonces,

$$\begin{aligned}
P(N_1 \leq s_1, N_2(t) \leq s_2) &= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \binom{k_1+k_2}{k_1} p_1^{k_1} p_2^{k_2} \frac{e^{-\lambda t}}{(k_1+k_2)!} \\
&= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \frac{1}{k_1! k_2!} p_1^{k_1} (\lambda t)^{k_1} p_2^{k_2} (\lambda t)^{k_2} e^{-\lambda(p_1+p_2)t} \\
&= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} (\lambda p_1 t)^{k_1} \frac{e^{-\lambda p_1 t}}{k_1!} (\lambda p_2 t)^{k_2} \frac{e^{-\lambda p_2 t}}{k_2!} \\
&= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} P(N_1 = k_1) P(N_2(t) = k_2) \\
&= \left(\sum_{k_1=0}^{s_1} P(N_1 = k_1) \right) \left(\sum_{k_2=0}^{s_2} P(N_2(t) = k_2) \right) \\
&= P(N_1 \leq s_1) P(N_2(t) \leq s_2).
\end{aligned}$$

Por lo tanto $N_1(t)$ y $N_2(t)$ son procesos de Poisson independientes con parámetros λp_1 y λp_2 . □

Definición 4.0.5 Sea Λ una v.a. positiva con distribución G y sea $N(t)$ sea un proceso de conteo tal que, dado $\Lambda = \lambda$, $N(t)$ es un proceso de Poisson con parámetro λ . Así,

$$P(N(t+s) - N(s) = n) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dG(\lambda).$$

El proceso $N(t)$ se llama un proceso de Poisson condicional, condicionado al evento de que $\Lambda = \lambda$, este es un proceso de Poisson con parámetro λ . Cabe resaltar que $N(t)$ no es un proceso de Poisson.

Calculemos la distribución condicional de Λ dado que $N(t) = n$. Para $d\lambda$ pequeño,

$$\begin{aligned}
P(\Lambda \in (\lambda, \lambda + d\lambda) | N(t) = n) &= \frac{P(N(t) = n | \Lambda \in (\lambda, \lambda + d\lambda)) P(\Lambda \in (\lambda, \lambda + d\lambda))}{P(N(t) = n)} \\
&= \frac{e^{-\lambda t} \frac{(\lambda t)^n}{n!} dG(\lambda)}{\int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dG(\lambda)},
\end{aligned}$$

se sigue entonces que la distribución condicional de Λ , dado que $N(t) = n$, está dada por

$$P(\Lambda \leq x | N(t) = n) = \frac{\int_0^x e^{-\lambda t} \frac{(\lambda t)^n}{n!} dG(\lambda)}{\int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dG(\lambda)}.$$

Definición 4.0.6 Un proceso estocástico $\{W(t), t \geq 0\}$ es un proceso compuesto de Poisson si se puede representar, para $t \geq 0$, por

$$W(t) = \sum_{i=1}^{N(t)} X_i,$$

donde $N(t)$ es un proceso de Poisson y $X_1, X_2, \dots$ es una familia de v.a.i.i.d. que es independiente del proceso $N(t)$. Por lo tanto, si $\{W(t), t \geq 0\}$ es un proceso compuesto de Poisson, entonces $W(t)$ es una variable aleatoria compuesta de Poisson con parámetro λ y distribución F .

La función generadora de momentos de W se obtiene condicionando sobre $N(t)$, de lo anterior tenemos,

$$\begin{aligned} E[e^{tW}] &= \sum_{n=0}^{\infty} E[e^{tW} | N(t) = n] P(N(t) = n) \\ &= \sum_{n=0}^{\infty} E[e^{t(X_1 + \dots + X_n)} | N(t) = n] \frac{e^{-\lambda t} (\lambda t)^n}{n!} \\ &= \sum_{n=0}^{\infty} E[e^{t(X_1 + \dots + X_n)}] \frac{e^{-\lambda t} (\lambda t)^n}{n!} \\ &= \sum_{n=0}^{\infty} E[e^{tX_1}]^n \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \end{aligned}$$

donde las igualdades anteriores se desprenden de la independencia de los $\{X_1, X_2, \dots\}$ y N , por lo tanto la función generadora de momentos de X_i está dada por,

$$m_X(t) = E[e^{tX_i}],$$

se sigue que

$$\begin{aligned} E[e^{tW}] &= \sum_{n=0}^{\infty} [m_X(t)]^n \frac{e^{-\lambda t} (\lambda t)^n}{n!} \\ &= e^{\lambda t(m_X(t)-1)}. \end{aligned}$$

La diferenciación produce que

$$\begin{aligned} E[W] &= \lambda E[X] \\ \text{Var}(W) &= \lambda E[X^2], \end{aligned}$$

donde X tiene distribución F .

Cuando F es una función de distribución de probabilidad discreta, se puede ver a W como una combinación lineal de v.a.i de Poisson. Supongamos que X_i son variables aleatorias discretas de tal manera que,

$$P(X_i = j) = p_j, \quad j = 1, \dots, k, \quad \sum_{j=1}^k p_j = 1.$$

Si N_i , denota el número de los X_i 's que son iguales a j , $j = 1, \dots, k$, se puede expresar a W como,

$$W = \sum_j j N_j$$

donde los N_j son v.a.i de Poisson con parámetros λp_j , $j = 1, \dots, k$ luego la media y la varianza de W están dadas por,

$$\begin{aligned} E[W] &= \sum_j j E[N_j] = \sum_j j \lambda p_j = \lambda E[X] \\ \text{Var}(W) &= \sum_j j^2 \text{Var}(N_j) = \sum_j j^2 \lambda p_j = \lambda E[X^2]. \end{aligned}$$

Capítulo 5

Algunas Aplicaciones del Proceso de Poisson

El proceso de Poisson es usado para modelar una gran cantidad de fenómenos aleatorios y los puntos de tiempo en los que esos ocurren en un intervalo de tiempo dado. También podemos observar como algunas importantes distribuciones surgen del proceso de Poisson, tales como la distribución de Poisson, la distribución exponencial y la distribución gamma, y como todas estas nos permiten construir otros procesos aleatorios más sofisticados.

Por tanto podemos concluir que la principal ventaja práctica del proceso de Poisson sobre otros modelos de proceso de conteo, es el hecho de que muchas de sus propiedades son explícitamente conocidas, ya que, en general es difícil o imposible obtener explícitamente la distribución de $N(t)$ para cualquier t si $N(t)$ fuera un proceso de conteo que no sea un proceso de Poisson. La propiedad de pérdida de memoria de los tiempos exponenciales entre llegadas, también es extremadamente conveniente al hacer cálculos que involucren el proceso de Poisson.

Algunos de los ejemplos característicos en los que podemos observar las ventajas prácticas de usar el proceso de Poisson para modelar los problemas (los cuales pueden verse con mayor profundidad en [6]), serán listados a continuación.

5.1. Desintegración radioactiva

Para entrar en contexto, la desintegración radioactiva es una propiedad de varios elementos naturales, así como de isótopos de elementos producidos artificialmente. La velocidad a la que se desintegra un elemento radiactivo se expresa en términos de su vida media; es decir, el tiempo requerido para que la mitad de cualquier cantidad dada del isótopo decaiga, dicha vida media puede variar desde una fracción de segundo hasta miles de millones de años.

El Cesio-137 (^{137}Cs) es un isótopo radiactivo de Cesio (Cs) que se forma por la fisión nuclear del Uranio-235 (^{235}U) y otros isótopos fisiónables en reactores nucleares y armas nucleares. Tiene una vida media de 27 años (se necesita este tiempo para que la mitad de los núcleos radiactivos se desintegren). La probabilidad de supervivencia para un núcleo está dada por $p_t = e^{-\lambda t}$ y la tasa de desintegración λ se estima a partir del tiempo de vida media de la siguiente forma:

$$\frac{1}{2} = e^{-27\lambda} \rightarrow \lambda = \frac{\ln 2}{27} \text{yr}^{-1} = 8,2 \times 10^{-10} \text{s}^{-1}$$

Sin embargo, consideremos una pequeña muestra de $1\mu\text{g}$ de $^{137}\text{Cs} \rightarrow N \approx 10^{15}$ núcleos. Entonces, la tasa media de desintegración es $N\lambda \approx 8,2 \times 10^5$ desintegraciones/s.

En consecuencia, cuál es la probabilidad de tener n de N eventos con núcleos en descomposición?

Considere N núcleos radiactivos en un momento $t = 0$, tal que $P(n, 0) = \delta_{n,N}$. Inicialmente, todos los N núcleos sobrevivieron a la desintegración: $P(N, 0) = 1$ y $P(N-1, 0) = P(N-2, 0) = \dots = P(0, 0) = 0$. Sin embargo, a medida que pasa el tiempo, tenemos una probabilidad distinta de cero de que solo una fracción del número total sobreviviera al decaimiento en un momento dado. Esa fracción se hace cada vez más pequeña y, finalmente, después de un tiempo suficientemente largo, esperamos que nadie sobreviva a la descomposición $P(0, t \rightarrow \infty) = 1$.

En un momento t entre estos dos extremos, se espera que haya una probabilidad finita $P(n, t)$ de que n de N núcleos sobrevivan a la desintegración. Queremos observar cómo esta probabilidad depende de la probabilidad en el momento anterior,

$$P(n, t + \Delta t) = P(n, t) + (n+1)q_{\Delta t}p_{\Delta t}^n P(n+1, t) - nq_{\Delta t}^{n-1}P(n, t),$$

Expandiendo Taylor alrededor de t :

$$\frac{\partial P(n, t)}{\partial t} = \frac{(1 - e^{\lambda \Delta t})e^{-n\lambda \Delta t}}{\Delta t}(n+1)P(n+1, t) - \frac{(1 - e^{\lambda \Delta t})e^{-(n-1)\lambda \Delta t}}{\Delta t}nP(n, t),$$

tomando el límite de $\Delta t \rightarrow 0$,

$$\frac{\partial P(n, t)}{\partial t} = \lambda(n+1)P(n+1, t) - \lambda nP(n, t), \quad n = 1, 2, \dots, N,$$

donde $\frac{q_{\Delta t}}{\Delta t} = \frac{1 - e^{\lambda \Delta t}}{\Delta t} \rightarrow \lambda$ es una tasa de descomposición fija.

Usando el método de la función generadora tenemos:

$$G(s, t) = \sum_{n=0}^N s^n P(n, t), \quad s < 1.$$

Aplicando en conjunto con la ecuación anterior, se tiene:

$$\begin{aligned}
\frac{\partial G(s, t)}{\partial t} &= \lambda \sum_{n=0}^N s^n [(n+1)P(n+1, t) - nP(n, t)] \\
\frac{\partial G}{\partial s} &= \sum_{n=0}^N s^{n-1} nP(n, t) = \sum_{n=1}^N s^{n-1} nP(n, t) = \sum_{n=0}^N s^n (n+1)P(n+1, t) \\
\frac{\partial G(s, t)}{\partial t} &= \lambda \left(\frac{\partial G}{\partial s} - s \frac{\partial G}{\partial s} \right), \\
\frac{\partial G(s, t)}{\partial t} &= \lambda(1-s) \frac{\partial G}{\partial s}, \quad G(s, t) = \sum_{n=0}^N s^n P(n, t), \quad s < 1,
\end{aligned}$$

Sustituimos $x = \ln(1-s)$,

$$\frac{\partial G(s, t)}{\partial t} + \frac{\partial G}{\partial x} = 0 \rightarrow G(s, t) = g(x - \lambda t),$$

donde $g(x - \lambda t)$ es una función arbitraria determinada a partir de la condición inicial para $P_n(0) = \delta_{n,N} \rightarrow G(s, 0) = s^N$, por lo tanto $g(x) = s^N = (1 - e^x)^N$ y en general $g(x - \lambda t) = (1 - e^{x - \lambda t})^N$.

Luego,

$$G(s, t) = (1 - e^{x - \lambda t})^N = [1 - (1-s)e^{-\lambda t}]^N = \sum_{n=0}^N s^n \frac{N!}{n!(N-n)!} e^{-n\lambda t} (1 - e^{-\lambda t})^{N-n},$$

obteniendo así, que la probabilidad de tener n núcleos sobrevivientes en el momento t esta dada por:

$$P(n, t) = \frac{N!}{n!(N-n)!} p_t^n (1 - p_t)^{N-n}, \quad p_t = e^{-\lambda t},$$

lo cual indica que tiene distribución binomial y en el límite $N \rightarrow \infty$, y $p_t \rightarrow 0$ tenemos que la probabilidad de tener n núcleos sobrevivientes tienen una distribución de Poisson con $\mu_t = Np_t$.

5.2. Clientes que compran un producto

Supongamos que los clientes llegan al parque de diversiones River View Park de acuerdo con un proceso de Poisson de parámetro λ ; cada cliente debe pagar \$2,500 pesos para ingresar el parque. Si el valor del precio de admisión se descuenta en el tiempo $t = 0$ según una distribución exponencial con tasa β , podemos determinar el valor total esperado del dinero recogido, al cual denotaremos por M . Si $S_1, S_2, \dots$ son los tiempos

de llegada, y $N(t)$ es el número de personas admitidas hasta el tiempo t , el valor esperado del dinero recaudado hasta el tiempo t está dado por $M = E \left[\sum_{k=1}^{N(t)} e^{-\beta S_k} \right]$. Para determinar este valor esperado, utilizamos la ley de la esperanza total:

$$M = \sum_{n=1}^{\infty} E \left[\sum_{k=1}^{N(t)} e^{-\beta S_k} | N(t) = n \right] P(N(t) = n).$$

Recordemos que si $U_1, U_2, \dots$ son v.a.i.i.d en el intervalo $(0, t]$, entonces la distribución condicional dada por $N(t) = n$ de $S_1, S_2, \dots, S_n$ es la misma que el de los estadísticos de orden $U_1, U_2, \dots$, entonces, dado que $\sum_{k=1}^n e^{-\beta U_k} = \sum_{k=1}^n e^{-\beta U_{(k)}}$, tenemos

$$\begin{aligned} \left[\sum_{k=1}^{N(t)} e^{-\beta S_k} | N(t) = n \right] &= E \left[\sum_{k=1}^n e^{-\beta U_{(k)}} \right] \\ &= E \left[\sum_{k=1}^n e^{-\beta U_k} \right] \\ &= n E[e^{-\beta U_1}] \\ &= \frac{n}{t} \int_0^t e^{-\beta u} du \\ &= \frac{n}{\beta t} [1 - e^{-\beta t}]. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} M = \sum_{n=1}^{\infty} \frac{n}{\beta t} [1 - e^{-\beta t}] P(N(t) = n) &= \frac{1 - e^{-\beta t}}{\beta t} \sum_{n=0}^{\infty} n P(N(t) = n) \\ &= \frac{1 - e^{-\beta t}}{\beta t} E[N(t)] \\ &= \frac{\lambda(1 - e^{-\beta t})}{\beta}. \end{aligned}$$

5.3. El problema de la colección de cupones

Consideremos el problema del coleccionista de cupones presentando en [7], de los cuales hay n tipos diferentes, numerados de 1 a n . Supóngase que los cupones se consiguen de uno a uno en una sucesión de v.a.i.i.d., cada una tomando el valor j con probabilidad $p_j \in (0, 1)$; $\sum_{j=1}^n p_j = 1$. Sea N el número mínimo de cupones necesario para conseguir la colección completa. El problema consiste en determinar: (i) la probabilidad de completar la colección cuando se consigue el k -ésimo cupón ($n \leq k$): $P(N = k)$, y, (ii) el número esperado de cupones que se necesitan para conseguir la colección completa: $E[N]$.

Una solución se puede plantear usando la distribución multinomial, como se explica en lo que sigue. Supóngase que hay n posibles resultados de un experimento aleatorio: $C_1, C_2, \dots, C_n$ y que $p_j = P(C_j); j = 1, 2, \dots, n, \sum_{j=1}^n p_j = 1$. Supongase que, de manera independiente, se repite el experimento aleatorio k veces: $k = 1, 2, \dots$. Sea ahora X_j el número de veces que ocurre el resultado C_j en las k repeticiones; $j = 1, 2, \dots, n; \sum_{j=1}^n x_j = k$. Entonces la v.a $\mathbf{X} = (X_1, X_2, \dots, X_n)$ tiene distribución multinomial de parámetros $k, p_1, p_2, \dots, p_n$; es decir:

$$p_{\mathbf{X}}(\mathbf{X}) = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] = \binom{k}{x_1 x_2 \dots x_n} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n},$$

donde $x_j \in 0, 1, \dots, n; j = 1, 2, \dots, n$ y $\sum_{j=1}^n x_j = k$.

En lo que sigue, S_j es el conjunto de los $\binom{n}{j}$ subconjuntos de tamaño j de $1, 2, \dots, n$. Para cada subconjunto C de $1, 2, \dots, n$, sea

$$p_c = \sum_{j \in C} P_j \text{ y } q_c := 1 - p_c.$$

Teorema 5.3.1 *Con las notaciones precedentes:*

1. $P(N > k) = \sum_{j=1}^n (-1)^{j+1} \sum_{c \in s_j} q_c^k; k = 0, 1, \dots$
2. $P(N = k) = \sum_{j=1}^n (-1)^{j+1} \sum_{c \in s_j} q_c^{k-1} p_c; k = 0, 1, \dots$
3. $m_N(t) := E[e^{tN}] = \sum_{j=1}^n (-1)^{j+1} \sum_{c \in s_j} \frac{p_c e^t}{1 - q_c e^t}, t < h, \text{ donde}$
 $h = \min\{-\ln(q_c), q_c \neq 0\}.$
4. $E(N) = \sum_{j=1}^n (-1)^{j+1} \sum_{c \in s_j} \frac{1}{p_c}.$

Demostración

1. Si el número de cupones necesarios para completar la colección es mayor que k ($k \leq 1$), es porque entre los primeros k cupones coleccionados falta alguno de la colección. En el contexto de la distribución multinomial, esto se escribe así:

$$P(N > k) = P\left(\bigcup_{j=1}^n \{X_j = 0\}\right).$$

Ahora, usando la fórmula de inclusión-exclusión:

$$\begin{aligned}
P\left(\bigcup_{j=1}^n \{X_j = 0\}\right) &= \sum_{\{i\} \in S_1} P(\{X_i = 0\}) - \sum_{\{i,j\} \in S_2} P(\{X_i = 0\} \cap \{X_j = 0\}) \\
&\quad + \cdots + (-1)^{n+1} P\left(\bigcap_{j=1}^n \{X_j = 0\}\right) \\
&= \sum_{\{i\} \in S_1} P(\{X_i = 0\}) - \sum_{\{i,j\} \in S_2} P(\{X_i = 0\} \cap \{X_j = 0\}) \\
&\quad + \cdots + (-1)^{n+1} P\left(\bigcap_{j=1}^n \{X_j = 0\}\right).
\end{aligned}$$

Ahora, para $c = \{i_1, i_2, \dots, i_j\} \in S_j$:

$$P(X_{i_1} = 0, X_{i_2} = 0, \dots, X_{i_j} = 0) = \sum \binom{k}{x_1 x_2 \dots x_n} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n},$$

donde la suma se hace sobre los $x_i \in \{0, 1, \dots, n\}$ tal que $x_{i_1} = x_{i_2} = \cdots = x_{i_j} = 0$ y $\sum_{j=1}^n x_j = k$. El resultado de esta suma es, por el teorema multinomial q_c^k . Es decir:

$$P(X_{i_1} = 0, X_{i_2} = 0, \dots, X_{i_j} = 0) = q_c^k$$

Esto termina la demostración de 1, cuando $k = 1, 2, \dots$.

En particular, del contexto del problema:

$$1 = P(N > 1) = P(N > 2) = \cdots = P(N > n - 1).$$

También $1 = P(N > 0)$, y de la identidad

$$1 - \prod_{j=1}^n (1 - x_j) = \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} x_c ; \quad x_1, x_2, \dots, x_n \in \mathbb{R},$$

donde $x_c = x_{i_1}, x_{i_2}, \dots, x_{i_j}$ cuando $c = \{i_1, i_2, \dots, i_j\} \subseteq \{1, 2, \dots, n\}$, se tiene

$$1 = \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} q_c^0.$$

2. Para demostrar este numeral, basta escribir

$$\begin{aligned}
P(N = k) &= P(N > k - 1) - P(N > k) ; \quad k = 1, 2, \dots \\
&= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} q_c^{k-1} - q_c^k \\
&= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} q_c^{k-1} p_c.
\end{aligned}$$

3.

$$\begin{aligned}
m_N(t) &= E[e^{tN}] = \sum_{k=1}^{\infty} e^{tk} P(N = k) \\
&= \sum_{k=1}^{\infty} e^{tk} \left(\sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} q_c^{k-1} p_c \right) \\
&= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} \sum_{k=1}^{\infty} q_c^{k-1} p_c e^{tk} \\
&= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} \frac{p_c e^t}{1 - q_c e^t} ; \quad t < h.
\end{aligned}$$

4.

$$\begin{aligned}
m_{N'}(t) &= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} \left[\frac{p_c e^t}{(1 - q_c e^t)} + \frac{p_c q_c e^{2t}}{(1 - q_c e^t)^2} \right] \\
m_{N'}(0) &= E[N] = \sum_{j=1}^n (-1)^{j+1} \frac{1}{p_c}.
\end{aligned}$$

Nota. Otra vez usaremos la identidad

$$1 - \prod_{j=1}^n (1 - x_j) = \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} x_c ; \quad x_1, x_2, \dots, x_n \in \mathbb{R},$$

donde $x_c = x_{i_1}, x_{i_2}, \dots, x_{i_j}$ cuando $c = \{i_1, i_2, \dots, i_j\} \subseteq \{1, 2, \dots, n\}$ y usando el hecho de que

$$\int_0^{\infty} e^{-p_c x} dx = \frac{1}{p_c},$$

podemos escribir

$$\begin{aligned}
E[N] &= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} \frac{1}{p_C} \\
&= \sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} \left(\int_0^\infty e^{-p_C x} dx \right) \\
&= \int_0^\infty \left(\sum_{j=1}^n (-1)^{j+1} \sum_{C \in S_j} e^{-p_C x} \right) dx \\
&= \int_0^\infty \left[1 - \prod_{j=1}^n (1 - e^{-p_j x}) \right] dx,
\end{aligned}$$

esto es,

$$E[N] = \int_0^\infty \left[1 - \prod_{j=1}^n (1 - e^{-p_j x}) \right] dx.$$

Esta última igualdad se relaciona con el proceso de Poisson como se explica en lo que sigue.

Supongamos que los cupones llegan de acuerdo con un proceso de Poisson de parámetro 1. Cabe aclarar que el evento del proceso de Poisson es de tipo j si el cupón es de tipo j . Ahora definimos una familia de procesos independientes de Poisson $\{N_j(t) : t \geq 0\}_{1 \leq j \leq m}$, donde el parámetro del j -ésimo proceso es p_j . Sea S_j el momento de la primera llegada del j -ésimo cupón, es decir, S_j es el valor más pequeño tal que $N_j(S_j) = 1$. El tiempo de interés es $S = \max\{S_j\}$ porque ese es el momento en que todos los tipos de cupones m habrán sido recogidos. Por el Teorema 4.0.3, S_j tiene función de densidad de probabilidad $f_{S_j}(t) = p_j e^{-p_j t}$, y cada S_j es independiente. Calculemos la función de distribución de S :

$$\begin{aligned}
P(S \leq t) &= P(\max\{S_j\} \leq t) = P(S_j \leq t, \forall t \leq m) \\
&= \prod_{j=1}^m P(S_j \leq t) \\
&= \prod_{j=1}^m (1 - P(S_j > t)) \\
&= \prod_{j=1}^m (1 - e^{-p_j t}).
\end{aligned}$$

Para una variable aleatoria no negativa Y se tiene, $E[Y] = \int_0^\infty (1 - F_Y(y)) dy$, entonces, $E[S] = \int_0^\infty (1 - F_S(t)) dt = \int_0^\infty (1 - \prod_{j=1}^m (1 - e^{-p_j t})) dt$. Hemos calculado el tiempo esperado hasta que se haya recolectado todos los tipos de cupones, mas

no el número esperado de cupones que necesitaríamos recolectar, para esto haremos lo siguiente. Sea I_i el i -ésimo tiempo entre llegada de cupones. Dado que, $\sum_{i=1}^M I_i = S$, entonces $E[\sum_{i=1}^M I_i] = E[S]$, pero como los I_i son independientes y exponenciales con parámetro $\lambda = 1$ y también son independientes de M , tenemos que $E[\sum_{i=1}^M I_i] = E[M]E[I_i] = E[M] \cdot 1 = E[M]$. Por lo tanto, el valor que calculamos para $E[S]$ también es el valor de $E[M]$.

Bibliografía

- [1] Adelson, R.M. (1966). *Compound Poisson Distributions*. Journal of the Operational Research Society. doi.org/10.1057/jors.1966.8
- [2] Billingsley, P. (1979). *Probability and Measure*. John Wiley and Sons, Inc., 1979.
- [3] Durret, R. (2010). *Probability: Theory and Examples*. Cambridge University Press.
- [4] Meyer, P.L. (1986). *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano.
- [5] Mood, A. M.; Graybill, F. A. and Boes, D. C.(1974). *Introduction to the Theory of Statistics*. Third Edition. McGraw-Hill, Inc.
- [6] Pinsky, M. A. and Karlin S. (1998). *An Introduction to Stochastic Modeling*. Third Edition. Amsterdam: Elsevier/Academic.
- [7] Ross, S. M. (2010). *Introduction to probability models*. Ten Edition. Academic Press.
- [8] Ross, S. M. (1996). *Stochastic Processes*. Second Edition. John Wiley & Sons, Inc.
- [9] Shiryaev, A. N. (1996). *Probability*. Second Edition. Springer.
- [10] Then, X. *IEOR 6711: Stochastic Models I*.