

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



FACULTAD DE CIENCIAS

INFERENCIA PARAMÉTRICA PARA EL PROCESO POISSON

2001

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

SIMÓN LUNAGÓMEZ CORIA

DIRECTORA DE TESIS:  
DRA. MARÍA ASUNCIÓN BEGOÑA FERNÁNDEZ FERNÁNDEZ

MÉXICO





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ESTA TESIS NO SALE  
DE LA BIBLIOTECA



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**MAT. MARGARITA ELVIRA CHÁVEZ CANO**  
Jefa de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis **INFERENCIA PARAMETRICA  
PARA EL PROCESO POISSON.**

realizado por **LUNAGOMEZ CORIA SIMON**

con número de cuenta **9561219-3**, pasante de la carrera de **ACTUARIA**

Dicho trabajo cuenta con nuestro voto aprobatorio

Atentamente

Director de Tesis **DRA. MA. ASUNCION BEGOÑA FERNANDEZ FERNANDEZ**  
Propietario

*Asunción Begoña*

Propietario **MAT. MARGARITA ELVIRA CHAVEZ CANO**

*Margarita Chávez*

Propietario **DR. JOSE ALFREDO LOPEZ MIMBEIA**

*Jose Alfredo Lopez*

Suplente **DRA. EKATERINA TODOROVA KOLKOVSKA**

*Ekaterina Todorova*

Suplente **M. EN I. DEL CARMEN HERANDNEZ AYUSO**

*M. del Carmen Herandnez*

Consejo Departamental de  
Matemáticas



*P.A. Lucha*

M. en C. **JOSE ANTONIO FLORES DIAZ**

FACULTAD DE CIENCIAS  
SECCION DE MATEMATICAS

DE  
MATEMATICAS

# Inferencia Paramétrica para el Proceso Poisson

Simón Lunagómez Coria

# Indice General

<b>1</b>	<b>Introducción</b>	<b>vii</b>
<b>2</b>	<b>Preliminares</b>	<b>1</b>
2.1	El Modelo Poisson . . . . .	2
2.2	Tiempos de espera . . . . .	6
2.2.1	La Distribución Geométrica . . . . .	6
2.2.2	La Distribución Binomial Negativa . . . . .	10
2.2.3	El Modelo Exponencial . . . . .	13
2.2.4	Los Modelos Gamma . . . . .	16
2.3	La Distribución Uniforme . . . . .	19
2.4	La distribución Normal . . . . .	20
2.5	Estadísticas de Orden . . . . .	22
2.6	Simulación . . . . .	22
<b>3</b>	<b>El Proceso Poisson</b>	<b>25</b>
3.1	El Proceso Poisson en la recta real . . . . .	26
3.2	Propiedades del Proceso Poisson . . . . .	32
3.2.1	Distribución de los tiempos entre llegadas . . . . .	32
3.2.2	Distribución condicional de los tiempos de llegada . . . . .	35
3.2.3	Ley de los Grandes Números . . . . .	37
3.3	Generalizaciones del Proceso Poisson . . . . .	38
3.3.1	El Proceso Poisson no Estacionario . . . . .	38
3.3.2	El Proceso Poisson Compuesto . . . . .	43
3.3.3	El Proceso Poisson en el Plano . . . . .	49
3.3.4	Procesos Puntuales . . . . .	51
3.4	Caracterizaciones fuertes del Proceso Poisson . . . . .	54

3.4.1	El funcional característico . . . . .	54
3.4.2	El Teorema de Renyi . . . . .	60
3.5	Representación del Proceso Poisson Compuesto . . . . .	63
3.6	Simulación del Proceso Poisson . . . . .	68
<b>4</b>	<b>El problema de estimación para el Proceso Poisson</b> . . . . .	<b>79</b>
4.1	Observación durante un intervalo de tiempo fijo . . . . .	79
4.1.1	El caso estacionario . . . . .	79
4.1.2	El proceso no estacionario . . . . .	81
4.1.3	Forma alternativa de obtener un estimador . . . . .	82
4.2	La cota inferior de Cramér y Rao para la estimación en tiempo fijo . . . . .	84
4.3	Observaciones continuas durante un tiempo aleatorio. . . . .	88
4.4	Intervalos de confianza . . . . .	89
4.5	Inspección continua dado un plan de muestreo . . . . .	91
4.5.1	El caso determinístico . . . . .	91
4.5.2	El caso aleatorio . . . . .	92
4.5.3	Tiempos de observación aleatorios con distribución Exponencial . . . . .	93
4.6	Realización discretizada . . . . .	112
4.7	Distribución asintótica de estimadores . . . . .	114
4.7.1	Tiempos de observación con distribución Exponencial . . . . .	115
4.7.2	Tiempo de observación aleatorio . . . . .	118
4.8	El problema de estimación para el Proceso Poisson Compuesto . . . . .	122
4.9	Algunas pruebas de hipótesis para el proceso Poisson . . . . .	126
4.9.1	Cocientes de verosimilitud . . . . .	126
4.9.2	Prueba para la estacionareidad del proceso . . . . .	128
<b>5</b>	<b>El punto de vista bayesiano</b> . . . . .	<b>129</b>
5.1	Densidades a priori conjugadas . . . . .	130
5.2	El caso estacionario . . . . .	131
5.2.1	Observación durante tiempo fijo . . . . .	131
5.2.2	Observación durante tiempo aleatorio . . . . .	134
5.3	El caso no estacionario . . . . .	136
5.3.1	Tiempos entre ocurrencias . . . . .	136
5.3.2	Ocurrencias en intervalos . . . . .	138
5.4	Tiempos de observación aleatorios . . . . .	139
5.5	Estimación por intervalo . . . . .	141
5.6	Pruebas de hipótesis . . . . .	142
<b>6</b>	<b>Comentarios finales</b> . . . . .	<b>149</b>
<b>A</b>	<b>Probabilidad</b> . . . . .	<b>151</b>
A.1	Espacios y medidas de probabilidad . . . . .	151
A.2	Lemas de clases monótonas . . . . .	152
A.3	Teorema de la Probabilidad Total . . . . .	152

A.4	Teoría de integración . . . . .	153
A.5	Función de Distribución . . . . .	154
A.6	Independencia . . . . .	155
A.7	Funciones de densidad y de masa . . . . .	155
A.8	Media y Varianza . . . . .	156
A.9	Probabilidad conjunta y marginal . . . . .	156
A.10	Convoluciones . . . . .	157
A.11	Esperanza y probabilidad condicional . . . . .	158
A.12	La Regla de Bayes . . . . .	159
A.13	Teoría Límite . . . . .	160
A.13.1	Tipos de Convergencia . . . . .	160
A.13.2	Leyes Asintóticas . . . . .	161
A.14	Función Generatriz de Momentos . . . . .	161
A.15	Función Característica . . . . .	162
<b>B</b>	<b>Inferencia Paramétrica</b> . . . . .	<b>165</b>
B.1	La Familia Exponencial . . . . .	165
B.2	El modelo estadístico paramétrico . . . . .	166
B.3	Verosimilitud . . . . .	166
B.4	Suficiencia . . . . .	169
B.5	Pruebas de hipótesis . . . . .	169
B.5.1	El caso general . . . . .	169
B.5.2	Teoría asintótica . . . . .	171
B.6	Teoría de estimación insesgada . . . . .	172
B.7	Resultado de normalidad asintótica para estimadores . . . . .	173
B.8	Intervalos de confianza . . . . .	174
<b>C</b>	<b>Sobre la Teoría Bayesiana</b> . . . . .	<b>175</b>
C.1	Generalidades . . . . .	175
C.2	Estimación . . . . .	176
C.3	Análisis conjugado . . . . .	177
C.4	Distribuciones a priori no informativas . . . . .	177
C.5	Criterio de información . . . . .	178
C.6	Estimación por intervalo . . . . .	179
<b>D</b>	<b>Guía de instrucciones de S-PLUS</b> . . . . .	<b>181</b>
	References . . . . .	187



# Introducción

El Proceso Poisson es un modelo matemático para fenómenos caracterizados por objetos dispuestos de manera aleatoria en un continuo. Cada objeto es representado de forma idealizada en el modelo como un punto, cuya ubicación en cierto espacio (por ejemplo  $\mathbb{R}^k$ ) se identifica con la posición del objeto. Es común referirse a dichos puntos como 'ocurrencias' del proceso.

*El Proceso Poisson posee muchas propiedades que justifican su papel central en la teoría de procesos estocásticos; dichas propiedades se basan en ciertos supuestos de independencia. Esto se discute con detalle en el Capítulo 2.*

Esta tesis está enfocada en gran medida a aquellos procesos tales que el espacio donde los puntos ocurren es el conjunto de los números reales no negativos. Aunque no haya razón de tipo matemático para ello, se referirá a dicho espacio como 'tiempo', ya que los fenómenos temporales predominan en muchos campos de aplicación. El estudio del Proceso Poisson en la recta real permite analizar muchas propiedades atribuidas a versiones más generales de este proceso. Sin embargo, procesos tales como los Procesos Poisson no estacionarios (Sección 3.3.1), y los Procesos Poisson Compuestos (Sección 3.3.2) son de interés tanto en el aspecto teórico como en el de modelación.

La implementación del Proceso Poisson (y de sus generalizaciones) como modelo matemático es un proceso que consta de varias etapas. Esto inicia cuando, basándose en ideas propias del campo de aplicación, el investigador propone al Proceso Poisson como modelo para describir al objeto de estudio. En este contexto, y bajo el supuesto de que se tiene un modelo razonable, el investigador será capaz de determinar que predicciones puede llevar a cabo sobre los eventos que él considere puedan ocurrir. Esta formulación es el único paso tal que esta basado totalmente en la Teoría de Probabilidad. Sin embargo, con el fin de desarrollar el modelo, el investigador debe ser capaz de incorporar información proveniente del mundo exterior. De esta manera el modelo podrá describir lo que ya ha sido observado como lo que pueda aprenderse de dichas observaciones. Es en este punto donde surge la necesidad de emplear técnicas de inferencia estadística.

Toda la teoría de inferencia estadística empleada en esta tesis se desarrolla dentro del siguiente escenario : el investigador ha considerado que cierto fenómeno puede modelarse mediante un proceso

Poisson o alguna de sus generalizaciones, de tal manera que el conjunto de modelos posibles se reduzca a una familia que puede ser parametrizada. Se trata el problema de estimación (puntual por intervalo) y el de construir pruebas de hipótesis paramétricas para estos procesos. Los problemas de inferencia son tipificados de acuerdo a dos criterios: tipo de proceso que se está observando y plan de muestreo. Se entiende por plan de muestreo a la elección de ciertas variables aleatorias relacionadas con el proceso en cuestión sobre las cuales se basará la inferencia. No se discutirá la validez del modelo (pruebas de bondad y ajuste).

La tesis está estructurada en cuatro capítulos:

El primer capítulo es una revisión de las propiedades básicas de las distribuciones de probabilidad involucradas en la definición y resultados referentes al *Proceso Poisson y sus generalizaciones*. También se discuten las distribuciones relacionadas con las técnicas de inferencia paramétrica aplicadas a dichos procesos.

En el segundo capítulo se define al *Proceso Poisson Estacionario* y se discuten sus propiedades principales, entre las que figura la distribución de los tiempos entre ocurrencias. Se presentan algunas de las generalizaciones de dicho proceso, esto con dos propósitos: primero, poder discutir modelos con aplicaciones más realistas; y segundo, acceder a ciertos resultados de tipo analítico. Los *procesos no estacionarios* y el *Proceso Poisson Compuesto* son expuestos de acuerdo a la primera motivación mientras que el *Proceso Poisson Puntual* a la segunda. Hay una discusión dedicada a características del *Proceso Poisson Puntual* y a resultados relativos a la representación del *Proceso Poisson Compuesto* como 'suma' de *Procesos Poisson Puntuales*.

El tercer capítulo está dedicado a problemas de inferencia del *Proceso Poisson* desde el punto de vista frecuentista. Se calculan estimadores de máxima verosimilitud para este tipo de procesos bajo distintos planes de muestreo. La parte central de este capítulo es una discusión sobre el comportamiento del estimador máximo verosímil de la intensidad del proceso, tomando un número fijo de tiempos aleatorios de distribución exponencial. La relación entre el parámetro de la distribución de dichos tiempos y la intensidad del proceso es crucial para el desempeño del estimador. El autor propone algoritmos que, bajo ciertas condiciones permiten aminorar este inconveniente. Se tratan algunos aspectos de pruebas de hipótesis paramétricas para estos procesos.

En el capítulo cuatro se retoman los problemas de inferencia paramétrica tratados en el capítulo anterior, ahora desde el punto de vista bayesiano. En este enfoque se supone la existencia de información *a priori* o inicial por parte del investigador. Con el propósito de representar dicha información, recurre al *Análisis Conjugado* y al *Análisis de Referencia*. En la sección dedicada a pruebas de hipótesis, se ilustra como estas pueden ser utilizadas en problemas que involucran toma de decisiones.

La simulación por computadora es un recurso muy utilizado para visualizar (en cierto sentido) el comportamiento de los modelos probabilísticos. Se trata además de una herramienta valiosa para la resolución de problemas estadísticos, ya que hace posible valorar el desempeño de los estimadores a partir de una perspectiva complementaria a la que proporcionan los resultados analíticos. En la tesis se exploran ambas posibilidades. La implementación de los algoritmos de simulación se lleva a cabo usando el paquete *S-PLUS*.

# 2

## Preliminares

### Introducción

El Proceso Poisson debe su nombre a la distribución de probabilidad propuesta por S. Poisson en 1837. Dicha distribución está presente tanto en la versión más sencilla del Proceso Poisson, como en las generalizaciones definidas en espacios abstractos [ver Kingman]. Por otra parte, varios de los Teoremas importantes para estos procesos involucran otras distribuciones univariadas conocidas. Debido a estas y otras razones, el estudio del Proceso Poisson presupone conocimientos en Teoría Distribucional.

En este primer capítulo se revisan algunos aspectos básicos de las distribuciones de probabilidad relacionadas con el Proceso Poisson y con técnicas de inferencia paramétrica. Se exponen los resultados analíticos requeridos en los siguientes capítulos y las ideas intuitivas detrás de cada distribución.

La Sección 2.1 está dedicada a la Distribución Poisson. La Sección 2.2 se refiere a las distribuciones utilizadas para describir situaciones que involucran tiempos de espera; se trata de las Distribuciones Exponencial, Gamma, Geométrica y Binomial Negativa. Se presenta una breve discusión acerca de la Distribución Uniforme (Sección 2.3), y otra sobre la Normal (Sección 2.4). En la Sección 2.5 se enuncian algunos resultados sobre estadísticas de orden. Se incluye una sección acerca de simulación de variables aleatorias (Sección 2.6).

Mediante el uso de ejemplos, se busca ilustrar como algunas de estas distribuciones pueden usarse para resolver problemas concretos, lo cual es un paso importante para entender el papel del Proceso Poisson como herramienta de modelación.

El material de este capítulo fué tomado de [Feller], [Johnson y Kotz], y [Feldman].

## 2.1 El Modelo Poisson

Sea  $X_1, X_2, \dots, X_n$  un conjunto de ensayos Bernoulli, todos con probabilidad de éxito  $p_n$ . El número total de éxitos tiene distribución  $\text{Bi}(n, p_n)$ . Supóngase que el experimento es tal que

$$\lim_{n \rightarrow \infty} np_n = \lambda,$$

con  $\lambda > 0$ .

Es decir, el número esperado de éxitos tiende a cierta constante positiva cuando  $n$  crece. Sea  $Y_n$  una variable aleatoria con distribución  $\text{Bi}(n, p_n)$ , entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr\{Y_n = k\} &= \lim_{n \rightarrow \infty} \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k! (n-k)!} (p_n)^k \left[1 - \frac{np_n}{n}\right]^n [1 - p_n]^{-k}, \end{aligned}$$

mientras  $n \rightarrow \infty$ ,  $\frac{n!}{k! (n-k)!} \rightarrow 1$  y  $[1 - p_n]^{-k} \rightarrow 1$ . Debido a que  $f_n(x) = \left(1 - \frac{x}{n}\right)^n \rightarrow e^{-x}$  uniformemente para cualquier vecindad compacta de  $\lambda$ , se puede afirmar que  $\left[1 - \frac{np_n}{n}\right]^n \rightarrow \exp(-\lim_{n \rightarrow \infty} np_n) = \exp(-\lambda)$ .

Por lo tanto, para  $p$  pequeña y  $n$  grande, se tiene que

$$\Pr(\# \text{ de éxitos} = k) \approx e^{-\lambda} \frac{\lambda^k}{k!},$$

donde  $\lambda = np$ .

A partir del desarrollo en serie de Taylor de  $h(\lambda) = e^\lambda$  se sigue  $\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1$ .

**Definición 2.1.1** Sea  $X$  una variable aleatoria discreta que toma valores en el conjunto de los números naturales. Se dice que  $X$  tiene Distribución Poisson con parámetro  $\lambda$ . Lo cual se denota con  $X \sim \text{Poi}(\lambda)$ , si su función de masa viene dada por

$$\Pr\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Es fácil verificar que: Si  $X$  tiene distribución  $\text{Poi}(\lambda)$  entonces

$$E[X] = \lambda \quad \text{y} \quad \text{Var}[X] = \lambda.$$

En las Figuras 2.1, 2.2 y 2.3 se presentan gráficas de la función de masa de la Distribución Poisson con  $\lambda = 1, 6$  y  $.5$  respectivamente.

La función generatriz de momentos de una variable aleatoria  $X$  con distribución  $\text{Poi}(\lambda)$  viene dada por

$$E[e^{tX}] = e^{\lambda(e^t - 1)}.$$

El uso de dicha función permite demostrar fácilmente que la Distribución Poisson es una distribución reproductiva. Es decir, si  $X$  y  $Y$  son variables aleatorias independientes, con distribución  $\text{Poi}(\lambda)$  y  $\text{Poi}(\mu)$  respectivamente, entonces  $X + Y$  tiene distribución  $\text{Poi}(\lambda + \mu)$ . Esto se explica en el siguiente resultado, el cual trata también el caso de sumas infinitas, y da las condiciones para la convergencia de dicha suma.

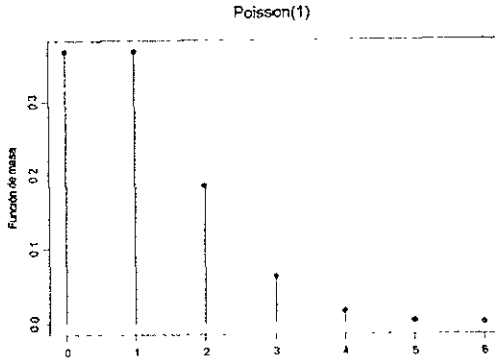


FIGURE 2.1

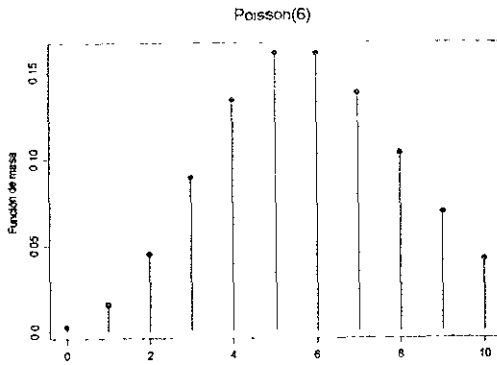


FIGURE 2.2.

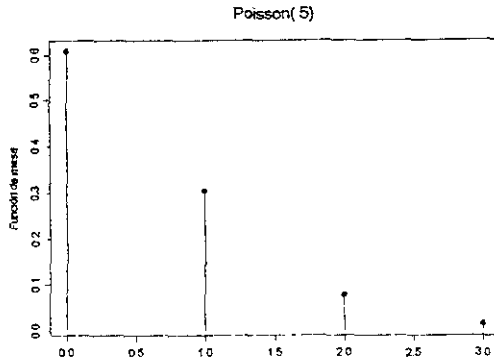


FIGURE 2.3.

**Teorema 2.1.2** Sea  $\{X_j : j \in \mathbb{N}\}$  una colección de variables aleatorias independientes, donde  $X_j$  tiene distribución  $\text{Poi}(\mu_j)$  para cada  $j$ . Entonces  $S_n = \sum_{j=1}^n X_j$  tiene distribución  $\text{Poi}(\sum_{j=1}^n \mu_j)$ . Si  $\sigma = \sum_{j=1}^{\infty} \mu_j$  converge, entonces  $S = \sum_{j=1}^{\infty} X_j$  converge con probabilidad 1, y  $S$  tiene distribución  $\text{Poi}(\sigma)$ . Si por otra parte  $\sigma$  diverge, entonces  $S$  diverge con probabilidad 1.

**Demostración.** Sea  $n \in \mathbb{N}$ , entonces la variable aleatoria

$$S_n = \sum_{j=1}^n X_j$$

tiene distribución  $\text{Poi}(\sigma_n)$ , donde

$$\sigma_n = \sum_{j=1}^n \mu_j.$$

Esto puede verse del hecho que la función generatriz de la suma de variables aleatorias independientes es el producto de las funciones generatrices de los sumandos [ver Ash]

$$\begin{aligned} \mathbb{E}[e^{tS_n}] &= \prod_{j=1}^n \mathbb{E}[e^{tX_j}] \\ &= \prod_{j=1}^n e^{\mu_j [e^t - 1]} \\ &= e^{\sigma_n [e^t - 1]}. \end{aligned}$$

Se tiene además que la convergencia de funciones generatrices implica convergencia en distribución. Ahora

$$\begin{aligned} E [e^{tS}] &= \prod_{j=1}^{\infty} E [e^{tX_j}] \\ &= e^{\sigma [e^t - 1]} \end{aligned}$$

corresponde a la función generadora de momentos de una variable aleatoria  $\text{Poi}(\sigma)$ .

Esta última expresión representa a una función generadora sólo en el caso  $\sigma < \infty$ . ■

En seguida se presentan algunas aplicaciones concretas para este modelo probabilista.

**Ejemplo 2.1.3** Supóngase que un libro tiene 400 páginas y que la probabilidad de que una página se encuentre libre de errores de impresión es .98. Es razonable suponer que el número de ocurrencias en páginas diferentes son independientes, de tal forma que el número de páginas libres de errores en el libro, el cual se representa con  $S_{400}$  es una variable aleatoria que tiene distribución  $\text{Bi}(400, .98)$ . Sea  $U_{400} = 400 - S_{400}$ , el número de páginas que necesita correcciones. Entonces, si  $X$  tiene una distribución  $\text{Bi}(400, .02)$  y  $Y$  se distribuye  $\text{Poi}(8)$ , se tiene que

$$\begin{aligned} \Pr \{U_{400} > 10\} &= 1 - \Pr \{X \leq 10\} \\ &\approx 1 - \Pr \{Y \leq 10\} \\ &= .184. \end{aligned}$$

El valor exacto es .182. ■

El siguiente resultado será de utilidad al hablar del Proceso Poisson Compuesto.

**Teorema 2.1.4** Sea  $N$  una variable aleatoria con distribución  $\text{Poi}(\lambda)$  y sea  $M$  una variable aleatoria tal que la distribución condicional de  $M$  dado  $N$  es  $\text{Bi}(N, p)$ , entonces  $M$  se distribuye  $\text{Poi}(\lambda p)$ .

**Demostración.** Aplicando el Teorema de Probabilidad Total

$$\begin{aligned} \Pr \{M = k\} &= \Pr \{M = k, N \geq k\} \\ &= \sum_{n=k}^{\infty} \Pr \{M = k \mid N = n\} \Pr \{N = n\} \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}. \end{aligned}$$

Simplificando esta última expresión, se obtiene

$$\begin{aligned} \Pr \{M = k\} &= e^{-\lambda} \frac{p^k}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} \lambda^n \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{j=0}^{\infty} \frac{[(1-p)\lambda]^j}{j!} \\ &= e^{-\lambda p} \frac{(\lambda p)^k}{k!}, \end{aligned}$$

donde  $k \in \{0, 1, 2, \dots\}$ . ■

La Distribución Poisson pertenece a la Familia Exponencial (ver Apéndice). El parámetro natural es  $\theta = \ln(\lambda)$ , por lo tanto el espacio paramétrico natural es  $\mathbf{R}$ .

## 2.2 Tiempos de espera

Un investigador inicia la observación del fenómeno que está estudiando en un instante al cual se denota con 0. Existe un suceso en especial que él desea observar. Es de interés la longitud del lapso entre 0 y el instante en que dicho suceso ocurre, a esto se le conoce como *tiempo de espera*.

Al hablar de 'tiempo' se está considerando cualquiera de las siguientes posibilidades:

**Tiempo discreto.**-El sistema solamente puede sufrir cambios en una cantidad numerable de instantes predeterminados.

**Tiempo continuo.**-Cuando se establece una biyección entre los 'instantes' y los puntos de  $\mathbf{R}^+$ .

Esto es totalmente intuitivo, pero no es necesario introducir terminología más complicada. En las siguientes secciones se presentan algunas distribuciones que sirven como modelos para tiempos de espera.

### 2.2.1 La Distribución Geométrica

**Definición 2.2.1** Sea  $X_1, X_2, \dots$  una sucesión de ensayos Bernoulli con probabilidad de éxito  $p > 0$ . Entonces

$$T_1 = \inf \{n > 0 \mid X_n = 1\}$$

es una variable aleatoria, ya que  $\{T_1 > k\} = \{X_1 = 0, X_2 = 0, \dots, X_k = 0\}$  es un evento. Se dice entonces que  $T_1$  tiene Distribución Geométrica con parámetro  $p$ , lo cual se denota con  $T_1 \sim \text{Geo}(p)$ .

Sea  $\{X_k : k \in \mathbf{N}\}$  una sucesión de ensayos Bernoulli, con probabilidad de éxito  $p$  y sean

$$\left\{ S_k = \sum_{j=1}^k X_j : k \in \mathbf{N} \right\}.$$

Entonces la función de distribución de  $T_1$  viene dada por



$$\begin{aligned} F_{T_1}(n) &= \Pr(T_1 \leq n) = \Pr(S_n \geq 1) \\ &= 1 - \Pr(S_n = 0) = 1 - [1-p]^n \quad \text{para } n = 1, 2, \dots \end{aligned}$$

y su función de masa es

$$\begin{aligned} f_{T_1}(n) &= \Pr(T_1 = n) = F_{T_1}(n) - F_{T_1}(n-1) \\ &= [1-p]^{n-1} - [1-p]^n = p[1-p]^{n-1} \quad \text{para } n = 1, 2, \dots \end{aligned}$$

De forma alternativa, la función de masa de  $T_1$  puede encontrarse observando que el evento  $\{T_1 = n\}$  ocurre si y sólo si los primeros  $n-1$  ensayos son fracasos y el  $n$ -ésimo ensayo es un éxito. Esto gracias a la independencia de los ensayos

Todos los momentos de  $T_1$  existen, como lo demuestra el siguiente argumento. Sea  $\alpha > 0$  fija y sea  $\lambda_\alpha = \sum_{n=1}^{\infty} n^\alpha (1-p)^{n-1}$ . Se considera la razón  $r_n$  de términos sucesivos:

$$r_n = \frac{[n+1]^\alpha (1-p)^n}{n^\alpha (1-p)^{n-1}} = \left(1 + \frac{1}{n}\right)^\alpha (1-p).$$

Debe ser claro que  $r_n \rightarrow (1-p) < 1$ , por lo tanto por el Criterio de Convergencia del cociente de D'Alembert [ver Widder], se tiene que  $\lambda_\alpha < \infty$ .

Sea  $Y$  una variable aleatoria con distribución  $\text{Geo}(p)$ . Entonces la función generatriz de momentos de  $Y$  es:

$$\begin{aligned} E(e^{tY}) &= \sum_{y=1}^{\infty} e^{ty} p [1-p]^{y-1} \\ &= \frac{p}{1-p} \sum_{y=1}^{\infty} (e^t [1-p])^y \\ &= \frac{p}{1-p} \left[ \frac{1}{1 - e^t [1-p]} - 1 \right] \\ &= \frac{e^t p}{1 - e^t [1-p]}. \end{aligned}$$

De donde se concluye que  $m_Y(t) = \frac{e^t p}{1 - e^t [1-p]}$ . La media de  $Y$  viene dada por

$$\begin{aligned} \frac{\partial}{\partial t} m_Y(\omega) &= \frac{e^t p}{1 - e^t [1-p]} + \frac{(e^t)^2 p (1-p)}{[1 - e^t [1-p]]^2} \\ \frac{\partial}{\partial t} m_Y(\omega) \Big|_{t=0} &= \frac{1}{p}. \end{aligned}$$

Esto se escribe  $E[Y] = \frac{1}{p}$ . Se puede demostrar que  $E[Y^2] = \frac{2-p}{p^2}$ , con lo cual  $\text{Var}[Y] = \frac{1-p}{p^2}$ .

La Distribución Geométrica posee una propiedad comúnmente conocida como *Pérdida de Memoria*, esto es, si  $T$  tiene distribución  $\text{Geo}(p)$ , entonces

$$\Pr\{T = x + j \mid T > j\} = \Pr\{T = x\},$$

donde  $j \in \{0, 1, 2, \dots\}$  y  $x \in \{1, 2, \dots\}$ . Esto se verifica de la siguiente manera

$$\begin{aligned} \Pr\{T = x + j \mid T > j\} &= \frac{\Pr\{T = x + j, T > j\}}{\Pr\{T > j\}} \\ &= \frac{\Pr\{T = x + j\}}{\Pr\{T > j\}} \\ &= \frac{p(1-p)^{(x+j)-1}}{(1-p)^j} \\ &= p(1-p)^{x-1} \\ &= \Pr\{T = x\}. \end{aligned}$$

Esta propiedad caracteriza la Distribución Geométrica (entre todas las distribuciones definidas sobre los enteros no negativos), de la misma forma que la propiedad correspondiente caracteriza a la Distribución Exponencial (ver Sección 2.2.3).

La demostración es la siguiente: A partir de la expresión

$$\Pr\{T = x + j \mid T > j\} = \Pr\{T = x\},$$

con  $j \in \{0, 1, 2, \dots\}$  y  $x \in \{1, 2, \dots\}$ , se sigue que

$$\Pr\{T = x + j\} = \Pr\{T = x\} \Pr\{T > j\}.$$

Ahora bien

$$\begin{aligned} 1 &= \sum_{k=1}^{\infty} \Pr\{T = k\} \\ &= \sum_{j=0}^{\infty} \Pr\{T = 1 + j\} \\ &= \sum_{j=0}^{\infty} \Pr\{T = 1\} \Pr\{T > j\}. \end{aligned}$$

Se define

$$p = \Pr\{T = 1\} = \left( \sum_{j=0}^{\infty} \Pr\{T > j\} \right)^{-1}.$$

Es claro que

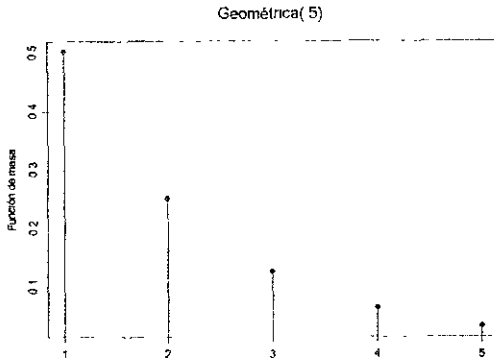


FIGURE 2.4

$$\begin{aligned}
 1 &= \Pr\{T = 1\} + \sum_{j=0}^{\infty} \Pr\{T = 2 + j\} \\
 &= p + \Pr\{T = 2\} p^{-1},
 \end{aligned}$$

entonces

$$\Pr\{T = 2\} = p(1 - p).$$

Procediendo de manera inductiva se demuestra que  $T$  tiene distribución  $\text{Geo}(p)$ .

En la Figura 2.4 se presenta la función de masa correspondiente a una distribución Geométrica de parámetro .5.

**Ejemplo 2.2.2** En una sucesión de tiradas de un dado justo (bajo el supuesto de independencia) la probabilidad de que el primer cuatro se observe en la sexta tirada está dada por

$$\Pr\{Y = 5\} = \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right) = .067,$$

donde  $Y$  denota el número de veces que no apareció el cuatro antes de la primera vez en la que es observado. La probabilidad de que se necesiten al menos seis tiradas para observar un cuatro viene dada por

$$\Pr\{Y \geq 5\} = \sum_{y=5}^{\infty} \left(\frac{5}{6}\right)^y \left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^5 = .402.$$

La probabilidad de que se necesiten a lo mas cinco intentos antes de obtener un cuatro esta dada por

$$\Pr\{Y \leq 4\} = \sum_{y=0}^4 \left(\frac{5}{6}\right)^y \left(\frac{1}{6}\right) = 1 - \left(\frac{5}{6}\right)^5. \quad \blacksquare$$

A veces resulta conveniente adoptar una definición alternativa para la Distribución Geométrica. Sea  $T'_1 = T_1 - 1$ . Entonces  $T'_1$  cuenta 'el número de fracasos antes del primer éxito' en una sucesión de ensayos Bernoulli con probabilidad de éxito  $p$ .

En este caso

$$E(T'_1) = \frac{1-p}{p}$$

y

$$\text{var}(T'_1) = \frac{1-p}{p^2}.$$

La función generatriz de momentos viene dada por

$$m_{T'_1}(t) = \frac{p}{1 - (1-p)e^t}.$$

### 2.2.2 La Distribución Binomial Negativa

Sea  $\{X_i : i \in \mathbf{N}\}$  una sucesión de ensayos Bernoulli con probabilidad de éxito  $p$ . Es de interés el número de ensayos que se debe llevar a cabo con el fin de observar el  $k$ -ésimo éxito, donde  $k$  es un entero positivo. Se denotará a este tiempo de espera con  $T_k$ . Para que  $\{T_k = r\}$  con  $r \geq k$ , debe ocurrir que:

- i) Hasta el tiempo  $r - 1$  ha habido  $k - 1$  éxitos.
- ii) El  $r$ -ésimo ensayo ha sido un éxito.

Por lo tanto, la probabilidad de que  $\{T_k = r\}$  para una sucesión de ensayos Bernoulli con probabilidad de éxito  $p$ , viene dada por

$$\begin{aligned} \Pr\{T_k = r\} &= \Pr\{X = k - 1\} p \\ &= \binom{r-1}{k-1} p^{k-1} q^{r-k} p \\ &= \binom{r-1}{k-1} p^k q^{r-k}, \end{aligned}$$

donde  $X$  tiene distribución  $\mathbf{Bi}(r - 1, p)$ .

**Definición 2.2.3** Se dice que una variable aleatoria discreta  $T_k$  tiene distribución Binomial Negativa con parámetros  $k$  y  $p$ , lo cual se denota con  $\overline{\mathbf{Bi}}(k, p)$ , si su función de masa viene dada por

$$\Pr\{T_k = r\} = \binom{r-1}{k-1} p^k q^{r-k}.$$

Donde  $k$  es un entero positivo fijo,  $p \in (0, 1)$  y  $r \geq k$ .

Si  $Y$  es una variable aleatoria con distribución  $\overline{\text{Bi}}(k, p)$ , entonces

$$\begin{aligned} E[e^{tY}] &= \sum_{y=k}^{\infty} e^{ty} \binom{y-1}{k-1} p^k q^{y-k} \\ &= e^{tk} \sum_{y=k}^{\infty} \binom{y-1}{k-1} p^k (qe^t)^{y-k} \\ &= \frac{e^{tk} p^k}{(1-qe^t)^k} \sum_{y=k}^{\infty} \binom{y-1}{k-1} (1-qe^t)^k (qe^t)^{y-k} \\ &= \frac{(e^t p)^k}{(1-qe^t)^k}, \end{aligned}$$

siempre que  $qe^t < 1$ .

Por lo tanto la función generatriz de momentos de  $Y$  viene dada por

$$m_Y(t) = \frac{(e^t p)^k}{(1-qe^t)^k}.$$

Debido a que la función generatriz de la suma de variables aleatorias independientes es el producto de las funciones generatrices de los sumandos, (ver apéndice) se puede concluir lo siguiente:

- i) Si  $X$  y  $Y$  son variables aleatorias independientes y con distribución  $\overline{\text{Bi}}(k_1, p)$  y  $\overline{\text{Bi}}(k_2, p)$  respectivamente, entonces  $X + Y$  se distribuye  $\overline{\text{Bi}}(k_1 + k_2, p)$
- ii) Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes y con distribución  $\text{Geo}(p)$ , entonces  $Y = \sum_{i=1}^n X_i$  es una variable aleatoria con distribución  $\overline{\text{Bi}}(n, p)$ .

De ii) se sigue que

$$\begin{aligned} E[T_k] &= kE[T_1] = \frac{k}{p} \\ \text{Var}[T_k] &= k\text{Var}[T_1] = \frac{kq}{p^2} \end{aligned}$$

donde  $T_1$  se distribuye  $\text{Geo}(p)$ .

En las Figuras 2.5 y 2.6 se presentan gráficos de las funciones de masa de distribuciones Binomiales Negativas de parámetros (3, .5) y (4, .8), respectivamente.

**Ejemplo 2.2.4** *Cierta matemático siempre carga con una caja de fósforos en su bolsillo izquierdo y otra en el derecho. Cuando quiere un fósforo, selecciona un bolsillo al azar, se dice que ha ocurrido un éxito si seleccionó el bolsillo izquierdo. Entonces, las elecciones consecutivas que hace, forman una sucesión de ensayos Bernoulli con  $p = \frac{1}{2}$ . Se hace el supuesto que inicialmente cada caja contiene exactamente  $N$  fósforos y se considera el momento en el que, por primera vez, el matemático se*

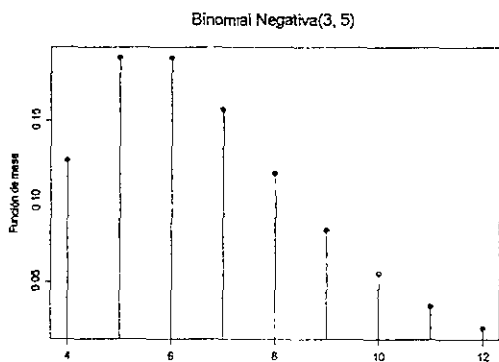


FIGURE 2.5.

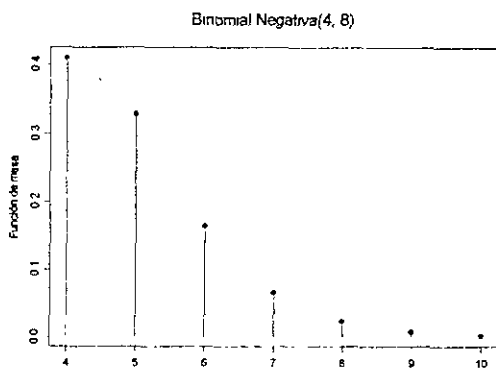


FIGURE 2.6.

encuentra con que una de las cajas esta vacía. En ese momento la otra caja debe contener  $r \in \{0, 1, 2, \dots, N\}$  fósforos, se denota la probabilidad de este evento con  $u_r$ . El matemático se encontrará con que la caja en el bolsillo izquierdo esta vacía, conservando aún  $r$  cerillos en la caja derecha si y sólo si, exactamente  $N - r$  fracasos precedieron el  $(N + 1)$ -ésimo éxito. La probabilidad de este evento viene dada por  $\Pr\{W = 2N - r + 1\}$ , donde  $W \sim \text{Bi}(N + 1, \frac{1}{2})$ . El mismo argumento se aplica para el bolsillo derecho, por lo tanto, la probabilidad buscada es

$$\begin{aligned} u_r &= 2 \Pr\{W = 2N - r + 1\} \\ &= \binom{2N - r}{N} 2^{-2N+r}. \quad \blacksquare \end{aligned}$$

**Ejemplo 2.2.5** De acuerdo a las reglas del beisbol, no hay un límite para el número de jugadores que han de batear en un juego de 9 entradas. Lo que si está determinado, es que deben llevarse a cabo 27 outs para que un equipo complete las nueve entradas. Se desea tener cierta idea acerca del número esperado de jugadores que van a batear. Se supondrá que cada bateador puede tomarse como un ensayo Bernoulli con  $q = .260$  de llegar a alguna base. Entonces  $T_{27}$  será el tiempo de espera (número de hombres al bat) para que el equipo complete las nueve entradas. Se tiene que

$$E\{T_{27} = .740\} = \frac{27}{.740} \approx 36.5$$

y

$$\text{Var}\{T_{27} = .740\} = \frac{(27)(.260)}{(.740)^2} \approx 12.8.$$

Por lo tanto, se espera un número de 36 o 37 bateadores, de los cuales 9 o 10 llagarán a alguna base. ■

### 2.2.3 El Modelo Exponencial

Sea  $F$  la función de distribución de la variable aleatoria que representa el tiempo de ocurrencia de algún evento; Por ejemplo, el arribo del siguiente cliente a una cola o la próxima llamada a una estación telefónica. Debido a que el tiempo de espera debe ser positivo, se supondrá que  $F(0) = 0$  y que  $F(x) < 1$  para toda  $x$ . Se hará la siguiente hipótesis

$$\frac{1 - F(x + y)}{1 - F(x)} = 1 - F(y), \quad x, y \geq 0. \quad (2.1)$$

El lado derecho de esta ecuación es la probabilidad de que el tiempo de espera exceda  $y$ . Mientras que el lado izquierdo es la probabilidad condicional de que el tiempo de espera exceda  $x + y$  dado que excede  $x$ .

Esto atribuye al mecanismo de tiempo de espera cierto tipo de pérdida de memoria: Si después de un lapso de  $x$  unidades de tiempo el evento aún no ha ocurrido, el tiempo que queda por esperar está condicionalmente distribuido de la misma forma que el tiempo de espera desde el inicio.

Esta condición determina por completo la forma funcional de  $F$ . Si  $U(x) = 1 - F(x)$ , entonces  $U(x+y) = U(x)U(y)$ . Esta es la forma de la Ecuación de Cauchy, y como  $U$  es acotada,  $U(x) = e^{-\beta x}$  para algún  $\beta$ . Debido a que  $\lim_{x \rightarrow \infty} U(x) = 0$ ,  $\beta$  debe ser positiva. Entonces la Expresión (2.1) implica que  $F$  tiene la forma exponencial

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\beta x} & \text{si } x > 0, \end{cases}$$

y recíprocamente.

**Definición 2.2.6** Se dice que una variable aleatoria  $X$  tiene distribución exponencial con parámetro  $\beta$  ( $\beta > 0$ ), lo cual se denota con  $X \sim \text{Exp}(\beta)$  si su función de densidad es de la forma

$$f_X(x) = \beta e^{-\beta x} \quad \text{con } x > 0.$$

Si  $X$  es una variable aleatoria con distribución  $\text{Exp}(\beta)$ , entonces

$$\begin{aligned} E[e^{tx}] &= \int_0^{\infty} e^{tx} f_X(x) dx \\ &= \frac{1}{1 - \frac{1}{\beta}t}. \end{aligned}$$

Por lo tanto la función generatriz de momentos de  $X$ , viene dada por

$$m_X(t) = \frac{1}{1 - \frac{1}{\beta}t}$$

De donde es inmediato que

$$E[X] = \frac{1}{\beta} \quad \text{y} \quad E[X^2] = \frac{2}{\beta^2},$$

por lo tanto  $\text{Var}[X] = \beta^{-2}$ .

En la Figura 2.7 se presenta un gráfico de la función de densidad de una  $\text{Exp}(1)$ .

**Ejemplo 2.2.7** Se considera una sucursal bancaria con dos cajeros, y se supone que cuando  $A$  entra al sistema se encuentra con que  $B$  está siendo atendido por uno de los cajeros y  $C$  por el otro. Esto de modo que  $A$  será atendido tan pronto  $B$  o  $C$  se vayan, ocupando el cajero que esté disponible. Si la cantidad de tiempo que un cajero necesita para atender a cada cliente se distribuye exponencialmente con parámetro  $\beta$ , cual es la probabilidad de que, de los tres clientes,  $A$  sea el último en salir del banco? Sean  $X_A, X_B$  y  $X_C$ , respectivamente las variables aleatorias que representen el tiempo que le toma al cajero atender a  $A, B$  y  $C$ . El enunciado “ $A$  será atendido tan pronto  $B$  o  $C$  se vayan” significa que el tiempo en el cual el cajero atiende a  $A$ , empieza a correr desde que la variable aleatoria  $\min(X_B, X_C)$  toma su valor. Entonces  $A$  es el último en salir del banco si y sólo si

$$X_A + \min(X_B, X_C) > \max(X_B, X_C).$$



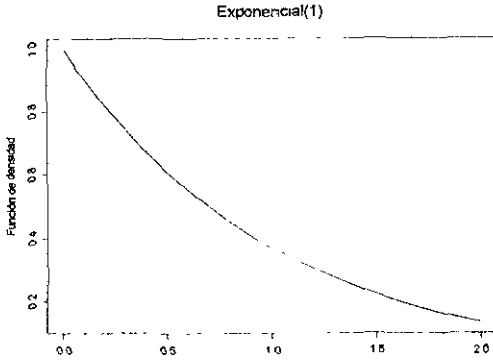


FIGURE 2.7.

Sean

$$W = \max(X_B, X_C) - \min(X_B, X_C) \quad \text{y} \quad N = 1_{\{X_B > X_C\}}.$$

Se tiene que

$$\begin{aligned} \Pr \{W > t, N = 0\} &= \int_0^\infty \left( \int_{x_B+t}^\infty \beta \exp[-\beta x_C] dx_C \right) \beta \exp[-\beta x_B] dx_B \\ &= \int_0^\infty \beta \exp[-\beta(x_B+t)] \exp[-\beta x_B] dx_B \\ &= \frac{1}{2} \exp[-\beta t]. \end{aligned}$$

Del mismo modo  $\Pr \{W > t, N = 1\} = \frac{1}{2} \exp[-\beta t]$ , entonces

$$\begin{aligned} \Pr \{W > t\} &= \Pr \{W > t, N = 0\} + \Pr \{W > t, N = 1\} \\ &= \exp[-\beta t]. \end{aligned}$$

Resulta inmediato que  $W$  tiene distribución  $\text{Exp}(\beta)$ . Gracias a que  $W$  sólo esta en función de  $X_B$  y  $X_C$ , esta variable es independiente de  $X_A$ . Del hecho que  $X_A$  y  $W$  son independientes e idénticamente distribuidas, se concluye que  $\Pr \{X_A > W\} = \frac{1}{2}$ . ■

**Ejemplo 2.2.8** Se supone que el tiempo de vida de cierto tipo de transistor tiene una distribución exponencial con una media de 1/500 horas. Sea  $X$  la variable aleatoria que represente el tiempo de vida del transistor. entonces

$$\Pr \{X > x\} = \int_x^\infty \frac{1}{500} e^{-\frac{x}{500}} dt = e^{-\frac{x}{500}}.$$

Bajo el supuesto que el transistor ha estado operando por 300 horas, la probabilidad condicional de que dicho transistor durará por lo menos otras 600 horas es

$$\begin{aligned}\Pr\{X > 900 \mid X > 300\} &= \frac{\Pr\{X > 900\}}{\Pr\{X > 300\}} \\ &= \frac{e^{-\frac{900}{600}}}{e^{-\frac{300}{600}}} \\ &= e^{-\frac{6}{6}},\end{aligned}$$

la cual es igual a  $\Pr\{X > 600\}$ . ■

### 2.2.4 Los Modelos Gamma

Antes de iniciar esta discusión, se presenta a la Función Gamma, la cual se denota con  $\Gamma$ , y cuyo valor para cualquier  $\alpha > 0$  viene dado por la siguiente expresión

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

La función  $\Gamma$  posee muchas propiedades, que son tanto útiles como interesantes, entre las cuales se encuentran las siguientes, que se enuncian sin demostración.

$$\begin{aligned}\Gamma(\alpha + 1) &= \alpha\Gamma(\alpha) \quad \text{para cualquier } \alpha > 0 \\ \Gamma(n + 1) &= n! \quad \text{donde } n \text{ es un número entero} \\ \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi} \\ \Gamma\left(n + \frac{1}{2}\right) &= \frac{(2n)!}{2^{2n}n!}\sqrt{\pi} \quad \text{para } n \text{ entero } \geq 1,\end{aligned}\tag{2.2}$$

y

$$\int_0^{\infty} x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{para cualesquiera } \alpha > 0, \beta > 0.$$

Las demostraciones pueden revisarse en [Bartle].

**Definición 2.2.9** Sea  $X$  una variable aleatoria continua que toma valores en el conjunto de los números reales no negativos. Se dice que  $X$  tiene distribución Gamma con parámetros  $\alpha$  y  $\beta$ , lo cual se denota con  $X \sim \text{Gam}(\alpha, \beta)$ , si su función de densidad es de la forma

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta},$$

con  $\alpha, \beta > 0$ .

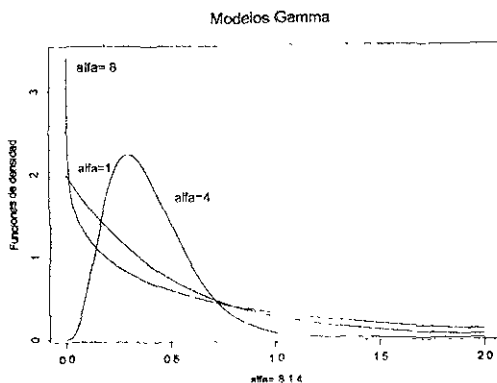


FIGURE 2.8.

**Comentario 2.2.10** Si se hace  $\alpha = 1$ , lo que se obtiene es una  $\text{Exp}(\beta)$ .

El parámetro  $\alpha$  es llamado el *parámetro de forma* de la distribución y a  $\beta$  se le conoce como el *parámetro de escala*. Estos nombres parecen naturales al ver la Figura 2.8, donde  $\alpha = .8, 1, 4$ .

Si  $Y$  es una variable aleatoria con distribución  $\text{Gam}(\alpha, 1)$ , entonces  $X = Y/\beta$  es una variable aleatoria con distribución  $\text{Gam}(\alpha, \beta)$  y los momentos de  $X$  pueden determinarse fácilmente a partir de los momentos de  $Y$ . Para cualquier  $r > 0$ ,

$$\begin{aligned} E(Y^r) &= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha+r-1} e^{-t} dt \\ &= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)}. \end{aligned}$$

Este cálculo muestra que todos los momentos de  $Y$  existen. En particular, si se hace uso de la igualdad (2.2)

$$E(Y) = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha,$$

y

$$E(Y^2) = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \alpha(\alpha+1).$$

Entonces

$$\text{Var}(Y) = \alpha(\alpha+1) - \alpha^2 = \alpha.$$

Por lo tanto, la media y la varianza de  $X$ , una variable aleatoria con distribución  $\text{Gam}(\alpha, \beta)$ , vienen dadas por

$$E(X) = \frac{1}{\beta} E(Y) = \frac{\alpha}{\beta}$$

$$\text{Var}(X) = \frac{1}{\beta^2} \text{Var}(Y) = \frac{\alpha}{\beta^2}.$$

En seguida se presenta una expresión para la función de distribución de una variable aleatoria  $\text{Gam}(k, \beta)$ , donde  $k$  es un entero.

**Teorema 2.2.11** *La función de distribución acumulada de una v.a.  $\text{Gam}(k, \beta)$  viene dada por*

$$F(x) = \sum_{i=k}^{\infty} e^{-x\beta} \frac{(\beta x)^i}{i!}.$$

**Demostración.** Se tiene que

$$\sum_{i=k}^{\infty} e^{-x\beta} \frac{(\beta x)^i}{i!} = 1 - e^{-x\beta} \sum_{i=0}^{k-1} \frac{(\beta x)^i}{i!},$$

entonces

$$\begin{aligned} \frac{d}{dx} F(x) &= \beta e^{-x\beta} \sum_{i=0}^{k-1} \frac{(\beta x)^i}{i!} - e^{-x\beta} \sum_{i=0}^{k-1} \frac{i(\beta)^i x^{i-1}}{i!} \\ &= e^{-x\beta} \sum_{i=0}^{k-1} \left[ \frac{(\beta)^{i+1} x^i}{i!} - \frac{(\beta)^i x^{i-1}}{(i-1)!} \right] \\ &= e^{-x\beta} \frac{(\beta)^k x^{k-1}}{(k-1)!}. \quad \blacksquare \end{aligned}$$

Sea  $X$  una variable aleatoria con distribución  $\text{Gam}(\alpha, \beta)$ , su función generatriz de momentos viene dada por

$$m_X(t) = \frac{1}{\left(1 - \frac{1}{\beta} t\right)^\alpha}.$$

La siguiente propiedad es una consecuencia inmediata de lo anterior: Si  $X_1$  y  $X_2$  son variables aleatorias con distribución  $\text{Gam}(\alpha_1, \beta)$  y  $\text{Gam}(\alpha_2, \beta)$  respectivamente, y dichas variables son independientes, entonces  $X_1 + X_2$  es una variable aleatoria con distribución  $\text{Gam}(\alpha_1 + \alpha_2, \beta)$ .

**Ejemplo 2.2.12** Se tiene que un aparato (por ejemplo una lámpara) consiste de un componente (por ejemplo un foco) cuyo tiempo total de vida, el cual se denota con  $L$  es una variable aleatoria exponencial con parámetro  $\beta$ . Cuando el componente falla es reemplazado inmediatamente por uno nuevo con la misma distribución de tiempo de vida. Conforme este proceso se lleva a cabo, los componentes agotados son reemplazados por componentes nuevos (cuyos tiempos de vida son independientes entre sí). La pregunta es: Para un entero fijo  $k$ , ¿Cuál es el modelo probabilístico apropiado para describir el tiempo  $T_k$  en el cual se lleva a cabo el  $k$ -ésimo reemplazo? Se denota con  $L_1, L_2, \dots$  a los tiempos de vida de los componentes que han sido reemplazados. La variable aleatoria  $L_1$  es el instante en el cual el primer reemplazo tiene lugar, y  $L_1 + L_2$  es el instante en el que el segundo reemplazo se lleva a cabo. Por lo tanto  $T_k = \sum_{i=1}^k L_i$ . Debido a que  $L_i \sim \text{Gam}(1, \beta)$ ,  $i \in \{1, 2, \dots, k\}$  y al supuesto de independencia, se concluye que  $\sum_{i=1}^k L_i$  tiene distribución  $\text{Gam}(k, \beta)$ . ■

La Distribución Gamma pertenece a la Familia Exponencial (ver Apéndice). El parámetro natural es  $\theta = (\alpha, \beta)$ , por lo tanto el espacio paramétrico natural es  $(0, \infty) \times (0, \infty)$ . En caso de que  $\alpha$  esté especificado, el parámetro natural es  $\theta = \beta$  y el espacio paramétrico correspondiente es  $(0, \infty)$ .

**Definición 2.2.13** Sea  $X$  una variable aleatoria con distribución  $\text{Gam}(\frac{v}{2}, \frac{1}{2})$  donde  $v \in \{1, 2, \dots\}$ . Se dice entonces que  $X$  tiene distribución Ji-cuadrada con  $v$  grados de libertad, lo cual se denota con  $X \sim \text{Xi}(v)$ .

Es claro que si  $X \sim \text{Xi}(v)$ , entonces  $E(X) = v$ , y  $\text{var}(X) = 2v$ . Esta distribución se empleará en la construcción de intervalos de confianza, y resulta útil como aproximación de tipo asintótico.

### 2.3 La Distribución Uniforme

**Definición 2.3.1** Una variable aleatoria  $U$  se distribuye uniformemente sobre  $(a, b)$ , donde  $a < b$ , si tiene la siguiente función de densidad

$$f_U(u) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq u \leq b \\ 0 & \text{en otro caso.} \end{cases}$$

Si  $U$  tiene distribución uniforme sobre  $(a, b)$ , lo cual se denota con  $U \sim U(a, b)$ , entonces su función de distribución viene dada por

$$F_U(x) = \begin{cases} 0 & \text{para } u \leq a \\ \frac{x-a}{b-a} & \text{para } a < x \leq b \\ 1 & \text{para } x > b. \end{cases}$$

Su media y varianza son, respectivamente,

$$E[U] = \frac{1}{2}(a + b) \quad \text{y} \quad \text{Var}[U] = \frac{(b - a)^2}{12}.$$

En la Figura 2.9 se presenta un gráfico de la función de distribución de una  $U(1, 2)$ .

Esta distribución generaliza la idea de experimentos donde todos los resultados son 'equiprobables', y será de especial utilidad al hacer inferencias de tipo bayesiano.

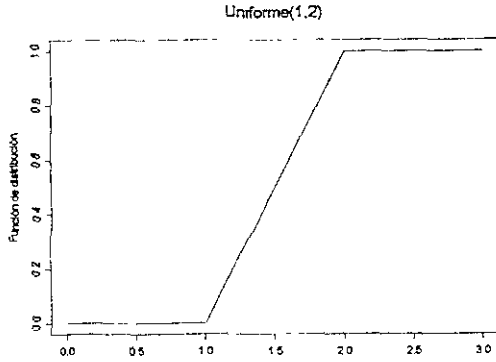


FIGURE 2.9.

## 2.4 La distribución Normal

**Definición 2.4.1** Una variable aleatoria  $X$  tiene distribución Normal con parámetros  $\xi$  y  $\sigma^2$ , si tiene función de densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2\right], \quad \sigma > 0, \quad \xi \in \mathbb{R}.$$

Esto se denota con  $X \sim N(\xi, \sigma^2)$ .

La función de densidad de  $U = (X - \xi) / \sigma$  es

$$f_U(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right],$$

la cual no depende de los parámetros  $\xi, \sigma^2$ . Esta es la *forma estándar* de la Distribución Normal. A la variable aleatoria  $U$  se le llama *variable normal estandarizada*. En la Figura 2.10 se presenta una gráfica de la función de densidad correspondiente.

Debido a que

$$\Pr\{X \leq x\} = \Pr\left[U \leq \frac{x-\xi}{\sigma}\right],$$

tales probabilidades pueden evaluarse a partir de tablas de la función de distribución acumulada de  $U$ , la cual viene dada por:

$$\Phi(u) = \Pr\{U \leq u\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx.$$

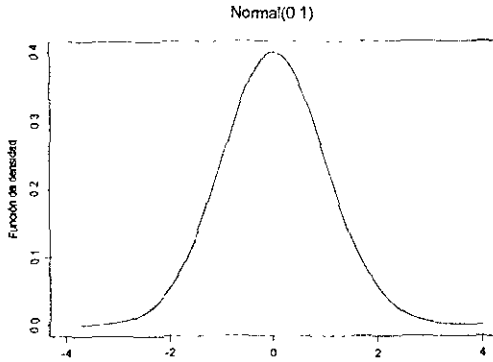


FIGURE 2.10.

La notación  $\Phi(\cdot)$  es ampliamente usada. Además, resulta conveniente tener una notación simple para los cuantiles de la distribución de  $U$ . Esta se define por

$$\Phi(z_\alpha) = \alpha$$

de modo que  $z_\alpha$  es el cuantil  $\alpha$  de la distribución  $N(0, 1)$ .

Si  $U$  tiene función de densidad normal estándar, y debido a que dicha distribución es simétrica alrededor del cero, se tiene que,

$$E[U] = 0,$$

del mismo modo

$$E[U^r] = 0.$$

si  $r$  es impar.

Si  $r$  es par

$$\begin{aligned} E[U^r] &= \sqrt{2/\pi} \int_0^\infty x^r e^{-x^2/2} dx \\ &= \sqrt{2/\pi} 2^{(r+1)/2} \int_0^\infty t^{(r-1)/2} e^{-t} dt \\ &= 2^{r/2} \Gamma\left(\frac{1}{2}(r+1)\right) / \sqrt{\pi} \\ &= (r-1)(r-3)\cdots 3 \cdot 1. \end{aligned}$$

Sea  $X$  una variable aleatoria con distribución  $N(\xi, \sigma^2)$ . La media y varianza de  $X$  vienen dadas por

$$E[X] = \xi \quad y \quad Var[X] = \sigma^2.$$

La función generatriz de momentos de esta variable es

$$m_X(t) = e^{t\mu + (t^2\sigma^2/2)}.$$

Esta distribución será útil al hacer inferencias de tipo asintótico (Secciones 4.7 y 4.9).

### 2.5 Estadísticas de Orden

**Definición 2.5.1** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias continuas e independientes, todas con función de distribución  $F(\cdot)$  y definidas en el mismo espacio de probabilidad  $(\Omega, \mathfrak{F}, P)$ . Sean  $Y_1 \leq Y_2 \leq \dots \leq Y_n$  las variables aleatorias que se obtienen al hacer  $Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)$  el conjunto  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  permutado de forma creciente, esto para toda  $\omega \in \Omega$ . A las variables  $Y_i$  se les denomina como las estadísticas de orden de  $X_1, X_2, \dots, X_n$ .

De acuerdo a los siguientes resultados, las distribuciones marginales, así como la distribución conjunta de  $Y_1, Y_2, \dots, Y_n$ , estarán en función de  $F$ .

**Teorema 2.5.2** Dada una muestra  $X_1, X_2, \dots, X_n$  proveniente de una distribución con densidad continua  $f(\cdot)$  entonces la densidad conjunta de las estadísticas de orden  $Y_1 \leq Y_2 \leq \dots \leq Y_n$  viene dada por

$$f(y_1, y_2, \dots, y_n) = n! \prod_{i=1}^n f(y_i), \quad y_1 < y_2 < \dots < y_n.$$

**Corolario 2.5.3** Si las  $X_i, i = 1, \dots, n$ , se distribuyen uniformemente sobre  $(0, t)$ , entonces la función de densidad conjunta de las estadísticas de orden  $Y_1, Y_2, \dots, Y_n$  es

$$f(y_1, y_2, \dots, y_n) = \frac{n!}{t^n}, \quad y_1 < y_2 < \dots < y_n < t.$$

**Teorema 2.5.4** Sean  $Y_1, Y_2, \dots, Y_n$  las estadísticas de orden de una muestra aleatoria,  $X_1, X_2, \dots, X_n$  tomadas de una distribución continua con función de distribución  $F$  y función de densidad  $f$ . Entonces la función de densidad de  $Y_j$  es

$$f_{Y_j}(y) = \frac{n!}{(j-1)!(n-j)!} f(y) [F(y)]^{j-1} [1-F(y)]^{n-j}.$$

Este material se puede revisar en [Mood, Graybill y Boes].

### 2.6 Simulación

Supóngase que es posible generar una sucesión de variables aleatorias independientes con distribución uniforme en  $(0, 1)$ , se denota a dicha sucesión con  $\{U_i : i \in \mathbf{N}\}$ . No se tratará el problema de generar dichas variables.



**Proposición 2.6.1** *Sea  $X$  una variable aleatoria con función de distribución  $F$ . Se hace el supuesto que  $F$  es continua y estrictamente creciente. Sea  $U$  una variable aleatoria con distribución uniforme en  $(0, 1)$ . Entonces  $F^{-1}(U)$  tiene distribución  $F$ .*

**Demostración.** Debido a que  $U$  es una variable aleatoria uniforme en  $(0, 1)$ ,  $\Pr\{U \leq x\} = x$ ,  $0 \leq x \leq 1$ . Entonces  $\Pr\{F^{-1}(U) \leq t\} = \Pr\{U \leq F(t)\} = F(t)$ . ■

Trivialmente, el Modelo Exponencial cumple con las hipótesis de este resultado, entonces, para generar una muestra aleatoria  $\{Y_i : i \in \mathbf{N}\}$  de una distribución exponencial de parámetro  $\beta$  a partir de  $\{U_i : i \in \mathbf{N}\}$ , basta hacer

$$Y_i = -\frac{1}{\beta} \ln \left( \frac{U_i}{\beta} \right).$$

Es claro que el Modelo Gamma también cumple con dichas hipótesis, sin embargo no es fácil en general obtener  $F^{-1}$  para este caso. Esta dificultad tiene una solución simple, si el parámetro de forma  $\alpha$  es un entero positivo. Lo que procede a hacerse es generar  $\alpha$  variables aleatorias exponenciales independientes con el parámetro  $\beta$  correspondiente, y luego sumar dichas variables, de esta forma la suma tendrá distribución **Gam** $(\alpha, \beta)$ .



# 3

## El Proceso Poisson

2

### Introducción

Un rasgo característico de muchos sistemas que se desarrollan en el tiempo es la presencia de incertidumbre, dicha incertidumbre puede deberse a un mecanismo aleatorio que genere una colección numerable de 'tiempos de ocurrencias' de modo que el número de 'ocurrencias' en un intervalo es una variable aleatoria. El Proceso Poisson Estacionario y sus generalizaciones (Procesos Poisson no estacionarios, compuestos y puntuales) son los modelos más sencillos para describir esta clase de fenómenos. Su simplicidad, así como sus propiedades más conocidas, son en gran parte consecuencia de ciertos supuestos de independencia estocástica.

El objetivo de este capítulo es el de definir al Proceso Poisson y presentar teoremas elementales referentes a este proceso y a sus generalizaciones; se busca además exponer los resultados necesarios para plantear adecuadamente los problemas de inferencia paramétrica que se tratarán en los capítulos siguientes.

En la Sección 3.1 se define al Proceso Poisson Estacionario (PPE) sobre la recta real. En la siguiente sección (3.2) se discuten tres propiedades importantes de este proceso: La distribución de los tiempos entre ocurrencias, la distribución condicional de los tiempos de llegada y la Ley de los grandes números para el proceso Poisson. La Sección 3.3 está dedicada a algunas generalizaciones de este proceso, se trata del Proceso Poisson no Estacionario (PPNE), el Proceso Poisson Compuesto (PPC) y el Proceso Poisson Puntual (PPP). Hay que mencionar que mientras el PPNE y el PPC son discutidos desde el punto de vista de aplicaciones (modelación), lo que motivó a incluir al PPP en este capítulo fué la posibilidad de acceder al aparato analítico necesario para demostrar los teoremas discutidos en las secciones subsiguientes. La Sección 3.4 se refiere a caracterizaciones del PPP. En la Sección 3.5 se discute cierta forma de descomposición de un PPC como suma de procesos Poisson Puntuales. En la última sección (3.6) se presentan algunos de los algoritmos más utilizados para la simulación de este tipo de procesos.

Muchos de los resultados teóricos de este capítulo fueron desarrollados tomando como base [Kingman], [Cinlar] y [Karlin y Taylor]. Los ejemplos de aplicaciones prácticas fueron tomados de [Snyder], [Karlin y Taylor] y [Nelson]. El material referente a simulación fue tomado de [Bratley].

### 3.1 El Proceso Poisson en la recta real

En la Sección 2.1 se definió la Distribución Poisson. Este modelo puede utilizarse para describir de manera probabilista el número de ocurrencias de cierto suceso de interés en un intervalo  $(0, t]$  con  $t$  fijo. Una forma de visualizar esto es la siguiente: Se considera una partición

$$\left\{ \left( \frac{kt}{N}, \frac{(k+1)t}{N} \right] : N \in \mathbf{N} \text{ y } k \in \{0, 1, 2, \dots, N-1\} \right\},$$

y sea  $A_{k,N}$  el evento

$$\left\{ \text{hubo al menos una ocurrencia en } \left( \frac{kt}{N}, \frac{(k+1)t}{N} \right] \right\}.$$

Sea  $p_N = \Pr \{A_{k,N}\}$  y se supone que los eventos

$$\{A_{k,N} : k \in \{0, 1, 2, \dots, N-1\}\}$$

son independientes. Entonces bajo el supuesto de que  $\lim_{n \rightarrow \infty} np_n = \lambda t$  para algún  $\lambda > 0$  se tiene que el número de ocurrencias en  $(0, t]$  sigue una distribución  $\text{Poi}(\lambda t)$ .

El siguiente paso consiste en proponer un modelo probabilista para un fenómeno que se desarrolla en el tiempo (idealizando al tiempo como  $\mathbf{R}^+$ ) y tal que para cada intervalo finito  $(a, b]$  se cumplen los supuestos antes mencionados.

Antes de hacer esto se presentan algunos conceptos básicos.

Un *proceso estocástico* es una familia de variables aleatorias  $\{Y_l : l \in \Lambda\}$  definidas en un mismo espacio de probabilidad  $(\Omega, \mathfrak{F}, \mathbf{P})$ , donde  $\Lambda$  es un conjunto de índices. Serán de especial interés los procesos donde  $\Lambda = \mathbf{R}^+$  o  $\mathbf{N}$ , esto con el propósito de dar a entender (en la mayoría de los casos) que cada variable aleatoria se encuentra asociada con un instante en el tiempo (discreto o continuo). Dado un proceso estocástico  $\{Y_l : l \in \Lambda\}$ , al conjunto de valores que pueden tomar las variables  $Y_l$  se le denomina *espacio de estados*.

**Definición 3.1.1** Se dice que un proceso estocástico  $\{N_t : t \geq 0\}$  es un proceso de contar si satisface las siguientes condiciones:

- i)  $N_t \geq 0$ .
- ii)  $N_t$  sólo toma valores en los enteros.
- iii) si  $s < t$ , entonces  $N_s \leq N_t$ .
- iv) Para  $s < t$ ,  $N_t - N_s$  representa el número de eventos que ha ocurrido en el intervalo  $(s, t]$ .

La definición es intuitivamente clara. El objetivo es describir matemáticamente el número de ocurrencias o eventos de interés desde un tiempo inicial al cual se denota con 0, hasta cierto tiempo  $t$ .

No se estudiará toda esta clase de modelos, únicamente se desarrollará la teoría referente a una subclase. Los elementos de dicha subclase poseen la llamada Propiedad de Incrementos Independientes. Esta propiedad será clave al momento de demostrar resultados.

**Definición 3.1.2** Sea  $\{X_t : t \geq 0\}$  un proceso estocástico con espacio de estados  $\Lambda = \mathbf{R}^k$  o  $\mathbf{N}$ , se dice que dicho proceso posee la Propiedad de Incrementos Independientes, si para cualesquiera  $0 \leq s_1 < t_1 \leq s_2 < t_2, \dots, s_k < t_k$  se tiene que

$$X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \dots, X_{t_k} - X_{s_k}$$

son variables aleatorias independientes. Esto para toda  $k \geq 2$ .

Más adelante se verá que esta característica es conveniente desde el punto de vista computacional. No todo proceso de contar posee la Propiedad de Incrementos Independientes.

**Ejemplo 3.1.3** Una población de bacterias se desarrolla sin la presencia de depredadores. Supóngase que en una unidad de tiempo cada individuo puede generar, independientemente de otros individuos,  $X$  descendientes,  $X$  tiene distribución  $\text{Bi}(k, p)$ . Sea  $Y_n$  el tamaño de la población en el tiempo  $n$ . Se supondrá que  $Y_0 = 2$ . Se tiene que

$$\Pr \{Y_2 - Y_1 = 4k \mid Y_1 - Y_0 = 0\} = 0$$

y

$$\Pr \{Y_2 - Y_1 = 4k \mid Y_1 - Y_0 = 2\} = (p^k)^4 > 0.$$

Por lo tanto  $\{Y_n : n \geq 0\}$  es un proceso de contar que no posee la Propiedad de Incrementos Independientes ■

Ahora se presenta la definición del modelo sobre el cual se centra esta discusión.

**Definición 3.1.4** Sea  $\{N_t : t \geq 0\}$  un proceso de contar tal que:

- i)  $N_0 = 0$ .
- ii) El proceso posee la Propiedad de Incrementos Independientes.
- iii) El número de eventos en cualquier intervalo de longitud  $t$  es una variable aleatoria con distribución  $\text{Poi}(\lambda t)$ . Es decir, para cualesquiera  $s, t \geq 0$

$$\Pr \{N_{t+s} - N_s = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad n = 0, 1, \dots$$

Se dice entonces que  $\{N_t : t \geq 0\}$  es un Proceso Poisson Estacionario con intensidad  $\lambda$ .

El adjetivo *estacionario* se debe al hecho de que la distribución de  $N_{t+s} - N_s$  no depende de  $s$ . Más tarde se removerá esta restricción. En la Figura 3.1 se presenta una trayectoria simulada de un proceso Poisson estacionario con  $\lambda = 1$ .

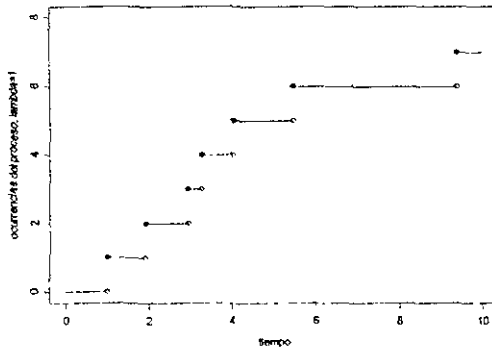


FIGURE 3.1.

**Ejemplo 3.1.5** Se supone que el número de clientes que se integran a una fila se puede describir mediante un Proceso Poisson Estacionario con intensidad  $\lambda = 2$ ; el tiempo se mide en minutos.

a) ¿Cuál es la probabilidad de que se integren 5 personas a la fila entre los minutos 7 y 10, y 4 personas entre los minutos 12 y 19?

Sea  $A$  el evento de interés, entonces gracias al punto ii) de la Definición 3.1.4

$$\begin{aligned} \Pr \{A\} &= \Pr \{N_{10} - N_7 = 5, N_{19} - N_{12} = 4\} \\ &= \Pr \{N_{10} - N_7 = 5\} \Pr \{N_{19} - N_{12} = 4\}, \end{aligned}$$

aplicando iii) se obtiene

$$\begin{aligned} \Pr \{A\} &= \Pr \{N_3 = 5\} \Pr \{N_7 = 4\} \\ &= e^{-(2)(3)} \frac{[(2)(3)]^5}{5!} e^{-(2)(7)} \frac{[(2)(7)]^4}{4!}. \quad \blacksquare \end{aligned}$$

Ahora se presenta una caracterización del Proceso Poisson Estacionario (PPE), esto se hace a partir de la llamada Ley de Eventos Raros.

**Definición 3.1.6** Se dice que un experimento cumple con la Ley de Eventos Raros si

i) Existe una constante positiva  $\lambda$  tal que la probabilidad de que ocurra exactamente un evento en un intervalo de longitud  $h$  viene dada por

$$\Pr \{\text{una ocurrencia en } (t, t + h]\} = \lambda h + o(h),$$

donde  $o(h)$  es una función tal que  $\lim_{x \rightarrow 0} \frac{o(x)}{x} = 0$ .

ii) La probabilidad de que ocurran dos o más eventos en un intervalo de longitud  $h$  es

$$\Pr \{\text{dos o más ocurrencias en } (t, t + h]\} = o(h).$$

iii) El número de eventos que ocurren en intervalos ajenos son variables aleatorias independientes, esto es, para cualquier entero  $m = 2, 3, \dots$  y puntos en el tiempo  $t_0 = 0 < t_1 < t_2 < \dots < t_m$ , las variables aleatorias

$$\{\# \text{ de ocurrencias en } (t_i, t_{i+1})\}$$

son independientes.

En seguida se muestra la relación entre este concepto y el Proceso Poisson.

**Teorema 3.1.7** Para un experimento en el que se cumple la Ley de Eventos Raros, el número de ocurrencias para un intervalo de la forma  $(s, t]$  es una variable aleatoria con distribución Poisson.

**Demostración.** Sean  $h, t > 0$ . Se define

$$p_n(s) = \Pr \{\text{ocurran exactamente } k \text{ eventos en un intervalo de longitud } s\}.$$

Entonces

$$\begin{aligned} p_0(t+h) &= \Pr \{\text{no hubo ocurrencias en } (0, t+h)\} \\ &= \Pr \{\text{no hubo ocurrencias ni en } (0, t] \text{ ni en } (t, t+h)\} \\ &= \Pr \{\text{no hubo ocurrencias en } (0, t]\} \\ &\quad \cdot \Pr \{\text{no hubo ocurrencias en } (t, t+h)\} \\ &= p_0(t) p_0(h); \end{aligned}$$

esto gracias a la condición iii) de la Definición 3.1.6.

Ahora bien

$$\begin{aligned} &\Pr \{\text{no hubo ocurrencias en } (t, t+h)\} \\ &= 1 - \Pr \{\text{una o más ocurrencias en } (t, t+h)\} \\ &= 1 - \Pr \{\text{una ocurrencia en } (t, t+h)\} \\ &\quad - \Pr \{\text{más de una ocurrencia en } (t, t+h)\} \\ &= 1 - \lambda h - o(h). \end{aligned}$$

Entonces  $p_0(t+h) = p_0(t)[1 - \lambda h - o(h)]$ , lo cual se puede expresar en la siguiente forma

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) - p_0(t) \frac{o(h)}{h}.$$

Haciendo tender  $h$  a cero se obtiene la ecuación diferencial

$$p_0'(t) = -\lambda p_0(t),$$

cuya solución, para la condición de frontera  $p_0(0) = 1$  viene dada por  $p_0(t) = e^{-\lambda t}$ . De manera similar  $p_1(t+h) = p_1(t)p_0(h) + p_0(t)p_1(h)$ , o, visto de otra forma

$$p_1(t+h) = p_1(t)[1 - \lambda h - o(h)] + p_0(t)[\lambda h + o(h)].$$

De donde se puede obtener la ecuación diferencial

$$p_1'(t) = -\lambda p_1(t) + \lambda p_0(t),$$

la cual equivale a resolver

$$e^{\lambda t} p_1'(t) + \lambda e^{\lambda t} p_1(t) = \lambda.$$

Visto de otra forma

$$\frac{d}{dt} [e^{\lambda t} p_1(t)] = \lambda.$$

Integrando ambos lados respecto a  $t$  se tiene que

$$e^{\lambda t} p_1(t) = \lambda t + c \quad c \in \mathbb{R}.$$

Entonces la solución es  $p_1(t) = \lambda t e^{-\lambda t}$ , debido a la condición inicial  $p_1(0) = 0$ . Procediendo de manera totalmente análoga, se llega a que

$$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad \text{para } n \in \{2, 3, \dots\}.$$

La expresión que satisface este sistema de ecuaciones diferenciales es

$$p_n(t) = (\lambda t)^n e^{-\lambda t} \frac{1}{n!},$$

esto se demuestra por inducción. Ya se verificó el resultado para  $n = 0, 1$ . Supóngase ahora que

$$p_{k-1}(t) = (\lambda t)^{k-1} e^{-\lambda t} \frac{1}{(k-1)!}.$$

Entonces

$$p_k'(t) = -\lambda p_k(t) + \lambda (\lambda t)^{k-1} e^{-\lambda t} \frac{1}{(k-1)!},$$

luego

$$e^{\lambda t} p_k'(t) + \lambda e^{\lambda t} p_k(t) = \lambda^k \frac{t^{k-1}}{(k-1)!}.$$

Se sigue que

$$e^{\lambda t} p_k(t) = \lambda^k \frac{t^k}{k!}. \quad \blacksquare$$

Por lo tanto, si un proceso estocástico  $\{N_t : t \geq 0\}$  cumple con las siguientes condiciones.

i) Es un proceso de contar tal que  $N_0 = 0$ .



ii) El proceso cumple con la Ley de Eventos Raros.

Entonces,  $\{N_t : t \geq 0\}$  es un PPE, y la intensidad de dicho proceso es la constante  $\lambda$  que se menciona en la Definición 3.1.6.

La propiedad de incrementos independientes es un elemento clave en la teoría del Proceso Poisson Estacionario. Dicha propiedad establece que el número de ocurrencias del proceso en intervalos ajenos determinísticos son variables aleatorias independientes. A continuación se demuestra que esto aún es válido bajo ciertas condiciones de aleatoriedad.

**Teorema 3.1.8** Sean  $0 = T_0 \leq T_1 \leq T_2 \leq \dots$  variables aleatorias finitas, de modo que

$$\{T_i - T_{i-1} : i \in \{1, 2, \dots\}\}$$

son variables aleatorias independientes, idénticamente distribuidas y con función de densidad. Sea  $\{N_t : t \geq 0\}$  es un PPE con intensidad  $\lambda$ . Se hace el supuesto que las  $T_i$  son independientes del proceso. Entonces para cualquier colección de números naturales

$$0 \leq i_1 < j_1 \leq i_2 < j_2 \leq \dots$$

las variables aleatorias

$$\{N_{T_{j_m}} - N_{T_{i_m}} : m \in \{1, 2, \dots\}\}$$

son independientes.

**Demostración.** Se considera el caso con dos intervalos aleatorios, los tiempos se denotan con  $T_1, \dots, T_4$ . Entonces, para  $k_1, k_2 \in \{0, 1, \dots\}$  se tiene

$$\begin{aligned} & \Pr \{N_{T_2} - N_{T_1} = k_1, N_{T_4} - N_{T_3} = k_2\} \\ &= \int \Pr \{N_{t_2} - N_{t_1} = k_1, N_{t_4} - N_{t_3} = k_2\} f_{\mathbf{T}}(t_1, t_2, t_3, t_4) dt \end{aligned}$$

donde  $f_{\mathbf{T}}$  es la densidad conjunta de los tiempos. Se integra sobre el soporte de la densidad. Se sigue

$$\begin{aligned} & \Pr \{N_{T_2} - N_{T_1} = k_1, N_{T_4} - N_{T_3} = k_2\} \\ &= \int \Pr \{N_{t_2} - N_{t_1} = k_1\} \Pr \{N_{t_4} - N_{t_3} = k_2\} f_{\mathbf{T}}(t_1, t_2, t_3, t_4) dt \\ &= \int \Pr \{N_{y_2+y_1} - N_{y_1} = k_1\} \\ & \quad \cdot \Pr \{N_{y_4+y_3+y_2+y_1} - N_{y_3+y_2+y_1} = k_2\} f_{\mathbf{Y}}(y_1, y_2, y_3, y_4) dy \\ &= \int \Pr \{N_{y_2} = k_1\} \Pr \{N_{y_4} = k_2\} f_{\mathbf{Y}}(y_1, y_2, y_3, y_4) dy, \end{aligned}$$

donde

$$Y_1 = T_1, Y_2 = T_2 - T_1, Y_3 = T_3 - T_2, Y_4 = T_4 - T_3.$$

Por hipótesis estas variables son independientes. Se denota con  $f_Y$  a la densidad marginal correspondiente. Entonces

$$\begin{aligned} & \int \Pr \{N_{y_2} = k_1\} \Pr \{N_{y_4} = k_2\} f_Y(y_1, y_2, y_3, y_4) dy \\ &= \int \Pr \{N_{y_2} = k_1\} f_{Y_2}(y_2) dy_2 \int \Pr \{N_{y_4} = k_2\} f_{Y_4}(y_4) dy_4 \\ &= \Pr \{N_{Y_2} = k_1\} \Pr \{N_{Y_4} = k_2\}. \end{aligned}$$

Por lo tanto

$$\begin{aligned} & \Pr \{N_{T_2} - N_{T_1} = k_1, N_{T_4} - N_{T_3} = k_2\} \\ &= \Pr \{N_{T_2 - T_1} = k_1\} \Pr \{N_{T_4 - T_3} = k_2\} \\ &= \int \Pr \{N_{t_2 - t_1} = k_1\} f_{T_1, T_2}(t_1, t_2) dt_1 dt_2 \\ & \quad \cdot \int \Pr \{N_{t_4 - t_3} = k_1\} f_{T_3, T_4}(t_3, t_4) dt_3 dt_4 \\ &= \int \Pr \{N_{t_2} - N_{t_1} = k_1\} f_{T_1, T_2}(t_1, t_2) dt_1 dt_2 \\ & \quad \cdot \int \Pr \{N_{t_4} - N_{t_3} = k_1\} f_{T_3, T_4}(t_3, t_4) dt_3 dt_4 \\ &= \Pr \{N_{T_2} - N_{T_1} = k_1\} \Pr \{N_{T_4} - N_{T_3} = k_2\}, \end{aligned}$$

lo cual concluye la prueba. El argumento puede extenderse a cualquier número finito de intervalos aleatorios. ■

Este Teorema será útil cuando se haga inferencias para el Proceso Poisson tomando un número finito de intervalos con longitud aleatoria de distribución exponencial.

## 3.2 Propiedades del Proceso Poisson

### 3.2.1 Distribución de los tiempos entre llegadas

Sea  $\{X_k : k \in \mathbf{N}\}$  una sucesión de variables aleatorias independientes con distribución  $\text{Exp}(\lambda)$ . Se considera ahora el proceso  $\{S_k : k \in \mathbf{N}\}$ , donde

$$S_k = \sum_{i=1}^k X_i, \quad k \in \mathbf{N}.$$

Se define

$$N_t = \sum_{k=1}^{\infty} 1_{[S_k < t]} \quad t \in \mathbf{R}^+.$$

**Teorema 3.2.1**  $\{N_t : t \geq 0\}$  definido de esta forma es un Proceso Poisson con intensidad  $\lambda$ .

**Demostración.** Es claro que

$$\{N_t \geq r\} = \{S_r < t\},$$

entonces

$$\begin{aligned} \Pr\{N_t \geq r\} &= \Pr\{S_r < t\} \\ &= \sum_{i=r}^{\infty} e^{-\lambda t} \frac{(\lambda t)^i}{i!}. \end{aligned}$$

Esto se sigue del hecho de que  $S_r$  tiene distribución **Gam**  $(r, \lambda)$  y de aplicar el Teorema 2.2.11. Se concluye que  $N_t$  se distribuye **Poi**  $(\lambda t)$ .

Ahora se examinará lo que ocurre para un intervalo  $(a, b)$  con  $0 \leq a < b < \infty$ .

Sea

$$X_{a,1} = S_{i_a} - a,$$

donde  $i_a = \min_{k \in \mathbf{N}} \{S_k > a\}$ . Y sea

$$X_{a,k} = S_{(i_a)+(k-1)} - S_{(i_a)+(k-2)}, \quad k \geq 2.$$

Aplicando la Propiedad de Pérdida de Memoria de la Distribución Exponencial, se sigue que  $X_{a,1}$  tiene distribución **Exp**  $(\lambda)$ , ya que

$$\begin{aligned} \Pr\{X_{a,1} > t\} &= \Pr\{X_{i_a} > t + (a - S_{(i_a)-1}) \mid X_{i_a} > (a - S_{(i_a)-1})\} \\ &= \Pr\{X_{i_a} > t\} \end{aligned}$$

y  $X_{i_a}$  es un elemento de  $\{X_k : k \in \mathbf{N}\}$ . Se considera ahora el proceso  $\{X_{a,k} : k \in \mathbf{N}\}$ . Usando el mismo razonamiento que se empleó para  $\{X_k : k \in \mathbf{N}\}$  y para  $(0, t]$ , se sigue que  $N_b - N_a$  tiene distribución **Poi**  $(\lambda(b-a))$ .

Falta verificar que  $\{N_t : t \geq 0\}$  posee la Propiedad de Incrementos Independientes.

Sea  $0 \leq a_1 < b_1 \leq a_2 < b_2 < \infty$ . Se construyen los procesos

$$\{M_t : t \geq 0\} \quad \text{y} \quad \{Q_t : t \geq 0\},$$

donde

$$M_t = \sum_{k=1}^{\infty} \mathbf{1}_{[S_k < t, S_k < b_1]} \quad t \in \mathbf{R}^+$$

y

$$Q_t = \sum_{k=1}^{\infty} \mathbf{1}_{[S_k < t, S_k > a_2]} \quad t \in \mathbf{R}^+.$$

Es claro que  $N_{b_1} - N_{a_1}$  puede expresarse como una función de  $\{M_t : t \geq 0\}$  y que  $N_{b_2} - N_{a_2}$  esta en función únicamente de  $\{Q_t : t \geq 0\}$ . Ya que

$$N_{b_1} - N_{a_1} = M_{b_1} - M_{a_1}$$

y

$$N_{b_2} - N_{a_2} = Q_{b_2}$$

Se tiene que  $\{M_t : t \geq 0\}$  es independiente de  $\{Q_t : t \geq 0\}$  por lo tanto  $N_{b_1} - N_{a_1}$  es independiente de  $N_{b_2} - N_{a_2}$ .

Este argumento se generaliza fácilmente para cualesquiera  $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k$ , con  $k \geq 2$  de donde se obtiene la Propiedad de Incrementos Independientes. ■

El recíproco del teorema anterior también es verdadero, esto brinda una nueva caracterización del PPE

**Teorema 3.2.2** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson estacionario de intensidad  $\lambda$ . Se define

$$S_k = \inf_{t \geq 0} \{N_t = k\}$$

y

$$X_k = S_k - S_{k-1}$$

Donde  $k \in \{1, 2, \dots\}$  y  $S_0 = 0$ . Entonces las variables aleatorias  $X_k$  son independientes y distribuidas exponencialmente con parámetro  $\lambda$ .

**Demostración.** Se sabe que

$$\Pr \{N_t = 0\} = e^{-\lambda t}.$$

Además

$$\{X_1 > t\} = \{N_t = 0\},$$

entonces

$$\Pr \{X_1 \leq t\} = 1 - e^{-\lambda t},$$

de donde se sigue que  $X_1$  se distribuye  $\text{Exp}(\lambda)$ .

Se procede a encontrar la distribución de  $X_2$  condicionado a  $X_1$

$$\begin{aligned} \Pr \{X_2 > t \mid X_1 = x_1\} &= \Pr \{X_2 > t \mid N_r = 0, r < x_1, N_{x_1} = 1\} \\ &= \Pr \{N_{x_1+t} - N_{x_1} = 0 \mid N_r = 0, r < x_1, N_{x_1} = 1\} \\ &= \Pr \{N_{x_1+t} - N_{x_1} = 0\}. \end{aligned}$$

Esta última igualdad se debe a la Propiedad de Incrementos Independientes. Entonces

$$\Pr\{X_2 > t \mid X_1 = x_1\} = e^{-\lambda t},$$

lo cual muestra que  $X_2$  es independiente de  $X_1$  y que  $X_2$  tiene distribución  $\text{Exp}(\lambda)$ . Se tiene además que

$$\begin{aligned} & \Pr\{X_3 > t \mid X_1 = x_1, X_2 = x_2\} \\ &= \Pr\{X_3 > t \mid N_r = 0, r < x_1, N_s = 1, x_1 \leq s < x_1 + x_2, N_{x_1+x_2} = 2\} \\ &= \Pr\{N_{x_1+x_2+t} - N_{x_1+x_2} = 0 \mid N_r = 0, r < x_1, \\ & \quad N_s = 1, x_1 \leq s < x_1 + x_2, N_{x_1+x_2} = 2\} \\ &= \Pr\{N_{x_1+x_2+t} - N_{x_1+x_2} = 0\} \\ &= e^{-\lambda t}. \end{aligned}$$

Aplicando de manera inductiva este razonamiento se concluye el resultado deseado. ■

**Corolario 3.2.3** *La siguiente identidad es válida*

$$N_t = \sum_{k=1}^{\infty} 1_{\{S_k < t\}}.$$

Este resultado aprovecha el hecho de que  $\{N_t = n\} = \{S_n < t \leq S_{n+1}\}$  y será útil a la hora de simular al Proceso Poisson.

**Ejemplo 3.2.4** *Existen buenas razones para suponer que los tiempos en los cuales una neurona lleva a cabo cada sinapsis se comportan como un Proceso Poisson con intensidad  $\lambda = \frac{1}{2}$ .*

*¿Cuál es la probabilidad de que se tenga que esperar más de 3 minutos antes de observar la primera sinapsis y 10 minutos a partir de la primera sinapsis para observar la segunda?*

*La probabilidad buscada es*

$$\begin{aligned} \Pr\{X_1 > 3, X_2 > 10\} &= \Pr\{X_1 > 3\} \Pr\{X_2 > 10\} \\ &= e^{-2(3)} e^{-2(10)}. \end{aligned}$$

*¿Cuál es el tiempo esperado para la ocurrencia de la cuarta sinapsis?*

*Se usa el hecho de que  $S_4$  tiene distribución  $\text{Gam}(4, 2)$ , entonces  $E[S_4] = (4)(2)$ . ■*

### 3.2.2 Distribución condicional de los tiempos de llegada

Supóngase que se tiene un Proceso Poisson estacionario  $\{N_t : t \geq 0\}$  y se sabe que exactamente una ocurrencia ha tenido lugar en el intervalo  $(0, t]$ . Es de interés conocer la distribución de la variable aleatoria que representa el tiempo exacto en el cual ocurrió el evento. Debido a la estacionariedad del proceso y a la Propiedad de Incrementos Independientes, parece razonable que para cualquier intervalo  $[s, w + s]$ , con  $s, w \geq 0$ ,  $s + w \leq t$ , la probabilidad de contener a la ocurrencia no cambie.

Lo anterior implicaría que el tiempo en el que ocurre el evento debe distribuirse como una variable aleatoria uniforme a lo largo de  $(0, t]$ .

Para verificar esto, sea  $v \leq t$ ,

$$\begin{aligned} \Pr\{X_1 < v \mid N_t = 1\} &= \frac{\Pr\{X_1 < v, N_t = 1\}}{\Pr\{N_t = 1\}} \\ &= \frac{\Pr\{N_v = 1, N((v, t]) = 0\}}{\Pr\{N_t = 1\}} \\ &= \frac{\Pr\{N_v = 1\} \Pr\{N((v, t]) = 0\}}{\Pr\{N_t = 1\}} \\ &= \frac{\lambda v e^{-\lambda v} e^{-\lambda(t-v)}}{\lambda t e^{-\lambda t}} \\ &= \frac{v}{t}. \end{aligned}$$

**Teorema 3.2.5** Si el Proceso Poisson  $\{N_t : t \geq 0\}$  se encuentra condicionado con el evento  $\{N_t = n\}$ , entonces las variables aleatorias  $S_1, S_2, \dots, S_n$  tienen la misma distribución conjunta que las estadísticas de orden correspondientes a  $n$  variables aleatorias uniformes sobre el intervalo  $(0, t)$ .

**Demostración.** Se debe calcular la función de densidad conjunta de  $S_1, S_2, \dots, S_n$  condicionada con  $\{N_t = n\}$ . Se consideran  $0 < t_1 < t_2 < \dots < t_{n+1} = t$  y sean  $h_i$  reales positivo lo suficientemente pequeños para que  $t_i + h_i < t_{i+1}$ ,  $i = 1, \dots, n$ . Entonces

$$\begin{aligned} &\Pr\{t_i < S_i \leq t_i + h_i, i = 1, 2, \dots, n \mid N_t = n\} \\ &= \frac{\Pr\{N((t_i, t_i + h_i]) = 1, i = 1, \dots, n, \text{ ningún evento en lo que queda de } (0, t]\}}{\Pr\{N_t = n\}} \\ &= \frac{\lambda h_1 e^{-\lambda h_1} \dots \lambda h_n e^{-\lambda h_n} e^{-\lambda(t-h_1-h_2-\dots-h_n)}}{e^{-\lambda t} (\lambda t)^n / n!} \\ &= \frac{n!}{t^n} h_1 \cdot h_2 \cdot \dots \cdot h_n. \end{aligned}$$

Por lo tanto

$$\frac{\Pr\{t_i < S_i \leq t_i + h_i, i = 1, 2, \dots, n \mid N_t = n\}}{h_1 \cdot h_2 \cdot \dots \cdot h_n} = \frac{n!}{t^n}.$$

Tomando el límite cuando  $h_i \rightarrow 0$ , se obtiene la densidad conjunta de  $S_1, S_2, \dots, S_n$  dado el evento  $\{N_t = n\}$ , la cual es

$$f_{S_1, S_2, \dots, S_n}(t_1, t_2, \dots, t_n) = \frac{n!}{t^n}, \quad 0 < t_1 < t_2 < \dots < t_n,$$

por lo que se da la prueba por concluida. ■

**Ejemplo 3.2.6** Una compañía de seguros modela la frecuencia del siniestro 'robo de automóvil' mediante un PPE de intensidad  $\lambda$ . Se sabe que ocurrieron 5 robos de este tipo en el mes de abril. Si se cuenta únicamente con esta información, de acuerdo al Teorema 3.2.5, no es posible hacer afirmación alguna acerca de los tiempos exactos de ocurrencia de dichos siniestros.

El Teorema 3.2.5 sugiere (Sección 3.6) algoritmos para la simulación del PPE.

### 3.2.3 Ley de los Grandes Números

El siguiente resultado es de gran importancia y es conocido como Ley de los Grandes Números para el Proceso Poisson.

**Teorema 3.2.7** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con intensidad  $\lambda$ , entonces el siguiente límite se satisface con probabilidad 1

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \lambda.$$

**Demostración.** La variable aleatoria  $N((0, t])$  tiene distribución  $\text{Poi}(\lambda t)$ , entonces

$$E[N_t] = \lambda t \quad \text{var}[N_t] = \lambda t.$$

De la desigualdad de Chebyshev se sigue que, para cualquier  $\varepsilon > 0$ ,

$$\Pr \left\{ \left| \frac{N_t}{t} - \lambda \right| \geq \varepsilon \right\} \leq \frac{\lambda}{\varepsilon^2 t}.$$

Tomando  $t_k = k^2$  en la expresión anterior, se tiene que

$$\sum_{k=1}^{\infty} \Pr \left\{ \left| \frac{N_{k^2}}{k^2} - \lambda \right| \geq \varepsilon \right\} < \infty,$$

por lo tanto, gracias al lema de Borel-Cantelli [ver Brémaud], la siguiente desigualdad se cumple con probabilidad 1:

$$\left| \frac{N_{k^2}}{k^2} - \lambda \right| \geq \varepsilon,$$

para a lo más un número finito de valores enteros de  $k$ . Como  $\varepsilon > 0$  es una cantidad arbitraria, esto muestra que

$$\lim_{k \rightarrow \infty} \frac{N_{k^2}}{k^2} = \lambda,$$

con probabilidad 1.

Para concluir la demostración, se toma a  $k$  como la parte entera de  $t^{\frac{1}{2}}$ , entonces para  $t > 1$ ,

$$N_{k^2} \leq N_t \leq N_{(k+1)^2}$$

y

$$k^2 \leq t < (k+1)^2.$$

A partir de que

$$\frac{(k+1)^2}{k^2} \rightarrow 1,$$

se sigue lo que se pretendía demostrar. ■

### 3.3 Generalizaciones del Proceso Poisson

#### 3.3.1 El Proceso Poisson no Estacionario

Gracias al trabajo realizado en las secciones anteriores, en este momento se dispone de varios resultados que permiten emplear al Proceso Poisson Estacionario como herramienta para solucionar cuestiones prácticas. Si bien una gran cantidad de fenómenos se ajustan a las condiciones requeridas en la Sección 3.1, el pedir que la distribución de las variables  $N_{t+s} - N_t$  ( $t, s > 0$ ) dependa únicamente de  $s$  es demasiado restrictivo en muchos casos. Por lo cual resulta natural ofrecer una versión más general del PPE.

Estas son algunas situaciones prácticas en las que se hace evidente la necesidad de desarrollar la teoría del Proceso Poisson en este sentido.

**Ejemplo 3.3.1** Una computadora de cierta universidad recibe mensajes de correo electrónico desde fuera del plantel. Se trata de un servidor que distribuye dichos mensajes a distintas computadoras y estaciones de trabajo dentro de la universidad. Se ha observado que en días hábiles la intensidad del tráfico para este servidor se incrementa gradualmente desde las 5 a.m. hasta las 8 p.m. y que de las 8 p.m. a la 1 a.m. aumenta a una tasa aún mayor, para luego disminuir drásticamente.

**Ejemplo 3.3.2** El Proceso Poisson Estacionario es un modelo muy utilizado para fenómenos relacionados con decaimiento radioactivo. Sin embargo, se ha comprobado que en la emisión de rayos gamma y en el proceso de aniquilación de protones, la frecuencia con la cual las partículas son emitidas, disminuye con el tiempo.

Del mismo modo, todos los problemas propuestos en las secciones precedentes pueden presentarse con variantes de este estilo.

Una forma de remover la condición de estacionariedad es tomar como punto de partida la Definición 3.1.4. Entonces se propone la siguiente extensión del concepto de PPE.

**Definición 3.3.3** Sea  $\{N_t : t \geq 0\}$  un proceso de contar tal que:

- i)  $N_0 = 0$ .
- ii) El proceso posee la Propiedad de Incrementos Independientes.
- iii) El número de eventos en cualquier intervalo  $(t, t + \Delta t]$  es una variable aleatoria que se distribuye Poisson con parámetro  $\int_t^{t+\Delta t} \lambda(s) ds$  para alguna función  $\lambda$  Riemann integrable y no negativa. Es decir, para cualesquiera  $t, \Delta t \geq 0$ ,

$$\Pr \{N_{t+\Delta t} - N_t = n\} = e^{-\int_t^{t+\Delta t} \lambda(s) ds} \frac{\left(\int_t^{t+\Delta t} \lambda(s) ds\right)^n}{n!} \quad n = 0, 1, \dots$$

Se dice entonces que  $\{N_t : t \geq 0\}$  es un Proceso Poisson con función de intensidad  $\lambda$ .

**Ejemplo 3.3.4** En el Ejemplo 3.3.2 se mencionó el fenómeno de emisión de rayos gamma. Es común suponer que [ver Snyder], esto puede modelarse mediante un proceso Poisson no estacionario con función de intensidad de la forma

$$\lambda(t) = \alpha_1 \exp(-\alpha_2 t) \quad t \geq 0,$$



donde  $x_1 > 0$  depende de la cantidad de material en la fuente radioactiva y  $x_2 > 0$  es el recíproco de la vida media de la fuente.

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda$ , se supondrá que dicha función es positiva salvo un conjunto de medida cero, y sea

$$\rho(t) = E[N_t].$$

Para  $t_2 > t_1 > 0$ , se tiene que

$$N_{t_2} = N_{t_1} + (N_{t_2} - N_{t_1}),$$

y

$$E[N_{t_2}] = E[N_{t_1}] + E[N_{t_2} - N_{t_1}]$$

Por la condición iii) de la Definición 3.3.3

$$E[N_{t_2}] = E[N_{t_1}] + \left[ \int_{t_1}^{t_2} \lambda(s) ds \right].$$

Debido a que  $\lambda > 0$  casi donde sea, se sigue que  $E[N_{t_2}] > E[N_{t_1}]$ , de donde se tiene que  $\rho$  es una función estrictamente creciente.

**Teorema 3.3.5** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson definido en  $(\Omega, \mathfrak{F}, \mathbf{P})$  con función de intensidad  $\lambda$ , la cual es positiva salvo en un conjunto de medida cero. Se define

$$M_t(\omega) = N_{\tau(t)}(\omega)$$

para toda  $t \geq 0$  y  $\omega \in \Omega$ , donde  $\tau$  es la función inversa de  $\rho$ . Entonces  $\{M_t : t \geq 0\}$  es un PPE con intensidad 1.

**Demostración.** El primer paso es mostrar que para toda  $t > 0$ ,  $M_t$  es una variable aleatoria con distribución **Poi**( $t$ ). Esto es una consecuencia del hecho que  $\rho$  es una función estrictamente creciente, gracias a esto el evento  $\{M_t = k\}$  es equivalente a  $\{N_{\tau(t)} = k\}$  para toda  $k \in \mathbf{N}$ . De aquí se sigue que  $M_t$  se comporta de acuerdo a una distribución Poisson con parámetro  $E[N_{\tau(t)}] = \rho(\tau(t)) = t$ .

Sean  $0 \leq s_1 < s_2 < t_1 < t_2$ , y se consideren las variables aleatorias  $M_{t_2} - M_{t_1}$  y  $M_{s_2} - M_{s_1}$ . Por definición, se tiene que

$$\begin{aligned} M_{t_2} - M_{t_1} &= N_{\tau(t_2)} - N_{\tau(t_1)} \\ M_{s_2} - M_{s_1} &= N_{\tau(s_2)} - N_{\tau(s_1)}. \end{aligned}$$

Se sigue de la monotonía estricta de  $\rho$  que  $(\tau(t_1), \tau(t_2))$  y  $(\tau(s_1), \tau(s_2))$  son ajenos. Se puede concluir que las variables  $M_{t_2} - M_{t_1}$  y  $M_{s_2} - M_{s_1}$  son independientes. Esto también es válido para uniones finitas arbitrarias de intervalos ajenos. ■

**Corolario 3.3.6** Sea  $\rho$  una función positiva y estrictamente creciente. Entonces  $T_1, T_2, \dots$  son los tiempos de ocurrencias de un Proceso Poisson no estacionario con  $E[N_t] = \rho(t)$  si y sólo si  $\rho(T_1), \rho(T_2), \dots$  son los tiempos de ocurrencias de un Proceso Poisson estacionario con intensidad 1.

**Demostración.** Una implicación es obvia a partir del Teorema 3.3.5. Mediante un argumento totalmente análogo se puede obtener un Proceso Poisson  $\{Q_t : t \geq 0\}$  con función de intensidad  $\lambda$ , a partir de un PPE  $\{N_t : t \geq 0\}$  con intensidad 1. Esto de la siguiente forma:

$$Q_t(\omega) = N_{\rho(t)}(\omega),$$

donde

$$\rho(t) = E[Q_t]. \quad \blacksquare$$

**Teorema 3.3.7** Sean  $S_1, S_2, \dots$  los tiempos de ocurrencias de un Proceso Poisson no estacionario  $\{Q_t : t \geq 0\}$  cuya función  $\rho$  sea estrictamente creciente. Entonces para toda  $n$ ,

$$\Pr\{S_{n+1} - S_n > t \mid S_1, \dots, S_n\} = \exp[-(\rho(S_n + t) - \rho(S_n))].$$

**Demostración.** Se tiene que

$$\begin{aligned} & \Pr\{S_{n+1} - S_n > t \mid S_1, \dots, S_n\} \\ &= \Pr\{Q_{S_n+t} - Q_{S_n} = 0 \mid Q_{r_i} = i - 1, S_{i-1} \leq r_i < S_i, i \in \{1, 2, \dots, n\}\}, \end{aligned}$$

donde  $S_0 = 0$ . De la Propiedad de Incrementos Independientes y del corolario anterior se sigue:

$$\begin{aligned} & \Pr\{S_{n+1} - S_n > t \mid S_1, \dots, S_n\} \\ &= \Pr\{Q_{S_n+t} - Q_{S_n} = 0\} \\ &= \Pr\{N_{\rho(S_n+t)} - N_{\rho(S_n)} = 0\} \\ &= \exp[-(\rho(S_n + t) - \rho(S_n))], \end{aligned}$$

donde  $\{N_t : t \geq 0\}$  es un PPE con intensidad 1.  $\blacksquare$

**Corolario 3.3.8** La densidad conjunta de las variables aleatorias

$$S_1, S_2 - S_1, \dots, S_n - S_{n-1}$$

es dada por

$$f_{S_1, S_2 - S_1, \dots, S_n - S_{n-1}}(x_1, \dots, x_n) = \rho'(x_1) \cdots \rho' \left( \sum_{i=1}^n x_i \right) \exp \left( -\rho \left( \sum_{i=1}^n x_i \right) \right).$$

**Demostración.** Se tiene la siguiente equivalencia de eventos

$$\begin{aligned}
 & \{S_1 = x_1, S_2 - S_1 = x_2, \dots, S_n - S_{n-1} = x_n\} \\
 &= \{Q_{r_i} = i - 1, S_{i-1} \leq r_i < S_i, i \in \{1, 2, \dots, n\}, Q_{S_n} = n\} \\
 &= \{N_{\rho(r_i)} = i - 1, S_{i-1} \leq r_i < S_i, i \in \{1, 2, \dots, n\}, N_{\rho(S_n)} = n\} \\
 &= \{N_{\rho(t_i)} - N_{\rho(S_{i-1})} = 0, S_{i-1} \leq t_i < S_i, \\
 & \quad N_{\rho(S_i)} - N_{\rho(S_{i-1})} = 1, i \in \{1, 2, \dots, n\},
 \end{aligned}$$

donde  $S_0 = 0$ . Entonces

$$\begin{aligned}
 & f_{S_1, S_2 - S_1, \dots, S_n - S_{n-1}}(x_1, \dots, x_n) \\
 &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon^n} \prod_{i=1}^n \Pr \{N_{\rho(S_i) - \epsilon} - N_{\rho(S_{i-1})} = 0\} \Pr \{N_{\rho(S_i)} - N_{\rho(S_i) - \epsilon} = 1\} \\
 &= \prod_{i=1}^n \Pr \{N_{\rho(S_i)} - N_{\rho(S_{i-1})} = 0\} \rho'(S_i) \tag{3.1}
 \end{aligned}$$

$$= \prod_{i=1}^n \Pr \{N_{\rho(S_i)} - N_{\rho(S_{i-1})} = 0\} \lambda(S_i), \tag{3.2}$$

donde  $\sum_{j=1}^i x_j = s_i$ . De esta última expresión se sigue el resultado. ■

**Corolario 3.3.9** Sea  $\{Q_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda$ , y sea  $r > 0$ . La densidad conjunta de  $Q_r$ , y los tiempos de ocurrencias en  $[0, r]$ , es de la forma

$$\begin{aligned}
 & f_{Q_r, S_1, S_2, \dots, S_n}(n, s_1, s_2, \dots, s_n) \\
 &= \left( \prod_{i=1}^n \lambda(s_i) \right) \exp \left[ - \int_0^r \lambda(u) du \right].
 \end{aligned}$$

**Demostración.** Se hará uso de la siguiente propiedad [ver Ash]

$$\begin{aligned}
 & f_{Q_r, S_1, S_2, \dots, S_n}(n, s_1, s_2, \dots, s_n) \\
 &= f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) \\
 & \quad \cdot \Pr \{Q_r = n \mid S_1 = s_1, S_2 = s_2, \dots, S_n = s_n\}.
 \end{aligned}$$

Para obtener la densidad conjunta de los tiempos de ocurrencia, se aplican técnicas de cambio de variable a la Expresión (3.1), de donde se sigue

$$f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) = \prod_{i=1}^n \Pr \{N_{s_i} - N_{s_{i-1}} = 0\} \lambda(s_i)$$

Ahora bien, para que  $\{Q_r = n\}$  dado el evento  $\{S_1 = s_1, S_2 = s_2, \dots, S_n = s_n\}$ , debe ocurrir que tras el tiempo  $s_n$  no se registre ocurrencia alguna en  $(0, r]$ . Entonces

$$\begin{aligned}
 & \Pr \{Q_r = n \mid S_1 = s_1, S_2 = s_2, \dots, S_n = s_n\} \\
 &= \Pr \{Q_r - Q_{s_n} = 0\}.
 \end{aligned}$$

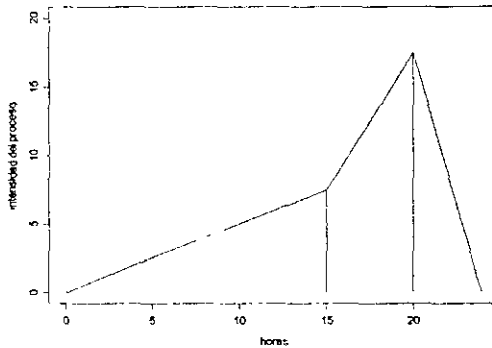


FIGURE 3.2.

Se puede concluir

$$f_{Q_r, s_1, s_2, \dots, s_n}(n, s_1, s_2, \dots, s_n) = \left( \prod_{i=1}^n \lambda(s_i) \right) \exp \left[ - \int_0^r \lambda(u) du \right]. \quad \blacksquare \quad (3.3)$$

**Ejemplo 3.3.10** Retomando el Ejemplo 3.3.1, y suponiendo que la función de intensidad es lineal por tramos, de modo que

$$\lambda(t) = \begin{cases} .5t & \text{si } t \in (0, 15] \\ 2t - 22.5 & \text{si } t \in (15, 20] \\ 105 - 4.375t & \text{si } t \in (20, 24]. \end{cases}$$

El tiempo se mide en horas, tomando como inicio las 5 a.m.

En la Figura 3.2 se presenta un gráfico de la función de intensidad  $\lambda$ .

Bajo estas condiciones, la probabilidad de recibir 15 mensajes entre las 7 y las 9:30 a.m. viene dada por

$$\Pr \{ N_{16.5} - N_{14} = 15 \} = \exp \left[ - \int_{14}^{16.5} \lambda(u) du \right] \frac{\left( \int_{14}^{16.5} \lambda(u) du \right)^{15}}{15!},$$

donde

$$\begin{aligned} \int_{14}^{16.5} \lambda(u) du &= \int_{14}^{15} .5u du + \int_{15}^{16.5} 2u - 22.5 du \\ &= 13.5. \end{aligned}$$

Entonces

$$\Pr \{N_{16.5} - N_{14} = 15\} = .0945.$$

Ahora se aplicará el Corolario 3.3.9 para obtener la densidad conjunta de  $N, S_1, \dots, S_N$  en el mismo intervalo. En este caso

$$\begin{aligned} & f_{N_{(14, 16.5]}, S_1, S_2, \dots, S_n} (n, s_1, s_2, \dots, s_n) \\ &= \left( \prod_{i=1}^n (.5s_i) 1_{(0,15]}(s_i) + (2s_i - 22.5) 1_{(15,20]}(s_i) \right) \\ & \cdot \exp \left[ - \int_{14}^{16.5} \lambda(u) du \right], \end{aligned}$$

donde  $1_{(0,15]}$  y  $1_{(15,20]}$  representan las funciones indicadoras para los intervalos correspondientes. ■

### 3.3.2 El Proceso Poisson Compuesto

El Proceso Poisson es una herramienta poderosa para modelar una gran variedad de fenómenos, pues brinda información acerca de los tiempos de ocurrencia de un suceso de interés y cuenta el número de ocurrencias para todo intervalo. En todos los modelos propuestos hasta este momento se trató a las ocurrencias como hechos indistinguibles entre sí (salvo en su posición). Sin embargo el investigador puede verse en la necesidad de ‘etiquetar’ con cierta característica extra a cada ocurrencia del evento. Si dichas etiquetas son representadas mediante vectores aleatorios en  $\mathbf{R}^d$ , entonces reciben el nombre de *marcas*.

Sea  $\mathcal{U}$  el conjunto conformado por todos los valores permisibles para las marcas del proceso en cuestión, dicha  $\mathcal{U}$  puede ser:

i) Una colección numerable  $\{V_1, V_2, \dots\}$  donde cada  $V_k$  es un vector perteneciente a  $\mathbf{R}^d$ . Un ejemplo típico es la colección  $\{1, 2, \dots\}$  de enteros positivos. La marca  $n$  en el  $i$ -ésimo punto puede indicar que  $n$  eventos ocurrieron de manera simultánea en el instante  $s_i$ . De esta forma ciertos procesos que no cumplen la Ley de Eventos Raros pueden verse como un Proceso Poisson con  $\mathcal{U} = \{1, 2, \dots\}$ .

ii) Una región de un espacio euclidiano finito-dimensional. Un ejemplo es el conjunto  $\mathbf{R}^+$  usado para representar las intensidades de descargas eléctricas que recibe un sistema a lo largo del tiempo.

Estas ideas sugieren dos posibles extensiones del Proceso Poisson.

**Definición 3.3.11** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con intensidad  $\lambda$ , y sea  $\{Y_k : k \in \mathbf{N}\}$  una sucesión de variables aleatorias independientes idénticamente distribuidas con distribución  $G$ , e independientes del Proceso Poisson. Entonces, el Proceso Poisson Compuesto de intensidad  $\lambda$  y con marcas  $\{Y_k : k \in \mathbf{N}\}$ , es el proceso  $\{Z_t : t \geq 0\}$ , donde

$$Z_t = \sum_{k=1}^{N_t} Y_k.$$

El Proceso Poisson Marcado de intensidad  $\lambda$  y con marcas  $\{Y_k : k \in \mathbf{N}\}$  es la sucesión de parejas  $(S_1, Y_1), (S_2, Y_2), \dots$  donde  $\{S_k : k \in \mathbf{N}\}$  son los tiempos en los que tienen lugar las ocurrencias del proceso  $\{N_t : t \geq 0\}$ .

Sea  $\mathcal{U}$  el espacio de estados del proceso  $\{Y_k : k \in \mathbf{N}\}$ . Se dice entonces que  $\mathcal{U}$  es el espacio de marcas del proceso  $\{Z_t : t \geq 0\}$ .

En seguida se presentan algunas aplicaciones de este modelo.

**Ejemplo 3.3.12** Se supone que el número de reclamaciones en una compañía de seguros a lo largo del tiempo se comporta como un Proceso Poisson  $\{N_t : t \geq 0\}$  de intensidad  $\lambda$ . Sea  $Y_k$  el monto de la  $k$ -ésima reclamación. Entonces  $Z_t = \sum_{k=1}^{N_t} Y_k$  representa el monto total de las reclamaciones hechas hasta tiempo  $t$ .

**Ejemplo 3.3.13** Sea  $N_t$  el número de impactos que recibe un sistema desde un tiempo inicial 0 hasta un tiempo  $t$ , y sea  $Y_k$  el daño producido al sistema por el  $k$ -ésimo impacto. Se da por hecho que el daño se mide en magnitudes positivas, esto es  $\Pr\{Y_k \geq 0\} = 1$ , y el daño se acumula aditivamente, de modo que  $Z_t = \sum_{k=1}^{N_t} Y_k$  representa el daño hecho al sistema hasta el tiempo  $t$ . Se supone además que el sistema se mantiene en operación siempre y cuando el daño acumulado no sobrepase cierto nivel  $\eta$ , y falla en caso contrario. Sea  $T$  el tiempo de falla del sistema. Es claro que

$$\{T > t\} \quad \text{si sólo si} \quad \{Z_t < \eta\}.$$

Se tiene que

$$\begin{aligned} \Pr\{Z_t \leq z\} &= \Pr\left\{\sum_{k=1}^{N_t} Y_k \leq z\right\} \\ &= \sum_{n=0}^{\infty} \Pr\left\{\sum_{k=1}^{N_t} Y_k \leq z \mid N_t = n\right\} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} G^{(n)}(z), \end{aligned}$$

donde  $G^{(n)}(y) = \Pr\{Y_1 + Y_2 + \dots + Y_n \leq y\}$ , esto gracias a que  $N_t$  es independiente de  $\{Y_k : k \in \mathbf{N}\}$ . Entonces

$$\Pr\{T > t\} = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} G^{(n)}(\eta).$$

Todos los sumandos son no negativos, de forma que es posible intercambiar la suma con la integral para obtener el tiempo medio de falla del sistema

$$\begin{aligned} E[T] &= \int_0^{\infty} \Pr\{T > t\} dt \\ &= \sum_{n=0}^{\infty} \left( \int_0^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} t dt \right) G^{(n)}(\eta) \\ &= \sum_{n=0}^{\infty} \left( \frac{\Gamma(n+1)}{\lambda n!} \right) G^{(n)}(\eta) \\ &= \lambda^{-1} \sum_{n=0}^{\infty} G^{(n)}(\eta). \end{aligned}$$

Esta expresión se simplifica de gran manera en el caso especial en el que las variables aleatorias  $\{Y_k : k \in \mathbf{N}\}$  se distribuyen  $\text{Exp}(\mu)$ . Entonces la suma  $Y_1 + Y_2 + \dots + Y_n$  tiene distribución Gamma, ya que

$$G^{(n)}(\eta) = 1 - \sum_{k=0}^{n-1} \frac{(\mu\eta)^k e^{-\mu\eta}}{k!} = \sum_{k=n}^{\infty} \frac{(\mu\eta)^k e^{-\mu\eta}}{k!},$$

entonces

$$\begin{aligned} \sum_{n=0}^{\infty} G^{(n)}(\eta) &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \frac{(\mu\eta)^k e^{-\mu\eta}}{k!} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\mu\eta)^k e^{-\mu\eta}}{k!} \\ &= \sum_{k=0}^{\infty} (1+k) \frac{(\mu\eta)^k e^{-\mu\eta}}{k!} \\ &= 1 + \mu\eta. \end{aligned}$$

De modo que, cuando  $\{Y_k : k \in \mathbf{N}\}$  se distribuyen exponencialmente, se tiene que

$$E[T] = \frac{1 + \mu\eta}{\lambda}. \quad \blacksquare$$

Ahora se obtendrá la media y varianza de  $Z_t$ . Para calcular  $E[Z_t]$ , primero se condiciona con  $N_t$ , de donde se tiene

$$E[Z_t] = E(E\{Z_t | N_t\}).$$

Ahora

$$\begin{aligned} E[Z_t | N_t = n] &= E\left[\sum_{i=1}^{N_t} Y_i \mid N_t = n\right] \\ &= E\left[\sum_{i=1}^n Y_i \mid N_t = n\right] \\ &= E\left[\sum_{i=1}^n Y_i\right] \\ &= nE[Y_1], \end{aligned}$$

donde se ha usado el hecho de que  $N_t$  es independiente de  $\{Y_k : k \in \mathbf{N}\}$ . Entonces

$$E\{Z_t | N_t\} = N_t E\{Y_1\},$$

por lo tanto

$$E[Z_t] = \lambda t E\{Y_1\}. \quad (3.4)$$

Para calcular  $\text{var} [Z_t]$  se hará uso de la fórmula de varianza condicional [ver Mood, Graybill, Boes]

$$\text{var} [Z_t] = E [\text{var} [Z_t | N_t]] + \text{var} [E [Z_t | N_t]], \quad (3.5)$$

entonces

$$\begin{aligned} \text{var} [Z_t | N_t = n] &= \text{var} \left[ \sum_{i=1}^{N_t} Y_i | N_t = n \right] \\ &= \text{var} \left[ \sum_{i=1}^n Y_i \right] \\ &= n \text{var} (Y_1). \end{aligned}$$

Luego

$$\text{var} [Z_t | N_t] = N_t \text{var} (Y_1). \quad (3.6)$$

Sustituyendo las Expresiones (3.4) y (3.6) en la Expresión (3.5) se obtiene

$$\begin{aligned} \text{var} [Z_t] &= E [N_t \text{var} [Y_1]] + \text{var} [N_t E [Y_1]] \\ &= \lambda t [\text{var} [Y_1] + (E [Y_1])^2] \\ &= \lambda t E [Y_1^2], \end{aligned}$$

recordando que  $\text{var} [N_t] = \lambda t$ .

**Ejemplo 3.3.14** Se acepta el supuesto que el número de familias que emigran a cierta área a lo largo del tiempo, se comporta de acuerdo a un Proceso Poisson estacionario con intensidad  $\lambda = 2$  (el tiempo se mide en semanas). Se considera que el número de integrantes en cada familia es una variable aleatoria que puede tomar los valores 1, 2, 3, 4 con probabilidad  $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$  respectivamente, entonces, se desea conocer el valor medio y la varianza del número total de individuos que emigran a esta área durante un periodo de 5 semanas.

Sea  $Y_i$  el número de personas en la  $i$ -ésima familia, se tiene entonces

$$E [Y_i] = \frac{5}{2} \quad \text{y} \quad E [Y_i^2] = \frac{43}{6},$$

luego

$$E [Z_5] = 25 \quad \text{y} \quad \text{var} [Z_5] = \frac{215}{3}. \quad \blacksquare$$

**Teorema 3.3.15** Sea  $\{Z_t : t \geq 0\}$  un Proceso Poisson Compuesto con marcas  $\{Y_1, Y_2, \dots\}$ . Dicho proceso posee la Propiedad de Incrementos Independientes.



**Demostración.** Sean  $0 \leq a_1 < b_1 \leq a_2 < b_2 < \infty$ , entonces

$$\begin{aligned} & \Pr \{ Z_{b_1} - Z_{a_1} \leq z_1, Z_{b_2} - Z_{a_2} \leq z_2 \} \\ &= \Pr \left\{ \sum_{j=N_{a_1}+1}^{N_{b_1}} Y_j \leq z_1, \sum_{j=N_{a_2}+1}^{N_{b_2}} Y_j \leq z_2 \right\} \\ &= \sum_{h_1, g_1, h_2, g_2 \in \mathbf{N}} \Pr \left\{ \sum_{j=N_{a_1}+1}^{N_{b_1}} Y_j \leq z_1, \sum_{j=N_{a_2}+1}^{N_{b_2}} Y_j \leq z_2 \mid A(h_1, g_1, h_2, g_2) \right\} \\ & \quad \cdot \Pr \{ A(h_1, g_1, h_2, g_2) \}, \end{aligned}$$

$$\begin{aligned} \text{donde } A(h_1, g_1, h_2, g_2) &= \{ N_{a_1} = h_1, N_{b_1} = g_1, N_{a_2} = h_2, N_{b_2} = g_2 \} \text{ y} \\ & 0 \leq h_1 \leq g_1 \leq h_2 \leq g_2 \quad h_1, g_1, h_2, g_2 \in \mathbf{N}. \end{aligned}$$

Del hecho que las variables aleatorias  $\{Y_1, Y_2, \dots\}$  son independientes entre sí, y del proceso  $\{N_t : t \geq 0\}$ , se sigue que

$$\begin{aligned} & \Pr \{ Z_{b_1} - Z_{a_1} \leq z_1, Z_{b_2} - Z_{a_2} \leq z_2 \} \\ &= \sum_{h_1, g_1, h_2, g_2 \in \mathbf{N}} \left( \prod_{r=1}^2 \Pr \left\{ \sum_{j=h_r+1}^{g_r} Y_j \leq z_r \right\} \right) \cdot \Pr \{ A(h_1, g_1, h_2, g_2) \}. \end{aligned}$$

Esto último puede expresarse de la siguiente manera

$$\begin{aligned} & \sum_{h_1, g_1, h_2, g_2 \in \mathbf{N}} \left( \prod_{r=1}^2 \Pr \left\{ \sum_{j=h_r+1}^{g_r} Y_j \leq z_r \right\} \right) \cdot \Pr \{ A(h_1, g_1, h_2, g_2) \} \\ &= \sum_{l_0, l_1, l_2, l_3 \in \mathbf{N}} \left( \prod_{r=1}^2 \Pr \left\{ \sum_{j=1}^{l_r-1} Y_j \leq z_r \right\} \right) \cdot \Pr \{ B(l_0, l_1, l_2, l_3) \}, \end{aligned}$$

donde

$$B(l_0, l_1, l_2, l_3) = \{ N_{a_1} = l_0, N_{b_1} - N_{a_1} = l_1, N_{a_2} - N_{b_1} = l_2, N_{b_2} - N_{a_2} = l_3 \}.$$

Es claro que

$$A(h_1, g_1, h_2, g_2) = B(h_1, g_1 - h_1, h_2 - g_1, g_2 - h_2)$$

Se aprovecha la Propiedad de Incrementos Independientes para  $\{N_t : t \geq 0\}$

$$\begin{aligned} & \sum_{l_0, l_1, l_2, l_3 \in \mathbf{N}} \left( \prod_{r \in \{1,3\}} \Pr \left\{ \sum_{j=1}^{l_r} Y_j \leq w_r \right\} \right) \cdot \Pr \{ B(l_0, l_1, l_2, l_3) \} \\ &= \sum_{l_0, l_1, l_2, l_3 \in \mathbf{N}} \left( \prod_{k \in \{0,2\}} \Pr \{ L_k = l_k \} \right) \left( \prod_{r \in \{1,3\}} \Pr \left\{ \sum_{j=1}^{l_r-1} Y_j \leq w_r \right\} \Pr \{ L_r = l_r \} \right) \\ &= \sum_{l_1, l_3 \in \mathbf{N}} \left( \prod_{r \in \{1,3\}} \Pr \left\{ \sum_{j=1}^{l_r-1} Y_j \leq w_r \right\} \Pr \{ L_r = l_r \} \right), \end{aligned}$$

donde  $L_0 = N_{a_1}, \dots, L_3 = N_{b_2} - N_{a_2}$ , y  $w_1 = z_1, w_3 = z_2$ . Entonces

$$\begin{aligned} & \sum_{l_0, l_1, l_2, l_3 \in \mathbf{N}} \left( \prod_{r \in \{1,3\}} \Pr \left\{ \sum_{j=1}^{l_r} Y_j \leq w_r \right\} \right) \cdot \Pr \{B(l_0, l_1, l_2, l_3)\} \\ &= \prod_{r \in \{1,3\}} \sum_{l_r \in \mathbf{N}} \Pr \left\{ \sum_{j=1}^{l_r} Y_j \leq w_r \right\} \Pr \{L_r = l_r\} \\ &= \prod_{k \in \{1,2\}} \sum_{i \in \mathbf{N}} \Pr \left\{ \sum_{j=0}^{N_{b_k} - N_{a_k} = i} Y_j \leq z_k \mid N_{b_k} - N_{a_k} = i \right\} \Pr \{N_{b_k} - N_{a_k} = i\} \\ &= \prod_{k \in \{1,2\}} \sum_{i \in \mathbf{N}} \Pr \left\{ \sum_{j=N_{a_k}+1}^{N_{b_k}} Y_j \leq z_k \mid N_{b_k} - N_{a_k} = i \right\} \Pr \{N_{b_k} - N_{a_k} = i\} \end{aligned}$$

con  $Y_0 = 0$ . Se puede concluir

$$\begin{aligned} & \Pr \{Z_{b_1} - Z_{a_1} \leq z_1, Z_{b_2} - Z_{a_2} \leq z_2\} \\ &= \prod_{k \in \{1,2\}} \Pr \left\{ \sum_{j=N_{a_k}+1}^{N_{b_k}} Y_j \leq z_k \right\}. \end{aligned}$$

El resultado se obtiene de generalizar estos cálculos para el caso  $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k < \infty$ , con  $k \in \mathbf{N}$  ■

Sea  $\{(S_k, Y_k) : k \in \mathbf{N}\}$  un Proceso Poisson Marcado de intensidad  $\lambda$  y sea  $\{Z_t : t \geq 0\}$  el Proceso Poisson Compuesto que se obtiene a partir de éste. Se supondrá que para  $p \in (0, 1)$  fija se tiene

$$\Pr \{Y_k = 1\} = p, \quad \Pr \{Y_k = 0\} = 1 - p.$$

Ahora se consideran los procesos formados por las ocurrencias marcadas con unos y por las ocurrencias marcadas con ceros. Estos procesos pueden definirse también como

$$Z_t^1 = \sum_{k=1}^{N_t} Y_k \quad \text{y} \quad Z_t^0 = \max_{k \in \mathbf{N}} \{S_k \leq t\} - Z_t^1.$$

$Z_t^1$  es un Proceso Poisson compuesto, y por lo tanto posee la Propiedad de Incrementos Independientes. Del Teorema 2.1.4 se sigue que  $Z_t^1$  se distribuye  $\text{Poi}(\lambda p t)$ . Esto implica que  $\{Z_t^1 : t \geq 0\}$  es un Proceso Poisson de intensidad  $\lambda p$ . Mediante un argumento totalmente análogo, se tiene que  $\{Z_t^0 : t \geq 0\}$  es un Proceso Poisson de intensidad  $\lambda(1 - p)$ . De hecho, estos dos procesos son independientes.

Se verificará que  $\Pr \{Z_t^0 = j, Z_t^1 = k\} \approx \Pr \{Z_t^0 = j\} \Pr \{Z_t^1 = k\}$ , para  $j, k \in \{1, 2, \dots\}$ .

Esta independencia de eventos se establece de la siguiente forma

$$\begin{aligned}
 \Pr \{Z_i^0 = j, Z_i^1 = k\} &= \Pr \{Z_i = j + k, Z_i^1 = k\} \\
 &= \Pr \{Z_i^1 = k \mid Z_i = j + k\} \Pr \{Z_i = j + k\} \\
 &= \frac{(j+k)!}{j!k!} p^k (1-p)^j \frac{(\lambda t)^{j+k} e^{-\lambda t}}{(j+k)!} \\
 &= \left[ \frac{e^{-\lambda p t} (\lambda p t)^k}{k!} \right] \left[ \frac{e^{-\lambda(1-p)t} (\lambda(1-p)t)^j}{j!} \right] \\
 &= \Pr \{Z_i^1 = k\} \Pr \{Z_i^0 = j\}.
 \end{aligned}$$

Esta idea se generaliza en la Sección 3.5.

### 3.3.3 El Proceso Poisson en el Plano

Un Proceso Poisson en el plano es un modelo probabilista para puntos dispuestos en el plano al azar. Queda descrito mediante una colección de variables aleatorias  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^2)\}$  indexada por los conjuntos Borelianos de  $\mathbb{R}^2$  y definida en cierto espacio de probabilidad  $(\Omega, \mathfrak{F}, \mathbf{P})$ , donde  $N_A$  es el número (aleatorio) de puntos (aleatorios) en  $A$ . Se supondrá que si  $A$  es acotado, es decir, si está contenido en un rectángulo de área finita, entonces  $N_A(\omega)$  es finito  $\mathbf{P}$ -casi seguramente. Además no se permite superposición de puntos, esto es, para todo  $x \in \mathbb{R}^2$ ,  $N_{\{x\}} = 0$  ó  $1$ ,  $\mathbf{P}$ -casi seguramente.

Se dice entonces que  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^2)\}$  es un *proceso puntual simple*.

**Definición 3.3.16** *Un Proceso Poisson Homogéneo en  $\mathbb{R}^2$  con intensidad  $\lambda > 0$  es un proceso puntual simple que cumple:*

- i) *Para cualquier colección finita  $A_1, A_2, \dots, A_n$  de Borelianos ajenos por parejas de  $\mathbb{R}^2$ , las variables aleatorias  $N_{A_1}, \dots, N_{A_n}$  son independientes*
- ii) *Para cualquier  $A$  boreliano acotado de  $\mathbb{R}^2$ ,*

$$\Pr \{N_A = k\} = e^{-\lambda|A|} \frac{(\lambda|A|)^k}{k!}$$

donde  $k \in \mathbb{N}$ , y  $|A|$  es la medida de Lebesgue de  $A$ .

La siguiente construcción ilustra varias analogías entre el modelo propuesto en la Definición 3.3.16 y el PPE. Entre dichas analogías se encuentra el hecho de que la distribución de las ocurrencias en un conjunto dado, condicionada al número de ocurrencias en dicho conjunto, es uniforme. Se tiene además una construcción dentro de la cual se aproxima a la Distribución Poisson mediante otra distribución discreta, donde esta distribución toma un número finito de valores (Resulta instructivo comparar esto con la discusión que da inicio a la Sección 3.1).

**Teorema 3.3.17** *Sea  $\Pi$  un conjunto numerable de puntos en el plano, sin puntos de acumulación finitos. Dichos puntos están distribuidos de manera aleatoria, de modo que, para todo boreliano acotado  $H$  de  $\mathbb{R}^2$ , cada punto en  $H \cap \Pi$  se distribuye de manera uniforme en  $H$ . Se requiere además que los puntos en  $H \cap \Pi$  sean independientes entre sí. Entonces, las variables aleatorias*

$$N_A = \# \{A \cap \Pi\}, \quad A \in \mathfrak{B}(\mathbb{R}^2)$$

forman un Proceso Poisson Homogéneo en  $\mathbb{R}^2$ , siempre y cuando la distribución de  $N_A$  sea invariante a traslaciones de  $A$ .

**Demostración.** Se consideran los cuadrados  $S_D$  con centro en  $\theta$  con lados paralelos a los ejes coordenados y de longitud  $D$ . Sea  $K$ , por el momento, un entero fijo, y se toman  $K$  vectores aleatorios independientes  $(X_i, Y_i)$ ,  $i = 1, \dots, K$  de modo que el punto  $M_i = (X_i, Y_i)$  se encuentra uniformemente distribuido en  $S_D$ , esto es

$$\Pr \{(X_i, Y_i) \in A\} = \frac{|A|}{D^2} \quad \text{para todo } A \in \mathfrak{B}(\mathbb{R}^2), A \subset S_D.$$

Esto modela  $K$  puntos colocados al azar en  $S_D$  de manera independiente.

Sean  $A_1, \dots, A_n \in \mathfrak{B}(\mathbb{R}^2)$  y contenidos en  $S_D$ . Se supone además que dichos conjuntos son ajenos. La distribución conjunta de  $N_{A_1}, \dots, N_{A_n}$  se obtiene de la siguiente forma

$$\begin{aligned} & \Pr \{N_{A_1} = k_1, \dots, N_{A_n} = k_n\} \\ &= \Pr \{N_{A_0} = k_0, N_{A_1} = k_1, \dots, N_{A_n} = k_n\}, \end{aligned}$$

donde

$$A_0 = S_D - \bigcup_{j=1}^n A_j \quad \text{y} \quad k_0 = K - \sum_{j=1}^n k_j.$$

Como cada punto tiene probabilidad  $p_i = |A_i|/D^2$  de caer en  $A_i$ , se obtiene una distribución multinomial de parámetros  $(K, n)$

$$\Pr \{N_{A_1} = k_1, \dots, N_{A_n} = k_n\} = \frac{K!}{k_0! \dots k_n!} p_0^{k_0} \dots p_n^{k_n}.$$

Por lo tanto, si  $D$  y  $K$  tienden simultáneamente a  $\infty$ , de tal manera que  $K/D^2 = \lambda > 0$ , se tiene que

$$\begin{aligned} & \lim_{K, D \rightarrow \infty} \frac{K!}{k_0! \dots k_n!} p_0^{k_0} \dots p_n^{k_n} \\ &= \lim_{K, D \rightarrow \infty} \frac{K!}{k_0! \dots k_n!} K^k \left[ \left( \frac{D^2 - \sum_{j=1}^n |A_j|}{D^2} \right) K \right]^{k_0} \\ & \quad \cdot \left( \frac{|A_1| K}{D^2} \right)^{k_1} \dots \left( \frac{|A_n| K}{D^2} \right)^{k_n} \end{aligned} \quad (3.7)$$

$$\begin{aligned} &= \lim_{K, D \rightarrow \infty} \frac{K!}{k_0! K^k} \left[ \left( 1 - \frac{\sum_{j=1}^n |A_j|}{D^2} \right) K \right]^{k_0} \frac{(|A_1| \lambda)^{k_1}}{k_1!} \dots \frac{(|A_n| \lambda)^{k_n}}{k_n!} \\ &= \lim_{K, D \rightarrow \infty} \frac{K^{K-k_0}}{K^{K-k_n}} \left( 1 - \frac{\lambda \sum_{j=1}^n |A_j|}{K} \right)^{K - \sum_{j=1}^n k_j} \frac{(|A_1| \lambda)^{k_1}}{k_1!} \dots \frac{(|A_n| \lambda)^{k_n}}{k_n!} \\ &= \exp \left\{ -\lambda \sum_{j=1}^n |A_j| \right\} \frac{(|A_1| \lambda)^{k_1}}{k_1!} \dots \frac{(|A_n| \lambda)^{k_n}}{k_n!} \\ &= \prod_{j=1}^n e^{-\lambda |A_j|} \frac{(|A_j| \lambda)^{k_j}}{k_j!}. \end{aligned} \quad (3.8)$$

Es claro que la Expresión (3.8) representa la función de masa conjunta de  $N_{A_1}, \dots, N_{A_n}$  para un proceso Poisson homogéneo en el plano con intensidad  $\lambda$ . ■

Este resultado puede usarse de la siguiente manera: Si se tiene un área muy grande ( $S_D$ ) sobre la cual se ha dispuesto una gran cantidad de puntos ( $K$ ) de manera aleatoria y uniforme, y si se desea calcular la cantidad

$$\Pr \{N_{A_1} = k_1, \dots, N_{A_n} = k_n\}$$

donde  $k_1 + \dots + k_n$  es mucho menor que  $K$ , y donde  $A_1, \dots, A_n$  se encuentran alejados de la frontera de  $S_D$ , entonces se puede aplicar el Teorema 3.3.17.

### 3.3.4 Procesos Puntuales

Sea  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio de probabilidad, y se toma  $\mathbf{R}^k$ , donde  $k \in \{1, 2, \dots\}$ , con la topología usual. Se denota con  $\mathfrak{B}(\mathbf{R}^k)$  a la  $\sigma$ -álgebra de Borel para este conjunto.

Se considera la siguiente familia de subconjuntos de  $\mathbf{R}^k$ :

$$\mathcal{N}(\mathbf{R}^k) := \{C \mid C \text{ es numerable}\}.$$

Dados estos objetos, a toda función  $\Pi: \Omega \rightarrow \mathcal{N}(\mathbf{R}^k)$  se le denomina *subconjunto aleatorio* de  $\mathbf{R}^k$  definido en  $(\Omega, \mathfrak{F}, \mathbf{P})$ .

**Definición 3.3.18** Sea  $\Pi$  un subconjunto aleatorio de  $\mathbb{R}^k$  definido en  $(\Omega, \mathfrak{F}, \mathbf{P})$ , tal que las funciones

$$N_A(\cdot) = \#(A \cap \Pi(\cdot))$$

sean variables aleatorias para todo  $A \in \mathfrak{B}(\mathbb{R}^k)$ . Donde  $\#(A \cap \Pi(\cdot))$  es la cardinalidad de  $\{A \cap \Pi(\cdot)\}$  si este conjunto es finito y se denota con  $\infty$  en caso contrario. Se dice entonces que  $\Pi$  es un proceso puntual sobre  $\mathbb{R}^k$  definido en  $(\Omega, \mathfrak{F}, \mathbf{P})$ .

Tomando a la Definición 3.3.18 como punto de partida, es posible acceder a conceptos y resultados que, si bien fueron concebidos para aplicarse a espacios más generales [ver Kingman], permiten (en su versión restringida al caso  $(\mathbb{R}^k, \mathfrak{B}(\mathbb{R}^k))$ ,  $k \in \{1, 2, \dots\}$ ) demostrar con relativa simplicidad, ciertas propiedades de los procesos discutidos en las secciones anteriores, particularmente del Proceso Poisson Compuesto.

La Definición 3.3.18 considera una clase muy amplia de procesos. Es claro que no todo proceso puntual es un proceso Poisson.

**Ejemplo 3.3.19** Sea  $S$  un cubo en  $\mathbb{R}^n$  con la topología usual y  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio medible. A cada  $\omega \in \Omega$  corresponde el conjunto  $\Pi(\omega) = \{X_1(\omega), X_2(\omega), \dots, X_k(\omega)\}$  de puntos de  $S$ . Se supondrá que cada  $X_i$  es un vector aleatorio con distribución uniforme en el cubo y  $X_n$  es independiente de  $X_m$  si  $n \neq m$ . Debido a que cada  $X_i$  es un vector aleatorio, el suceso  $\{X_i \in A\}$  es un evento, para todo  $A \in \mathfrak{B}(\mathbb{R}^n)$ . Por lo tanto  $\{N_A = \tau\}$  con  $\tau \in \mathbf{N}$  también lo es. Entonces  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^n)\}$  es un proceso puntual.

**Ejemplo 3.3.20** Sea  $\{T_n : n \in \mathbf{N}\}$  una sucesión de variables aleatorias con valores reales y  $0 \leq T_1 \leq T_2 \leq \dots$  y  $T_n \rightarrow \infty$  con probabilidad 1. Entonces las variables aleatorias

$$N_A = \sum_{n=1}^{\infty} \mathbf{1}_A(T_n) \quad A \in \mathfrak{B}(\mathbb{R}^+)$$

conforman un proceso puntual sobre  $\mathbb{R}^+$ .

Ahora se generalizarán conceptos relativos a procesos no estacionarios, dichos procesos fueron presentados en la Sección 3.3.1.

**Definición 3.3.21** Sea  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^k)\}$  un proceso puntual sobre  $\mathbb{R}^k$ . La medida de intensidad de  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^k)\}$  se define como

$$\psi(A) = E[N_A], \quad A \in \mathfrak{B}(\mathbb{R}^k).$$

**Teorema 3.3.22** La medida de intensidad de un proceso puntual es una medida.

**Demostración.** Resulta obvio que  $\psi$  es una función no negativa y que  $\psi(\emptyset) = 0$ .

Si  $A_1, A_2, \dots$  son ajenos por parejas y  $A = \bigcup_{k=1}^{\infty} A_k$

$$N_A = \sum_{k=1}^{\infty} N_{A_k}.$$

Tomando esperanzas

$$\psi(A) = \sum_{k=1}^{\infty} \psi(A_k),$$

lo cual concluye la prueba. ■

En este trabajo sólo se consideran aquellas medidas de intensidad que pueden obtenerse a partir de una *intensidad*. Una *intensidad* es una función Borel-medible  $\lambda$ , que es positiva. Esto permite hablar de la integral de Lebesgue de  $\lambda$ , generando medidas del tipo

$$\psi(A) = \int_A \lambda(x) dx.$$

Es de gran importancia el caso en el que  $\lambda$  es constante, entonces

$$\psi(A) = \lambda |A|,$$

donde  $|A|$  es la medida de Lebesgue de  $A$ .

**Definición 3.3.23** Un proceso puntual  $\{N_A : A \in \mathfrak{B}(\mathbf{R})\}$  sobre  $\mathbf{R}$  es un Proceso Poisson Puntual con intensidad  $\lambda$  si:

- i)  $\{N_A : A \in \mathfrak{B}(\mathbf{R})\}$  tiene incrementos independientes, esto en el sentido de que  $N_{B_1}, \dots, N_{B_n}$  son independientes para  $B_1, \dots, B_n$  borelianos ajenos por parejas.
- ii) Para cada  $B \in \mathfrak{B}(\mathbf{R})$

$$\Pr\{N_B = n\} = \exp(-\psi(B)) \frac{(\psi(B))^n}{n!} \quad n = 0, 1, \dots$$

Es necesario exigir que  $\psi$  no posea átomos (es decir que para todo  $x \in \mathbf{R}$ ,  $\psi(\{x\}) = 0$ ), ya que de lo contrario, si existiera  $\tau_0 \in \mathbf{R}$  con  $\psi(\{\tau_0\}) > 0$  se tendría  $\Pr(N_{\{\tau_0\}} \geq 2) > 0$ , lo cual contradice la idea inicial de que sólo puede haber una ocurrencia a la vez.

**Teorema 3.3.24** Sea  $\{N_A : A \in \mathfrak{B}(\mathbf{R})\}$  un proceso Poisson puntual sobre  $\mathbf{R}$ . Entonces, la familia de variables aleatorias

$$\{N_{(0,t)} : t > 0\}$$

forman un proceso Poisson (Definición 3.1.4). Dado un proceso Poisson, es posible construir, de manera única, un proceso Poisson puntual.

**Demostración** Se mostrará primero que la Definición 3.3.23 implica a la 3.1.4.

Sea  $\{N_A : A \in \mathfrak{B}(\mathbf{R})\}$  un proceso Poisson puntual. Se considera a la subfamilia  $\{N_{(0,t)} : t > 0\}$ . Las condiciones ii) y iii) de la Definición 3.1.4 se cumplen trivialmente para este nuevo proceso estocástico, ya que todo intervalo semiabierto es un Boreliano. Del hecho que  $\psi$  es no atómica se sigue la condición i) de la Definición 3.1.4.

Es fácil verificar que  $N_{(0,t]}$  es un proceso de contar, ya que la función

$$N_{(0,t]}(\cdot) = \# \{ \Pi(\cdot) \cap (0, t] \}$$

es positiva, sólo toma valores en los enteros y es monótona creciente respecto a  $t$ .

Se procederá a demostrar que a partir de un proceso Poisson  $\{N_t : t \geq 0\}$  se puede construir de forma única un proceso que satisfaga la Definición 3.3.23.

Se propone

$$\Pi(\cdot) = \bigcup_{k=1}^{\infty} \{S_k(\cdot)\},$$

donde  $S_k$  denota al instante en el cual tiene lugar la  $k$ -ésima ocurrencia. Es claro que dicho conjunto es numerable. Se sabe que las funciones  $S_k(\cdot)$  son variables aleatorias, por lo tanto la función

$$f(\cdot) = 1_{|A|}(S_k(\cdot)) \quad A \in \mathfrak{B}(\mathbb{R}) \quad , k \in \mathbb{N}$$

es una variable aleatoria, ya que toda función indicadora correspondiente a un boreliano es medible y la composición de funciones medibles es medible.

Se sigue que

$$\begin{aligned} N(A)(\cdot) &= \# \{ \Pi(\cdot) \cap A \} \\ &= \sum_{k=1}^{\infty} 1_{|S_k \in A|}(\cdot) \\ &= \sum_{k=1}^{\infty} 1_{|A|}(S_k(\cdot)) \end{aligned}$$

es una variable aleatoria.

Por lo tanto se puede construir de manera única un proceso Poisson puntual a partir de  $\{N_t : t \geq 0\}$ . Las condiciones i) y ii) son consecuencia de aplicar resultados referentes al Lema de Clases Monótonas [ver Ash]. ■

### 3.4 Caracterizaciones fuertes del Proceso Poisson

#### 3.4.1 El funcional característico

Del mismo modo que la función característica y la función generatriz de momentos permiten llevar a cabo ciertos cálculos referentes a la distribución de variables aleatorias (tales como sumas), el funcional característico de un proceso estocástico es una herramienta poderosa para realizar operaciones que involucren al proceso en cuestión. En esta sección se obtendrán expresiones para el funcional característico del Proceso Poisson Puntual y para el Proceso Poisson Compuesto.



El caso general

**Teorema 3.4.1** Sea  $\{N(A) : A \in \mathfrak{B}(\mathbb{R}^n)\}$  un Proceso Poisson sobre  $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$  con medida de intensidad  $\mu$ , y sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función Borel-medible. Entonces la suma

$$\Phi = \sum_{X \in \Pi} f(X) \quad (3.9)$$

es absolutamente convergente con probabilidad 1 si y sólo si

$$\int \min(|f(x)|, 1) \mu(dx) < \infty. \quad (3.10)$$

Si esta condición se cumple, entonces

$$E[e^{\theta \Phi}] = \exp \left[ \int (e^{\theta f(x)} - 1) \mu(dx) \right], \quad (3.11)$$

para cualquier número complejo  $\theta$  tal que haga que la integral de la derecha converja, y en particular siempre que  $\theta$  sea imaginario puro.

**Demostración.** El primer paso de la demostración consiste en verificar que la Igualdad (3.11) se cumple para todo número complejo  $\theta$ , si  $f$  es una función que toma un número finito de valores y se anula fuera de un conjunto de  $\mu$ -medida finita. Esto es a lo que se conoce con el nombre de *función simple* [Bartle].

Sea  $f$  una función simple positiva que toma los valores  $f_1, f_2, \dots, f_k$ , y sean  $A_i = \{x : f(x) = f_i\}$  conjuntos medibles tales que  $m_j = \mu(A_j) < \infty$ . Claramente los conjuntos  $A_i$  son ajenos entre sí. Entonces las variables aleatorias  $N_j = N(A_j)$  son independientes, y para cada  $j$  se tiene que  $N_j$  tiene distribución  $\text{Poi}(m_j)$ . Sea

$$\Phi = \sum_{X \in \Pi} f(X) = \sum_{j=1}^k f_j N_j.$$

Para todo número complejo  $\theta$  se tiene que,

$$\begin{aligned} E[e^{\theta \Phi}] &= \prod_{j=1}^k E[e^{\theta f_j N_j}] \\ &= \prod_{j=1}^k \exp[(e^{\theta f_j} - 1) m_j] \\ &= \exp \left[ \sum_{j=1}^k \int_{A_j} (e^{\theta f_j(x)} - 1) \mu(dx) \right] \\ &= \exp \left[ \int (e^{\theta f(x)} - 1) \mu(dx) \right], \end{aligned}$$

debido a que  $f = 0$  en el complemento de la unión de los  $A_j$ .

Cualquier función medible positiva  $f$  puede expresarse como el límite de una sucesión creciente  $\{f_i : i \in \mathbf{N}\}$  de funciones simples no negativas. Entonces, si se toma  $\theta = -u$  un número real negativo, y se definen las sumas  $\Phi_j = \sum_{X \in \Pi} f_j(X)$  se tiene

$$\begin{aligned} E[e^{-u\Phi}] &= \lim_{j \rightarrow \infty} E[e^{-\theta\Phi_j}] \\ &= \lim_{j \rightarrow \infty} \exp \left[ - \int (1 - e^{-uf_j(x)}) \mu(dx) \right] \\ &= \exp \left[ - \int (1 - e^{-uf(x)}) \mu(dx) \right], \end{aligned}$$

para  $u > 0$ , por el Teorema de Convergencia Monótona [Ash]. Si la Expresión (3.10) se cumple, la última integral converge y tiende a 0 cuando  $u \rightarrow 0$ , lo cual muestra que  $\Phi$  es una variable aleatoria finita. Esto se explicará con mayor detalle. Sean  $B_1 = \{x : f(x) \geq 1\}$  y  $B_2 = \{x : f(x) < 1\}$ , entonces

$$\int \min(|f(x)|, 1) \mu(dx) = \int_{B_1} \mu(dx) + \int_{B_2} f(x) \mu(dx).$$

Como la Expresión (3.10) se cumple, ambas integrales del lado derecho convergen. Entonces

$$\int_{B_1} (1 - e^{-uf(x)}) \mu(dx) < \int_{B_1} \mu(dx) < \infty,$$

y gracias a que  $y > 1 - e^{-y}$  para toda  $y$  se tiene

$$\int_{B_2} (1 - e^{-uf(x)}) \mu(dx) < \int_{B_2} uf(x) \mu(dx) < \infty.$$

Entonces ambos miembros de la Ecuación (3.11) son funciones analíticas de  $\theta$  en  $\text{Re } \theta \leq 0$ , de donde se sigue que la Expresión (3.11) es válida en el dominio  $\text{Re } \theta \leq 0$ . Por otra parte, si (3.10) no se cumple, la última integral diverge para toda  $u > 0$ , lo que implica  $E[e^{-u\Phi}] = 0$ , por lo tanto  $\Phi = \infty$  con probabilidad 1.

Por el momento el teorema está demostrado para  $f \geq 0$ . Para terminar la prueba, se aplica el teorema a las funciones positivas

$$f^+ = \max(f, 0), \quad f^- = \max(-f, 0).$$

La suma converge absolutamente si y sólo si

$$\Phi_+ = \sum_{X \in \Pi} f^+(X) = \sum_{X \in \Pi_+} f(X),$$

y

$$\Phi_- = \sum_{X \in \Pi} f^-(X) = - \sum_{X \in \Pi_-} f(X)$$

convergen, donde  $\Pi_+$  y  $\Pi_-$  son restricciones de  $\Pi$  en

$$G_+ = \{x : f(x) > 0\} \quad G_- = \{x : f(x) < 0\}.$$

Esto demuestra que la Expresión (3.10) es una condición necesaria y suficiente para la convergencia. Si esta se da y  $\theta$  es un imaginario puro, se tiene que

$$\begin{aligned} E [e^{\theta\Phi}] &= E [e^{\theta\Phi_+ - \theta\Phi_-}] \\ &= E [e^{\theta\Phi_+}] E [e^{\theta\Phi_-}] \\ &= \exp \left[ \int (e^{\theta f^+} - 1) d\mu \right] \exp \left[ \int (e^{-\theta f^-} - 1) d\mu \right] \\ &= \exp \left[ \int (e^{\theta f^+} + e^{-\theta f^-} - 2) d\mu \right] \\ &= \exp \left[ \int (e^{\theta f} - 1) d\mu \right], \end{aligned}$$

usando el hecho de que el número de ocurrencias del proceso en conjuntos ajenos son variables aleatorias independientes.

Por lo tanto la Igualdad (3.11) se cumple para todo número imaginario puro  $\theta$ , y su extensión a todos los números complejos tales que la integral converge es un problema de continuación analítica.

Un caso particular de interés del Teorema 3.4.1 es cuando se toma  $f \geq 0$  y  $\theta = -1$ , entonces la Expresión (3.11) establece que

$$E [e^{-\Phi_f}] = \exp \left[ - \int (1 - e^{-f(x)}) \mu(dx) \right], \tag{3.12}$$

donde

$$\Phi_f = \sum_{X \in \Pi} f(X) \tag{3.13}$$

El lado izquierdo de la Expresión (3.12), vista como un operador cuyo argumento es una función arbitraria  $f$  es un ejemplo de lo que se conoce como el *funcional característico* de un proceso estocástico. Este nombre se debe a que, la Ecuación (3.12) caracteriza al Proceso Poisson. Supóngase que  $\Pi$  es cualquier subconjunto aleatorio numerable de  $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$  y se define  $\Phi_f \leq \infty$  mediante la Expresión (3.13). Se hace además el supuesto que, para alguna medida  $\mu$ , la Ecuación (3.12) se cumple para una cierta clase  $\mathcal{F}$  de funciones. Si  $\mathcal{F}$  es lo suficientemente abundante como para contener a todas aquellas funciones que toman solamente un número finito de valores diferentes  $y_1, y_2, \dots, y_k$  la Ecuación (3.12) implica

$$E [e^{-\Phi_f}] = \exp \left[ - \sum_{j=1}^k (1 - e^{-y_j}) m_j \right],$$

donde

$$m_j = \mu \{x : f(x) = y_j\}.$$

Si se toma  $z_j = e^{-f_j}$  y  $A_j = \{x : f(x) = y_j\}$ ,

$$E \left[ z_1^{N(A_1)} z_2^{N(A_2)} \dots z_k^{N(A_k)} \right] = \prod_{j=1}^k e^{m_j(z_j-1)}.$$

Como esto se cumple para  $0 < z_j < 1$ , las variables  $N(A_j)$  son independientes con distribución  $\text{Poi}(m_j)$ . Por lo tanto

$$\{N(A) = \#(A \cap \Pi) : A \in \mathfrak{B}(\mathbf{R}^n)\}$$

es un Proceso Poisson con medida de intensidad  $\mu$ .

De lo anterior se concluye que, para demostrar que un proceso puntual es un Proceso Poisson, es suficiente verificar que la Expresión (3.12) se cumple para una clase lo suficientemente grande de funciones.

El funcional característico para el Proceso Poisson Compuesto

Sea  $\{N_t : t \geq 0\}$  un proceso Poisson con función de intensidad  $\lambda$ , y sea  $\{U_k : k \in \mathbf{N}\}$  una sucesión de vectores aleatorios independientes e idénticamente distribuidos, se supone además que son independientes del proceso. Se define

$$Z_t \begin{cases} \sum_{j=1}^{N_t} U_j & \text{si } N_t \geq 1 \\ 0 & \text{si } N_t = 0 \end{cases}$$

para  $t \geq 0$ . Entonces  $\{Z_t : t \geq 0\}$  es un proceso Poisson compuesto con marcas  $\{U_k : k \in \mathbf{N}\}$ .

**Teorema 3.4.2** *El funcional característico asociado con las variables aleatorias  $\{Z_t : t \geq 0\}$  está dado por la siguiente esperanza*

$$\varphi_Z(F) = E \left[ \exp \left( i \int_0^t F^T(t) dx(t) \right) \right], \quad (3.14)$$

donde la  $T$  denota al operador transpuesta. Aquí  $\{F(t) : t \geq 0\}$  es una función Borel-medible arbitraria  $F : \mathbf{R} \rightarrow \mathbf{R}^m$ , y  $m$  es la dimensión de  $U_1$ . La integral debe interpretarse como la integral de Riemann-Stieltjes que toma los siguientes valores

$$\int_0^t F^T(t) dx(t) = \begin{cases} 0, & N_t = 0 \\ \sum_{j=1}^{N_t} F^T(s_j) U_j, & N_t \geq 1, \end{cases}$$

donde  $U_j$  es la marca asignada a la ocurrencia que tuvo lugar en el instante  $s_j$ .

**Demostración.** En seguida se calcula la esperanza que aparece en la Expresión (3.14). Haciendo uso de las propiedades de la esperanza condicional, se tiene que

$$\begin{aligned}\varphi_Z(F) &= \sum_{n=0}^{\infty} \Pr\{N_t = n\} E \left[ \exp \left( i \int_0^t F^T(z) d\tau(z) \right) \mid N_t = n \right] \\ &= \Pr\{N_t = 0\} + \sum_{n=1}^{\infty} \Pr\{N_t = n\} E \left[ \exp \left( i \sum_{j=1}^{N_t} F^T(s_j) U_j \right) \mid N_t = n \right].\end{aligned}\quad (3.15)$$

Condicionando con los tiempos de ocurrencias y utilizando el hecho de que las marcas son independientes e idénticamente distribuidas, se obtiene

$$\begin{aligned}& E \left[ \exp \left( i \sum_{j=1}^{N_t} F^T(S_j) U_j \right) \mid N_t = n \right] \\ &= E \left\{ E \left[ \exp \left( i \sum_{j=1}^{N_t} F^T(S_j) U_j \right) \mid N_t = n; S_1, S_2, \dots, S_{N_t} \right] \mid N_t = n \right\} \\ &= E \left[ \prod_{j=1}^{N_t} \varphi_u(F(S_j)) \mid N_t = n \right],\end{aligned}$$

donde  $\varphi_u(\alpha)$  es la función característica de las marcas.

$$\varphi_u(\alpha) = E \left( e^{\alpha^T u} \right).$$

Del Corolario 3.3.9 resulta inmediato que

$$f_{S_1, S_2, \dots, S_{N_t} \mid N_t = n}(s_1, s_2, \dots, s_n) = \frac{\prod_{j=1}^n \lambda(s_j)}{\left[ \int_0^t \lambda(u) du \right]^n}. \quad (3.16)$$

Usando el hecho de que las variables  $\{S_i : i \in \mathbf{N}\}$  y  $\{U_i : i \in \mathbf{N}\}$  son independientes, y aprovechando la Expresión (3.16), se verifica que

$$\begin{aligned}& E \left[ \exp \left( i \sum_{j=1}^{N_t} F^T(S_j) U_j \right) \mid N_t = n \right] \\ &= E \left[ E \left( \exp \left( i \sum_{j=1}^{N_t} F^T(S_j) U_j \right) \mid S_1, \dots, S_{N_t} \right) \mid N_t = n \right] \\ &= \int \left[ \int \exp \left( i \sum_{j=1}^{N_t} F^T(S_j) U_j \right) dF_U \right] dF_{S_1, S_2, \dots, S_{N_t} \mid N_t} \\ &= \left[ \frac{\int_0^t \lambda(t) \varphi_u(F(t)) dt}{\int_0^t \lambda(t) dt} \right]^n.\end{aligned}$$

donde  $F_U$  es la función de distribución conjunta de  $(U_1, U_2, \dots, U_{N_i})$ , y  $F_{S_1, S_2, \dots, S_{N_i} | N_i}$  es la función de distribución condicional de los tiempos de ocurrencias en  $(0, l]$ , dado  $N_i$ .

Sustituyendo esta Expresión en (3.15) y gracias a que  $N_i$  tiene distribución Poisson, se concluye que el funcional característico de un proceso Poisson compuesto esta dado por

$$\varphi_Z(F) = \exp \left[ \int_0^l \lambda(t) (\varphi_u(F(t)) - 1) dt \right]. \quad \blacksquare \quad (3.17)$$

Un caso de interés es cuando el espacio de marcas  $\mathcal{U} = \{V_1, V_2, \dots\}$  es numerable y la probabilidad de asignar la marca  $V_k$  es  $p_k$  para  $k \in \mathbb{N}$ . Bajo estas circunstancias se tiene que

$$\varphi_u(\alpha) = \sum_{k=1}^{\infty} p_k e^{\alpha^T V_k},$$

y la Expresión (3.17) se vuelve

$$\varphi_Z(F) = \exp \left[ \int_0^l \lambda(t) \left( \sum_{k=1}^{\infty} p_k e^{F(t)^T V_k} - 1 \right) dt \right]. \quad (3.18)$$

Estos resultados serán de utilidad al obtener la Representación del Proceso Poisson Compuesto.

### 3.4.2 El Teorema de Renyi

A lo largo de este capítulo, el suponer una Propiedad de Incrementos Independientes ha sido crucial en la demostración de resultados. El siguiente Teorema permite caracterizar al Proceso Poisson Puntual sin hacer uso de supuestos de independencia; en vez de ello, se requiere que la probabilidad de que no se registren ocurrencias del proceso en uniones finitas de rectángulos tenga cierta forma funcional.

**Teorema 3.4.3** *Sea  $\mu$  una medida no atómica en  $\mathbb{R}^d$  tal que sea finita para conjuntos acotados. Sea  $\Pi$  un subconjunto aleatorio numerable de  $\mathbb{R}^d$  tal que, siempre que  $A$  sea una unión finita de rectángulos*

$$\Pr \{ \Pi \cap A = \emptyset \} = e^{-\mu(A)}$$

*entonces  $\{N(A) = \#(A \cap \Pi) : A \in \mathfrak{B}(\mathbb{R}^d)\}$  es un Proceso Poisson con medida de intensidad  $\mu$ .*

**Demostración.** Para fines de esta prueba se considera que un  $k$ -cubo es un rectángulo  $(a, b]$  para el cual

$$a_i = v_i 2^{-k} \quad b_i = (v_i + 1) 2^{-k} \quad (1 \leq i \leq d),$$

donde los  $v_i$  son enteros. Para cualquier  $k \in \mathbb{N}$  los  $k$ -cubos forman una disección de  $\mathbb{R}^d$ . Para cualquier  $k$ -cubo  $C$ , se denota con  $Y(C)$  al evento  $\{C \cap \Pi = \emptyset\}$ . Tómesse  $k$  fijo y sean  $C_1, C_2, \dots, C_n$

$k$ -cubos distintos y por tanto ajenos, entonces

$$\begin{aligned} \Pr \left\{ \bigcap_{r=1}^n Y(C_r) \right\} &= \Pr \{ C_r \cap \Pi = \emptyset : r \in \{1, 2, \dots, n\} \} \\ &= \Pr \left\{ \Pi \cap \left( \bigcup_{r=1}^n C_r \right) = \emptyset \right\} \\ &= \exp \left\{ -\mu \left( \bigcup_{r=1}^n C_r \right) \right\} \\ &= \prod_{r=1}^n e^{-\mu(C_r)} \end{aligned}$$

Por lo tanto los eventos  $Y(C_i)$  para los  $k$ -cubos  $C_i$ ,  $i \in \{1, \dots, n\}$  son independientes con

$$\Pr \{Y(C_i)\} = e^{-\mu(C_i)}$$

Se puede usar este hecho para calcular la distribución de

$$N(G) = \# \{ \Pi \cap G \}$$

para cualquier conjunto  $G$  abierto y acotado. Cualquier  $x \in G$  está contenido, para  $k$  lo suficientemente grande, en algún  $k$ -cubo subconjunto de  $G$ . Es más, si  $x, y \in G$  y  $x \neq y$  entonces a partir de cierta  $k$ ,  $x$  y  $y$  se encuentran contenidos en  $k$ -cubos distintos. Entonces

$$N(G) = \lim_{k \rightarrow \infty} N_k(G) \tag{3.19}$$

donde  $N_k(G)$  es el número de  $k$ -cubos  $C$  tales que están contenidos en  $G$  y  $Y(C)$  no ocurre. Claramente  $N_k(G)$  se incrementa con  $k$ .

Debido a que los eventos  $Y(C)$  son independientes, la función generatriz de momentos de  $N_k(G)$  es

$$\begin{aligned} E(z^{N_k(G)}) &= \prod [\Pr \{Y(C)\} + (1 - \Pr \{Y(C)\}) z] \\ &= \prod [e^{-\mu(C)} + (1 - e^{-\mu(C)}) z] \end{aligned}$$

siempre que  $|z| < 1$ . El producto se toma sobre todos los  $k$ -cubos  $C \subset G$ . En particular, si se toma  $0 \leq z < 1$ , de la Expresión (3.19) y del Teorema de Convergencia Monótona [Bartle] se sigue

$$E(z^{N(G)}) = \lim_{k \rightarrow \infty} \prod [z + (1 - z) e^{-\mu(C)}], \tag{3.20}$$

con la convención  $z^\infty = 0$ .

Para tales  $z$ , y para todo  $\mu \geq 0$ ,

$$z + (1 - z)e^{-\mu} \geq e^{-(1-z)\mu}$$

Entonces

$$\begin{aligned} E(z^{N(G)}) &\geq \lim_{k \rightarrow \infty} \prod e^{-(1-z)\mu(C)} \\ &\geq e^{-(1-z)\mu(G)} \end{aligned}$$

gracias a que los  $k$ -cubos  $C \subset G$  son ajenos. Haciendo  $z \rightarrow 1$  se deduce que  $N(G)$  es finita con probabilidad 1.

Para calcular el límite en la Expresión (3.20), es necesario usar el hecho de que, para  $k$ -cubos en una región acotada,  $\mu(C)$  tiende uniformemente a cero cuando  $k \rightarrow \infty$ . Esto es una consecuencia de la falta de átomos de  $\mu$ , y se probará por contradicción. Basta mostrar que esto se cumple para los subconjuntos de un 0-cubo fijo  $C_0$ . Si el resultado fuera falso, entonces existiría  $\delta > 0$  y  $k$ -cubos  $C \subset C_0$  para cualquier  $k$ , con la propiedad  $\mu(C) > \delta$ . Se dice que un  $k$ -cubo  $C$  es de tipo 1 si para toda  $l > k$ ,  $C$  contiene un  $l$ -cubo  $C'$  tal que  $\mu(C') > \delta$ . Por hipótesis  $C_0$  es de tipo 1. Se sigue que de los  $2^d$  1-cubos en  $C_0$ , al menos uno (se denotará  $C_1$ ) es de tipo 1. Procediendo de esta manera se obtiene una sucesión  $C_0 \supset C_1 \supset C_2 \supset \dots$  donde  $C_k$  es un  $k$ -cubo de tipo 1. Un  $k$ -cubo de este tipo contiene a un cubo  $C'$  con  $\mu(C') > \delta$  y por lo tanto  $\mu(C) > \delta$ . Esto implica que  $\mu(C_k) > \delta$  para toda  $k$ . Pero como  $C_k$  es un  $k$ -cubo con lados de longitud  $2^{-k}$ , la intersección de todos los  $C_k$  es vacía o consta de un solo punto, cuya medida tendría que ser mayor o igual a  $\delta$ . Esta es la contradicción que se buscaba.

Por lo tanto regresando a  $G$ , se puede usar el hecho de que  $\mu(C) \leq \delta$  para todos los  $k$ -cubos en  $G$  y con  $k$  lo suficientemente grande. Si  $\mu \leq \delta$  y  $z$  está fijo

$$e^{-(1-z)\mu} \leq z + (1 - z)e^{-\mu} \leq e^{-(1-z)\psi(\delta)\mu}$$

donde  $\psi(\delta) \rightarrow 1$  cuando  $\delta \rightarrow 0$ . Entonces

$$e^{-(1-z)\mu(G)} \leq E(z^{N(G)}) \leq e^{-(1-z)\psi(\delta)\mu(G)}$$

y haciendo  $\delta \rightarrow 0$ ,

$$E(z^{N(G)}) = e^{-(1-z)\mu(G)}.$$

Esto se cumple para toda  $z$  en  $0 \leq z < 1$ , de donde se sigue que  $N(G)$  tiene distribución Poisson con parámetro  $\mu(G)$ . Ahora sean  $G_1, G_2, \dots, G_m$  conjuntos abiertos, acotados y ajenos. Ningún  $k$ -cubo puede estar contenido en más de un  $G_j$ , entonces las variables

$$N_k(G_1), N_k(G_2), \dots, N_k(G_m)$$



son independientes para  $k$  fijo. Haciendo  $k \rightarrow \infty$  se concluye que las variables

$$N(G_1), N(G_2), \dots, N(G_m)$$

son independientes.

Para completar la prueba, se debe extender estos resultados de conjuntos abiertos acotados a la familia de conjuntos de Borel. Esto se logra aplicando el Lema de Clases Monótonas y usando el hecho de que cualquier función medible es el límite de funciones simples. ■

### 3.5 Representación del Proceso Poisson Compuesto

Primero se desarrollará la representación cuando el espacio de marcas es numerable. Sea  $\{Z_t : t \geq 0\}$  un Proceso Poisson Compuesto con espacio de marcas  $\mathcal{U} = \{V_1, V_2, \dots\}$ , donde la  $k$ -ésima marca  $V_k$  ocurre con probabilidad  $p_k$ . Sea

$$\{0 = t_{0,n} < t_{1,n} < t_{2,n} < \dots < t_{n,n} = t : n \in \{1, 2, \dots\}\}$$

una sucesión de particiones de  $[0, t)$  tal que

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} (t_{j,n} - t_{j-1,n}) = 0.$$

Se definen los procesos  $\{Z(t, V_k) : t \geq 0\}$ , con  $k = 1, 2, \dots$  de la siguiente manera

$$Z(t, V_k) = \lim_{n \rightarrow \infty} \sum_{j=1}^n [Z(t_{j,n}) - Z(t_{j-1,n})] 1_{\{V_k\}}(Z(t_{j,n}) - Z(t_{j-1,n})), \quad (3.21)$$

donde  $1_{\{V\}}(x)$  es la función indicadora para la marca, definida por

$$1_{\{V\}}(x) = \begin{cases} 1, & x = V \\ 0, & x \neq V. \end{cases}$$

Entonces  $Z(t, V_k)$  es la suma de las marcas que ocurren durante  $[0, t)$  y que tienen valor  $V_k$ . Esto implica que la siguiente expresión se cumple con probabilidad 1:

$$Z_t = \sum_{k=1}^{\infty} Z(t, V_k) = \sum_{k=1}^{\infty} V_k N(t, V_k), \quad (3.22)$$

donde  $N(t, V_k)$  es el número de ocurrencias que tuvieron lugar en  $[0, t)$  y tienen marca  $V_k$ . La expresión (3.22) sugiere que  $\{Z_t : t \geq 0\}$  puede, de cierta forma, descomponerse en los procesos de contar

$$\{N(t, V_1) : t \geq 0\}, \{N(t, V_2) : t \geq 0\}, \dots, \{N(t, V_k) : t \geq 0\}, \dots$$

y que  $\{Z_t : t \geq 0\}$  puede reconstruirse en términos de dichos procesos. Esto se explica con mayor detalle en el siguiente resultado.

**Teorema 3.5.1** Sea  $\{Z_t : t \geq 0\}$  un Proceso Poisson Compuesto con espacio de marcas numerable  $\{V_1, V_2, \dots\}$ . Si los tiempos de ocurrencia se comportan como un Proceso Poisson con función de intensidad  $\lambda$ , y la probabilidad de que la marca  $V_k$  sea asignada a un punto en particular es  $p_k$ . Entonces los procesos  $\{N(t, V_k) : t \geq 0\}$ , con  $k \in \{1, 2, \dots\}$  son Procesos Poisson independientes entre sí, y el  $k$ -ésimo proceso tiene por función de intensidad  $p_k \lambda$ . En estos términos se tiene la siguiente representación de  $Z_{(\cdot)}$ :

$$Z_t = \sum_{k=1}^{\infty} V_k N(t, V_k). \quad (3.23)$$

**Demostración.** El Proceso Poisson puede verse como un caso especial de Proceso Poisson Compuesto con espacio de marcas  $\mathcal{U} = \{1\}$ . En este caso  $\varphi_x(jv) = \exp[j\alpha]$ , y la Expresión (3.17) se reduce al funcional característico de un Proceso Poisson. Un proceso de contar que cumple con las condiciones de la Definición 3.1.4 salvo que la magnitud de sus saltos es  $V$ , es equivalente a un Proceso Poisson Compuesto con espacio de marcas  $\mathcal{U} = \{V\}$ . Para este proceso se tiene que

$$\varphi_Z(F) = \exp \left[ \int \lambda(s) \left( e^{jF^T(s)V} - 1 \right) ds \right]. \quad (3.24)$$

Ahora se demostrará que los procesos definidos en la Expresión (3.21) son Procesos Poisson independientes entre sí, similares a los descritos en el párrafo anterior; de modo que  $\{Z(t, V_k) : t \geq 0\}$  experimenta saltos de magnitud  $V_k$  y  $\{N(t, V_k) : t \geq 0\}$  tiene función de intensidad  $p_k \lambda$ . La Expresión (3.23) es una consecuencia inmediata de esto. Por lo tanto se construyen procesos  $\{Z^*(t, V_k) : t \geq 0\}$  para  $k \in \{1, 2, \dots\}$  tales que sean independientes y cuyo funcional característico este dado por

$$\exp \left[ \int_0^t p_k \lambda(s) \left( e^{v^T(s)V_k} - 1 \right) ds \right].$$

Entonces, de la Expresión (3.24) se sigue que  $\{Z^*(t, V_k) : t \geq 0\}$  es un Proceso Poisson Compuesto con espacio de marcas  $V_k$  y función de intensidad  $p_k \lambda$ . Esto implica que  $\{Z^*(t, V_k) : t \geq 0\}$  equivale a  $\{Z(t, V_k) : t \geq 0\}$  para cada  $k$  por separado, debido a que tienen el mismo funcional característico. El objetivo es ahora demostrar que esta igualdad se cumple de manera conjunta para todas las  $k$  simultáneamente. Con este propósito se define

$$Z_t^* = \sum_{k=1}^{\infty} Z(t, V_k)$$

Por construcción y como consecuencia de 3.18, es evidente que  $\{Z^*(t, V_k) : t \geq 0\}$  tiene el mismo funcional característico en  $[0, t]$  que  $\{Z(t, V_k) : t \geq 0\}$  para toda  $t > 0$ . Entonces  $\{Z^*(t, V_k) : t \geq 0\}$  y  $\{Z(t, V_k) : t \geq 0\}$  tienen las mismas distribuciones si se restringen a cualquier intervalo semiabierto. Pero, por construcción

$$Z^*(t, V_k) = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} [Z^*(t_{jn}) - Z^*(t_{j-1,n})] 1_{\{V_k\}}(Z^*(t_{jn}) - Z^*(t_{j-1,n}))$$

entonces  $\{Z^*(t, V_k) : t \geq 0\}$  es la misma función sobre  $\{Z_t^* : t \geq 0\}$  que  $\{Z(t, V_k) : t \geq 0\}$  sobre  $\{Z_t : t \geq 0\}$  para toda  $k$ . Se sigue que los procesos

$$\{Z_t^* : t \geq 0\}, \{Z(t, V_k) : t \geq 0\}, k \in \{1, 2, \dots\}$$

tienen la misma distribución conjunta que los procesos

$$\{Z_t^* : t \geq 0\}, \{Z^*(t, V_k) : t \geq 0\}, k \in \{1, 2, \dots\}. \quad \blacksquare$$

El siguiente paso será remover la restricción de que el espacio de marcas sea numerable. El objetivo actual es obtener una expresión análoga a (3.23) para el caso continuo. Sea  $\{Z_t : t \geq 0\}$  un Proceso Poisson Compuesto con espacio de marcas  $\mathcal{U}$ . Se supondrá que  $\mathcal{U}$  es una región (no necesariamente acotada) de un espacio Euclidiano finito-dimensional; entonces cada marca  $V \in \mathcal{U}$  es un vector de números reales.

Los saltos de  $\{Z_t : t \geq 0\}$  tales que sus amplitudes se encuentren en regiones ajenas de  $\mathcal{U}$  determinan como ciertos Procesos Poisson Compuestos independientes contribuyen al proceso original. Esto puede verse de la siguiente forma: Se toma una partición  $B_1, B_2, \dots, B_k, \dots$  de  $\mathcal{U}$  y una sucesión de particiones

$$\{0 = t_{0,n} < t_{1,n} < t_{2,n} < \dots < t_{n,n} = t : n \in \{1, 2, \dots\}\}$$

de  $[0, t)$  tal que

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} (t_{j,n} - t_{j-1,n}) = 0.$$

Se definen los procesos  $\{Z(t, B_k) : t \geq 0\}$  para  $k \in \{1, 2, \dots\}$  de la siguiente forma:

$$Z(t, B_k) = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} \{Z(t_{j,n}) - Z(t_{j-1,n})\} 1_{\{B_k\}}(Z(t_{j,n}) - Z(t_{j-1,n})), \quad (3.25)$$

donde  $1_{\{B_k\}}(x)$  es la función indicadora del conjunto  $B_k$ . Estas definiciones ofrecen una versión generalizada de la Expresión (3.21). El proceso  $\{Z(t, B_k) : t \geq 0\}$  es un proceso acumulador de marcas para aquellos puntos del proceso original a los cuales se les asignó una marca perteneciente a  $B_k$ . Para  $t \geq 0$  la siguiente expresión se cumple con probabilidad 1:

$$Z_t = \sum_{k=1}^{\infty} Z(t, B_k). \quad (3.26)$$

Mediante un argumento análogo al usado en la demostración del teorema anterior, se sigue que

$$\{Z(t, B_k) : t \geq 0\}, k \in \{1, 2, \dots\}$$

son Procesos Poisson Compuestos independientes entre sí, cuyas intensidades respectivas son  $p_k \lambda$ , donde  $p_k$  es la probabilidad de que la marca correspondiente a un punto en especial tome un valor en  $B_k$ .

Se toma una partición  $B_1, B_2, \dots, B_k, \dots$  de  $\mathcal{U}$  de tal forma que el diámetro de ningún  $B_k$  sea mayor que  $\delta$ . Sea  $N(t, B_k)$  el número de ocurrencias en  $[0, t)$  que tienen marcas en  $B_k$ . Por el argumento antes expuesto, los procesos  $\{N(t, B_k) : t \geq 0\}$  son Procesos Poisson independientes cuyas funciones de intensidad correspondientes son  $p_k \lambda$ . Sea  $V_k \in B_k$  y sea  $\|V_k\| \approx \sup_{Y \in B_k} |V_k - Y|$ . Entonces de  $\|V_k\| < \delta$  y de la Expresión (3.26) se sigue que

$$\left| Z_t - \sum_{k=1}^{\infty} V_k N(t, B_k) \right| \leq \sum_{k=1}^{\infty} \|V_k\| N(t, B_k) \leq \delta N_t,$$

donde  $N_t$  es el número total de ocurrencias en  $[0, t)$ . Es claro que

$$E[\delta N_t] = \delta \int_0^t \lambda(s) ds \quad \text{y} \quad \text{var}[\delta N_t] = \delta^2 \int_0^t \lambda(s) ds.$$

Por lo tanto, para toda sucesión de particiones

$$\{B_{1,n}, B_{2,n}, \dots, B_{k,n}, \dots : n \in \mathbb{N}\}$$

de  $\mathcal{U}$  tal que  $\delta \rightarrow 0$  conforme  $n \rightarrow \infty$ , se cumple

$$\lim_{n \rightarrow \infty} E \left( \left| Z_t - \sum_{k=1}^{\infty} V_k N(t, B_{k,n}) \right|^2 \right) = 0.$$

Entonces, la suma  $\sum_{k=1}^{\infty} V_k N(t, B_{k,n})$  converge a un límite (en  $\mathcal{L}^2$ ) conforme  $n$  tiende a infinito. Se denotará a este límite con  $\int_{\mathcal{U}} V N(t, dV)$ . De este modo, un Proceso Poisson Compuesto con espacio de marcas no numerable puede representarse mediante

$$Z_t = \int_{\mathcal{U}} V N(t, dV) \tag{3.27}$$

Lo cual provee de un resultado que generaliza al Teorema 3.5.1.

La interpretación del término  $N(t, dV)$  en 3.27 es la siguiente:  $\{N(t, dV) : t \geq 0\}$  es un Proceso Poisson que cuenta el número de ocurrencias en  $[0, t)$  cuyas marcas pertenecen al volumen  $[V, V + dV]$ . La función de intensidad correspondiente es  $[P_v(V + dV) - P_v(V)] \lambda(t)$ , donde  $P_v(V)$  es la función de distribución de la marca. Los procesos

$$\{N(t, dV_j) : t \geq 0\} \quad , j = 1, 2, \dots$$

son independientes para volúmenes ajenos. Entonces

$$N(t, B_k) = \int_{B_k} N(t, dV) \tag{3.28}$$

es un proceso de contar asociado a las ocurrencias que les fueron asignadas marcas en  $B_k$ . La función de intensidad de  $\{N(t, B_k) : t \geq 0\}$  es

$$\lambda(t) \int_{B_k} P_v(dV) = \lambda(t) p_k.$$

Otra interpretación de la Expresión (3.27) puede darse en términos de Procesos Poisson en  $\mathbf{R}^n$ . Desde este punto de vista los procesos marcados pueden tratarse como procesos cuyas ocurrencias tienen lugar en  $\mathbf{R}^+ \times \mathcal{U}$ ; cada punto tiene una coordenada referente al tiempo en  $\mathbf{R}^+$  y una coordenada espacial en  $\mathcal{U}$ . Sea  $\Lambda$  un intervalo de tiempo en  $\mathbf{R}^+$  y  $B$  un elemento de  $\mathfrak{B}(\mathcal{U})$ . Se denotará con  $M(\Lambda, B)$  al número de ocurrencias del proceso en  $\Lambda \times B$ . Se sigue de la discusión anterior que

$$\{M(\Lambda, B) : \Lambda \text{ intervalo de } \mathbf{R}^+, B \in \mathfrak{B}(\mathcal{U})\}$$

es un proceso Poisson sobre el espacio  $\mathbf{R}^+ \times \mathcal{U}$ . Esto es, el número de ocurrencias en conjuntos ajenos de  $\mathbf{R}^+ \times \mathcal{U}$  son independientes, y el número de eventos que ocurren en una región  $\Lambda \times B$  es una variable aleatoria Poisson con parámetro

$$\int_{\Lambda} \int_B \lambda(t) dt P_v(dV).$$

Entonces el proceso  $\{M(\Lambda, B) : \Lambda \text{ intervalo de } \mathbf{R}^+, B \in \mathfrak{B}(\mathcal{U})\}$  es un proceso Poisson multidimensional con función de intensidad  $\lambda(t, u)$ , la cual puede factorizarse de la siguiente forma:  $\lambda(t, u) = \mu(t) \phi(u)$ . Si la variable aleatoria que determina a las marcas es continua con densidad  $p_V(V)$ , entonces la función de intensidad de este proceso es  $\mu(t) p_V(V)$ , además

$$N(t, dV) = \int_0^t M(ds, dV)$$

Por lo tanto, la Expresión (3.27) también puede escribirse como

$$Z_t = \int_0^t \int_{\mathcal{U}} VM(ds, dV).$$

Se sigue que un proceso Poisson compuesto puede representarse en términos de un proceso Poisson que se desarrolla en cierto espacio-tiempo. Entonces, queda demostrado el siguiente resultado

**Teorema 3.5.2** Sea  $\{Z_t : t \geq 0\}$  un proceso Poisson compuesto con un espacio de marcas no numerable  $\mathcal{U}$ . El proceso  $\{N_t : t \geq 0\}$  tiene intensidad  $\lambda(\cdot)$ , y la función de distribución de las marcas es  $P_V(V)$ . Entonces  $Z_t$  puede representarse de la siguiente forma

$$Z_t = \int_0^t \int_{\mathcal{U}} VM(ds, dV),$$

donde  $\{M(\Lambda, B) : \Lambda \text{ intervalo de } \mathbf{R}^+, B \in \mathfrak{B}(\mathcal{U})\}$  es un proceso Poisson en  $\mathbf{R}^+ \times \mathcal{U}$  cuya función de intensidad es de la forma  $\mu(t) p_V(V)$ , de modo que

$$E[M(\Lambda, B)] = \int_{\Lambda} \int_B \lambda(t) dt P_v(dV)$$

donde  $p_V(V)$  es la función de densidad de las marcas.

## 3.6 Simulación del Proceso Poisson

El objetivo de esta sección es presentar los algoritmos básicos para simular al Proceso Poisson. De acuerdo a la Proposición 2.6.1, si  $U \sim U(0, 1)$ , entonces la variable aleatoria

$$Y = -\ln(1 - U) / \lambda \quad (3.29)$$

tiene distribución  $\text{Exp}(\lambda)$ .

Del análisis hecho en la Sección 3.2.1 se desprende una relación entre las distribuciones Poisson y Exponencial. Si los tiempos entre ocurrencias de un Proceso Poisson son variables aleatorias exponenciales con parámetro 1, entonces el número de ocurrencias del proceso en el intervalo  $(0, \theta)$  tiene distribución  $\text{Poi}(\theta)$ . A partir de este hecho es fácil obtener un método para generar variables aleatorias con distribución Poisson. Si  $\{Y_i : i \in \mathbf{N}\}$  es una sucesión de variables aleatorias independientes con distribución  $\text{Exp}(1)$ , se desea encontrar un número  $X$  tal que el  $X$ -ésimo evento tenga lugar antes del tiempo  $\theta$  pero tal que el  $(X + 1)$ -ésimo ocurra en un tiempo posterior a  $\theta$ ; esto es,

$$\sum_{i=1}^X Y_i \leq \theta < \sum_{i=1}^{X+1} Y_i.$$

Aprovechando la Expresión (3.29), se sigue

$$-\sum_{i=1}^X \ln U_i \leq \theta < -\sum_{i=1}^{X+1} \ln U_i,$$

o de manera equivalente

$$\prod_{i=1}^X U_i \geq e^{-\theta} > \prod_{i=1}^{X+1} U_i.$$

Se tiene entonces el siguiente resultado.

**Proposición 3.6.1** Sea  $\{U_i : i \in \mathbf{N}\}$  una sucesión de variables aleatorias independientes con distribución uniforme en  $(0, 1)$ , y sea  $X$  el número entero más pequeño tal que

$$\prod_{i=1}^{X+1} U_i < e^{-\lambda}.$$

Entonces  $X$  tiene distribución  $\text{Poi}(\lambda)$ .

El razonamiento que conduce a este resultado provee de un algoritmo para simular variables aleatorias con distribución  $\text{Poi}(\lambda)$  a partir de variables con distribución  $U(0, 1)$ .

**Algoritmo 3.6.2** Simulación de una variable aleatoria con distribución  $\text{Poi}(\lambda)$ .

i) Hacer: **ocurrencias** ← 0, **producto** ← 1.

ii) Generar una observación  $u$  a partir de una distribución  $U(0, 1)$ , y actualizar el valor de **producto**: **producto** ← **producto** \*  $u$ .

iii) Si **producto** ≥  $e^{-\lambda}$ , hacer **ocurrencia** ← **ocurrencia** + 1 e ir a ii).

iii') Si **producto** <  $e^{-\lambda}$ , la salida del programa es el valor actual de **ocurrencias**.

**Comentario 3.6.3** Si bien es posible simular variables aleatorias con distribución Exponencial o Poisson haciendo uso de la Proposición 2.6.1, el paquete S-PLUS permite hacer esto de forma directa mediante los comandos *rexp* y *rpois* (ver Apéndice).

Tomando como punto de partida el Algoritmo 3.6.2, resulta simple simular el número de ocurrencias en un intervalo  $(0, t]$ , de un PPE con intensidad  $\lambda$ , así como los tiempos en que ocurren los eventos. A continuación se presenta una rutina para el paquete S-PLUS que lleva a cabo esta tarea.

```
# Inicialización de variables
lambda<-intensidad del proceso
tiempo<-longitud del intervalo
s1<-c(0)
suma<-0
ocurre<-0
# Obtención de los tiempos de ocurrencia
while(suma<tiempo)
{ suma<-suma+rexp(1,lambda)
  if(suma<tiempo)
  { ocurre<-ocurre+1
    s1<-c(s1,suma) }
}
s<-s1[-1]
s<-sort(s)
s2<-c(0,s,tiempo)
# Graficación del proceso simulado
w1<-seq(ocurre)
w2<-c(0,seq(ocurre-1))
plot(s,w1,type='p',pch=16,xlim=c(0,tiempo),
ylim=c(0,ocurre+1),xlab='tiempo',ylab='ocurrencias
del proceso,lambda=1')
points(s,w2,type='p',pch=1)
for(i in 1:(ocurre+1))
{ x<-seq(s2[i],s2[i+1],length=50)
  y<-rep(i-1,50)
  lines(x,y,type='l') }
```

En las Figuras 3.3 y 3.4 se presentan simulaciones de un proceso Poisson estacionario con intensidades 1 y .5 respectivamente.

Otra forma de simular al Proceso Poisson Estacionario con intensidad  $\lambda$  es la siguiente. Supóngase que el proceso va a ser restringido a un intervalo predeterminado  $(0, t]$  con  $t > 0$ . Se genera una variable  $K$  con distribución Poisson de parámetro  $\lambda$ . Por el Teorema 3.2.5 la posición de cada ocurrencia posee una distribución uniforme en  $(0, t]$ , la cual es fácil de generar.

Esta idea se implementa en el siguiente programa

```
# Inicialización de variables
```

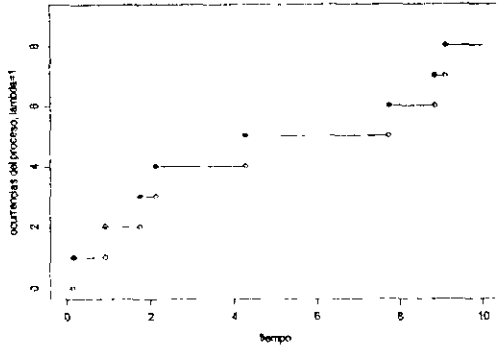


FIGURE 3.3.

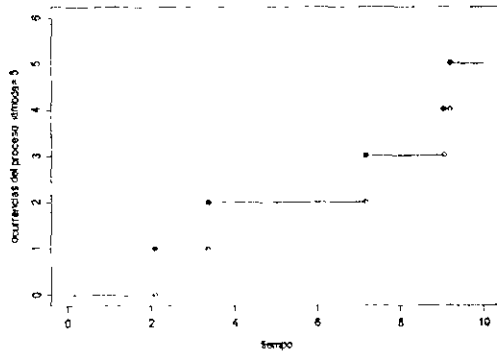


FIGURE 3.4.



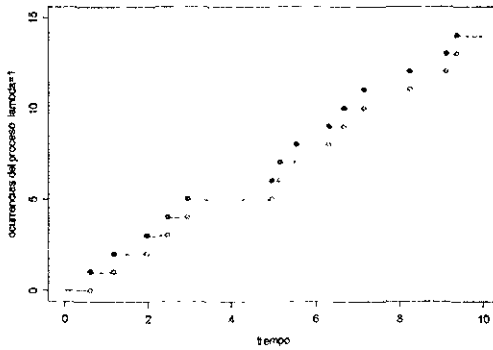


FIGURE 3.5.

```

lambda<-intensidad del proceso
tiempo<-longitud del intervalo
# Simulación del número de ocurrencias
ocurre<-rpois(1,lambda*tiempo)
# Obtención del vector de tiempos de ocurrencia
s<-runif(ocurre,0,tiempo)
s<-sort(s)
s2<-c(0,s,tiempo)
# Graficación del proceso simulado
w1<-seq(ocurre)
w2<-c(0,seq(ocurre-1))
plot(s,w1,type='p',pch=16,xlim=c(0,tiempo),
ylim=c(0,ocurre+1),xlab='tiempo',
ylab='ocurrencias del proceso, lambda=1')
points(s,w2,type='p',pch=1)
for(i in 1:(ocurre+1))
{ x<-seq(s2[i],s2[i+1],length=50)
  y<-rep(i-1,50)
  lines(x,y,type='l') }

```

En las Figuras 3.5 y 3.6 se ilustra lo que se obtuvo al ejecutar este programa, tomando intensidades de 1 y .25, respectivamente.

Un algoritmo muy usado para simular un Proceso Poisson estacionario de intensidad  $\lambda$  en el plano consiste en

**Algoritmo 3.6.4** Simulación de un PPE en  $R^2$

- i) Delimitar un rectángulo de área  $h$ .

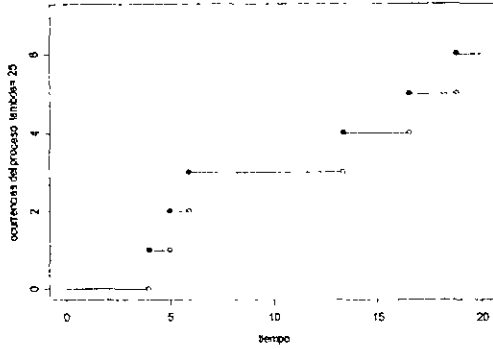


FIGURE 3.6.

- ii) Generar una variable  $N$  con distribución  $\text{Poi}(\lambda h)$ .
- iii) Generar  $N$  variables uniformes independientes, distribuidas sobre el rectángulo.

Para generar un punto de manera uniforme sobre un rectángulo de vértices  $a, b, c, d$ , basta con generar una pareja  $(X, Y)$ , donde  $X, Y$  se distribuyen  $U(a, b)$  y  $U(a, d)$  respectivamente ( $a$  es el vértice en la parte inferior izquierda, los demás se enumeran en sentido levógiro).

Esta es la rutina que ejecuta este algoritmo.

```
# Inicialización de variables
lambda<-intensidad del proceso
largo<-largo del rectángulo
ancho<-ancho del rectángulo
# Simulación del número de ocurrencias
n<-rpois(1,lambda)
# Obtención de las posiciones de las ocurrencias
x<-runif(n,0,ancho)
y<-runif(n,0,largo)
# Graficación del proceso simulado
plot(x,y)
```

En las Figuras 3.7 y 3.8 se presentan simulaciones de Procesos Poisson estacionarios con intensidades 2 y 10 respectivamente.

Este procedimiento puede a su vez aprovecharse para simular un Proceso Poisson no estacionario. El argumento es el siguiente: Sea  $\{N_A : A \in \mathfrak{B}(\mathbf{R}^2)\}$  un proceso Poisson estacionario en  $\mathbf{R}^2$  con intensidad 1, y sea  $\lambda : \mathbf{R}^+ \rightarrow \mathbf{R}$  una función integrable no negativa tal que  $\int_0^t \lambda(x) dx < \infty$  para toda  $t \in \mathbf{R}^+$ . Se asocia una variable  $N_t$  a cada  $t \in \mathbf{R}^+$ , dicha  $N_t$  se define como el número de puntos

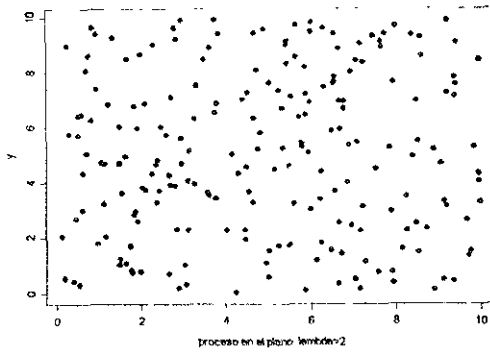


FIGURE 3.7.

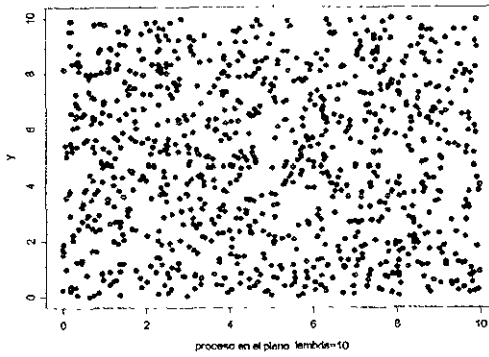


FIGURE 3.8.

del proceso  $\{N_A : A \in \mathfrak{B}(\mathbb{R}^2)\}$  que caen en el conjunto

$$A_t = \{(x, y) \in \mathbb{R}^2 \mid x \in (0, t], 0 \leq y \leq \lambda(x)\}$$

De la teoría desarrollada en la Sección 3.3.3 se sigue que  $N_t$  se distribuye  $\text{Poi}\left(\int_0^t \lambda(x) dx\right)$ , y que la familia de variables aleatorias  $\{N_t : t \geq 0\}$  posee la Propiedad de Incrementos Independientes.

La rutina que lleva a cabo las simulaciones es la siguiente (para el caso particular  $\lambda(x) = \ln(\psi * x)$ ,  $\psi > 0$ ).

```
# Inicialización de variables
psi<-parámetro asociado a la intensidad del proceso
tiempo<-longitud del intervalo
# Cálculo del máximo de la función de intensidad en el intervalo
altura<-log(tiempo*psi)
# Simulación del proceso de intensidad 1 en el rectángulo
k<-rpois(1, altura*tiempo)
poin<-cbind(runif(k,0,tiempo),runif(k,0,altura))
# Obtención de los tiempos de ocurrencias para el proceso no estacionario
ocurre<-0
for(j in 1:k)
{ if(poin[j,2]<=log(psi*poin[j,1])&ocurre==0)
  {
  v<-c(poin[j,1])
  ocurre<-1 }
  if(poin[j,2]<=log(psi*poin[j,1])&ocurre>0)
  {
  v<-c(v,poin[j,1])
  ocurre<-ocurre+1 } }
s<-sort(v)
s2<-c(0,s,tiempo)
# Graficación del proceso simulado
w1<-seq(ocurre)
w2<-c(0,seq(ocurre-1))
plot(s,w1,type='p',pch=16,xlim=c(0,tiempo),
ylim=c(0,ocurre+1),xlab='tiempo',
ylab='ocurrencias del proceso, psi=1')
points(s,w2,type='p',pch=1)
for(i in 1:(ocurre+1))
{ x<-seq(s2[i],s2[i+1],length=50)
  y<-rep(i-1,50)
  lines(x,y,type='l') }
```

En las Figuras 3.9 y 3.10 pueden verse dos simulaciones obtenidas usando este procedimiento y donde  $\lambda(x) = .3x$  y  $\lambda(x) = \ln x$  respectivamente.

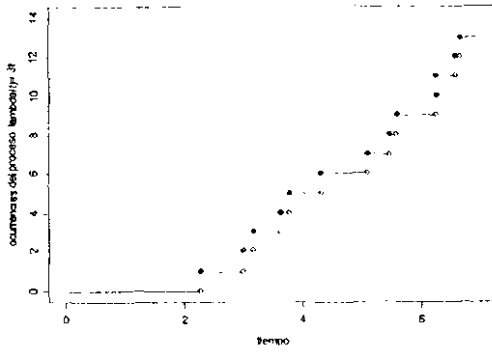


FIGURE 3.9

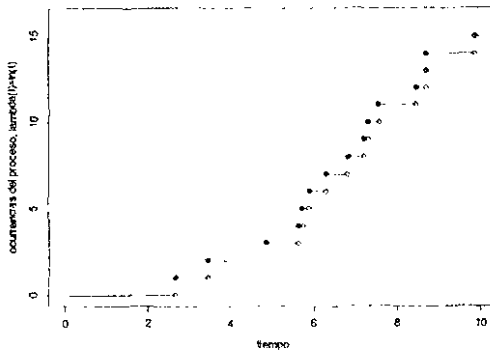


FIGURE 3.10.

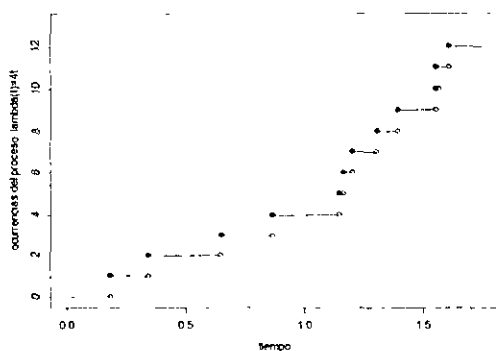


FIGURE 3.11.

Otra forma de simular un proceso Poisson no estacionario es haciendo uso del Teorema 3.3.5. Esto se hace aprovechando el siguiente hecho: Si  $\{M_t : t \geq 0\}$  es un proceso Poisson estacionario de intensidad 1, y  $\{N_t : t \geq 0\}$  es un proceso Poisson con función de intensidad  $\lambda$ , entonces  $M_{\rho(t)}$  posee la misma distribución que  $N_t$ , esto para toda  $t \geq 0$ , donde  $\rho(t) = E(N_t)$ .

Para implementar esta técnica basta con aplicar una transformación monótona creciente (de hecho se trata de  $\tau$ , la función inversa de  $\rho$ ) al vector compuesto por los tiempos de ocurrencias (se llamó  $s$  a dicho vector), esto para cualquiera de las rutinas usadas para simular un PPE con intensidad 1. El caso particular que se consideró fue  $\lambda(x) = 4x$  y  $t = 1$ . En la Figura 3.11 se presenta la simulación correspondiente.

A partir de estos resultados es claro el camino a seguir para simular un proceso Poisson compuesto en un intervalo  $(0, t]$ . Se generan dos sucesiones de variables  $\{S_k : k \in \mathbf{N}\}$  y  $\{Y_i : i \in \mathbf{N}\}$ , donde las variables  $S_k$  representan los tiempos de ocurrencias para un proceso Poisson con función de intensidad  $\lambda$  y las  $Y_i$  tienen función de distribución  $F_Y$ . Estas últimas reproducen a las marcas asociadas a cada ocurrencia. Entonces la variable

$$Z_t = \sum_{i=1}^{N_t} Y_i \quad t \in \mathbf{R}^+,$$

donde  $N_t$  es el mayor entero que cumple

$$S_{N_t} < t.$$

posee la misma distribución que  $M_t$ , donde  $\{M_t : t \geq 0\}$  es un proceso Poisson compuesto de intensidad  $\lambda$  y marcas con distribución  $F_Y$ .

**Comentario 3.6.5** La función de distribución de  $Y_i$  puede no cumplir con las condiciones de la Proposición 2.6.1, por lo tanto será necesario aplicar otras técnicas para generar a dichas variables.

La siguiente rutina en S-PLUS sirve para simular este tipo de procesos; Se considera el caso particular en que las marcas tienen distribución exponencial.

```
# Inicialización de variables
lambda<-intensidad del proceso
tiempo<-longitud del intervalo
dseta<-parámetro de la distribución de las marcas
# Obtención de los tiempos de ocurrencia
ocurre<-rpois(1,lambda*tiempo)
s<-runif(ocurre,0,tiempo)
s<-sort(s)
s2<-c(0,s,tiempo)
# Simulación de las marcas
w<-rexp(ocurre-1,dseta)
w1<-cumsum(c(w, rexp(1,dseta)))
w2<-cumsum(c(0,w))
w3<-c(0,w1)
# Graficación del proceso simulado
plot(s,w1,type='p',pch=16,xlim=c(0,tiempo),
ylim=c(0,w1[ocurre]+1),xlab='tiempo',
ylab='ocurrencias del proceso, lambda=1')
points(s,w2,type='p',pch=1)
for(i in 1:(ocurre+1))
{ x<-seq(s2[i],s2[i+1],length=50)
  y<-rep(w3[i],50)
  lines(x,y,type='l') }
```

En las figuras 3.12 y 3.13 se presentan simulaciones correspondientes a procesos de Poisson Compuestos de intensidad 1 y marcas con distribución  $\text{Exp}(2)$  y  $N(4, 1)$ , respectivamente.

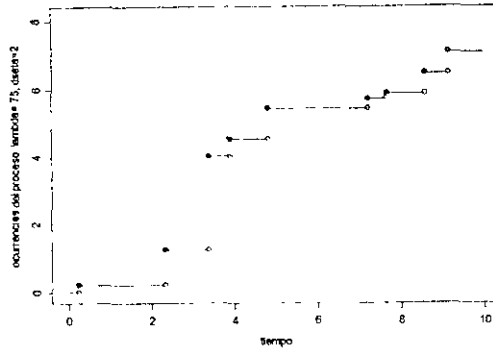


FIGURE 3.12.

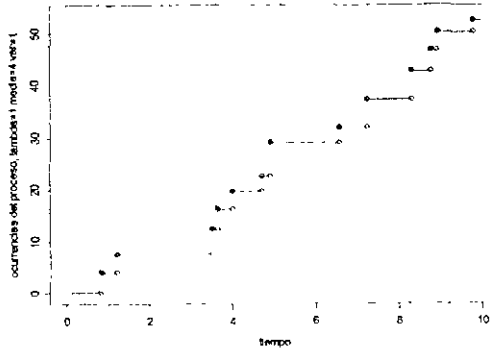


FIGURE 3.13



# El problema de estimación para el Proceso Poisson

## Introducción

El presente capítulo está dedicado al estudio de algunos problemas de inferencia paramétrica para el Proceso Poisson, de acuerdo a la Teoría de Máxima Verosimilitud y adoptando los criterios de optimalidad propuestos por la Teoría de Estimación Insesgada. Dichos problemas son clasificados de acuerdo a dos criterios: plan de muestreo y tipo de proceso observado.

Los problemas de estimación puntual son tratados en las Secciones 4.1-4.6. En la Sección 4.1 se discute el caso en el que el tiempo de observación es fijo. En la Sección 4.2 se ofrece una cota para la matriz de error cuadrático medio para esta forma de observación. El caso en el que el tiempo de observación es aleatorio es tratado en la Sección 4.3. El problema de estimación basado en el número de ocurrencias del proceso en un número finito de intervalos de longitud aleatoria es analizado con detalle en la Sección 4.5. Algunos problemas de estimación para el Proceso Poisson Compuesto son tratados en la Sección 4.8. Más adelante se discuten algunas propiedades asintóticas de los estimadores. La última sección está dedicada a presentar algunos aspectos de la teoría de pruebas de hipótesis para este tipo de procesos, según el paradigma propuesto por J. Neyman y K. Pearson.

El contenido de este capítulo está basado en gran parte en [Snyder] y en el Capítulo 6 de [Basawa]. Los resultados de Teoría Asintótica pueden consultarse en [Lehmann] y [Serfling].

## 4.1 Observación durante un intervalo de tiempo fijo

### 4.1.1 El caso estacionario

Supóngase que se puede observar un Proceso Poisson Estacionario  $\{N_t : t \geq 0\}$  definido en  $(\Omega, \mathfrak{F}, \mathbf{P})$  con intensidad  $\lambda$ , durante un intervalo determinado  $(0, \tau]$ . El número de ocurrencias del proceso en dicho intervalo es una variable aleatoria. De acuerdo a la notación propuesta en el capítulo anterior, esta variable se denota mediante el símbolo  $N_\tau$ . Recordando el análisis hecho en la Sección 3.2.1 es

ESTA TESIS NO SALE  
DE LA BIBLIOTECA

posible obtener las variables

$$S_n = \sum_{i=1}^n X_i, \quad n \in \{1, \dots, N_r\},$$

donde  $S_n$  representa el instante exacto en el que se dió la  $n$ -ésima ocurrencia del proceso  $\{N_t : t \geq 0\}$ , y las variables aleatorias  $X_i$  representan al tiempo transcurrido entre el  $(i-1)$ -ésimo evento y el  $i$ -ésimo. Por el Teorema 3.2.2 las variables  $X_i$  son independientes y con distribución  $\mathbf{Exp}(\lambda)$ .

Cada realización del proceso provee una muestra de tamaño  $N_r$  de una población que se rige por una ley de distribución  $\mathbf{Exp}(\lambda)$ . Lo que procede a hacerse es obtener el estimador máximo verosímil para  $\lambda$ .

La función de verosimilitud originada por el tipo de observación es:

$$L_\lambda \propto \prod_{k=1}^{n_r} [\lambda \exp[-\lambda x_k]] \Pr\{X_{n_r+1} > (r - s_{n_r})\}.$$

Donde  $\{X_{n_r+1} > (r - s_{n_r})\}$  representa el hecho de que tras la ocurrencia  $n_r$ -ésima, no se registran más eventos en el intervalo  $(0, \tau]$ .

Esta expresión se simplifica de la siguiente manera.

$$\Pr\{X_{n_r+1} > \alpha\} = \int_\alpha^\infty \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_\alpha^\infty = e^{-\lambda \alpha}.$$

Entonces

$$\begin{aligned} L_\lambda &\propto \lambda^{n_r} \exp\left(-\lambda \sum_{k=1}^{n_r} x_k\right) \exp[-\lambda (r - s_{n_r})] \\ &= \lambda^{n_r} \exp(-\lambda r). \end{aligned}$$

Luego

$$\begin{aligned} L_\lambda &\propto \lambda^{n_r} \exp(-\lambda r) \\ \ln L_\lambda &= n_r \ln \lambda - \lambda r + c \\ \frac{d}{d\lambda} \ln L_\lambda &= \frac{n_r}{\lambda} - r, \end{aligned}$$

donde  $c$  es constante respecto a  $\lambda$ . Igualando a cero y despejando se tiene que

$$\hat{\lambda} = \frac{N_r}{r}.$$

Como  $N_r$  tiene distribución Poisson con media  $\lambda r$ , se sigue que  $E(\hat{\lambda}) = \lambda$  y  $\text{var}(\hat{\lambda}) = \lambda/r$ . Debido a que la Distribución Poisson pertenece a la Familia Exponencial y su espacio paramétrico natural [ver Lindsey] es todo  $\mathbf{R}$ , se tiene que  $\hat{\lambda}$  es suficiente y completo, y por ser insesgado se puede aplicar el Teorema de Lehmann-Scheff [ver Casella y Berger], de donde se concluye que  $\hat{\lambda}$  es el estimador insesgado de varianza mínima para  $\lambda$ .

### 4.1.2 El proceso no estacionario

En caso de que las observaciones provengan de un Proceso Poisson no estacionario, se sigue un razonamiento similar, pero hay que recordar que el Teorema 3.2.2 sólo es válido para procesos estacionarios. Por lo tanto debe encontrarse la función de verosimilitud basada en el evento

$$\{N_r = n; X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}.$$

El cual equivale a

$$\{N_r = n; S_1 = s_1, S_2 = s_2, \dots, S_n = s_n\},$$

donde  $s_1 = x_1, s_2 = x_1 + x_2, \dots, s_n = x_1 + x_2 + \dots + x_n$ .

La distribución conjunta de estas variables fue obtenida en el Corolario 3.3.9. Se denota con  $\lambda(\cdot, \theta)$  a la función de intensidad del proceso. Aquí  $\theta$  representa al vector de parámetros desconocidos.

Se tiene entonces el siguiente resultado: La función de verosimilitud de  $\theta$  para  $n$  observaciones en el intervalo  $(0, r]$  en los tiempos  $s_1 < s_2 < \dots < s_n$  es

$$L_\theta \propto \left( \prod_{i=1}^n \lambda(s_i, \theta) \right) \exp \left[ - \int_0^r \lambda(u, \theta) du \right]. \quad (4.1)$$

Por lo tanto el estimador máximo verosímil para  $\theta$  viene dado por el máximo de  $\ln L_\theta$  dada la función  $\lambda(\cdot, \theta)$ .

**Ejemplo 4.1.1** Se observa un Proceso Poisson de acuerdo a este plan de muestreo. La función de intensidad es de la forma  $\lambda(x, \theta) = \exp[x\theta]$ , entonces

$$\begin{aligned} L_\theta &\propto \left( \prod_{i=1}^n \lambda(s_i, \theta) \right) \exp \left[ - \int_0^r \lambda(u, \theta) du \right] \\ &= \exp \left[ \theta \sum_{i=1}^{n_r} s_i \right] \exp \left[ - \frac{1}{\theta} (e^{r\theta} - 1) \right]. \end{aligned}$$

La log-verosimilitud viene dada por

$$\ln L_\theta = \theta \sum_{i=1}^{n_r} s_i - \frac{1}{\theta} (e^{r\theta} - 1) + c,$$

donde  $c$  no depende del parámetro. Esta función no es fácil de optimizar de forma analítica. En la Figura 4.1 se presenta una gráfica de la log-verosimilitud, tomando  $r = 3$ ,  $\sum_{i=1}^{n_r} s_i = 96$  y  $c = 0$ . Para este caso particular el estimador máximo verosímil es  $\hat{\theta} = 1.347$ . ■

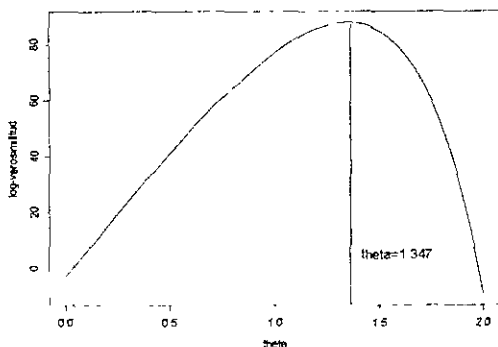


FIGURE 4.1.

4.1.3 Forma alternativa de obtener un estimador

El aprovechar la propiedad del proceso estacionario  $\{N_t : t \geq 0\}$  de que los tiempos entre ocurrencias sean variables aleatorias independientes con distribución exponencial no es la única forma de obtener un estimador para la intensidad  $\lambda$ . Otra forma de resolver el problema de estimación, sujeto a la restricción de observar solamente en el lapso  $(0, \tau]$ , consiste en dividir a dicho intervalo en subintervalos  $(0, t_1], (t_1, t_2], \dots, (t_{m-2}, t_{m-1}], (t_{m-1}, \tau]$  y hacer uso del punto iii) de la Definición 3.1.4, lo cual se traduce en observar a las variables aleatorias

$$N_{t_i} - N_{t_{i-1}} \quad i \in \{0, 1, \dots, m\}.$$

Entendiéndose que  $t_0 = 0$  y  $t_m = \tau$ . Estas variables aleatorias son independientes. De donde se concluye que al seleccionar  $t_1 = \frac{\tau}{m}, t_2 = \frac{2\tau}{m}, \dots, t_{m-1} = \frac{(m-1)\tau}{m}$  y tomar en cuenta las variables aleatorias antes mencionadas, se está obteniendo una muestra correspondiente a una población que se rige por una distribución  $Poi\left(\frac{\lambda\tau}{m}\right)$ .

La función de verosimilitud para este caso es

$$\begin{aligned} L_\lambda &\propto \prod_{i=1}^m \Pr\{N_{t_i} - N_{t_{i-1}} = n_i\} \\ &= \prod_{i=1}^m e^{-\lambda \frac{\tau}{m}} \frac{[\lambda \frac{\tau}{m}]^{n_i}}{n_i!} \\ &= e^{-\lambda \tau} \prod_{i=1}^m \frac{[\lambda \frac{\tau}{m}]^{n_i}}{n_i!}. \end{aligned} \tag{4.2}$$

entonces

$$\begin{aligned}\ln L_\lambda &= -\lambda r + \sum_{i=1}^m \left[ n_i \ln \lambda + n_i \ln \left( \frac{r}{n} \right) - \ln(n_i!) \right] + c \\ \frac{\partial}{\partial \lambda} \ln L_\lambda &= -r + \sum_{i=1}^m \frac{n_i}{\lambda} \\ &= -r + \frac{n_r}{\lambda}.\end{aligned}$$

Esto implica que el estimador máximo verosímil viene dado por

$$\hat{\lambda} = \frac{N_r}{r}$$

Resulta muy simple generalizar el procedimiento anterior para el caso no estacionario, debido a que no se hizo ningún supuesto respecto a la intensidad del proceso. Lo único que hay que tomar en cuenta es que la distribución de las variables  $N_i - N_{i-1}$ , se modifica de la siguiente manera

$$Pr \{ N_i - N_{i-1} = n_k \} = \exp \left[ - \int_{t_k}^{t_{k+1}} \lambda(u, \theta) du \right] \frac{\left( \int_{t_k}^{t_{k+1}} \lambda(u, \theta) du \right)^{n_k}}{n_k!}.$$

Donde  $\lambda(\cdot, \theta)$  es la función de intensidad del proceso y  $\theta$  es el vector de parámetros desconocidos. Esto implica que la función de verosimilitud es

$$\begin{aligned}L_\theta &\propto \prod_{i=1}^n \exp \left[ - \int_{t_{i-1}}^{t_i} \lambda(u, \theta) du \right] \frac{\left( \int_{t_{i-1}}^{t_i} \lambda(u, \theta) du \right)^{n_i}}{n_i!} \\ &= \exp \left[ - \int_0^r \lambda(u, \theta) du \right] \prod_{i=1}^n \frac{\left( \int_{t_{i-1}}^{t_i} \lambda(u, \theta) du \right)^{n_i}}{n_i!},\end{aligned}\tag{4.3}$$

Tampoco es necesario el suponer que los subintervalos sean de igual longitud. Para encontrar el estimador máximo verosímil, hay que maximizar la Expresión (4.3), dada la forma funcional de  $\lambda(\cdot, \theta)$ .

**Ejemplo 4.1.2** Se observó el número de ocurrencias de un proceso Poisson durante el lapso  $[0, 3)$ , los datos registrados son

Intervalo	# de ocurrencias
$[0, 1)$	1
$[1, 2)$	3
$[2, 3)$	1

Se sabe que la función de intensidad es de la forma  $\lambda(x; \theta) = x\theta$ . Entonces

$$\begin{aligned} L_{\theta}(\mathbf{x}) &= \exp \left[ - \int_0^3 \lambda(u, \theta) du \right] \left( \int_0^1 \lambda(u, \theta) du \right) \frac{\left( \int_1^2 \lambda(u, \theta) du \right)^3}{6} \left( \int_2^3 \lambda(u, \theta) du \right) \\ &= \exp \left[ - \frac{9}{2} \theta \right] \left( \frac{1}{2} \right) \left( \frac{3}{2} \right)^3 \left( \frac{5}{2} \right) \theta^5 \\ &= c \exp \left[ - \frac{9}{2} \theta \right] \theta^5, \end{aligned}$$

luego

$$\begin{aligned} \ln L_{\theta}(\mathbf{x}) &= 5 \ln \theta + \ln(c) - \frac{9}{2} \theta \\ \frac{\partial}{\partial \theta} \ln L_{\theta}(\mathbf{x}) &= \frac{5}{\theta} - \frac{9}{2}. \end{aligned}$$

Se concluye que  $\hat{\theta} = \frac{10}{9}$ . ■

## 4.2 La cota inferior de Cramér y Rao para la estimación en tiempo fijo

El objetivo de esta sección es establecer una cota inferior para la matriz de error cuadrático medio para un  $\hat{\theta}$  del vector  $\theta$  de parámetros desconocidos. En la Subsección 4.1.3 se consideró el caso en el que las observaciones consistían en contar el número de ocurrencias en cada uno de los subintervalos  $(t_i, t_{i+1}]$  que conformaban una partición del tiempo total de observación  $(0, \tau]$ . Para el caso analizado en las Subsecciones 4.1.1 y 4.1.2 las variables aleatorias de interés eran los tiempos entre ocurrencias  $\{X_1, X_2, \dots, X_{N_r}\}$ .

La cota inferior para la matriz de error correspondiente a los problemas antes mencionados que se presentará enseguida es un caso especial de una cota más general conocida como la cota inferior de Cramér y Rao [ver Mood, Graybill y Boes].

**Teorema 4.2.1** Sea  $\{N_t, t \geq 0\}$  un Proceso Poisson no estacionario con función de intensidad  $\lambda(\cdot, \theta)$ , donde  $\theta$  es el vector de parámetros desconocidos. Sea  $\hat{\theta}$  un estimador cualquiera para  $\theta$  obtenido a partir de observaciones basadas en los tiempos entre ocurrencias (ver Subsección 4.1.2) o en el número de ocurrencias en subintervalos (del mismo modo que el caso tratado en la Subsección 4.1.3). Entonces, la matriz de error cuadrático medio

$$\Sigma(\theta) = E_{\theta} \left[ (\hat{\theta} - \theta) (\hat{\theta} - \theta)^T \right],$$

satisface

$$\Sigma(\theta) \geq b(\theta) b^T(\theta) + \left[ J + \frac{\partial b(\theta)}{\partial \theta} \right] F^{-1}(\theta) \left[ J + \frac{\partial b(\theta)}{\partial \theta} \right]^T. \quad (4.4)$$

donde  $b(\theta) = E_{\theta} [\hat{\theta} - \theta]$  es el sesgo de  $\hat{\theta}$ ,  $\partial b(\theta) / \partial \theta$  es la matriz Jacobiana del sesgo (el  $(i, j)$ -ésimo elemento  $\partial b_i(\theta) / \partial \theta_j$ ),  $I$  denota a la matriz identidad y  $F(\theta)$  es la matriz de información de Fisher. Para el plan de muestreo basado en contar el número de ocurrencias en cada subintervalo, esta matriz viene dada por

$$F(\theta) = \sum_{i=1}^k \left[ \int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt \right]^{-1} \left[ \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial \theta} dt \right] \left[ \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial \theta} dt \right]^T, \quad (4.5)$$

y para el muestreo basado en los tiempos entre ocurrencias por

$$F(\theta) = \int_0^t \frac{1}{\lambda(t, \theta)} \left[ \frac{\partial \lambda(t, \theta)}{\partial \theta} \right] \left[ \frac{\partial \lambda(t, \theta)}{\partial \theta} \right]^T dt. \quad (4.6)$$

En las Expresiones (4.5) y (4.6)  $\partial \lambda(t, \theta) / \partial \theta$  denota al vector gradiente de  $\lambda(t, \theta)$  con respecto a  $\theta$ . La igualdad en la Expresión (4.4) se tiene si y sólo si el estimador  $\hat{\theta}$  satisface

$$\hat{\theta} = \theta + b(\theta) + \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right] F^{-1}(\theta) \left[ \frac{\partial \ln L(\theta)}{\partial \theta} \right], \quad (4.7)$$

para todo  $\theta$ .

**Demostración.** Se define

$$z(\theta) = \left[ \begin{array}{c} \hat{\theta} - \theta - b(\theta) \\ \frac{\partial \ln L_{\theta}}{\partial \theta} \end{array} \right],$$

donde  $L_{\theta}$  es la función de verosimilitud definida en (4.1) para el caso tratado en la Subsección 4.1.2 o es la dada por la Expresión (4.3) para la otra forma de observación. La cota dada en la Expresión (4.4) se obtiene a partir de  $E_{\theta} [z(\theta) z^T(\theta)]$ . Para valuar esta esperanza, es necesario primero calcular

$$E_{\theta} \left\{ \left[ \hat{\theta} - \theta - b(\theta) \right] \left[ \frac{\partial \ln L_{\theta}}{\partial \theta} \right]^T \right\}$$

y

$$E_{\theta} \left\{ \left[ \frac{\partial \ln L_{\theta}}{\partial \theta} \right] \left[ \frac{\partial \ln L_{\theta}}{\partial \theta} \right]^T \right\},$$

esto para ambas formas de observación. Se denota con  $P_{\theta}(n_1, n_2, \dots, n_k)$  a la función de masa conjunta para el número de ocurrencias en cada subintervalo. Entonces, para este caso

$$\begin{aligned} & E_{\theta} \left\{ \hat{\theta} - \theta - b(\theta) \right\} \\ &= \sum_{n_1} \sum_{n_2} \dots \sum_{n_k} \left[ \hat{\theta}(n_1, n_2, \dots, n_k) - \theta - b(\theta) \right] P_{\theta}(n_1, n_2, \dots, n_k) \\ &= 0. \end{aligned} \quad (4.8)$$

Por lo tanto, tras calcular el gradiente de la Expresión (4.8) con respecto a  $\theta$  y usando la relación  $\partial P/\partial\theta = P \cdot \partial \ln L/\partial\theta$ , se sigue que

$$\begin{aligned} & \sum_{n_1} \sum_{n_2} \dots \sum_{n_k} \left[ -I - \frac{\partial b(\theta)}{\partial\theta} \right] P_\theta(n_1, n_2, \dots, n_k) \\ & + \sum_{n_1} \sum_{n_2} \dots \sum_{n_k} \left[ \hat{\theta}(n_1, n_2, \dots, n_k) - \theta - b(\theta) \right] \left[ \frac{\partial \ln L_\theta}{\partial\theta} \right]^T P_\theta(n_1, n_2, \dots, n_k) \\ & = 0, \end{aligned}$$

entonces,

$$\begin{aligned} & E_\theta \left\{ \left[ \hat{\theta} - \theta - b(\theta) \right] \left[ \frac{\partial \ln L_\theta}{\partial\theta} \right]^T \right\} \\ & = I + \frac{\partial b(\theta)}{\partial\theta}. \end{aligned} \tag{4.9}$$

De manera análoga a la Expresión (4.8), se tiene que, para la observación basada en los tiempos entre ocurrencias

$$\begin{aligned} & E_\theta \left\{ \left[ \hat{\theta} - \theta - b(\theta) \right] \right\} \\ & = \left[ \hat{\theta}(0) - \theta - b(\theta) \right] \Pr_\theta \{ N_\tau = 0 \} \\ & + \sum_{n=1}^\infty \int \dots \int \left[ \hat{\theta}(n, S_1, S_2, \dots, S_n) - \theta - b(\theta) \right] f_S(S, N_\tau = n) dS_1 \dots dS_n \\ & = 0. \end{aligned} \tag{4.10}$$

Donde  $f_S(S, N_\tau = n)$  representa a la densidad conjunta del número de ocurrencias en  $(0, \tau]$  y las posiciones de las mismas (se integra sobre el conjunto  $0 \leq S_1 \leq S_2 \leq \dots \leq S_n \leq \tau$ ). El calcular el gradiente de esta expresión respecto a  $\theta$  conduce nuevamente a la Ecuación (4.9).

De la Expresión (4.3) se obtiene

$$\frac{\partial \ln L_\theta}{\partial\theta} = \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial\theta} dt \left[ -1 + \frac{N_{t_i} - N_{t_{i-1}}}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt} \right],$$

entonces

$$\begin{aligned} & \left[ \frac{\partial \ln L_\theta}{\partial\theta} \right] \left[ \frac{\partial \ln L_\theta}{\partial\theta} \right]^T \\ & = \sum_{i=1}^k \left[ -1 + \frac{N_{t_i} - N_{t_{i-1}}}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt} \right]^2 \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial\theta} dt \int_{t_{i-1}}^{t_i} \left( \frac{\partial \lambda(t, \theta)}{\partial\theta} \right)^T dt \\ & + \sum_{i \neq j} \left[ -1 + \frac{N_{t_i} - N_{t_{i-1}}}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt} \right] \left[ -1 + \frac{N_{t_j} - N_{t_{j-1}}}{\int_{t_{j-1}}^{t_j} \lambda(t, \theta) dt} \right] \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial\theta} dt \int_{t_{j-1}}^{t_j} \left( \frac{\partial \lambda(t, \theta)}{\partial\theta} \right)^T dt. \end{aligned}$$



Sea  $G_i = N_i - N_{i-1}$ , Aprovechando la siguientes igualdades

$$E_{\theta} \{G_i G_j\} = \begin{cases} \int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt \int_{t_{j-1}}^{t_j} \lambda(t, \theta) dt & i \neq j \\ \int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt + \left[ \int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt \right]^2 & i = j, \end{cases}$$

se obtiene

$$E \left( \left[ -1 + \frac{N_i - N_{i-1}}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt} \right]^2 \right) = \frac{1}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt}$$

y

$$E \left( \left[ -1 + \frac{N_i - N_{i-1}}{\int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt} \right] \left[ -1 + \frac{N_j - N_{j-1}}{\int_{t_{j-1}}^{t_j} \lambda(t, \theta) dt} \right] \right) = 0,$$

para  $i \neq j$ .

Se puede afirmar entonces que

$$\begin{aligned} F(\theta) &= E_{\theta} \left\{ \left[ \frac{\partial \ln L_{\theta}}{\partial \theta} \right] \left[ \frac{\partial \ln L_{\theta}}{\partial \theta} \right]^T \right\} \\ &= \sum_{i=1}^k \left[ \int_{t_{i-1}}^{t_i} \lambda(t, \theta) dt \right]^{-1} \left[ \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial \theta} dt \right] \left[ \int_{t_{i-1}}^{t_i} \frac{\partial \lambda(t, \theta)}{\partial \theta} dt \right]^T, \end{aligned} \tag{4.11}$$

esto para el plan de muestreo basado en observar el número de ocurrencias en subintervalos.

Para el caso en que las observaciones se basaron en los tiempos entre ocurrencias, el mismo procedimiento lleva a (4.11), con  $F(\theta)$  dada de acuerdo a la Expresión (4.6). Juntando estos resultados se tiene que, para ambos casos

$$\begin{aligned} \Lambda(\theta) &= E_{\theta} [z(\theta) z^T(\theta)] \\ &= \begin{bmatrix} [\Sigma(\theta) - b(\theta) b^T(\theta)] & \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right] \\ \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right]^T & F(\theta) \end{bmatrix}. \end{aligned}$$

Gracias a la linealidad del operador esperanza, y a que  $z(\theta) z^T(\theta)$  es una matriz positiva definida, para cualquier vector  $v$  se cumple  $v^T \Lambda(\theta) v = E_{\theta} \left[ (v^T z(\theta))^2 \right] \geq 0$ . Esta desigualdad es válida en particular para

$$v = \begin{bmatrix} v_1 \\ -F^{-1}(\theta) \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right]^T v_1 \end{bmatrix}$$

donde  $\mathbf{v}_1$  es arbitrario. Para esta elección se concluye que

$$\begin{aligned} & \mathbf{v}_1^T \Sigma(\theta) \mathbf{v}_1 \\ & \geq \mathbf{v}_1^T \left\{ b(\theta) b^T(\theta) + \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right] F^{-1}(\theta) \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right]^T \mathbf{v}_1 \right\}. \end{aligned}$$

Debido a la arbitrariedad con la cual se escogió  $\mathbf{v}_1$ , la desigualdad 4.4 queda demostrada. Por esta misma razón, se sigue que

$$\begin{aligned} & E_\theta \left[ (\mathbf{v}_1^T z(\theta))^2 \right] \\ & = E_\theta \left[ \left( \mathbf{v}_1^T \left\{ \hat{\theta} - \theta - b(\theta) - \left[ I + \frac{\partial b(\theta)}{\partial \theta} \right] F^{-1}(\theta) \left[ \frac{\partial \ln L_\theta}{\partial \theta} \right]^T \mathbf{v}_1 \right\} \right)^2 \right] \end{aligned}$$

será cero si y sólo si (4.7) se cumple para todo  $\theta$ . Lo cual concluye la prueba. ■  
Para más detalles, se recomienda consultar [Snyder].

### 4.3 Observaciones continuas durante un tiempo aleatorio.

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson Estacionario definido en  $(\Omega, \mathfrak{F}, P)$ . El proceso es observado de forma ininterrumpida hasta el momento en que se registra la  $k$ -ésima ocurrencia. La muestra  $x_1, x_2, \dots, x_k$  consta de los tiempos entre eventos. Este problema es muy similar al tratado en la Sección 4.1.1. Hay que notar que, a diferencia del caso tratado en la Sección 4.1.1, aquí el tamaño de muestra es fijo, mientras que el tiempo total de observación es una variable aleatoria. Esto implicará un cambio en la distribución del estimador máximo verosímil.

Sea  $\lambda$  la intensidad del proceso, por lo tanto la función de verosimilitud será

$$\begin{aligned} L_\lambda & \propto \prod_{i=1}^k \lambda \exp(-\lambda x_i) \\ & = \lambda^k \exp\left(-\lambda \sum_{i=1}^k x_i\right) \\ & = \lambda^k \exp(-\lambda s_k), \end{aligned}$$

donde  $s_k = \sum_{i=1}^k x_i$ . El estimador máximo verosímil es  $\hat{\lambda} = k/S_k$ . Se sabe que  $S_k$  se distribuye como una variable aleatoria  $\text{Gam}(k, \lambda)$ .

Como  $E(1/S_k) = \lambda/k - 1$ ,  $k > 1$ , se sigue que  $E(\hat{\lambda}) = (k/(k-1))\lambda$ . El sesgo en el estimador máximo verosímil puede eliminarse si se toma  $\tilde{\lambda} = ((k-1)/k)\hat{\lambda}$  como estimador. Del hecho que el espacio paramétrico natural de la distribución  $\text{Exp}(\lambda)$  es  $(0, \infty)$ , se concluye que  $S_k$  es una estadística completa y suficiente para  $\lambda$ . Por el Teorema de Lehmann-Scheffé  $\tilde{\lambda}$  es el estimador insesgado de varianza mínima para  $\lambda$ . Se tiene además que  $\text{var}(\tilde{\lambda}) = \lambda^2/(k-2)$ ,  $k > 2$ .

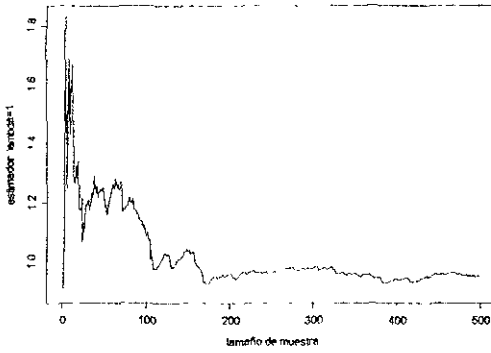


FIGURE 4.2.

Este estimador también es fuertemente consistente. Sea

$$\Pi(\omega) = \{S_1(\omega), S_2(\omega), \dots\}$$

y se considera la partición de  $\mathbf{R}^+$  dada por  $\{(k, k + 1] : k \in \mathbf{N}\}$ . Se tiene que  $N((k, k + 1]) \sim \text{Poi}(\lambda)$ , por lo tanto el número de ocurrencias en  $(k, k + 1]$  es finito con probabilidad 1. Esto implica que  $\Pi$  no posee puntos de acumulación finitos. Del Teorema 2.1.2 se sigue que, con probabilidad 1,  $\Pi$  es infinito, entonces

$$\Pr \left\{ \lim_{k \rightarrow \infty} S_k = \infty \right\} = 1.$$

Se toma ahora la sucesión  $\left\{ \frac{k}{S_k} : k \in \mathbf{N} \right\}$ . Del Teorema 3.2.7 se concluye que el estimador es fuertemente consistente.

En las Figuras 4.2 y 4.3 se presentan trayectorias simuladas para este estimador, con  $\lambda = 1, 15$ .

Hay que mencionar que esto no se da en los casos analizados en las Subsecciones 4.1.1 y 4.1.2, ya que toda sucesión de estimadores obtenida de incrementar el tamaño de muestra siempre puede representarse como una sucesión constante. Para el caso estudiado en la Subsección 4.1.3, el ‘aumentar el tamaño de muestra’ carece de sentido, lo que debe hacerse entonces es aumentar el tiempo de observación, de modo que el Teorema 3.2.7 pueda aplicarse.

#### 4.4 Intervalos de confianza

Debido a que la Distribución Poisson es discreta, no es posible construir intervalos de confianza para  $\lambda$  con total exactitud para un determinado nivel de confianza  $1 - \alpha$ ,  $\alpha \in (0, 1)$ . Aproximar los límites de confianza para  $\lambda$  con nivel  $1 - \alpha$  dado un valor observado para  $X$ , donde  $X$  tiene distribución

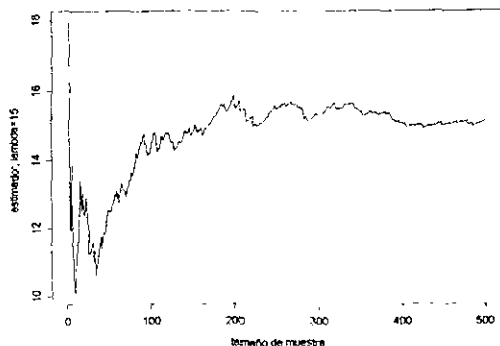


FIGURE 4.3.

$Poi(\lambda)$  se concreta a resolver las siguientes ecuaciones

$$\exp(-\lambda_I) \sum_{j=x}^{\infty} \frac{\lambda_I^j}{j!} = \frac{\alpha}{2} \tag{4.12}$$

$$\exp(-\lambda_S) \sum_{j=0}^{\tau} \frac{\lambda_S^j}{j!} = \frac{\alpha}{2}, \tag{4.13}$$

para  $\lambda_I$ ,  $\lambda_S$ , respectivamente. Esto puede utilizarse para los planes de muestreo presentados en las Secciones 4.1.1 y 4.1.3. Se calcula  $(\lambda_I, \lambda_S)$  a partir del número de ocurrencias en  $(0, \tau]$ , éste es un intervalo para  $\lambda_I$  (el intervalo para  $\lambda$  resulta de dividir  $(\lambda_I, \lambda_S)$  entre  $\tau$ ).

Del Teorema 2.2.11 se sigue que  $\Pr\{Y > y\} = \Pr\{X < \zeta\}$  donde  $X$  tiene distribución  $Poi(y)$  y  $Y$  se distribuye  $Gam(\zeta, 1)$ . Esto transforma el problema de resolver las Ecuaciones (4.12) y (4.13) en el de encontrar cuantiles de una distribución Gamma, los cuales pueden obtenerse utilizando una computadora (el comando para S-PLUS es `qgamma`). Si  $\zeta$  es de la forma  $v/2$ , donde  $v$  es un entero par, se tiene que  $\Pr\{X < v/2\} = \Pr\{W > 2y\}$ , donde  $W \sim \chi^2(v)$ . Sea  $x$  el número de ocurrencias observado. Entonces las Ecuaciones (4.12) y (4.13) pueden escribirse de la siguiente forma

$$\begin{aligned} \lambda_I &= .5c_{\alpha/2} \\ \lambda_S &= 5c_{1-\alpha/2} \end{aligned}$$

Donde  $c_{\alpha/2}$  es el cuantil  $\alpha/2$  de una distribución  $\chi^2(2x)$ , y  $c_{1-\alpha/2}$  corresponde al cuantil  $1 - \alpha/2$  de una distribución  $\chi^2(2(x+1))$ .

En cuyo caso los valores de  $\lambda_I$  y  $\lambda_S$  pueden hallarse mediante interpolación (con respecto al número de grados de libertad) en tablas de percentiles para la función de distribución de la Ji-cuadrada Central.

Para el caso contemplado en la Sección 4.3, se pueden construir cantidades pivotaes [ver Lindsey] a partir de la variable aleatoria  $S_k$ . Aquí  $\lambda$  sirve como parámetro de escala, entonces  $\lambda S_k$  se distribuye  $\text{Gam}(k, 1)$ . Por la misma razón  $2\lambda S_k$  tiene distribución  $\text{Xi}(2k)$ . Ahora, sean  $a$  y  $b$  puntos tales que

$$\Pr\{W < a\} = \frac{\alpha}{2} \quad \text{y} \quad \Pr\{W > b\} = \frac{\alpha}{2},$$

donde  $\alpha$  es el nivel de confianza deseado y  $W$  es una variable aleatoria con distribución  $\text{Xi}(2k)$ . Entonces

$$\Pr\left(\frac{a}{2S_k} < \lambda < \frac{b}{2S_k}\right) = 1 - \alpha,$$

por lo tanto  $(a/2S_k, b/2S_k)$  es un intervalo de confianza al  $1 - \alpha$  para  $\lambda$ .

## 4.5 Inspección continua dado un plan de muestreo

### 4.5.1 El caso determinístico

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson Estacionario de intensidad  $\lambda$ . Bajo el supuesto de que sólo es permitido conocer el estado del proceso en una cantidad finita de tiempos  $v_1, v_2, \dots, v_m$ , se dispone sólo de la información contenida en  $N_{v_1}, N_{v_2}, \dots, N_{v_m}$ . Las variables aleatorias observables de interés serán el número de ocurrencias en cada uno de los lapsos  $(0, v_1], (v_1, v_2], \dots, (v_{m-1}, v_m]$ , estas variables se denotan con  $Z_1, Z_2, \dots, Z_m$ .

De este modo

$$Z_k = N_{v_k} - N_{v_{k-1}} \quad k \in \{1, 2, \dots, m\}.$$

Esto permite obtener una función de verosimilitud bastante simple, esta es

$$L_\lambda \propto \prod_{k=1}^m \frac{e^{-\lambda \psi_k} (\lambda \psi_k)^{z_k}}{z_k!}, \quad (4.14)$$

donde  $\psi_k = v_k - v_{k-1}$ .

Esto se verifica a continuación:

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con intensidad  $\lambda$ , entonces para cualesquiera  $s, t \geq 0$ ,

$$\begin{aligned} \Pr\{N_{t+s} - N_t = k \mid N_u; u \leq t\} &= \Pr\{N_{t+s} - N_t = k\} \\ &= \frac{e^{-\lambda s} (\lambda s)^k}{k!} \quad k = 0, 1, \dots \end{aligned}$$

La primera igualdad resulta del hecho de que se trata de eventos en intervalos ajenos, mientras que la segunda es consecuencia de la estacionariedad del proceso.

Por lo tanto

$$L_\lambda \propto \lambda^n e^{-\lambda t} \prod_{k=1}^m \frac{\psi_k^{z_k}}{z_k!},$$

donde  $n = \sum_{k=1}^m z_k$  es el número total de eventos ocurridos hasta el tiempo  $t = \sum_{k=1}^m \psi_k$ . Esto implica que el estimador máximo verosímil es

$$\hat{\lambda} = \frac{N_t}{t},$$

que es el mismo que se había obtenido anteriormente.

4.5.2 El caso aleatorio

El siguiente paso será hacer que las cantidades  $v_k$  sean variables aleatorias. Se supondrá que las longitudes  $\psi_k$  son variables aleatorias independientes e idénticamente distribuidas. Se hará el supuesto adicional de que dichas variables sean independientes del proceso  $\{N_t : t \geq 0\}$ . Entonces al observar las variables  $N_{v_1}, N_{v_2}, \dots, N_{v_m}$  se cae en un caso similar al descrito en la subsección anterior. Sin embargo ha ocurrido un cambio, ahora el muestreo se lleva a cabo en dos etapas: primero,  $m$  observaciones  $\{\Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_m = \psi_m\}$  son obtenidas de forma aleatoria a partir de cierta densidad  $f_\Psi(\psi)$ , entonces en los instantes  $v_1 = \psi_1, v_2 = \psi_1 + \psi_2, \dots, v_m = \psi_1 + \psi_2 + \dots + \psi_m$  el proceso  $\{N_t : t \geq 0\}$  es observado.

La verosimilitud se basa en el evento

$$\{Z_k = z_k \mid V_1 = v_1, V_2 = v_2, \dots, V_m = v_m\},$$

el cual equivale a

$$\begin{aligned} & \{N_{V_k} - N_{V_{k-1}} = z_k \mid V_1 = v_1, V_2 = v_2, \dots, V_m = v_m\} \\ = & \{N_{v_k} - N_{v_{k-1}} = z_k \mid V_1 = v_1, V_2 = v_2, \dots, V_m = v_m\}. \end{aligned}$$

Debido a que las variables aleatorias  $V_k$  son independientes del proceso  $\{N_t : t \geq 0\}$ , se tiene que

$$\begin{aligned} & \{Z_k = z_k \mid \Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_m = \psi_m\} \\ = & \{N_{v_k} - N_{v_{k-1}} = z_k\}. \end{aligned}$$

Por lo tanto, la función de verosimilitud es

$$\begin{aligned} L_\lambda & \propto \prod_{k=1}^m P_\Gamma \{Z_k = z_k \mid \Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_m = \psi_m\} \\ & = \prod_{k=1}^m P_\Gamma \{N_{v_k} - N_{v_{k-1}} = z_k\} \\ & = \prod_{k=1}^m \frac{e^{-\lambda\psi_k} (\lambda\psi_k)^{z_k}}{z_k!}. \end{aligned}$$

El estimador máximo verosímil de  $\lambda$  (condicionando con  $T = t$  que es el tiempo total de observación) es  $\hat{\lambda} = N_t/t$  el mismo que se había obtenido anteriormente.

Se procede ahora a calcular el estimador sin condicionar con el vector

$$\{\Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_m = \psi_m\}.$$

Se tiene la igualdad [ver Ash]

$$\begin{aligned} & f_{Z_k, \Psi_1, \Psi_2, \dots, \Psi_m}(z_k, \psi_1, \psi_2, \dots, \psi_m) \\ &= f_{\Psi_1, \Psi_2, \dots, \Psi_m}(\psi_1, \psi_2, \dots, \psi_m) \\ & \cdot \Pr\{Z_k = z_k \mid \Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_m = \psi_m\}. \end{aligned}$$

De donde resulta inmediato que

$$\begin{aligned} L_\lambda &\propto \prod_{k=1}^m f_{\Psi_1, \Psi_2, \dots, \Psi_m}(\psi_1, \psi_2, \dots, \psi_m) \Pr\{N_{v_k} - N_{v_{k-1}} = z_k\} \\ &= \prod_{k=1}^m \Pr\{N_{v_k} - N_{v_{k-1}} = z_k\}. \end{aligned}$$

Es claro que el estimador máximo verosímil es nuevamente  $\hat{\lambda} = N_T/T$ .

Se ha usado el hecho de que, aún para tiempos aleatorios, el Proceso Poisson posee la Propiedad de Incrementos Independientes (ver Teorema 3.1.8).

La distribución no condicionada de  $\hat{\lambda}$  depende de la distribución de  $N_T$ , así como de la distribución de  $T$ . Esto implica que los atributos de  $\hat{\lambda}$  como estimador dependerán de la elección de la distribución de  $\Psi$ .

A continuación se presenta un análisis de las propiedades de este estimador para el caso  $\Psi \sim \text{Gam}(\phi)$ .

#### 4.5.3 Tiempos de observación aleatorios con distribución Exponencial

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson de intensidad  $\lambda$ . Se considera el plan de muestreo y el estimador  $\hat{\lambda} = N_T/T$  discutidos en la Subsección 4.5.2. Se supone que los tiempos entre observaciones siguen una ley  $\text{Exp}(\phi)$ , con  $\phi > 0$ , entonces  $T$  tiene distribución  $\text{Gam}(n_0, \phi)$ , donde  $n_0$  es el tamaño de muestra.

Este plan de muestreo es el adecuado cuando el proceso es observado usando un dispositivo cuyo tiempo de vida es exponencial.

**Ejemplo 4.5.1** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson de intensidad  $\theta$  desconocida. Las observaciones consisten en el número de ocurrencias del proceso registradas por cierto dispositivo electrónico. Se hace el supuesto de que este aparato tiene tiempo de vida útil con distribución  $\text{Exp}(\phi)$ . Cuando el dispositivo ‘falla’ es reemplazado inmediatamente por otro con las mismas características. Es razonable suponer que los tiempos de vida son independientes entre sí y del número de ocurrencias del

proceso. Se cuenta con el siguiente conjunto de datos:

dispositivo	tiempo de vida	ocurrencias
1	.2193	4
2	.3842	5
3	.1182	9
4	.4621	6

El valor del estimador máximo verosímil es:  $\hat{\lambda} = 20.2736$ . ■

**Teorema 4.5.2** El estimador máximo verosímil para  $\lambda$ , basado en observar el número de ocurrencias del proceso en  $n_0$  intervalos de longitud aleatoria con distribución  $\text{Exp}(\phi)$ , cumple

$$E(\hat{\lambda}) = E(E(\hat{\lambda} | T)) = \lambda$$

y

$$\begin{aligned} \text{var}(\hat{\lambda}) &= E(\text{var}(\hat{\lambda} | T)) + \text{var}(E(\hat{\lambda} | T)) \\ &= E\left(\frac{\lambda}{T}\right) \\ &= \left(\frac{\phi}{n_0 - 1}\right)\lambda. \end{aligned} \tag{4.15}$$

**Demostración.** Es claro que

$$\begin{aligned} E(\hat{\lambda} | T = t) &= \sum_{n=0}^{\infty} \frac{n}{t} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= \frac{1}{t} \sum_{n=0}^{\infty} n e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= \frac{1}{t} (\lambda t), \end{aligned}$$

entonces

$$\begin{aligned} E(E(\hat{\lambda} | T)) &= E(\lambda) \\ &= \lambda. \end{aligned}$$

Para la varianza, se tiene que

$$\text{var}(\hat{\lambda}) = E(\text{var}(\hat{\lambda} | T)) + \text{var}(E(\hat{\lambda} | T)).$$

Como  $\hat{\lambda}$  es insesgado para cualquier valor de  $T$ , el segundo sumando es cero; basta entonces calcular

$$\text{var}(\hat{\lambda} | T) = E(\hat{\lambda}^2 | T) - [E(\hat{\lambda} | T)]^2.$$



Trabajando en el primer sumando del lado derecho, obtiene

$$\begin{aligned}
 E(\widehat{\lambda}^2 \mid T = t) &= \sum_{n=0}^{\infty} \frac{n^2}{t^2} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
 &= \frac{1}{t^2} \sum_{n=0}^{\infty} n^2 e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
 &= \frac{1}{t^2} [\lambda t + \lambda^2 t^2] \\
 &= \frac{\lambda}{t} + \lambda^2.
 \end{aligned}$$

Por lo tanto

$$\text{var}(\widehat{\lambda} \mid T = t) = \frac{\lambda}{t},$$

entonces

$$\begin{aligned}
 \text{var}(\widehat{\lambda}) &= E\left(\frac{\lambda}{T}\right) + \text{var}(E(\widehat{\lambda} \mid T)) \\
 &= E\left(\frac{\lambda}{T}\right) + \text{var}(\lambda) \\
 &= E\left(\frac{\lambda}{T}\right).
 \end{aligned}$$

Sólo resta mostrar que  $E\left(\frac{\lambda}{T}\right)$  es de la forma deseada, ésto se sigue del siguiente desarrollo:

$$\begin{aligned}
 E\left(\frac{\lambda}{T}\right) &= \int_0^{\infty} \frac{\lambda \phi (\phi t)^{n_0-1} e^{-\phi t}}{t (n_0-1)!} dt \\
 &= \frac{\lambda \phi}{(n_0-1)!} \int_0^{\infty} (\phi t)^{n_0-2} e^{-\phi t} \phi dt \\
 &= \frac{\lambda \phi}{(n_0-1)!} \Gamma(n_0-1) \\
 &= \frac{\lambda \phi}{n_0-1},
 \end{aligned}$$

lo cual concluye la prueba. ■

Se procede ahora a encontrar la función de distribución de  $\widehat{\lambda}$ , para ello primero se obtendrá la distribución de  $N_T$ .

Aplicando el Teorema de la Probabilidad Total, se sigue que

$$\begin{aligned}
 \Pr(N_{V_k} - N_{V_{k-1}} = z) &= \Pr(N_{V_k} - N_{V_{k-1}} = z, \Psi_k \in (0, \infty)) \\
 &= \Pr(N_{V_k} - N_{V_{k-1}} = z, \cup_{t \in \mathbb{R}^+ \cup \{0\}} \Psi_k = t) \\
 &= \int_0^\infty \Pr(N_{V_k} - N_{V_{k-1}} = z, \Psi_k = t) dt \\
 &= \int_0^\infty \Pr(N_{V_k} - N_{V_{k-1}} = z \mid \Psi_k = t) f_{\Psi_k}(t) dt \\
 &= \int_0^\infty \frac{e^{-t\lambda} (t\lambda)^z}{z!} (\phi e^{-\phi t}) dt \\
 &= \int_0^\infty e^{-t(\phi+\lambda)} t^z \frac{\phi \lambda^z}{z!} dt \\
 &= \frac{\phi \lambda^z}{z! (\lambda + \phi)^{z+1}} \int_0^\infty e^{-t(\lambda+\phi)} [t(\lambda + \phi)]^z (\lambda + \phi) dt \\
 &= \frac{\phi \lambda^z}{z! (\lambda + \phi)^{z+1}} \Gamma(z + 1).
 \end{aligned}$$

Por lo tanto

$$\Pr(N_{V_k} - N_{V_{k-1}} = z) = (1 - p)p^z, \tag{4.16}$$

donde  $p = \lambda/(\phi + \lambda)$ . Se sigue que  $N_{V_k} - N_{V_{k-1}} \sim \text{Geo}(p)$  (En esta parametrización se cuenta el número de fracasos antes del primer éxito). Por lo tanto  $N_T$  se distribuye  $\overline{\text{Bi}}(n_0, p)$ .

Aplicando la Regla de Bayes, se obtiene:

$$\begin{aligned}
 f_{T|N_T}(t; n) &= \frac{\Pr\{N_T = n \mid T = t\} f_T(t)}{\int_0^\infty \Pr\{N_T = n \mid T = t\} f_T(t) dt} \\
 &= \frac{e^{-\lambda t} \frac{(\lambda t)^n}{n!} \frac{\phi^{n_0}}{\Gamma(n_0)} t^{n_0-1} e^{-\phi t}}{\binom{n_0+n-1}{n} \left(\frac{\phi}{\phi+\lambda}\right)^{n_0} \left(\frac{\lambda}{\phi+\lambda}\right)^n} \\
 &= \frac{(\phi + \lambda)^{n_0+n}}{\Gamma(n_0 + n)} t^{n_0+n-1} e^{-(\phi+\lambda)t}.
 \end{aligned} \tag{4.17}$$

Es decir,  $T$  condicionado con  $\{N_T = n\}$  tiene una distribución  $\mathbf{Gam}(n_0 + n, \phi + \lambda)$ , entonces, haciendo uso del Teorema de la Probabilidad Total se tiene que

$$\begin{aligned} \Pr \left\{ \frac{N_T}{T} \leq x \right\} &= \sum_{n=0}^{\infty} \Pr \left\{ \frac{N_T}{T} \leq x \mid N_T = n \right\} \Pr \{N_T = n\} \\ &= \sum_{n=0}^{\infty} \Pr \left\{ \frac{n}{x} \leq T \mid N_T = n \right\} \Pr \{N_T = n\} \\ &= \sum_{n=0}^{\infty} \left( 1 - F_{D_n} \left( \frac{n}{x} \right) \right) \Pr \{N_T = n\} \\ &= \sum_{n=0}^{\infty} \left( 1 - F_{D_n} \left( \frac{n}{x} \right) \right) \binom{n_0 + n - 1}{n} \left( \frac{\phi}{\phi + \lambda} \right)^{n_0} \left( \frac{\lambda}{\phi + \lambda} \right)^n, \end{aligned} \quad (4.18)$$

donde  $D_n$  es una variable aleatoria con distribución  $\mathbf{Gam}(n_0 + n, \phi + \lambda)$ .

Es claro que

$$\begin{aligned} &\sum_{k=0}^{\infty} \binom{k + n_0 - 1}{k} \left( \frac{\phi}{\phi + \lambda} \right)^{n_0} \left( \frac{\lambda}{\phi + \lambda} \right)^k \left[ 1 - F_{D_k} \left( \frac{k}{x} \right) \right] \\ &< \sum_{k=0}^{\infty} \binom{k + n_0 - 1}{k} \left( \frac{\phi}{\phi + \lambda} \right)^{n_0} \left( \frac{\lambda}{\phi + \lambda} \right)^k \\ &= 1, \end{aligned}$$

para cualquier valor de  $x$ . Se sigue de la Prueba M de Weierstrass [Widder] que esta serie converge de manera uniforme. Por ser cada sumando de la serie una función continua, y gracias a la convergencia uniforme, se tiene que la función de distribución de  $\hat{\lambda}$  también lo es. Desafortunadamente, no es válido derivar la Expresión (4.18) término a término, por lo cual no se dispone de la función de densidad del estimador. Por otra parte, es posible calcular la distribución asintótica para este caso, esto se tratará en la Sección 4.7.

Aquí se muestra una gráfica (Figura 4.4) de la función de distribución de este estimador cuando  $\phi = \lambda = 1$  y  $n_0 = 4$ .

Siguiendo un razonamiento análogo al hecho en la Sección 4.3, sólo que ahora

$$\Pi(\omega) = \{V_1(\omega), V_2(\omega), \dots\} \quad \text{donde } V_k = \sum_{i=1}^k \Psi_i,$$

se concluye que el estimador máximo verosímil es fuertemente consistente.

La función de densidad (en el sentido general, ver apéndice) para un solo tiempo de observación es

$$f_{N,T}(n, t) = (\lambda + \phi) t^n e^{-n \ln(\lambda + \phi) - t(\lambda + \phi)}.$$

Por lo tanto, la densidad conjunta de  $n$  y  $t$  pertenece a la Familia Exponencial. Aquí el parámetro natural y la estadística correspondiente vienen dados por

$$W(n, t) = (n, t)$$

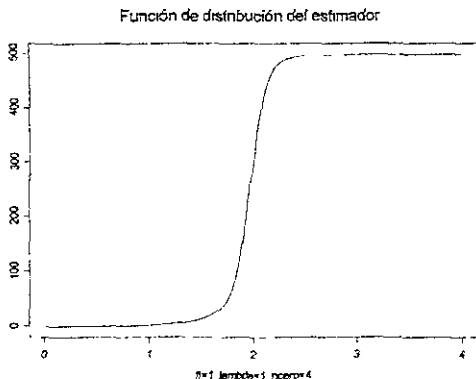


FIGURE 4.4.

y

$$\theta = (\ln(\phi + \lambda), \phi + \lambda).$$

Entonces el espacio paramétrico natural para este caso es  $\Theta = \mathbb{R} \times (0, \infty)$ , el cual carece de puntos frontera, por lo tanto  $W$  es una estadística completa. Del Lema de Factorización de Neyman-Fisher (ver apéndice) resulta inmediato que  $W$  es una estadística suficiente. Es claro que el estimador máximo verosímil para  $\lambda$  es una función de  $W$ . Aplicando el Teorema Lehmann-Scheffé se concluye que el estimador máximo verosímil es el estimador insesgado de varianza mínima para  $\lambda$ .

Con el fin de desarrollar cierta intuición acerca del comportamiento de este estimador se hará uso de simulaciones.

Esta es una rutina programada en S-PLUS que simula un Proceso Poisson de intensidad  $\lambda$  y calcula el estimador para un valor predeterminado de  $\phi$ .

```
# Se proponen  $\phi$ ,  $\lambda$  y el tamaño de muestra.
fi<-valor de  $\phi$ 
lambda<-valor de  $\lambda$ 
n<-tamaño de muestra
# Se genera el tiempo de observación.
tiempo<-rgamma(1,n,fi)
ocurre<-0
suma<-0
# Se observa el número de ocurrencias del proceso.
while(suma<tiempo)
{ incr<-rexp(1,lambda)
  suma<-suma+incr
  ocurre<-ocurre+1
```

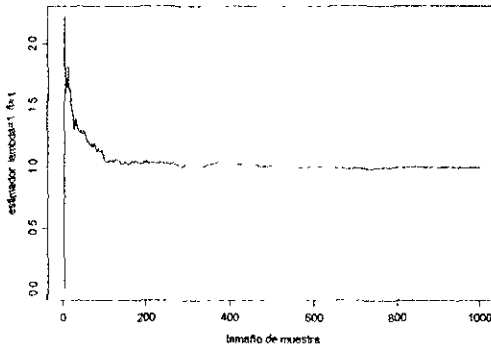


FIGURE 4.5.

```

}
# Se calcula el estimador.
estimador<-(ocurre-1)/tiempo

```

En las Figuras 4.5, 4.6 y 4.7 se muestran las trayectorias simuladas del estimador con  $\lambda = 1$  y  $\phi = 1, 1$  y 250, respectivamente. Cuando se toma  $\phi = 250$ , se observa el siguiente comportamiento: para tamaño de muestra pequeño ( $n_0 < 50$ ), se tiene que  $\hat{\lambda} = 0$ , a partir del momento en el que se registra la primera ocurrencia, el estimador oscila entre valores mayores o iguales al del parámetro. El comparar estas gráficas, sugiere que existe una relación inversa entre  $\phi$  y la velocidad de convergencia ( $\lambda$  fijo), y que dicha relación puede llegar a ser significativa.

Otro experimento consiste en simular el proceso y calcular el estimador para distintos tamaños de muestra ( $\lambda$  y  $\phi$  fijos). En el caso  $\phi = 100$ ,  $\lambda = 1$  se observó que, para tamaño de muestra entre 1 y 100, o bien se tenía  $\hat{\lambda} = 0$  o se obtenía una sobreestimación del parámetro. Las discrepancias entre las estimaciones (para distintos tamaños de muestra) disminuyen al hacer decrecer  $\phi$ . En las Figuras 4.8, 4.9 y 4.10 se muestran los resultados de las simulaciones.

En la Figura 4.11 se ilustra lo que se obtuvo al simular el proceso tomando  $n_0$  y  $\phi$  fijos, y haciendo variar  $\lambda$ . Se observó que había una mayor variabilidad conforme se incrementa  $\lambda$ , sin embargo esto ocurre lentamente, incluso para  $n_0$  pequeño ( $n_0 < 50$ ).

A pesar de que  $\phi$  y  $\lambda$  juegan un papel simétrico en la Expresión (4.15), es muy distinto el comportamiento que se observa al hacer simulaciones con  $\lambda$  y  $n_0$  fijos y considerando distintos valores de  $\phi$  respecto a lo obtenido en el experimento anterior. Los resultados de estas simulaciones pueden verse en las Figuras 4.12, 4.13 y 4.14, donde se tomó  $\lambda = 1$ ,  $n = 10$ ,  $n = 50$  y  $n = 5000$ .

Este estudio de simulación sirve como evidencia empírica para la siguiente afirmación: El emplear  $\phi$  mucho mayor a  $\lambda$  en este plan de muestreo tiene un impacto negativo en el desempeño del estimador en términos de convergencia y de error cuadrático medio. Esto adquiere mayor relevancia cuando el tamaño de muestra es pequeño, lo cual se explica en parte, desde el punto de vista analítico, a partir

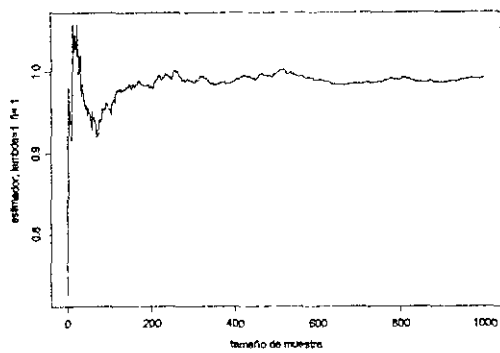


FIGURE 4.6.

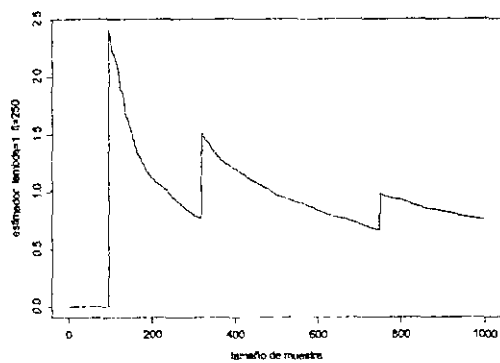


FIGURE 4.7.

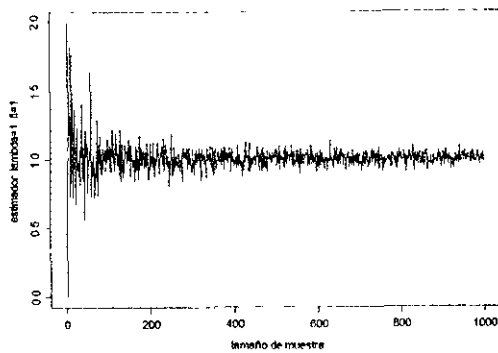


FIGURE 4.8.

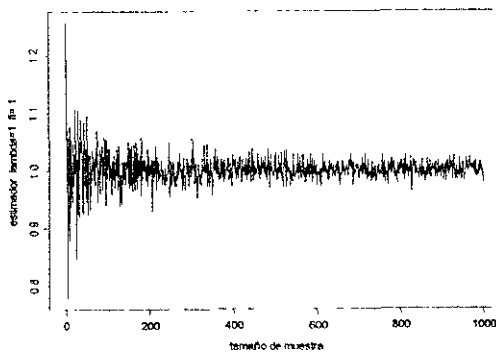


FIGURE 4.9.

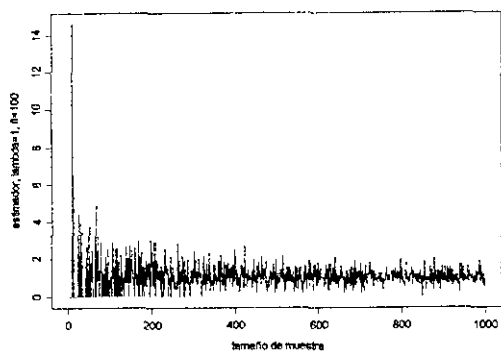


FIGURE 4.10.

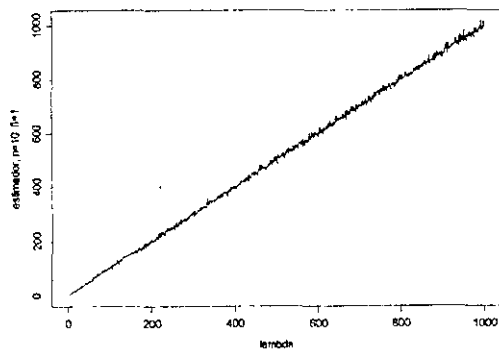


FIGURE 4.11.



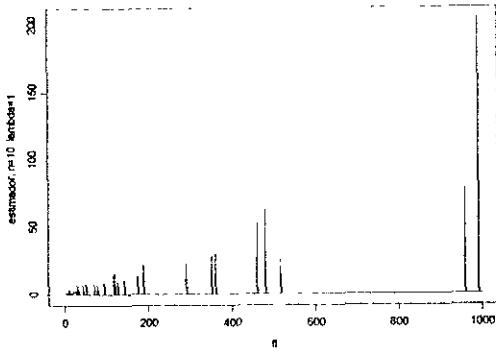


FIGURE 4.12.

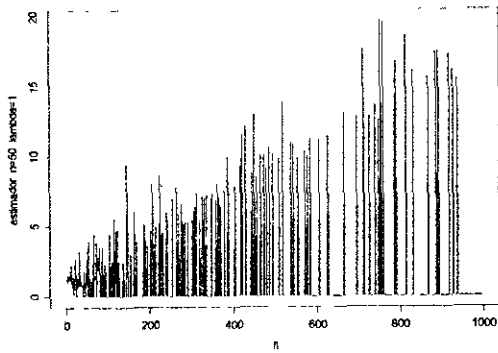


FIGURE 4.13.

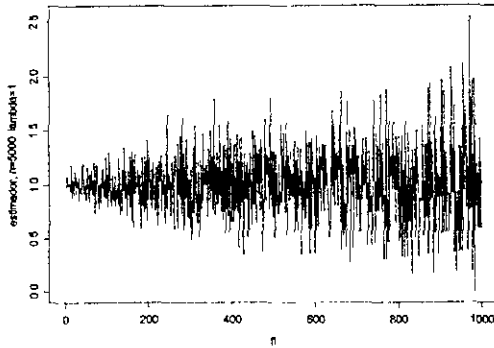


FIGURE 4.14.

de la Figura 4.15, aquí puede verse que la varianza del estimador (vista como función de  $\phi$  y  $n_0$ ) se incrementa drásticamente cuando se toma  $\phi$  grande y  $n_0$  pequeño. En la Figura 4.16 se presenta una gráfica en curvas de nivel de la varianza (como función de  $\phi$  y  $\lambda$ ), tomando  $n_0 = 10$ .

Con el siguiente algoritmo se pretende aminorar este efecto:

**Algoritmo 4.5.3** -Procedimiento en 2 fases para estimar  $\lambda$ , usando tiempos aleatorios de observación con distribución Exp.-

- i) Se propone un  $\phi$  inicial (éste se denota con  $\phi_1$ ).
- ii) Se toman  $K$  observaciones en tiempos aleatorios con distribución  $\text{Exp}(\phi_1)$ , donde

$$K = \inf_{n \in \{1, 2, \dots\}} \{N_{T_n} > 0\}.$$

Aquí  $T_n$  denota el tiempo transcurrido desde el inicio del experimento, hasta el  $n$ -ésimo tiempo aleatorio.

- iii) Sea  $\phi_2 = \frac{\phi_1}{K}$ .
- iv) Se observa al proceso de acuerdo al plan de muestreo original, usando  $\phi_2$  como parámetro para los tiempos de observación, se toman  $n_0$  tiempos aleatorios.
- v) Se calcula el estimador  $\hat{\lambda} = \frac{N_{T_2}}{T_2}$ , donde  $N_{T_2}$  y  $T_2$  son el número de ocurrencias, y el tiempo invertido en esta segunda etapa.

Se están haciendo los siguientes supuestos:

- i) El investigador controla  $\phi$ .
- ii) Durante la primera parte del algoritmo, se puede observar el proceso durante una cantidad indefinida de tiempos aleatorios.

Una forma de interpretar ésto, es que se trata de un experimento que se lleva acabo en dos etapas, donde la primera etapa es de tipo 'exploratorio', y sirve para obtener una primera estimación de  $\lambda$ .

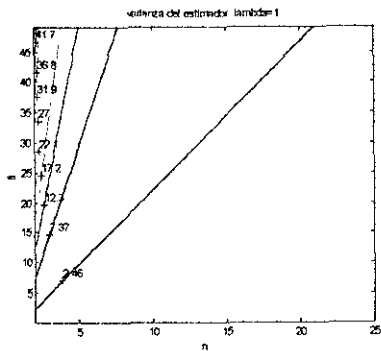


FIGURE 4.15.

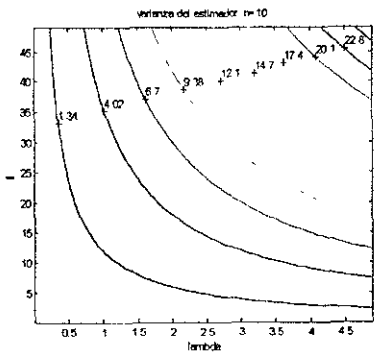


FIGURE 4.16

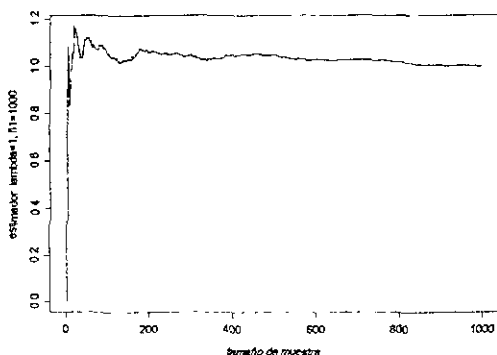


FIGURE 4.17.

A partir de dicha estimación se elige una nueva  $\phi$ , la cual se espera, permita obtener una segunda estimación del parámetro de modo que se disminuya la varianza y que aún con pocos tiempos de observación para esta segunda etapa, dicha estimación será cercana al verdadero valor de  $\lambda$ .

En las Figuras 4.17 y 4.18 se presentan trayectorias simuladas de este tipo de estimadores. Sólo se considera el comportamiento del estimador para la segunda parte del experimento (se procede de igual modo en el resto de las simulaciones). Se usó una  $\phi_1 = 1$ , 1000, y  $\lambda = 1$ .

Se realizaron también simulaciones con  $\phi_1$  y  $\lambda$  fijos, y tomando distintos tamaños de muestra, esto puede verse en las Figuras 4.19 y 4.20.

Otro experimento consiste efectuar simulaciones del estimador tomando  $\lambda = 1$  y  $n = 10$  y variando  $\phi_1$ . Los resultados se muestran en la Figura 4.21.

De acuerdo a las simulaciones, el estimador modificado acelera la convergencia y disminuye la varianza. Una justificación para este comportamiento se obtiene del siguiente resultado:

**Teorema 4.5.4** Sea  $\hat{\lambda} = \frac{N T_2}{T_2}$  como se describió en el algoritmo, entonces, la varianza del estimador viene dada por la siguiente expresión.

$$var(\hat{\lambda}) = \frac{\lambda^2}{(n_0 - 1)} \ln\left(\frac{\lambda + \phi_1}{\lambda}\right).$$

**Demostración.** Debido a que, para cualquier elección de  $\phi_1$ , el estimador  $\hat{\lambda}$  es insesgado, se tiene la siguiente igualdad:

$$\begin{aligned} var(\hat{\lambda}) &= E(var(\hat{\lambda} | K)) \\ &= \frac{\lambda \phi_1}{(n_0 - 1)} E\left(\frac{1}{K}\right). \end{aligned}$$

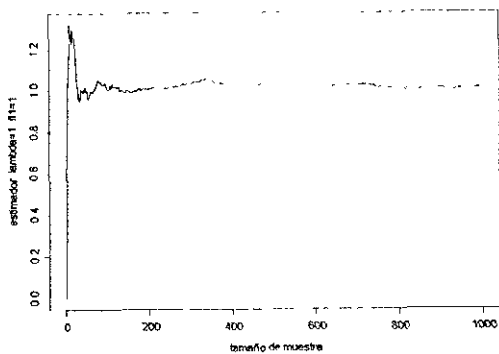


FIGURE 4.18.

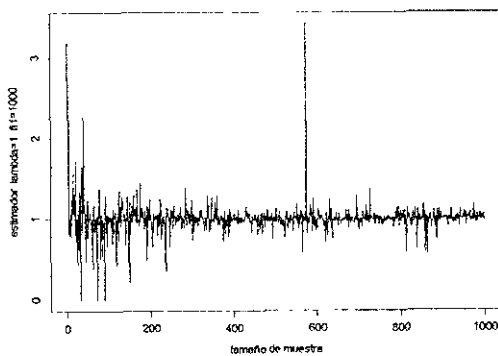


FIGURE 4.19.

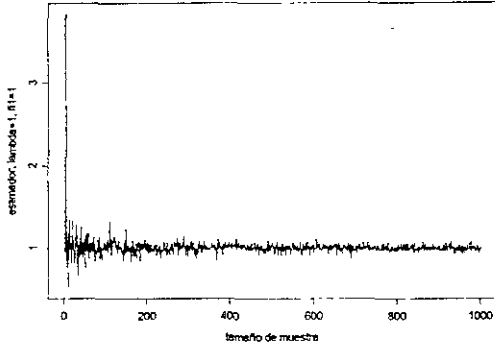


FIGURE 4.20.

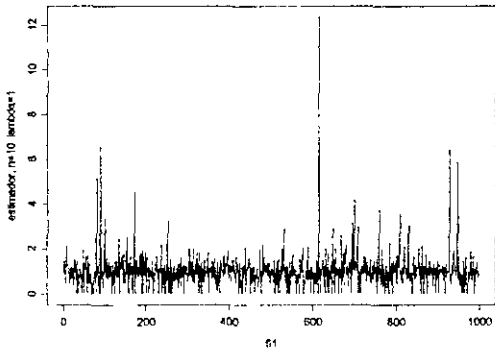


FIGURE 4.21.

De acuerdo a la definición de varianza condicional [Casella y Berger]

$$\begin{aligned}
 \text{var}(\hat{\lambda} \mid K) &= E(\hat{\lambda}^2 \mid K) - (E(\hat{\lambda} \mid K))^2 \\
 &= E(\hat{\lambda}^2 \mid K) - \lambda^2 \\
 &= E(E(\hat{\lambda}^2 \mid T) \mid K) - \lambda^2 \\
 &= E\left(\frac{\lambda}{T} + \lambda^2 \mid K\right) - \lambda^2 \\
 &= \frac{\lambda\phi_1}{K(n_0 - 1)}.
 \end{aligned}$$

Se procede ahora a calcular esta esperanza

$$\begin{aligned}
 E\left(\frac{1}{\bar{K}}\right) &= \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{\phi_1}{\lambda + \phi_1}\right)^{k-1} \frac{\lambda}{\lambda + \phi_1} \\
 &= \frac{\lambda}{\phi_1} \sum_{k=1}^{\infty} \frac{\left(\frac{\phi_1}{\lambda + \phi_1}\right)^k}{k} \\
 &= \frac{\lambda}{\phi_1} \ln\left(\frac{\lambda + \phi_1}{\lambda}\right).
 \end{aligned}$$

Se concluye que la varianza del estimador viene dada por

$$\text{var}(\hat{\lambda}) = \frac{\lambda^2}{(n_0 - 1)} \ln\left(\frac{\lambda + \phi_1}{\lambda}\right),$$

que es lo que se quería demostrar. ■

Este teorema tiene la siguiente implicación: mientras que para el plan de muestreo original la varianza del estimador es una función lineal de  $\phi$ , para el muestreo modificado la varianza se incrementa de manera logarítmica respecto a  $\phi_1$ .

También se exploró la posibilidad de mejorar el estimador mediante un algoritmo similar, cuando  $\lambda$  es mucho mayor a  $\phi$ . Nuevamente el objetivo es disminuir la varianza y aumentar la velocidad de convergencia. A continuación se presenta el algoritmo con el que se pretende lograr esto:

**Algoritmo 4.5.5** -Corrección del estimador cuando  $\lambda$  es mucho mayor a  $\phi$ .

- i) proponer un  $\phi_1$  inicial.
- ii) Observar el proceso durante un tiempo aleatorio  $T_1$  con distribución  $\text{Exp}(\phi_1)$ .
- iii) Sea  $\rho = \frac{N_{T_1}}{T_1}$ , y hacer  $\phi_2 = \frac{1}{\rho}$  si  $N_{T_1} \geq 1$  y  $\phi_2 = \phi_1$  en otro caso.
- iv) Estimar  $\lambda$  de acuerdo al plan de muestreo original, usando tiempos entre observaciones con distribución  $\text{Exp}(\phi_2)$ .

A continuación se enlista el programa en S-PLUS que realiza estas simulaciones.

**#Proponer  $\phi_1$ , el vector de intensidades y el tamaño de muestra.**

```

fi1<-valor de  $\phi_1$ 
lambda<-vector de valores de  $\lambda$ 
n<-tamaño de muestra
m<-length(lambda)
#Se genera el primer tiempo de observación
tiempo1<-rgamma(1,1,fi1)
ocurre1<-0
suma1<-0
#Se observa el número de ocurrencias del proceso
while(suma1<tiempo1)
{ incr<-rexp(1,lambda)
  suma1<-suma1+incr
  ocurre1<-ocurre1+1
}
# Se obtiene  $\phi_2$ 
rho<-(ocurre1-1)/tiempo1
if(rho==0)
{
fi2<-fi1
}
else
{
fi2<-1/rho
}
# Se lleva a cabo la rutina para calcular  $\hat{\lambda}$  presentada anteriormente,
sólo que haciendo  $\phi = \phi_2$ .

```

Se denota con  $\hat{\lambda}_i$  al estimador de  $\lambda$  obtenido al usar  $\phi_i$ , ( $i \in \{1, 2\}$ ) para generar los tiempos de observación.  $\hat{\lambda}_2$  es el estimador que se usará finalmente. Se procederá a obtener su varianza.

**Teorema 4.5.6** *La varianza del estimador  $\hat{\lambda}_2$  viene dada por la siguiente expresión:*

$$\frac{\lambda \phi_1^2}{(n_0 - 1)(\lambda + \phi_1)} + \left( \frac{\lambda}{(\lambda + \phi_1)(n_0 - 1)} \right) \left[ \left( \frac{\lambda}{\lambda + \phi_1} \right) + \left( \frac{\phi_1}{\lambda + \phi_1} \right) \ln \left( \frac{\phi_1 + \lambda}{\phi_1} \right) \right].$$

**Demostración.** Se tiene

$$\begin{aligned} \text{var}(\hat{\lambda}_2) &= E \left[ \text{var}(\hat{\lambda}_2 \mid T_1 = t, N_{T_1} = n) \right] \\ &\quad + \text{var} \left[ E(\hat{\lambda}_2 \mid T_1 = t, N_{T_1} = n) \right] \\ &= E \left[ \text{var}(\hat{\lambda}_2 \mid T_1 = t, N_{T_1} = n) \right], \end{aligned}$$

ya que  $E(\hat{\lambda}_2 \mid T_1 = t, N_{T_1} = n) = \lambda$ , para todo  $t$  y  $n$ .



De acuerdo a las Expresiones (4.16) y (4.17), la densidad conjunta de  $(T_1, N_{T_1})$  viene dada por

$$\begin{aligned} f_{T_1, N_{T_1}}(t, n) &= f_{T_1 | N_{T_1}}(t, n) \Pr \{N_{T_1} = n\} \\ &= \frac{(\phi + \lambda)^{n+1}}{\Gamma(n+1)} \cdot t^n e^{-(\phi + \lambda)t} \\ &\quad \cdot \left(\frac{\phi}{\phi + \lambda}\right) \left(\frac{\lambda}{\phi + \lambda}\right)^n. \end{aligned}$$

Se sigue que

$$\begin{aligned} &E \left[ \text{var} \left( \hat{\lambda}_2 \mid T_1 = t, N_{T_1} = n \right) \right] \\ &= \frac{\lambda \phi_1}{n_0 - 1} \int_0^\infty f_{T_1, N_{T_1}}(t, 0) dt + \sum_{n=1}^\infty \int_0^\infty \frac{\lambda t}{(n_0 - 1)n} f_{T_1, N_{T_1}}(t, n) dt \\ &= \frac{\lambda \phi_1}{n_0 - 1} \int_0^\infty f_{T_1, N_{T_1}}(t, 0) dt \\ &\quad + \sum_{n=1}^\infty \frac{\lambda}{(n_0 - 1)n} \left[ \int_0^\infty t f_{T_1 | N_{T_1}}(t, n) dt \right] \Pr \{N_{T_1} = n\} \\ &= \frac{\lambda \phi_1}{n_0 - 1} \int_0^\infty f_{T_1, N_{T_1}}(t, 0) dt \\ &\quad + \sum_{n=1}^\infty \frac{\lambda}{(n_0 - 1)n} E[T_1 \mid N_{T_1} = n] \Pr \{N_{T_1} = n\}. \end{aligned}$$

Entonces

$$\text{var}(\hat{\lambda}_2) = \frac{\lambda \phi_1}{n_0 - 1} \Pr \{N_{T_1} = 0\} + \sum_{n=1}^\infty \frac{\lambda}{(\lambda + \phi)(n_0 - 1)} \left(1 + \frac{1}{n}\right) \Pr \{N_{T_1} = n\}.$$

Usando

$$\sum_{x=1}^\infty \frac{\left(\frac{\lambda}{\lambda + \phi}\right)^x}{x} = \ln \left(\frac{\phi + \lambda}{\phi}\right)$$

y

$$\sum_{x=1}^\infty \left(\frac{\lambda}{\lambda + \phi}\right)^x \left(\frac{\phi}{\lambda + \phi}\right) = \left(\frac{\lambda}{\lambda + \phi}\right),$$

se sigue que la varianza de  $\hat{\lambda}_2$  viene dada por

$$\frac{\lambda \phi_1^2}{(n_0 - 1)(\lambda + \phi_1)} + \left(\frac{\lambda}{(\lambda + \phi_1)(n_0 - 1)}\right) \left[ \left(\frac{\lambda}{\lambda + \phi_1}\right) + \left(\frac{\phi_1}{\lambda + \phi_1}\right) \ln \left(\frac{\phi_1 + \lambda}{\phi_1}\right) \right],$$

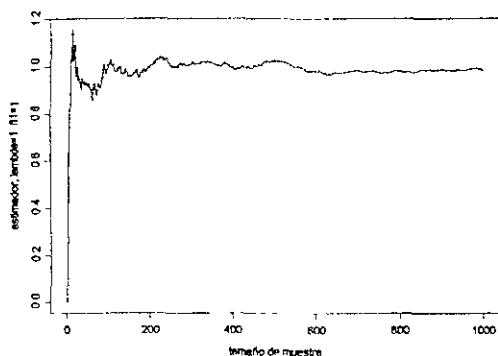


FIGURE 4.22.

lo cual concluye la prueba. ■

Para el caso particular  $\phi_1 = 1$ , se tiene que

$$\begin{aligned} \text{var}(\hat{\lambda}_2) &= \frac{\lambda}{(n_0 - 1)(\lambda + 1)} \\ &\quad + \left( \frac{\lambda}{(\lambda + 1)(n_0 - 1)} \right) \left[ \left( \frac{\lambda}{\lambda + 1} \right) + \left( \frac{1}{\lambda + 1} \right) \ln(1 + \lambda) \right] \\ &< \frac{1}{n_0 - 1} [2 + \ln(1 + \lambda)]. \end{aligned}$$

Es decir  $\text{var}(\hat{\lambda}_2)$  crece en relación a  $\lambda$  al menos tan lentamente como  $\ln(1 + \lambda)$ .

Para este estimador modificado se realizó el mismo estudio de simulación que para el caso anterior. Las trayectorias simuladas del estimador se presentan en las Figuras 4.22 y 4.23. La Figura 4.24 muestra lo que se obtuvo de repetir el experimento con  $\lambda$  y  $\phi_1$  fijos, y haciendo variar el tamaño de muestra. Lo que resultó del experimento que consiste en tomar a  $n_0$  y  $\lambda$  fijos y variar  $\phi_1$  se presenta en la Figura 4.25.

#### 4.6 Realización discretizada

El análisis en la Subsección 4.5.2 se obtuvo a partir de observaciones del proceso en tiempos  $V_1, V_2, \dots, V_{n_0}$  tales que los tiempos entre observaciones sucesivas  $\Psi_k = V_k - V_{k-1}$  eran variables aleatorias independientes e idénticamente distribuidas. Se supuso que los intervalos  $\Psi_k$  eran variables aleatorias observables. Si no se tiene esto, el procedimiento en dos fases no puede aplicarse.

Sea  $N_{V_1}, N_{V_2}, \dots$  las observaciones hechas en tiempos aleatorios  $V_1 < V_2 < \dots < V_{n_0}$  y de tal modo que  $\Psi_k = V_k - V_{k-1}$  sean independientes y distribuidas exponencialmente con parámetro  $\rho$ .

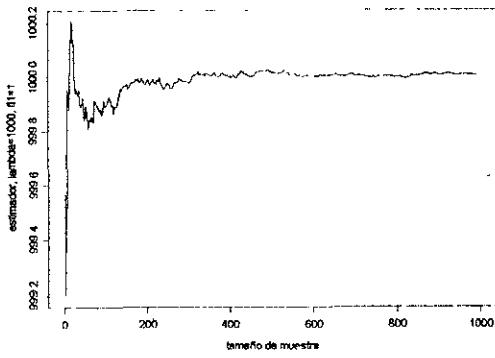


FIGURE 4.23.

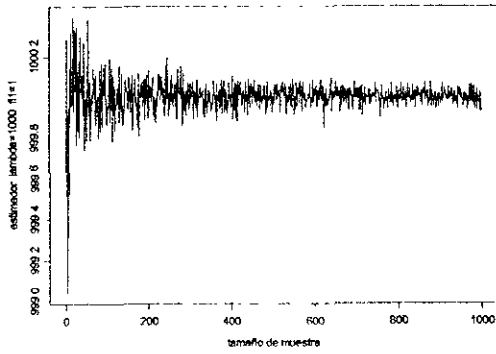


FIGURE 4.24.

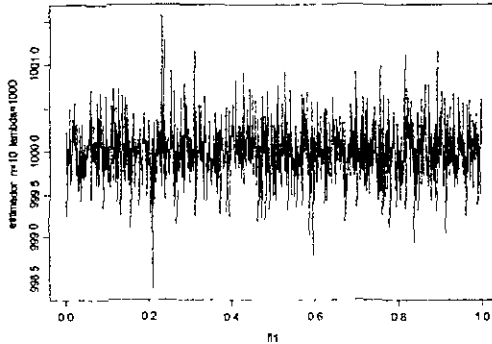


FIGURE 4.25.

Se denota con  $Z_k = N_{V_k} - N_{V_{k-1}}$  al número de eventos ocurridos en  $(V_{k-1}, V_k]$ . Ahora se buscará un estimador para  $\lambda$  basado únicamente en las observaciones  $(z_1, z_2, \dots, z_n)$ . Como  $\{N_t : t \geq 0\}$  es un Proceso Poisson, puede verificarse que  $\{Z_1, Z_2, \dots\}$  son variables independientes e idénticamente distribuidas. La función de densidad de  $Z_k$  se obtuvo en la sección anterior, por lo tanto de la Expresión (4.16) se sigue que la función de verosimilitud es:

$$L_\lambda \propto \left(\frac{\rho}{\rho + \lambda}\right)^{n_0} \left(\frac{\lambda}{\rho + \lambda}\right)^N,$$

donde  $N = \sum_{k=1}^{n_0} Z_k$ , el número total de eventos observado. Entonces el estimador máximo verosímil es  $\hat{\lambda} = \rho N / n_0$ . Se tiene que  $N$  tiene una distribución  $\overline{\text{Bi}}(n_0, p)$ , esto implica

$$E(N) = \frac{n_0 \lambda}{\rho},$$

y

$$var(N) = \frac{n_0 \lambda (\rho + \lambda)}{\rho^2}.$$

Por lo tanto  $E(\hat{\lambda}) = \lambda$  y  $var(\hat{\lambda}) = \frac{\lambda(\rho + \lambda)}{n_0}$ .

#### 4.7 Distribución asintótica de estimadores

La propiedad de asintoticidad normal de un estimador es deseable, ya que simplifica muchos aspectos matemáticos de la inferencia estadística. Se cuenta con un resultado (ver apéndice) que garantiza,

bajo ciertas condiciones de regularidad, la asintoticidad normal de un estimador máximo verosímil. A continuación se enlistan dichas condiciones de regularidad:

**Condición 1.**-Las derivadas  $\partial \ln f_\theta / \partial \theta$ ,  $\partial^2 \ln f_\theta / \partial \theta^2$  y  $\partial^3 \ln f_\theta / \partial \theta^3$  existen, para casi toda  $x$  contenida en un intervalo  $A$  tal que el verdadero valor del parámetro  $\theta_0 \in A$ .

**Condición 2.**-En el punto  $\theta_0$

$$E_{\theta_0} \left[ \frac{\partial f_\theta / \partial \theta}{f_\theta} \right] = 0, \quad E_{\theta_0} \left[ \frac{\partial^2 f_\theta / \partial \theta^2}{f_\theta} \right] = 0$$

**Condición 3.**-Para toda  $\eta \in A$ ,

$$\left| \frac{\partial^3 \ln f_\theta}{\partial \theta^3} \right| < M(x), \quad E_\eta [M(x)] < K$$

donde  $K$  no depende de  $\theta$ .

Aquí  $f_\theta$  denota a la función de densidad de donde provienen las observaciones, y  $\theta$  representa al parámetro de interés.

En esta sección se verifica que algunos estimadores que se han obtenido, cumplen con estas condiciones.

#### 4.7.1 Tiempos de observación con distribución Exponencial

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson de intensidad  $\lambda$ . Para el plan de muestreo especificado en la Sección 4.5.3, se tiene que

$$\begin{aligned} L_\theta &\propto \prod_{k=1}^n \frac{e^{-\theta \psi_k} (\theta \psi_k)^{z_k}}{z_k!} \\ &= e^{-\theta T} \theta^{N_T} \prod_{k=1}^n \frac{\psi_k^{z_k}}{z_k!}. \end{aligned}$$

Entonces

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln L_\theta &= -T + \frac{N_T}{\theta} \\ \frac{\partial^2}{\partial \theta^2} \ln L_\theta &= -\frac{N_T}{\theta^2} \\ \frac{\partial^3}{\partial \theta^3} \ln L_\theta &= 2\frac{N_T}{\theta^3}, \end{aligned}$$

por lo tanto la condición 1 se cumple. Además

$$\begin{aligned} \frac{\partial}{\partial \theta} e^{-\theta T} \theta^{N_T} &= e^{-\theta T} N_T \theta^{N_T-1} + \theta^{N_T} (-T) e^{-\theta T} \\ \frac{\partial^2}{\partial \theta^2} e^{-\theta T} \theta^{N_T} &= e^{-\theta T} N_T (N_T - 1) \theta^{N_T-2} + 2N_T \theta^{N_T-1} (-T) e^{-\theta T} \\ &\quad + \theta^{N_T} T^2 e^{-\theta T}. \end{aligned}$$

Entonces, para el verdadero valor del parámetro ( $\lambda$ ),

$$\begin{aligned} E_\lambda \left[ \frac{\frac{\partial}{\partial \theta} \mathbf{L}_\theta}{\mathbf{L}_\theta} \Big|_{\theta=\lambda} \right] &= E_\lambda \left[ \frac{e^{-\lambda T} N_T \lambda^{N_T-1} + \lambda^{N_T} (-T) e^{-\lambda T}}{e^{-\lambda T} \lambda^{N_T}} \right] \\ &= E_\lambda \left[ \frac{N_T}{\lambda} - T \right] \\ &= E \left[ E \left( \frac{N_T}{\lambda} - T \mid T \right) \right] \\ &= 0, \end{aligned}$$

y

$$\begin{aligned} E_\lambda \left[ \frac{\frac{\partial^2}{\partial \theta^2} \mathbf{L}_\theta}{\mathbf{L}_\theta} \Big|_{\theta=\lambda} \right] &= E_\lambda \left[ \frac{e^{-\lambda T} N_T (N_T - 1) \lambda^{N_T-2} + 2 N_T \lambda^{N_T-1} (-T) e^{-\lambda T} + \lambda^{N_T} T^2 e^{-\lambda T}}{e^{-\lambda T} \lambda^{N_T}} \right] \\ &= E_\lambda \left[ \frac{N_T^2 - N_T}{\lambda^2} - 2 \frac{T N_T}{\lambda} + T^2 \right] \\ &= E_\lambda \left[ E \left( \frac{N_T^2 - N_T}{\lambda^2} - 2 \frac{T N_T}{\lambda} + T^2 \mid T \right) \right] \\ &= E_\lambda \left[ \frac{(\lambda T + \lambda^2 T^2) - \lambda T}{\lambda^2} - 2 \frac{\lambda T (T)}{\lambda} + T^2 \right] \\ &= 0. \end{aligned}$$

Se sigue que la condición 2 se satisface. Es fácil ver lo que ocurre con la tercera condición, ya que si  $A = (a, b) \subset (0, \infty)$ , donde  $\lambda \in (a, b)$ , entonces

$$\frac{\partial^3}{\partial \theta^3} \ln \mathbf{L}_\theta = 2 \frac{N_T}{\theta^3}$$

y

$$\left| \frac{\partial^3}{\partial \theta^3} \ln \mathbf{L}_\theta \right| < \frac{2 N_T}{a^3}$$

para todo  $\theta \in A$ . Para concluir

$$\begin{aligned} E_\lambda \left[ \frac{2 N_T}{a^3} \right] &= \frac{2}{a^3} E_\lambda [E(N_T \mid T)] \\ &= \frac{2 \lambda}{a^3} E[T] \\ &< \frac{2 b n_0}{a^3 \phi}. \end{aligned}$$

Por lo tanto, se puede aplicar el resultado. La información esperada de Fisher para este caso viene dada por

$$\begin{aligned} F &= E_{\lambda} \left[ -\frac{\partial^2}{\partial \theta^2} \ln L_{\theta} \Big|_{\theta=\lambda} \right] \\ &= E_{\lambda} \left[ \frac{N_T}{\lambda^2} \right] \\ &= \frac{n_0}{\lambda \phi}. \end{aligned}$$

Esto implica que la distribución asintótica de  $\sqrt{n}(\hat{\lambda} - \lambda)$  es  $N(0, \lambda\phi)$ .

Se obtendrá un intervalo de confianza con nivel asintótico  $\alpha$ , aprovechando el hecho de que  $\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda\phi}}$  es, asintóticamente, una cantidad pivotal. Entonces, el intervalo

$$\begin{aligned} & \left\{ -z_{1-\alpha} < \frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\lambda\phi}} < z_{1-\alpha} \right\} \\ &= \left\{ -z_{1-\alpha} \sqrt{\lambda\phi} + \sqrt{n}\lambda < \sqrt{n}\hat{\lambda} < z_{1-\alpha} \sqrt{\lambda\phi} + \sqrt{n}\lambda \right\} \\ &= \left\{ \left( \sqrt{\lambda} - \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2 < \hat{\lambda} + \left( \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2 < \left( \sqrt{\lambda} + \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2 \right\} \\ &= \left\{ -\sqrt{\hat{\lambda} + \left( \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2} - \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} < -\sqrt{\lambda} < -\sqrt{\hat{\lambda} + \left( \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2} + \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right\} \\ &= \left\{ \sqrt{\sqrt{\hat{\lambda} + \left( \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2} + \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}}} > \lambda > \sqrt{\sqrt{\hat{\lambda} + \left( \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}} \right)^2} - \frac{z_{1-\alpha}\sqrt{\phi}}{2\sqrt{n}}} \right\}, \end{aligned}$$

tiene la probabilidad de cobertura buscada (al hacer  $n \rightarrow \infty$ ). Aquí  $z_{1-\alpha}$  denota al cuantil  $(1 - \alpha)$  de una distribución Normal estándar.

Una vez obtenidos estos resultados, surge una pregunta de interés práctico. ¿A partir de que valor de  $n$  resulta razonable emplear la Aproximación Normal?. Un experimento que da una respuesta parcial a esto consiste en utilizar *papel de probabilidad normal* para distintos valores de los parámetros. Este recurso gráfico [ver Johnson y Wichern] permite visualizar que tan cercana (en términos de convergencia uniforme) es la distribución empírica de los datos a una distribución  $N(\mu, \sigma^2)$ . En el eje  $y$  de la gráfica se indican los valores de las estadísticas de orden de los datos, mientras que el eje  $x$  corresponde a los cuantiles de una distribución  $N(0, 1)$ . Estas gráficas se interpretan de la siguiente forma: entre más colineales son los puntos de la gráfica, se dice que más se asemeja la distribución de los datos a una distribución normal.

Para datos simulados del estimador, los resultados fueron los siguientes: En el caso  $\lambda = 1$ ,  $\phi = 1$  y  $n = 10$  (ver Figura 4.26) el experimento mostró que era factible hacer uso de la Aproximación

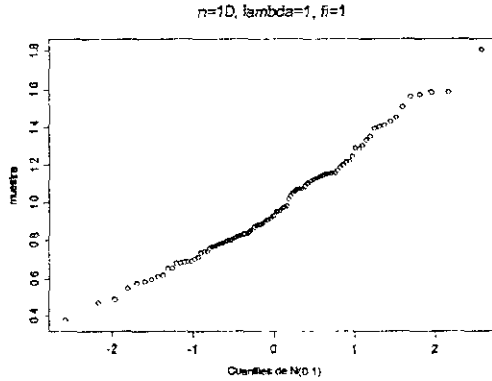


FIGURE 4.26.

Normal. Al hacer  $\lambda = 1$ ,  $\phi = 10$  y  $n = 10$ , se observó un comportamiento (ver Figura 4.27) que sugiere que aún no es válido recurrir al resultado asintótico. Es necesario incrementar el tamaño de muestra a un valor cercano a 100 (ver Figura 4.28) para que haya evidencia empírica que sustente el uso de la aproximación.

Nuevamente, el incrementar el valor de  $\phi$  tiene efectos negativos en las propiedades del estimador. Al hacer  $\phi = 50$ ,  $\lambda = 1$  se tiene que, es necesario un tamaño de muestra cercano a 250 (ver Figura 4.29) para que se observe colinealidad en los puntos.

Por otra parte, en el caso  $\phi = .05$  y  $\lambda = 1$  se observó que bastaba con  $n = 5$  para poder hacer uso de la Aproximación Normal. El resultado de esta simulación puede verse en la Figura 4.30.

4.7.2 Tiempo de observación aleatorio

Para el caso desarrollado en la Sección 4.3 la verosimilitud es

$$L_{\theta} \propto \theta^k \exp(-\theta S_k),$$

entonces

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln L_{\theta} &= \frac{k}{\theta} - S_k \\ \frac{\partial^2}{\partial \theta^2} \ln L_{\theta} &= -\frac{k}{\theta^2} \\ \frac{\partial^3}{\partial \theta^3} \ln L_{\theta} &= \frac{2k}{\theta^3}. \end{aligned}$$



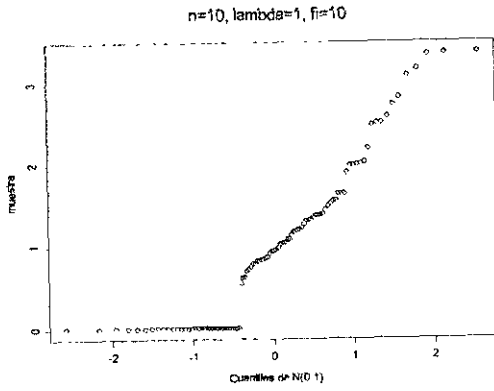


FIGURE 4.27.

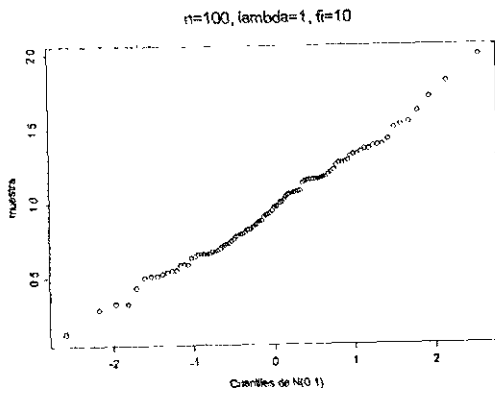


FIGURE 4.28.

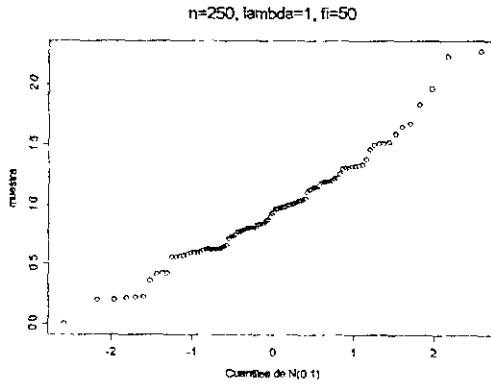


FIGURE 4.29.

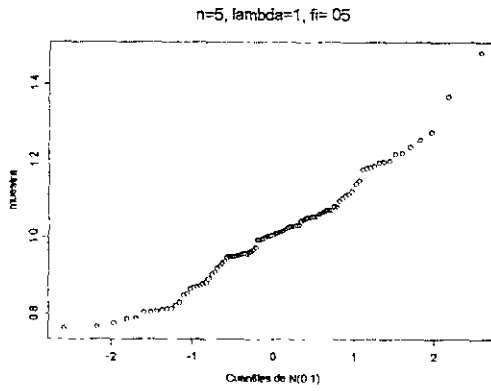


FIGURE 4.30.

por lo tanto, la condición 1) se cumple. Se tiene que

$$\begin{aligned}\frac{\partial}{\partial \theta} \mathbf{L}_\theta &= \theta^k (-S_k) \exp(-\theta S_k) + k\theta^{k-1} \exp(-\theta S_k) \\ \frac{\partial^2}{\partial \theta^2} \mathbf{L}_\theta &= \theta^k S_k^2 \exp(-\theta S_k) - 2k\theta^{k-1} S_k \exp(-\theta S_k) \\ &\quad + k(k-1)\theta^{k-2} \exp(-\theta S_k).\end{aligned}$$

Entonces, para el verdadero valor del parámetro

$$\begin{aligned}E_\lambda \left[ \frac{\frac{\partial}{\partial \theta} \mathbf{L}_\theta}{\mathbf{L}_\theta} \Big|_{\theta=\lambda} \right] &= E_\lambda \left[ -S_k + \frac{k}{\lambda} \right] \\ &= E_\lambda (-S_k) + \frac{k}{\lambda} \\ &= 0,\end{aligned}$$

ya que  $S_k$  tiene distribución  $\text{Gam}(\frac{1}{\lambda}, k)$ , además

$$\begin{aligned}E_\lambda \left[ \frac{\frac{\partial^2}{\partial \lambda^2} \mathbf{L}_\lambda}{\mathbf{L}_\lambda} \Big|_{\theta=\lambda} \right] \\ &= E_\lambda \left[ S_k^2 - 2\frac{k}{\lambda} S_k + \frac{k^2 - k}{\lambda^2} \right] \\ &= 0,\end{aligned}$$

de donde se sigue la condición 2) Es fácil ver que

$$\frac{\partial^3}{\partial \theta^3} \ln \mathbf{L}_\theta = \frac{2k}{\theta^3}.$$

Entonces, para  $\theta$  en  $A = (a, b) \subset (0, \infty)$ , donde  $\lambda \in (a, b)$ , se tiene

$$\left| \frac{2k}{\theta^3} \right| < \frac{2k}{a^3}.$$

Por lo tanto, se cumplen las condiciones de regularidad. Para este caso la información observada de Fisher es

$$\begin{aligned}E_\lambda \left[ -\frac{\partial^2}{\partial \theta^2} \ln \mathbf{L}_\theta \Big|_{\theta=\lambda} \right] &= E_\lambda \left[ \frac{k}{\lambda^2} \right] \\ &= \frac{k}{\lambda^2}.\end{aligned}$$

Se concluye que la distribución asintótica de  $\sqrt{n}(\hat{\lambda} - \lambda)$  es  $\mathbf{N}(0, \lambda^2)$ .

## 4.8 El problema de estimación para el Proceso Poisson Compuesto

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda(\cdot, \theta)$  y sean  $\{Y_k : k \in \mathbf{N}\}$  variables aleatorias independientes, con distribución  $F(\cdot; \theta)$ , y tales que sean independientes del Proceso Poisson. Entonces  $\{Z_t : t \geq 0\}$  es un Proceso Poisson Compuesto, donde

$$Z_t = \begin{cases} \sum_{k=1}^{N_t} Y_k & \text{si } N_t \geq 1 \\ 0 & \text{si } N_t = 0. \end{cases}$$

Se hará el supuesto adicional de que las variables  $\{Y_k : k \in \mathbf{N}\}$  poseen densidad, la cual se denota con  $f(\cdot; \theta)$ .

Primero se analizará el caso en el que las variables  $\{N_t : t \geq 0\}$  son observables y además se conoce la marca (la  $Y_k$  correspondiente) asociada a cada ocurrencia.

Se observa el proceso  $\{Z_t : t \geq 0\}$  de manera ininterrumpida en el intervalo  $(0, r]$ , entonces

$$\begin{aligned} f_{Y_1, Y_2, \dots, Y_n, N_t}(y_1, y_2, \dots, y_n, n) &= f_{N_t | Y_1, Y_2, \dots, Y_n}(n, y_1, y_2, \dots, y_n) \\ &\quad \cdot f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n) \\ &= \Pr\{N_t = n\} \prod_{i=1}^n f_{Y_i}(y_i). \end{aligned}$$

Por lo tanto, si se conoce el número de ocurrencias del proceso  $\{N_t : t \geq 0\}$  en

$$\{(v_i, v_{i+1}] \quad i \in \{0, 1, \dots, m-1\}\}$$

donde  $v_0 = 0$ ,  $v_m = r$ , y si se observan además las variables  $\{Y_1, Y_2, \dots, Y_{N_r}\}$ , se obtiene la siguiente función de verosimilitud

$$\begin{aligned} L_\theta &\propto \prod_{j=1}^m \left( \Pr_\theta \{N_{v_j} - N_{v_{j-1}} = n_j\} \prod_{l=N_{v_{j-1}}+1}^{N_{v_j}} f_{Y_l}(y_l; \theta) \right) \\ &= \left( \prod_{j=1}^m \Pr_\theta \{N_{v_j} - N_{v_{j-1}} = n_j\} \right) \prod_{l=1}^{N_r} f_{Y_l}(y_l; \theta) \\ &= \left( \exp \left[ - \int_0^r \lambda(u; \theta) du \right] \right) \left( \prod_{j=1}^m \frac{\left( \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right)^{n_j}}{n_j!} \right) \prod_{l=1}^{N_r} f_{Y_l}(y_l; \theta) \end{aligned} \quad (4.19)$$

si  $N_r \geq 1$ , de otra forma  $L_\theta \propto \Pr_\theta \{N_r = 0\}$ . Por lo tanto el estimador máximo verosímil para  $\theta$  se obtiene maximizando la Expresión (4.19) dada la forma funcional de  $f_{Y_i}$  y de  $\lambda(\cdot; \theta)$ .

**Ejemplo 4.8.1** Supóngase que, bajo estas condiciones se observa  $\{N_t : t \geq 0\}$  en el intervalo  $(0, r]$  y que  $Y_i \sim \text{Poi}(\theta)$ ,  $i \in \{1, 2, \dots, N_r\}$ , entonces

$$\begin{aligned} L_\theta &\propto \exp \left[ - \int_0^r \lambda(u; \theta) du \right] \left( \prod_{j=1}^m \frac{\left( \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right)^{n_j}}{n_j!} \right) \\ &\quad \cdot \exp[-N_r \theta] \prod_{l=1}^{N_r} \frac{\theta^{y_l}}{y_l!}, \end{aligned}$$

por lo tanto

$$\begin{aligned} \ln L_\theta &= - \int_0^r \lambda(u; \theta) du + \sum_{j=1}^m \left( n_j \ln \left( \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right) - \ln(n_j!) \right) \\ &\quad - N_r \theta + \sum_{i=1}^{N_r} (y_i \ln(\theta) - y_i!) + c \end{aligned}$$

y

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln L_\theta &= - \frac{\partial}{\partial \theta} \int_0^r \lambda(u; \theta) du + \sum_{j=1}^m n_j \frac{\partial}{\partial \theta} \ln \left( \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right) \\ &\quad - N_r + \frac{\sum_{i=1}^{N_r} y_i}{\theta}. \end{aligned}$$

Si  $\lambda(u; \theta) = \theta$ , se tiene que

$$\frac{\partial}{\partial \theta} \ln L_\theta = -r + \frac{\sum_{j=1}^m n_j}{\theta} - N_r + \frac{\sum_{i=1}^{N_r} y_i}{\theta},$$

de donde se sigue que el estimador máximo verosímil es

$$\hat{\theta} = \frac{\sum_{j=1}^m n_j + \sum_{i=1}^{N_r} y_i}{r + N_r}. \quad \blacksquare$$

Se supondrá ahora que las variables observables son los tiempos entre ocurrencias  $\{X_k : k \in \mathbb{N}\}$  y las marcas asociadas a dichas ocurrencias  $\{Y_k : k \in \mathbb{N}\}$ . El proceso es observado hasta que se registra la  $k$ -ésimo evento  $k \geq 1$ . Entonces

$$\begin{aligned} &f_{Y_1, Y_2, \dots, Y_k, X_1, X_2, \dots, X_k}(y_1, y_2, \dots, y_k, x_1, x_2, \dots, x_k) \\ &= \left( \prod_{i=1}^k f_{Y_i}(y_i) \right) \left( \prod_{i=1}^k \lambda e^{-\lambda x_i} \right), \end{aligned}$$

si  $\lambda(t) = \lambda$  para  $t \in (0, \infty)$ . De otra forma se aplica el Corolario 3.3.8, de modo que

$$\begin{aligned} &f_{Y_1, Y_2, \dots, Y_k, X_1, X_2, \dots, X_k}(y_1, y_2, \dots, y_k, x_1, x_2, \dots, x_k) \\ &= \left( \prod_{i=1}^k f_{Y_i}(y_i) \right) \rho'(x_1) \cdots \rho' \left( \sum_{i=1}^k x_i \right) \exp \left( -\rho \left( \sum_{i=1}^k x_i \right) \right), \end{aligned}$$

donde  $\rho(t) = E[N_t]$ .

Esto implica que la función de verosimilitud es

$$L_\lambda \propto \left( \prod_{i=1}^k f_{Y_i}(y_i; \lambda) \right) \lambda^k \exp \left[ -\lambda \sum_{i=1}^k x_i \right],$$

para el caso estacionario, y

$$L_{\theta} \propto \left( \prod_{i=1}^k f_{Y_i}(y_i; \theta) \right) \left( \prod_{i=1}^k \rho' \left( \sum_{i=1}^i x_i; \theta \right) \right) \exp \left( -\rho \left( \sum_{i=1}^k x_i; \theta \right) \right),$$

para el caso no estacionario.

Por lo tanto, para obtener una expresión para el estimador máximo verosímil, es necesario conocer la forma funcional de  $\lambda(\cdot)$  y de  $f_Y(\cdot)$ .

**Ejemplo 4.8.2** Sea  $\lambda(\cdot) = \lambda$  y  $Y_i \sim \text{Poi}(\lambda)$ ,  $i \in \{1, 2, \dots, k\}$ , entonces

$$\begin{aligned} L_{\lambda} &\propto \left( \prod_{i=1}^k e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} \right) \lambda^k \exp \left[ -\lambda \sum_{i=1}^k x_i \right] \\ &= \lambda^{k + \sum y_i} \exp \left[ -\lambda \left( k + \sum_{i=1}^k x_i \right) \right] \prod_{i=1}^k \frac{1}{y_i!}, \end{aligned}$$

de donde se sigue

$$\begin{aligned} \ln L_{\lambda} &= \left( k + \sum_{i=1}^k y_i \right) \ln \lambda - \lambda \left( k + \sum_{i=1}^k x_i \right) + \ln \left( \prod_{i=1}^k \frac{1}{y_i!} \right) + c \\ \frac{\partial}{\partial \lambda} \ln L_{\lambda} &= \frac{k + \sum_{i=1}^k y_i}{\lambda} - \left( k + \sum_{i=1}^k x_i \right), \end{aligned}$$

por lo tanto, el estimador máximo verosímil viene dado por

$$\hat{\lambda} = \frac{k + \sum_{i=1}^k y_i}{k + \sum_{i=1}^k x_i}. \quad \blacksquare$$

Otra forma de abordar problemas de estimación paramétrica para el Proceso Poisson Compuesto es haciendo uso del Teorema 3.5.1; dicho resultado establece que los procesos

$$\{N(t, u_k) : t \geq 0\} \quad k \in N$$

son Procesos Poisson independientes, donde  $\{u_1, u_2, \dots\}$  son los posibles valores de las marcas.

Bajo el supuesto de que los procesos  $\{N(t, u_k) : t \geq 0\}$  son observados de manera continua durante  $(0, r]$ , se tiene

$$\begin{aligned} &\Pr \{N_r = n, N(r, u_1) = n_1, N(r, u_2) = n_2, \dots\} \\ &= \Pr \{N_r = n \mid N(r, u_1) = n_1, N(r, u_2) = n_2, \dots\} \\ &\quad \cdot \Pr \{N(r, u_1) = n_1, N(r, u_2) = n_2, \dots\} \\ &= \prod_{i=1}^{\infty} \Pr \{N(t, u_i) = n_i\}, \end{aligned}$$

donde  $\sum_{i=1}^{\infty} n_i = n$ . Se sigue que

$$\begin{aligned} L_{\theta} &\propto \prod_{i=1}^{\infty} \Pr \{N(t, u_i) = n_i; \theta\} \\ &= \prod_{i=1}^{\infty} \exp[-\theta q_i(\theta) t] \frac{(\theta q_i(\theta) t)^{n_i}}{n_i!}, \end{aligned}$$

donde  $q_i$  es la probabilidad correspondiente a la marca  $u_i$ . Esto se extiende al caso presentado en la Sección 4.5.3 de la siguiente forma.

$$\begin{aligned} L_{\theta} &\propto \prod_{i=1}^{n_0} \frac{e^{-\theta \psi_i} (\theta \psi_i)^{z_i}}{z_i!} \\ &= \prod_{i=1}^{n_0} \left( \prod_{j=1}^{\infty} \exp[-\theta q_j(\theta) \psi_i] \frac{(\theta q_j(\theta) \psi_i)^{n_{ij}}}{n_{ij}!} \right) \\ &= \prod_{j=1}^{\infty} \left( \prod_{i=1}^{n_0} \exp[-\theta q_j(\theta) \psi_i] \frac{(\theta q_j(\theta) \psi_i)^{n_{ij}}}{n_{ij}!} \right) \end{aligned}$$

donde  $n_{ij} = n(u_i, u_j) - n(v_{i-1}, u_j)$  y  $\psi_i = u_i - v_{i-1}$ .

Hasta ahora se ha supuesto que se conoce tanto el número de ocurrencias que tuvieron lugar en el tiempo (fijo o aleatorio) como las marcas asignadas a cada una de las mismas. Puede ocurrir que solamente las variables  $\{Z_t : t \geq 0\}$  y  $\{N_t : t \geq 0\}$  sean observables.

Se observa el proceso  $\{Z_t : t \geq 0\}$  durante el lapso  $(0, r]$ , entonces

$$\begin{aligned} f_{Z_t, N_t}(z, n) &= f_{Z_t | N_t}(z, n) \Pr \{N_t = n\} \\ &= g^{(n)}(z) \Pr \{N_t = n\}, \end{aligned}$$

donde  $g^{(n)}$  es la función de densidad de  $\sum_{i=1}^n Y_i$  si  $n \geq 1$  y  $g^{(0)}(x) = \delta_0(x)$ . Por lo tanto, si se conoce el número de ocurrencias del proceso en los subintervalos

$$\{(v_i, v_{i+1}] \quad i \in \{0, 1, \dots, m-1\}\},$$

y las marcas correspondientes, se tiene que

$$\begin{aligned} L_{\theta} &\propto \prod_{i=1}^m g^{(n_i)}(z_i; \theta) \Pr_{\theta} \{N_{v_i} - N_{v_{i-1}} = n_i\} \\ &= \prod_{i=1}^m g^{(n_i)}(z_i; \theta) \left( \prod_{j=1}^m \exp \left[ - \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right] \frac{\left( \int_{v_{j-1}}^{v_j} \lambda(u; \theta) du \right)^{n_j}}{n_j!} \right), \end{aligned} \tag{4.20}$$

donde  $n_i$  es el número de ocurrencias de  $\{N_t : t \geq 0\}$  en  $(v_{i-1}, v_i]$  y  $z_i = z_{v_i} - z_{v_{i-1}}$ . Esto se aplica al plan de muestreo propuesto en la Sección 4.5.3 de la siguiente forma

$$L_{\theta} \propto \prod_{i=1}^m g^{(n_i)}(z_i; \theta) \exp[-\psi_i \theta] \frac{(\psi_i \theta)^{n_i}}{n_i!}.$$

Una vez establecida la forma funcional de la intensidad y de  $g$ , se procede a maximizar la Expresión (4.20).

## 4.9 Algunas pruebas de hipótesis para el proceso Poisson

### 4.9.1 Cocientes de verosimilitud

Se considera la hipótesis simple

$$H_0 : \lambda = \lambda_0 \quad H_1 : \lambda = \lambda_1$$

para la forma de observación del proceso propuesta en la Sección 4.1. Para este caso la función de verosimilitud (dado  $\lambda \in \{\lambda_1, \lambda_0\}$ ) viene dada por

$$L_{\lambda_i} \propto \lambda_i^{n_r} \exp[-r\lambda_i] \quad i \in \{0, 1\}.$$

Se aplicará el Lema de Neyman-Pearson para esta prueba. Se supondrá  $\lambda_1 > \lambda_0$ , entonces

$$\begin{aligned} \frac{L_{\lambda_0}}{L_{\lambda_1}} &= \left(\frac{\lambda_0}{\lambda_1}\right)^{n_r} \exp[-r(\lambda_0 - \lambda_1)] \\ \ln \frac{L_{\lambda_0}}{L_{\lambda_1}} &= n_r \ln \left(\frac{\lambda_0}{\lambda_1}\right) + r(\lambda_1 - \lambda_0). \end{aligned}$$

La región crítica para este caso es

$$\left\{ \omega : N_r(\omega) \ln \left(\frac{\lambda_0}{\lambda_1}\right) + r(\lambda_1 - \lambda_0) \leq k_\alpha \right\},$$

lo cual equivale a decir

$$\left\{ \omega : N_r(\omega) \geq \frac{k_\alpha - r(\lambda_1 - \lambda_0)}{\ln \left(\frac{\lambda_0}{\lambda_1}\right)} \right\}.$$

Por lo tanto la PUMP para viene dada por

$$\text{Se rechaza } H_0 \text{ si } N_r \geq k_\alpha^*,$$

donde  $k_\alpha^*$  es el cuantil  $(1 - \alpha)$  correspondiente a una distribución  $\text{Poi}(\lambda_0 r)$ . Debido a que la forma de la prueba no depende de  $\lambda_1$ , con la misma regla de rechazo se tiene la PUMP para  $H_0 : \lambda = \lambda_0$  contra  $H_1 : \lambda > \lambda_0$ .

Si el proceso se está observando de acuerdo a las condiciones impuestas en la Sección 4.3, se tiene que

$$L_{\lambda_i} \propto \lambda_i^k \exp[-\lambda_i s_k] \quad i \in \{0, 1\},$$



entonces

$$\ln \frac{L_{\lambda_0}}{L_{\lambda_1}} = k \ln \left( \frac{\lambda_0}{\lambda_1} \right) + s_k (\lambda_1 - \lambda_0).$$

Por lo tanto la región crítica viene dada por

$$\left\{ \omega : S_k(\omega) \leq \frac{k_\alpha - k \ln \left( \frac{\lambda_0}{\lambda_1} \right)}{(\lambda_1 - \lambda_0)} \right\}.$$

Se sigue que la PUMP para este caso es

$$\text{Se rechaza } H_0 \text{ si } S_k \leq k_\alpha^*,$$

donde  $k_\alpha^*$  es el cuantil  $\alpha$  correspondiente a una distribución **Gam**( $k, \lambda_0$ ). De nuevo, esta regla de decisión puede generalizarse al caso  $H_0 : \lambda = \lambda_0$  contra  $H_1 : \lambda > \lambda_0$  para obtener la PUMP.

El siguiente paso consiste en desarrollar este procedimiento cuando se observa al proceso de acuerdo al plan de muestreo planteado en la Sección 4.5.3, para este caso

$$L_{\lambda_1} \propto \prod_{k=1}^n \frac{e^{-\lambda_1 \psi_k} (\lambda_1 \psi_k)^{z_k}}{z_k!}.$$

Entonces el cociente de verosimilitud es

$$\begin{aligned} \frac{L_{\lambda_0}}{L_{\lambda_1}} &= \prod_{k=1}^n e^{-\psi_k(\lambda_0 - \lambda_1)} \left( \frac{\lambda_0}{\lambda_1} \right)^{z_k} \\ &= e^{-t(\lambda_0 - \lambda_1)} \left( \frac{\lambda_0}{\lambda_1} \right)^{n_T} \\ \ln \frac{L_{\lambda_0}}{L_{\lambda_1}} &= -t(\lambda_0 - \lambda_1) + n_T \ln \left( \frac{\lambda_0}{\lambda_1} \right). \end{aligned}$$

Se sigue que la región crítica para una prueba de tamaño  $\alpha$  viene dada por

$$\left\{ \omega : T(\omega)(\lambda_1 - \lambda_0) + N_T(\omega) \ln \left( \frac{\lambda_0}{\lambda_1} \right) \leq k_\alpha \right\}.$$

Sean  $\eta_1 = \lambda_1 - \lambda_0$  y  $\eta_2 = \ln \left( \frac{\lambda_0}{\lambda_1} \right)$ . Es necesario dar cierta descripción del estadístico  $\eta_1 T + \eta_2 N_T$ . El cálculo hecho para obtener la función de distribución de dicha variable aleatoria es muy similar al realizado para encontrar la distribución de  $\frac{N_T}{T}$ , esto es

$$\begin{aligned} \Pr \{ \eta_1 T + \eta_2 N_T \leq x \} &= \sum_{k=0}^{\infty} \Pr \{ \eta_1 T + \eta_2 N_T \leq x \mid N_T = k \} \Pr \{ N_T = k \} \\ &= \sum_{k=0}^{\infty} \Pr \left\{ T \leq \frac{x - \eta_2 N_T}{\eta_1} \mid N_T = k \right\} \Pr \{ N_T = k \} \\ &= \sum_{k=0}^{\infty} F_{D_k} \left( \frac{x - k\eta_2}{\eta_1} \right) \binom{k + n_0 - 1}{k} \frac{\lambda^k \phi^{n_0}}{(\lambda + \phi)^{k+n_0}}, \end{aligned}$$

donde  $D_n$  es una variable aleatoria con distribución  $\text{Gam}(n_0 + N_T, \phi + \lambda)$ .

Se considera ahora una prueba asintótica para este caso. Sea  $W_k$  el estimador máximo verosímil para  $\lambda$ , tomando  $k$  tiempos aleatorios de observación. Una prueba con nivel asintótico  $\alpha$  se obtiene de una región crítica del tipo

$$W_k \geq c_k,$$

donde

$$c_k = \lambda_0 + \frac{\sqrt{\lambda_0 \phi} z_\alpha}{\sqrt{n}}.$$

Esto de acuerdo a la teoría desarrollada en [Lehmann].

#### 4.9.2 Prueba para la estacionariedad del proceso

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda(t) = e^{\alpha + \beta t}$ , donde  $(\alpha, \beta)$  es el vector de parámetros desconocidos. Se desea probar la hipótesis  $H_0 : \beta = 0$ , es decir que el proceso es estacionario contra hipótesis alternativas bajo las cuales  $\beta \neq 0$ . Si el proceso se observa de manera continua durante un intervalo predeterminado  $(0, \tau]$ , el análisis hecho en la Sección 3.3.1 (ver Expresión (3.3)), resulta inmediato que, condicionando con  $\{N_\tau = n_0\}$ , la densidad conjunta de  $(S_1, S_2, \dots, S_{n_0})$  está dada por

$$f_{S_1, S_2, \dots, S_{n_0} | N_0 = n_0}(s_1, s_2, \dots, s_{n_0}) = \frac{n_0! \prod_{i=1}^{n_0} e^{\alpha + \beta s_i}}{\left(\int_0^\tau e^{\alpha + \beta u} du\right)^{n_0}}.$$

Se sigue que la verosimilitud condicionada para este caso es

$$\begin{aligned} L_\beta^* &\propto \frac{\prod_{i=1}^{n_0} e^{\alpha + \beta s_i}}{\left(\int_0^\tau e^{\alpha + \beta u} du\right)^{n_0}} \\ &= \frac{e^{n_0 \alpha} \exp\left[\beta \sum_{i=1}^{n_0} s_i\right]}{\left(\frac{1}{\beta} e^\alpha (e^{\beta \tau} - 1)\right)^{n_0}} \\ &= \frac{\beta^{n_0} \exp\left[\beta \sum_{i=1}^{n_0} s_i\right]}{(e^{\beta \tau} - 1)^{n_0}}. \end{aligned}$$

De modo similar se obtiene  $L_0^* = \frac{1}{\tau^{n_0}}$ . Aplicando el Lema de Neyman-Pearson se obtiene

$$\ln \frac{L_\beta^*}{L_0^*} = n_0 \ln(e^{\beta \tau} - 1) - n_0 \ln(\beta \tau) - \beta \sum_{i=1}^{n_0} s_i.$$

Esto implica que la región crítica para la PUMP de tamaño  $\alpha$  es

$$\left\{ \omega : \sum_{i=1}^{n_0} S_i(\omega) \geq \frac{n_0 \ln(e^{\beta \tau} - 1) - n_0 \ln(\beta \tau) - k_\alpha}{\beta} \right\}.$$

# 5

## El punto de vista bayesiano

### Introducción

La Teoría Bayesiana, llamada así en honor a Thomas Bayes, incorpora el concepto de distribución *a priori*. Se trata de una densidad que sintetiza el conocimiento con el que cuenta el investigador antes de llevar a cabo las observaciones. Según este enfoque, sólo es posible aprender de los datos a través de la función de verosimilitud. La integración de ambas fuentes de información se efectúa mediante el Teorema de Bayes. Lo que se obtiene es una distribución *a posteriori*, la cual se toma como el punto de partida para cualquier inferencia.

En este capítulo se tratan los problemas de estimación y pruebas de hipótesis para el Proceso Poisson usando el método Bayesiano. Se consideran los mismos planes de muestreo discutidos en el capítulo anterior.

Primero se calcula las densidades conjugadas para algunos casos de interés (Sección 5.1). La metodología referente a estimación puntual es discutida en las Secciones 5.2, 5.3 y 5.4. Mediante el uso de ejemplos se ilustra como hacer inferencias de este tipo incorporando información *a priori*, para ello se emplean distribuciones conjugadas. También se hará uso del análisis no informativo o de referencia, esto permite establecer cierto punto de comparación con los métodos frecuentistas. Hay una breve discusión (Sección 5.5) acerca de como calcular regiones de alta densidad posterior para estos casos. En la Sección 5.6 se trata lo referente a pruebas de hipótesis para el Proceso Poisson. La idea es ilustrar como las técnicas bayesianas permiten emplear el concepto de prueba de hipótesis como herramienta para resolver problemas de toma de decisiones.

Los conceptos relacionados con estimación pueden revisarse en [O'Hagan]. Lo referente a aproximaciones y análisis no informativo fue tomado de [Bernardo y Smith]. El material que trata el problema de pruebas de hipótesis puede ser consultado en [Berger]. Algunos comentarios ejemplos y aplicaciones a Teoría de Decisiones fueron basados en [Leonard] y [Buck, Cavanagh, Litton].

## 5.1 Densidades a priori conjugadas

En muchos de los casos analizados en el capítulo anterior, la muestra  $x_1, \dots, x_n$  proviene de una variable aleatoria con distribución **Poi** o **Exp**. Se calcula la densidad *a priori* conjugada para cada caso, esto se hace aprovechando el hecho de que ambas distribuciones pertenecen a la Familia Exponencial.

La verosimilitud para una muestra correspondiente a una distribución **Poi** ( $\theta$ ),  $\theta > 0$  tiene la forma

$$\begin{aligned} p(x_1, \dots, x_n \mid \theta) &= \prod_{i=1}^n \frac{\theta^{x_i} \exp(-\theta)}{x_i!} \\ &= \left( \prod_{i=1}^n x_i! \right)^{-1} \exp(-n\theta) \exp\left(\ln \theta \sum_{i=1}^n x_i\right). \end{aligned}$$

Del resultado general para la Familia Exponencial (ver apéndice) se sigue que la densidad *a priori* conjugada para  $\theta$  esta dada por

$$p(\theta \mid \tau_0, \tau_1) = \frac{1}{K(\tau_0, \tau_1)} \exp(-\tau_0 \theta) \exp(\tau_1 \ln \theta).$$

Para  $\tau_0 > 0$  y  $\tau_1 > -1$  la siguiente integral siempre existe (ver Sección 2.2.4)

$$K(\tau_0, \tau_1) = \int_0^{\infty} \theta^{\tau_1} \exp(-\tau_0 \theta) d\theta.$$

Haciendo  $\alpha = \tau_1 + 1$  y  $\beta = \tau_0$  se concluye que la familia  $\{\text{Gam}(\alpha, \beta) : \alpha, \beta > 0\}$  es conjugada a la distribución **Poi** ( $\theta$ ).

La función de verosimilitud correspondiente a una muestra  $x_1, \dots, x_n$  proveniente de una distribución **Exp** ( $\theta$ ) esta dada por

$$p(x_1, \dots, x_n \mid \theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right).$$

Entonces

$$\begin{aligned} K(\tau_0, \tau_1) &= \int_0^{\infty} \theta^{\tau_0} \exp(-\theta \tau_1) d\theta \\ &= \frac{\Gamma(\tau_0 + 1)}{\tau_1^{\tau_0 + 1}} \end{aligned}$$

para cualesquiera  $\tau_0 > -1$  y  $\tau_1 > 0$ , de donde es fácil concluir que

$$p(\theta \mid \tau_0, \tau_1) = \frac{\tau_1^{\tau_0 + 1}}{\Gamma(\tau_0 + 1)} \theta^{\tau_0} \exp(-\theta \tau_1).$$

Entonces la distribución conjugada natural para este caso es una **Gam** ( $\tau_0 + 1, \tau_1$ ).

## 5.2 El caso estacionario

### 5.2.1 Observación durante tiempo fijo

Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson estacionario de intensidad  $\theta$ . Se supone que dicho proceso se puede observar de manera ininterrumpida en el intervalo  $(0, r]$ .

Si las observaciones consisten en los tiempos entre ocurrencias, del análisis hecho en la Sección 4.1.1 se sigue que la función de verosimilitud es

$$\begin{aligned} p(\mathbf{x} | \theta) &\propto \theta^{n_r} \exp\left(-\theta \sum_{i=1}^{n_r} x_i\right) \exp(-\theta(r - S_{n_r})) \\ &= \theta^{n_r} \exp(-\theta r), \end{aligned}$$

donde  $x_i$  denota al tiempo transcurrido entre las ocurrencias indexadas con  $(i - 1)$  e  $i$ . La distribución *a priori* conjugada natural para este caso es

$$p(\theta | \tau_0, \tau_1) \propto \frac{\tau_1^{\tau_0}}{\Gamma(\tau_0)} \theta^{\tau_0} \exp(-\theta \tau_1).$$

Entonces

$$p(\theta | \mathbf{x}) \propto \theta^{n_r + \tau_0} \exp(-\theta(r + \tau_1)),$$

por lo tanto  $\theta | \mathbf{x}$  se distribuye **Gam**  $(n_r + \tau_0 + 1, r + \tau_1)$ .

El estimador bayesiano viene dado por

$$E(\theta | \mathbf{x}) = \frac{n_r + \tau_0 + 1}{r + \tau_1}$$

Si no se desea hacer uso de información *a priori*, basta hacer  $\tau_0 \rightarrow -1$  y  $\tau_1 \rightarrow 0$  para trabajar así con distribuciones no informativas, al hacer esto se obtiene un estimador que coincide con el estimador máximo verosímil.

**Ejemplo 5.2.1** *Bajo este esquema de muestreo se observaron 10 ocurrencias en 3 unidades de tiempo. Se representó el conocimiento inicial sobre la intensidad  $\theta$  mediante una **Gam**  $(2, 1)$ . Los tiempos entre ocurrencias fueron*

$$\mathbf{x} = (.07, .57, .59, .67, 1.14, 1.15, 1.94, 2.62, 2.64, 2.93).$$

*En la Figura 5.1 se presenta una gráfica que permite comparar las distribuciones *a priori* y *a posteriori*. El estimador bayesiano es  $E(\theta | \mathbf{x}) = 3$ , por otra parte, el estimador máximo verosímil  $\hat{\theta} = 10/3$ . ■*

Otra forma de abordar el problema de estimar  $\theta$ , restringiendo el proceso a  $(0, r]$  se sigue de observar el número de ocurrencias en  $k$  subintervalos predeterminados, por el momento se supondrá que dichos subintervalos son de igual longitud.

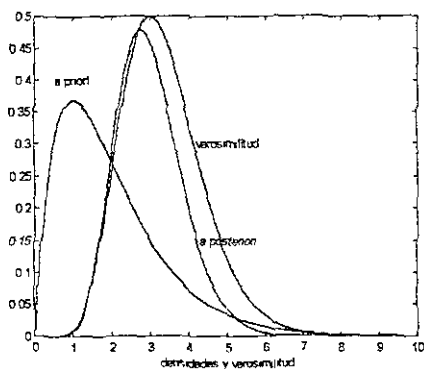


FIGURE 5.1.

De acuerdo a lo hecho en la Sección 4.1.3

$$\begin{aligned} p(\mathbf{x} | \theta) &\propto \theta^{\sum_{i=1}^k r_i} \exp(-r\theta) \\ &= \theta^{n_r} \exp(-r\theta). \end{aligned}$$

La densidad *a priori* natural para este caso es

$$p(\theta | \tau_0, \tau_1) \propto \theta^{\tau_0} \exp(-\tau_1 \theta),$$

entonces

$$p(\theta | \mathbf{x}) \propto \theta^{n_r + \tau_0} \exp(-\theta(r + \tau_1)).$$

Por lo tanto el estimador bayesiano es  $E(\theta | \mathbf{x}) = \frac{n_r + \tau_0 + 1}{r + \tau_1}$ . Haciendo  $\tau_0 \rightarrow -1$  y  $\tau_1 \rightarrow 0$  se obtiene un estimador que coincide con el máximo verosímil. Esto permite obtener una distribución *a priori* proporcional a una constante. Sin embargo, se recomienda [ver Bernardo y Smith] utilizar la distribución de Jeffreys para hacer el análisis de referencia. A continuación se calcula dicha distribución. Se tiene

$$\ln p(\mathbf{x} | \theta) = -r\theta + n_r \ln \theta - \ln \prod_{i=1}^n x_i!,$$

entonces, las derivadas de primer y segundo orden del logaritmo de la verosimilitud son

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x} | \theta)}{\partial \theta} &= -r + \frac{n_r}{\theta} \\ \frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta^2} &= -\frac{n_r}{\theta^2}. \end{aligned}$$

Por lo tanto, la información esperada de Fisher viene dada por

$$\begin{aligned} F(\theta) &= E_{\mathbf{x}|\theta} \left[ \frac{N_r}{\theta^2} \right] \\ &= \frac{r}{\theta} \end{aligned}$$

y la densidad *a priori* no informativa es

$$p(\theta) \propto \left( \frac{1}{\theta} \right)^{\frac{1}{2}}.$$

Se trata de una distribución impropia. La densidad *a posteriori* para este caso es

$$p(\theta | \mathbf{x}) \propto \theta^{n_r - \frac{1}{2}} \exp(-r\theta).$$

esto es  $\theta | \mathbf{x} \sim \text{Gam}(n_r + \frac{1}{2}, r)$ . La transformación que conduce a la *a priori* uniforme es

$$\phi \propto \int_0^\theta u^{-\frac{1}{2}} du = 2u^{\frac{1}{2}} \Big|_0^\theta \propto \theta^{\frac{1}{2}}.$$

Ahora se discutirá la aproximación a la distribución final mediante una Densidad Normal cuando se recurre a una densidad *a priori* conjugada

Se requiere que la moda de la distribución *a priori* se alcance en un punto interior de su soporte, por tanto se supondrá que  $\tau_0 > 0$ . Entonces la moda de  $p(\theta)$  es  $m = \frac{\tau_0}{\tau_1}$ . Por otra parte

$$\frac{\partial^2}{\partial \theta^2} \ln(p(\theta)) = -\frac{\tau_0}{\theta^2},$$

entonces

$$\begin{aligned} H_0 &= - \frac{\partial^2}{\partial \theta^2} \ln(p(\theta)) \Big|_{\theta=m} \\ &= \frac{\tau_0^2}{\tau_0}. \end{aligned}$$

La información observada de Fisher es

$$\frac{\partial^2}{\partial \theta^2} \ln(p(\mathbf{x}_n | \theta)) = -\frac{n_r}{\theta^2},$$

donde  $\mathbf{x}_n$  es el vector de observaciones y  $n_r$  es el número de ocurrencias en  $(0, \tau]$ . Se denota con  $\hat{\theta}_n$  al estimador máximo verosímil para este plan de muestreo, se sigue

$$\begin{aligned} H(\hat{\theta}_n) &= - \frac{\partial^2}{\partial \theta^2} \ln(p(\mathbf{x}_n | \theta)) \Big|_{\theta=\hat{\theta}_n} \\ &= \frac{n_r^2}{n_r}. \end{aligned}$$

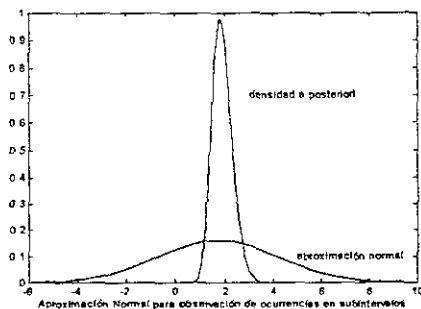


FIGURE 5.2.

Entonces, para  $n$  grande,  $p(\theta | \mathbf{x}_n)$  se aproxima con la densidad de una  $N(\mathbf{m}_n, \mathbf{H}_n)$ , donde

$$\begin{aligned} \mathbf{H}_n &= H_0 + H(\hat{\theta}_n) \\ &= \frac{\tau_1^2 n_r + \tau_0 r^2}{\tau_0 n_r} \\ \mathbf{m}_n &= \mathbf{H}_n^{-1} (H_0 m + H(\hat{\theta}_n) \hat{\theta}_n) \\ &= \frac{\tau_0 n_r}{\tau_1^2 n_r + \tau_0 r^2} (\tau_1 + r). \end{aligned}$$

Es claro que estas expresiones sólo están en función de  $\tau$  y del número total de ocurrencias del proceso en  $(0, \tau]$ .

**Ejemplo 5.2.2** *Se observó un Proceso Poisson Estacionario durante 10 unidades de tiempo, registrándose un total de 19 ocurrencias. Se propuso como distribución a priori una  $\text{Gam}(2, 1)$ . En este caso  $\mathbf{H}_n = 6.26$  y  $\mathbf{m}_n = 1.75$ . El estimador bayesiano es  $E(\theta | \mathbf{x}) = 21/11$ . En la Figura 5.2 se presentan ambas densidades.* ■

### 5.2.2 Observación durante tiempo aleatorio

Otro caso de interés consiste en observar al Proceso Poisson de manera continua hasta el momento en el que tiene lugar la  $k$ -ésima ocurrencia ( $k \geq 1$ ). Si las variables aleatorias que se toman en cuenta para las observaciones son los tiempos entre ocurrencias  $x_1, x_2, \dots, x_k$ , entonces, la función de verosimilitud es

$$p(\mathbf{x} | \theta) = \theta^k \exp\left(-\theta \sum_{i=1}^k x_i\right).$$



Mediante un tratamiento totalmente análogo al hecho para el problema de estimación con observaciones restringidas a  $(0, \tau]$ , se tiene que el estimador bayesiano viene dado por

$$E(\theta | \mathbf{x}) = \frac{k + \tau_0 + 1}{\sum_{i=1}^k x_i + \tau_1},$$

donde  $(\tau_0, \tau_1)$  es el vector de parámetros de la distribución *a priori*. Ahora se obtiene la distribución *a priori* de Jeffreys. Es fácil ver que

$$-\frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta^2} = -\frac{k}{\theta^2},$$

entonces, la información esperada de Fisher para este caso es  $F(\theta) = E_{\mathbf{x}|\theta} \left[ \frac{k}{\theta^2} \right] = \frac{k}{\theta^2}$ . Esto implica que, para este plan de muestreo la distribución *a priori* no informativa de Jeffreys es

$$p(\theta) \propto \frac{1}{\theta},$$

que es una distribución impropia. La densidad *a posteriori* viene dada por

$$p(\theta | \mathbf{x}) \propto \theta^{k-1} \exp \left( -\theta \sum_{i=1}^k x_i \right).$$

**Ejemplo 5.2.3** Se observa un Proceso Poisson Estacionario de acuerdo a este plan de muestreo, se toma  $k = 1$ . El tiempo de observación registrado fue  $x_1 = 5$ . Se consideraron dos distribuciones iniciales: una  $\text{Gam}(2, 1)$ , y la *a priori* no informativa de Jeffreys. En la Figura 5.3 se presentan las distribuciones finales que se obtuvieron para cada caso. Se observó que la densidad *a posteriori* obtenida a partir de Análisis Conjugado es muy semejante a la función de verosimilitud, para la otra distribución inicial, esto no fue así. ■

Se procede a encontrar la Aproximación Normal para la distribución final. Si la densidad *a priori* es de la forma

$$p(\theta) \propto \theta^{\tau_0} e^{-\tau_1 \theta},$$

se sigue que

$$m = \frac{\tau_0}{\tau_1} \quad \text{y} \quad H_0 = \frac{\tau_1}{\tau_0}.$$

Se tiene además que

$$\frac{\partial^2}{\partial \theta^2} \ln(p(\mathbf{x}_n | \theta)) = -\frac{k}{\theta^2},$$

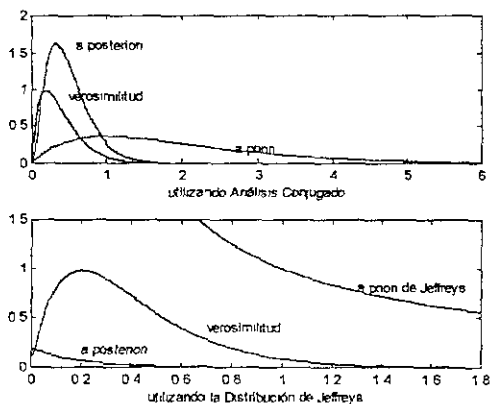


FIGURE 5.3.

donde  $k$  denota el número de observaciones. El estimador máximo verosímil para este caso es  $\hat{\theta}_k = \frac{k}{\sum_{i=1}^k x_i}$ , entonces

$$\begin{aligned} H(\hat{\theta}_n) &= \left. -\frac{\partial^2}{\partial \theta^2} \ln(p(\mathbf{x}_n, \theta)) \right|_{\theta=\hat{\theta}_n} \\ &= \frac{(\sum_{i=1}^n x_i)^2}{k}. \end{aligned}$$

Entonces, para  $n$  grande  $\theta \mid \mathbf{x}_n$  posee una densidad que puede aproximarse por una  $N(m_n, H_n)$ , donde

$$\begin{aligned} H_n &= \frac{\tau_1 k + \tau_0 (\sum_{i=1}^n x_i)^2}{\tau_0 k} \\ m_n &= \frac{\tau_0 k}{\tau_1 k + \tau_0 (\sum_{i=1}^n x_i)^2} \left( 1 + \sum_{i=1}^n x_i \right). \end{aligned}$$

### 5.3 El caso no estacionario

#### 5.3.1 Tiempos entre ocurrencias

Se considera ahora el problema de estimación con tiempo de observación fijo para el Proceso Poisson  $\{N_t : t \geq 0\}$  con función de intensidad  $\lambda(\cdot, \theta)$ . Si se restringe el proceso al intervalo  $(0, r]$  y las variables aleatorias observables  $(X_1, X_2, \dots, X_N)$  son los tiempos entre ocurrencias, entonces la función de verosimilitud es

$$p(\mathbf{s} \mid \theta) = \left( \prod_{i=1}^{n_r} \lambda(s_i, \theta) \right) \exp \left[ - \int_0^r \lambda(u, \theta) du \right], \tag{5.1}$$

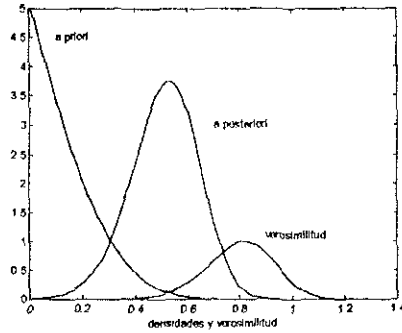


FIGURE 5.4.

donde  $s_i = x_1 + x_2 + \dots + x_n$ . Bajo estas condiciones, se tiene que esta función de verosimilitud no siempre puede ser factorizada de forma que el resultado de análisis conjugado para la Familia Exponencial (ver apéndice) pueda aplicarse. En seguida se presenta un caso particular en el que esto sí puede hacerse.

**Ejemplo 5.3.1** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda(x, \theta) = \exp[x\theta]$ . Se observa el número de ocurrencias en  $(0, \tau]$ , entonces

$$p(\mathbf{s} | \theta) = \exp \left[ \theta \sum_{i=1}^{n_r} s_i \right] \exp \left[ -\frac{1}{\theta} (e^{r\theta} - 1) \right].$$

donde  $s_i$  denota al instante en el que se registró la  $i$ -ésima ocurrencia. Haciendo  $g(\theta) = \exp \left[ -\frac{1}{\theta} (e^{r\theta} - 1) \right]$ ,  $f(s) = 1$ ,  $\phi(\theta) = \theta$ , se tiene

$$p(\theta | \tau) \propto \exp[\theta\tau_1] \exp \left[ -\frac{\tau_0}{\theta} (e^{r\theta} - 1) \right].$$

Se sigue que la distribución a posteriori viene dada por

$$p(\theta | \mathbf{s}) \propto \exp \left[ \theta \left( \tau_1 + \sum_{i=1}^{n_r} s_i \right) \right] \exp \left[ -\frac{\tau_0 + 1}{\theta} (e^{r\theta} - 1) \right].$$

Se observó un Proceso Poisson bajo este esquema, con  $r = 3$ . Los valores observados fueron  $n = 15$ ,  $\sum_{i=1}^{n_r} s_i = 26.7367$ . Se propuso una a priori conjugada con  $\tau_0 = \tau_1 = 1$ . En la Figura 5.4 se muestran las gráficas de las de las distribuciones inicial y final para este caso. Las constantes normalizadoras se calcularon numéricamente. ■

A falta de un resultado general, el investigador debe buscar por otros medios una familia conjugada para la función de verosimilitud dada en la Expresión (5.1). Bajo el supuesto

$$\lambda(t, \theta) = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_v t^v,$$

donde  $\theta = (\beta_0, \beta_1, \dots, \beta_v)$  y  $\lambda(\cdot, \theta) \geq 0$ , con  $\theta$  fijo. Se propone la siguiente familia de distribuciones *a priori*:

$$p(\theta) \propto g(\theta) \exp[\beta_0 \tau_0 + \beta_1 \tau_1 + \beta_2 \tau_2 + \dots + \beta_v \tau_v],$$

donde  $g(\theta)$  es un función no negativa sobre las variables  $(\beta_0, \beta_1, \dots, \beta_v)$ , de la forma

$$g(\theta) = \beta_0 \tau_{v+1} + \beta_1 \tau_{v+2} + \beta_2 \tau_{v+3} + \dots + \beta_v \tau_{2v+1}.$$

Para decidir sobre el grado del polinomio a usar, se puede recurrir al Criterio de Información de Akaike [ver Leonard].

Proceder de esta forma permite calcular de forma simple la distribución final, con un gran rango de posibilidades para representar el conocimiento *a priori* para  $\theta$ . De hecho la teoría no exige que  $\lambda(\cdot, \theta)$  sea un polinomio, pueden usarse expresiones del tipo

$$\lambda(t, \theta) = \alpha_0 + \beta_1 \text{sen}(t) + \beta_1 \cos(t) + \beta_2 \text{sen}(t^2) + \dots + \beta_v \cos(t^v),$$

o algún otra familia de aproximadores.

### 5.3.2 Ocurrencias en intervalos

En caso de que las cantidades observadas  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  consistan en el número de ocurrencias en subintervalos, es decir  $x_i = N_{v_i} - N_{v_{i-1}}$ , se sabe que

$$\begin{aligned} p(\mathbf{x} | \theta) &= \prod_{i=1}^n \exp \left[ - \int_{v_{i-1}}^{v_i} \lambda(u, \theta) du \right] \frac{\left( \int_{v_{i-1}}^{v_i} \lambda(u, \theta) du \right)^{x_i}}{x_i!} \\ &= \left( \prod_{i=1}^n \frac{1}{x_i!} \right) \exp \left[ - \int_0^r \lambda(u, \theta) du \right] \exp \left\{ \sum_{i=1}^n x_i \ln \left( \int_{v_{i-1}}^{v_i} \lambda(u, \theta) du \right) \right\}, \end{aligned}$$

de modo que haciendo  $f(x) = \frac{1}{x!}$ ,  $g(\theta) = \exp \left[ - \frac{1}{n} \int_0^r \lambda(u, \theta) du \right]$ ,  $\phi_i(\theta) = \ln \left( \int_{v_{i-1}}^{v_i} \lambda(u, \theta) du \right)$  y  $h(x_i) = x_i$  la densidad *a priori* conjugada natural puede encontrarse de forma inmediata.

**Ejemplo 5.3.2** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson con función de intensidad  $\lambda(x, \theta) = \theta e^x$ . Se observe el vector  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , donde  $x_i = N(v_i - 1, v_i]$  y  $v_{i-1} < v_i$  para toda  $i$ , entonces

$$\begin{aligned} p(\mathbf{x} | \theta) &\propto \prod_{k=1}^n e^{-\int_{v_{k-1}}^{v_k} \theta e^s ds} \frac{\left[ \int_{v_{k-1}}^{v_k} \theta e^s ds \right]^{x_k}}{x_k!} \\ &= \exp \{ -\theta (e^{v^n} - 1) \} \theta^{\sum_{k=1}^n x_k}. \end{aligned}$$

Se sigue que la densidad *a priori* conjugada es

$$p(\theta | \tau) \propto \exp \{ -\theta \tau_0 \} \theta^{\tau_1}.$$

Por tanto, la densidad *a posteriori* viene dada por

$$p(\theta | \mathbf{x}) \propto \exp \{ -\theta [(e^{v^n} - 1) + \tau_0] \} \theta^{\tau_1 + \sum_{i=1}^n x_i}.$$

es decir,  $\theta | \mathbf{x}$  posee una distribución **Gamma**  $(\tau_1 + \sum_{i=1}^n x_i + 1, (e^{v^n} - 1) + \tau_0)$ . ■

Hasta este momento se ha manejado a  $\theta$  como una variable unidimensional, el hacer  $\theta = (\theta_1, \theta_2)$  incrementa la complejidad de los cálculos, sin embargo los resultados se aplican exactamente de la misma forma.

**Ejemplo 5.3.3** *Un Proceso Poisson de intensidad  $\lambda(y, \theta) = \theta_1 + 2\theta_2 y$  ( $(\theta_1, \theta_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ ) es observado de forma ininterrumpida en  $(0, t]$ . Se registraron  $n = 2$  ocurrencias. La función de verosimilitud viene dada por*

$$p(n | \theta) \propto \left[ e^{-\int_0^t \lambda(y, \theta) dy} \right] \exp \left\{ n \ln \left( \int_0^t \lambda(y, \theta) dy \right) \right\}.$$

Se sigue que la densidad a priori conjugada es

$$p(\theta | \tau) \propto \left[ e^{-\tau_0 \int_0^t \lambda(y, \theta) dy} \right] \exp \left\{ \tau_1 \ln \left( \int_0^t \lambda(y, \theta) dy \right) \right\}.$$

Se supondrá  $\tau_1 = 1$ ,  $\tau_0 > 0$ . Resulta inmediato que

$$p(\theta | n) \propto e^{-(1+\tau_0)(\theta_1 t + \theta_2 t^2)} (\theta_1^3 t^3 + 3\theta_1^2 \theta_2 t^4 + 3\theta_1 \theta_2^2 t^5 + \theta_2^3 t^6).$$

Obteniendo la densidad posterior marginal para  $\theta_2$

$$\begin{aligned} p(\theta_2 | n) &\propto \int_0^\infty p(\theta | n) d\theta_1 \\ &= e^{-(1+\tau_0)\theta_2 t^2} \left[ t^{-1} \frac{\Gamma(4)}{(1+\tau_0)^4} + 3\theta_2 t \frac{\Gamma(3)}{(1+\tau_0)^3} + \right. \\ &\quad \left. 3\theta_2^2 t^3 \frac{\Gamma(2)}{(1+\tau_0)^2} + \theta_2^3 t^5 \frac{\Gamma(1)}{(1+\tau_0)} \right]. \end{aligned}$$

La constante normalizadora para este caso es

$$K(\tau_0) = (1 + \tau_0)^{-5} t^{-3} (2\Gamma(4)\Gamma(1) + 6\Gamma(2)\Gamma(3)). \quad \blacksquare$$

## 5.4 Tiempos de observación aleatorios

Un Proceso Poisson de intensidad  $\theta$  es observado de acuerdo al plan de muestreo propuesto en la Sección 4.5.3. Se denota con  $\{\Psi_1, \Psi_2, \dots, \Psi_{n_0}\}$  a las longitudes de los intervalos entre los tiempos de observación. Se supondrá que dichas variables son independientes, con distribución exponencial, e independientes del Proceso Poisson. Se tiene por hipótesis que las variables  $\{\Psi_1, \Psi_2, \dots, \Psi_{n_0}\}$  son observables.

Entonces la función de verosimilitud viene dada por

$$\begin{aligned} p(\mathbf{z} | \theta) &= \prod_{k=1}^{n_0} e^{-\theta \psi_k} \frac{(\theta \psi_k)^{z_k}}{z_k!} \\ &\propto \exp \left[ -\theta \sum_{k=1}^{n_0} \psi_k \right] \exp \left[ \ln \theta \sum_{k=1}^{n_0} z_k \right], \end{aligned}$$

si  $\{\Psi_1 = \psi_1, \Psi_2 = \psi_2, \dots, \Psi_{n_0} = \psi_{n_0}\}$ .

Haciendo

$$g(\theta) = \exp \left[ -\theta \sum_{k=1}^{n_0} \psi_k \right] \quad y \quad \phi(\theta) = \ln(\theta),$$

se tiene que una densidad *a priori* conjugada es

$$p(\theta | \tau) \propto \exp[-\theta \tau_1] \theta^{\tau_0}.$$

Por lo tanto

$$p(\theta | \mathbf{z}) \propto \exp \left[ -\theta \left( \tau_1 + \sum_{k=1}^{n_0} \psi_k \right) \right] \theta^{\tau_0 + \sum_{k=1}^{n_0} z_k},$$

es decir,  $\theta | \mathbf{z}$  se distribuye  $\text{Gam}(\tau_0 + \sum_{k=1}^{n_0} z_k + 1, \tau_1 + \sum_{k=1}^{n_0} \psi_k)$ .

Entonces el estimador bayesiano para este caso es

$$E(\theta | \mathbf{z}) = \frac{\tau_0 + \sum_{k=1}^{n_0} z_k + 1}{\tau_1 + \sum_{k=1}^{n_0} \psi_k}.$$

Otra forma de resolver este problema de estimación consiste en hacer inferencia para el vector de parámetros  $(\theta, \beta)$  donde  $\beta^{-1}$  es el parámetro correspondiente a la distribución de las variables  $\{\Psi_1, \Psi_2, \dots, \Psi_{n_0}\}$ , es decir, se toma

$$\begin{aligned} p(\mathbf{z} | \theta, \beta) &= \prod_{k=1}^{n_0} e^{-\theta \psi_k} \frac{(\theta \psi_k)^{z_k}}{z_k!} \beta e^{-\psi_k \beta} \\ &\propto \beta^{n_0} \exp \left[ -(\theta + \beta) \sum_{k=1}^{n_0} \psi_k \right] \exp \left[ \ln \theta \sum_{k=1}^{n_0} z_k \right], \end{aligned}$$

entonces la densidad *a priori* conjugada natural viene dada por

$$\begin{aligned} p(\theta, \beta | \tau) &\propto \exp[-(\theta + \beta) \tau_1 + \ln \theta \tau_2] \beta^{\tau_0} \\ &= \exp[-(\theta + \beta) \tau_1] \theta^{\tau_2} \beta^{\tau_0}. \end{aligned}$$

Se sigue que la densidad *a posteriori* de  $(\theta, \beta)$  es

$$p(\theta, \beta | \mathbf{z}) \propto \exp \left[ -(\theta + \beta) \left( \tau_1 + \sum_{k=1}^{n_0} \psi_k \right) \right] \theta^{\tau_2 + \sum_{k=1}^{n_0} z_k} \beta^{n_0 + \tau_0},$$

por lo tanto

$$p(\theta | \mathbf{z}) \propto \exp \left[ -\theta \left( \tau_1 + \sum_{k=1}^{n_0} \psi_k \right) \right] \theta^{\tau_2 + \sum_{k=1}^{n_0} z_k},$$

y

$$p(\beta | \mathbf{z}) \propto \exp \left[ -\beta \left( \tau_1 + \sum_{k=1}^{n_0} \psi_k \right) \right] \beta^{n_0 + \tau_0},$$

esto es  $\theta | \mathbf{z}$  y  $\beta | \mathbf{z}$  siguen una distribución Gamma con parámetros  $(\tau_2 + z_k + 1, \gamma)$  y  $(n_0 + \tau_0 + 1, \gamma)$  respectivamente, donde  $\gamma = \tau_1 + \sum_{k=1}^{n_0} \psi_k$ . Es claro que para  $\theta$ , se tiene la misma distribución *a posteriori* que se había obtenido antes. Es decir, el especificar o no a  $\beta$ , no tiene efecto alguno en las inferencias para  $\theta$ .

## 5.5 Estimación por intervalo

En muchos de los casos que se han tratado, la distribución *a posteriori* pertenece a la familia  $\{\text{Gam}(\alpha, \beta) : \alpha, \beta > 0\}$ ; si esto ocurre, el parámetro de forma  $\alpha$  determina como será la región de probabilidad para el parámetro de interés, el cual se denota con  $\theta$ . Si  $\alpha \leq 1$ , la región al de mayor densidad posterior al  $1 - \gamma$  para  $\theta$  está dada por

$$\{\theta : 0 < \theta \leq F_{\alpha, \beta}^{-1}(1 - \gamma)\},$$

donde  $F_{\alpha, \beta}$  es la función de distribución correspondiente a una  $\text{Gam}(\alpha, \beta)$ . Esto cambia cuando  $\alpha > 1$ , para ello se propone el siguiente algoritmo.

Es necesario que proponer cierto valor de tolerancia. Aquí  $f_{\alpha, \beta}$  denota a la función de densidad de una  $\text{Gam}(\alpha, \beta)$ .

**Algoritmo 5.5.1** -Encontrar región de mayor densidad posterior al  $1 - \gamma$  para una  $\text{Gam}(\alpha, \beta)$ ,  $\alpha > 1$ .

i) Crear una malla de puntos en  $(0, \frac{\alpha-1}{\beta})$ , y ordenarlos de menor a mayor.

El  $i$ -ésimo elemento de la malla, se denota con  $x_i$ .

$i=1$

Mientras  $F_{\alpha, \beta}(x_i) + (1 - \gamma) < 1$

ii) Se obtiene  $y_i := F_{\alpha, \beta}^{-1}(F_{\alpha, \beta}(x_i) + (1 - \gamma))$ .

iii) Si  $|f_{\alpha, \beta}(x_i) - f_{\alpha, \beta}(y_i)| < \text{tolerancia}$ , la salida del algoritmo es  $(x_i, y_i)$ .

iii') Si  $|f_{\alpha, \beta}(x_i) - f_{\alpha, \beta}(y_i)| \geq \text{tolerancia}$ , hacer  $i = i + 1$  e ir a ii).

El algoritmo aprovecha el hecho de que la única moda de la densidad se encuentra en  $\frac{\alpha-1}{\beta}$ . Se supone que la partición de  $(0, \frac{\alpha-1}{\beta})$  originada por la malla es lo suficientemente fina para que el algoritmo halle una solución.

**Ejemplo 5.5.2** Se aplicó este algoritmo a una densidad  $\text{Gam}(2, 1)$ , con el fin de encontrar la región de mayor densidad posterior al .85. El intervalo obtenido es el (0.130, 3.441). Esto puede verse en la Figura 5.5. ■

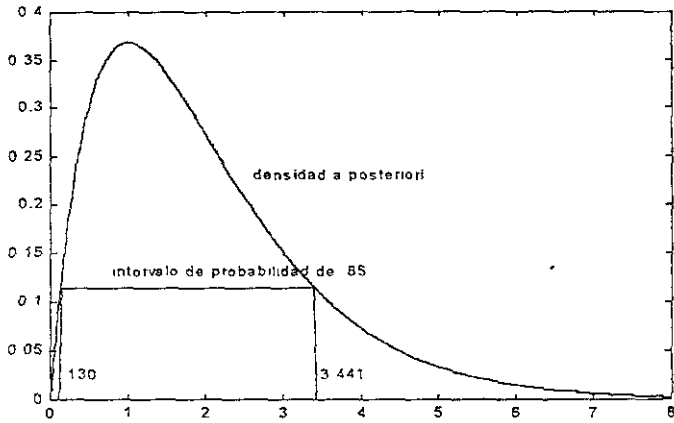


FIGURE 5.5.

### 5.6 Pruebas de hipótesis

Para resolver el problema de probar  $H_0 : \theta \in \Theta_0$  contra  $H_1 : \theta \in \Theta_1$  la Teoría Bayesiana propone el siguiente procedimiento. En vez de considerar niveles de significancia, lo que se hace es definir una función de pérdida de la forma

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
se acepta $H_0$	0	$l(d_0, \Theta_1)$
se acepta $H_1$	$l(d_1, \Theta_0)$	0

donde  $l(d_i, \Theta_j)$  denota a la pérdida asociada a haber aceptado la hipótesis  $i$  dado que  $H_j$  es verdadera, con  $l(d_i, \Theta_j) > 0$  si  $i \neq j$ . Para decidir en favor de  $H_0$  o de  $H_1$ , se compara la pérdida esperada para las distintas opciones, es decir, una vez obtenidas las cantidades

$$r_0 = p(\theta \in \Theta_1 | \mathbf{x}) l(d_0, \Theta_1)$$

$$r_1 = p(\theta \in \Theta_0 | \mathbf{x}) l(d_1, \Theta_0)$$

la acción óptima será  $d_i$ , si  $r_i = \min(r_0, r_1)$ ,  $i \in \{0, 1\}$ . Muchas veces se toma la distribución *priori*  $p(\theta)$  de modo que  $p(\theta \in \Theta_1) + p(\theta \in \Theta_0) = 1$ .

**Definición 5.6.1** Al cociente  $p(\theta \in \Theta_0 | \mathbf{x}) / p(\theta \in \Theta_1 | \mathbf{x})$  se le conoce como cociente de momios *posteriori*, mientras que a  $p(\theta \in \Theta_0) / p(\theta \in \Theta_1)$  se le llama cociente de momios *apriori*. A la cantidad

$$B = \frac{p(\theta \in \Theta_0 | \mathbf{x}) / p(\theta \in \Theta_1 | \mathbf{x})}{p(\theta \in \Theta_0) / p(\theta \in \Theta_1)}$$

se le denomina cociente de Bayes a favor de  $H_0$ .



El cociente de Bayes suele interpretarse como 'la ventaja de  $H_0$  respecto a  $H_1$  producida por los datos'. Esto resulta claro cuando las hipótesis son simples, esto es  $\Theta_0 = \{\theta_0\}$  y  $\Theta_1 = \{\theta_1\}$  con  $\theta_0 \neq \theta_1$ , ya que

$$p(\theta_0 | \mathbf{x}) = \frac{p(\theta_0) p(\mathbf{x} | \theta_0)}{p(\theta_0) p(\mathbf{x} | \theta_0) + p(\theta_1) p(\mathbf{x} | \theta_1)}$$

$$p(\theta_1 | \mathbf{x}) = \frac{p(\theta_1) p(\mathbf{x} | \theta_1)}{p(\theta_0) p(\mathbf{x} | \theta_0) + p(\theta_1) p(\mathbf{x} | \theta_1)}$$

entonces

$$\frac{p(\theta_0 | \mathbf{x})}{p(\theta_1 | \mathbf{x})} = \frac{p(\theta_0) p(\mathbf{x} | \theta_0)}{p(\theta_1) p(\mathbf{x} | \theta_1)}$$

y

$$B = \frac{p(\theta_0) p(\theta_1 | \mathbf{x})}{p(\theta_1) p(\theta_0 | \mathbf{x})}$$

$$= \frac{p(\mathbf{x} | \theta_0)}{p(\mathbf{x} | \theta_1)}$$

es decir,  $B$  es precisamente el cociente de verosimilitudes de  $H_0$  sobre  $H_1$ .

**Ejemplo 5.6.2** Sea  $\{N_t : t \geq 0\}$  un Proceso Poisson estacionario con intensidad  $\theta$ . Se desea probar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ . No se dispone de información a priori. Sin embargo, las creencias del investigador se concentran en estos dos valores, por lo tanto es razonable hacer  $p(\theta_0) = p(\theta_1) = \frac{1}{2}$ . Se propone la siguiente función de pérdida

	$\theta = \theta_0$	$\theta = \theta_1$
$d_0$	0	2
$d_1$	1	0

En este caso el investigador considera que "pierde el doble" si acepta que  $\theta = \theta_0$  cuando  $\theta = \theta_1$ , que al aceptar  $\theta = \theta_1$  siendo en verdad  $\theta = \theta_0$ .

Si el proceso se observa de manera ininterrumpida en  $(0, t]$  y la muestra consiste en los tiempos entre ocurrencias, se tiene que

$$p(\theta_0 | \mathbf{x}) = \frac{p(\theta_0) p(\mathbf{x} | \theta_0)}{p(\theta_0) p(\mathbf{x} | \theta_0) + p(\theta_1) p(\mathbf{x} | \theta_1)}$$

$$= \frac{p(\mathbf{x} | \theta_0)}{p(\mathbf{x} | \theta_0) + p(\mathbf{x} | \theta_1)}$$

$$= \frac{\theta_0^{N_t} \exp(-\theta_0 t)}{\theta_0^{N_t} \exp(-\theta_0 t) + \theta_1^{N_t} \exp(-\theta_1 t)}$$

Entonces la acción a tomarse tiene determinada por el mínimo entre

$$r_0 = \frac{2\theta_1^{N_t} \exp(-\theta_1 t)}{\theta_0^{N_t} \exp(-\theta_0 t) + \theta_1^{N_t} \exp(-\theta_1 t)}$$

y

$$r_1 = p(\theta_0 | \mathbf{x}).$$

Por otra parte, si las observaciones hubieran consistido en el número de ocurrencias en subintervalos (ver Sección 4.1.3) la probabilidad a posteriori sería

$$\begin{aligned} p(\theta_0 | \mathbf{x}) &= \frac{\exp(-\theta_0 t) \prod_{i=1}^n \left(\frac{\theta_0 t}{n}\right)^{n_i} / n_i!}{\exp(-\theta_0 t) \prod_{i=1}^n \left(\frac{\theta_0 t}{n}\right)^{n_i} / n_i! + \exp(-\theta_1 t) \prod_{i=1}^n \left(\frac{\theta_1 t}{n}\right)^{n_i} / n_i!} \\ &= \frac{\theta_0^{N_t} \exp(-\theta_0 t)}{\theta_0^{N_t} \exp(-\theta_0 t) + \theta_1^{N_t} \exp(-\theta_1 t)}. \end{aligned}$$

Es decir, a pesar de que se tomó un plan de muestreo distinto, la inferencia sobre  $\theta$  no cambia. ■

Resulta fácil generalizar este procedimiento al caso en el que se tienen más de dos hipótesis, esto es,  $H_0 : \theta \in \Theta_0, H_1 : \theta \in \Theta_1, \dots, H_m : \theta \in \Theta_m$ , donde  $\Theta_i \cap \Theta_j = \emptyset$  si  $i \neq j$ . El adoptar esta idea permite resolver algunos problemas que involucran toma de decisiones.

**Ejemplo 5.6.3** Es de interés calcular la edad de cierta pieza arqueológica, para ello se observa el decaimiento de un isótopo radioactivo contenido en dicha pieza. Este fenómeno puede modelarse usando un proceso Poisson estacionario  $\{N_t : t \geq 0\}$  de intensidad  $\theta$ . De acuerdo a la intensidad del proceso, es posible clasificar a la pieza arqueológica dentro de alguno de los siguientes períodos cronológicos

intensidad	período cronológico
$\theta \in (0, 1]$	anterior a 1000 años a.c.
$\theta \in (1, 2]$	entre 500 y 1000 años a.c.
$\theta > 2$	posterior a 500 años a.c.

El problema se traduce a comparar las hipótesis  $H_0 : \theta \in (0, 1], H_1 : \theta \in (1, 2]$  y  $H_2 : \theta > 2$ . Se tiene la siguiente función de pérdida

	$\theta \in (0, 1]$	$\theta \in (1, 2]$	$\theta \in (2, \infty)$
$d_0$	0	1	2
$d_1$	.5	0	1
$d_2$	.75	.5	0

En el contexto del problema, esto mide el 'costo por mala clasificación'. Las observaciones constan de los tiempos entre ocurrencias hasta el  $k$ -ésimo evento. La distribución a priori es  $\text{Gam}(2, \frac{1}{3})$ .

En este caso

$$\begin{aligned}
 p(\mathbf{x} | \Theta_0) &= \int_{\Theta_0} p(\mathbf{x} | \theta) p(\theta) d\theta \\
 &= \int_0^1 \theta^k \exp(-\theta S_k) \frac{3^2}{\Gamma(2)} \theta \exp(-3\theta) d\theta \\
 &= 9 \int_0^1 \theta^{k+1} \exp(-\theta(S_k + 3)) d\theta \\
 &= \frac{9}{(S_k + 3)^{k+2}} \int_0^1 z^{k+1} \exp(-z) dz \\
 &= \frac{9}{(S_k + 3)^{k+2}} \Gamma(k + 2) F_Y(1),
 \end{aligned}$$

donde  $Y \sim \text{Gam}\left(k + 2, \frac{1}{S_k + 3}\right)$ . Resulta fácil ver que

$$\begin{aligned}
 p(\mathbf{x} | \Theta_1) &= \frac{9}{(S_k + 3)^{k+2}} \Gamma(k + 2) (F_Y(2) - F_Y(1)) \\
 p(\mathbf{x} | \Theta_2) &= \frac{9}{(S_k + 3)^{k+2}} \Gamma(k + 2) (1 - F_Y(2)).
 \end{aligned}$$

Entonces, para obtener la probabilidad a posteriori de la hipótesis  $H_0$  es

$$\begin{aligned}
 &p(\Theta_0 | \mathbf{x}) \\
 &= \frac{F_\theta(1) p(\mathbf{x} | \Theta_0)}{F_\theta(1) p(\mathbf{x} | \Theta_0) + (F_\theta(2) - F_\theta(1)) p(\mathbf{x} | \Theta_1) + (1 - F_\theta(2)) p(\mathbf{x} | \Theta_2)} \\
 &= \frac{F_\theta(1) F_Y(1)}{F_\theta(1) F_Y(1) + (F_\theta(2) - F_\theta(1)) (F_Y(2) - F_Y(1)) + (1 - F_\theta(2)) (1 - F_Y(2))},
 \end{aligned}$$

por lo tanto

$$\begin{aligned}
 &p(\Theta_1 | \mathbf{x}) \\
 &= \frac{(F_\theta(2) - F_\theta(1)) (F_Y(2) - F_Y(1))}{F_\theta(1) F_Y(1) + (F_\theta(2) - F_\theta(1)) (F_Y(2) - F_Y(1)) + (1 - F_\theta(2)) (1 - F_Y(2))},
 \end{aligned}$$

y

$$p(\Theta_2 | \mathbf{x}) = 1 - (p(\Theta_0 | \mathbf{x}) + p(\Theta_1 | \mathbf{x})).$$

donde  $F_\theta$  es la función de distribución (a priori) de  $\theta$ .

Se sigue que las pérdidas esperadas para las hipótesis son

$$\begin{aligned}
 r_0 &= p(\Theta_1 | \mathbf{x}) + 2p(\Theta_2 | \mathbf{x}) \\
 r_1 &= .5p(\Theta_0 | \mathbf{x}) + p(\Theta_2 | \mathbf{x}) \\
 r_2 &= .75p(\Theta_0 | \mathbf{x}) + .5p(\Theta_1 | \mathbf{x}).
 \end{aligned}$$

Por lo tanto la pieza será clasificada dentro de la opción que tenga la menor pérdida esperada. ■

Otro caso de interés consiste en probar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta \neq \theta_0$ , donde  $\theta$  es un parámetro de la función de intensidad del Proceso Poisson. Esto introduce una dificultad técnica si la distribución *a priori*  $p(\theta)$  es continua,  $H_0$  nunca es aceptada, ya que  $p(\theta_0) = 0$ . Comúnmente se hace uso de alguno de los siguientes procedimientos para atacar este problema.

i) En vez de considerar  $H_0 : \theta = \theta_0$ , se prueba la hipótesis  $H_0 : \theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$  donde  $\varepsilon$  es una constante pequeña mayor que cero, de modo que el investigador pueda considerar 'indistinguibles' de  $\theta_0$  a los elementos de este intervalo. Existe el problema de que  $\varepsilon$  no tiene una interpretación probabilística, y por lo tanto no hay forma de dar una  $\varepsilon$  'óptima' en algún sentido.

ii) Se modifica  $p(\theta)$ , obteniendo así una  $p^*(\theta)$ , del siguiente modo

$$p^*(\theta) \begin{cases} \delta & \text{si } \theta = \theta_0 \\ (1 - \delta)p(\theta) & \text{si } \theta \neq \theta_0, \end{cases}$$

donde  $0 < \delta < 1$ . Es decir, se le asigna una probabilidad o masa a  $\theta_0$ , tratando en lo posible de conservar la forma de la distribución *a priori* original. Esto es congruente con la idea bayesiana de probabilidad, ya que, el hecho de plantear una prueba donde  $H_0 : \theta = \theta_0$  indica de manera implícita cierta creencia de que  $\theta$  en efecto toma ese valor.

**Ejemplo 5.6.4** *El gobierno de la ciudad de Londres ha llevado a cabo diversos estudios para determinar si debe cambiar los cruces peatonales tipo 'cebra' (frangas pintadas en la carretera) por cruces tipo 'pelicano' (con semáforo). De acuerdo a una investigación reciente, si el flujo de peatones se comporta de acuerdo a un proceso Poisson estacionario, conviene realizar el cambio. Existen buenas razones para suponer que el flujo de peatones se comporta de acuerdo a un proceso Poisson con función de intensidad  $\psi(t) = 20t + \lambda$ .*

*Es de interés probar  $H_0 : \theta = 0$  contra  $H_0 : \theta > 0$ . Aceptar  $H_0$  equivale a aceptar que el proceso es estacionario. Se observa el número de ocurrencias en  $(0, t]$ ,  $t > 0$ .*

*La función de verosimilitud es (ver Sección 4.1.3)*

$$p(\mathbf{x} | \theta) = \exp[-(\theta t^2 + \lambda t)] \frac{(\theta t^2 + \lambda t)^{N_t}}{N_t!}.$$

*La distribución a priori conjugada para  $\theta$  (bajo el supuesto que  $\lambda$  es especificado) viene dada por  $p(\theta | \tau) \propto \exp[-(\theta t^2 \tau_0)] (\theta t^2 + \lambda t)^{\tau_1}$  con  $\theta \geq 0$ ,  $\tau_0, \tau_1 > 0$ . Por lo tanto la distribución a posteriori es  $p(\theta | \mathbf{x}) \propto \exp[-\theta t^2(\tau_0 + 1)] (\theta t^2 + \lambda t)^{\tau_1 + N_t}$ . Si se asigna una probabilidad a priori de  $\frac{1}{3}$  al evento  $\{\theta = 0\}$ , entonces la densidad a priori modificada es de la forma*

$$p^*(\theta) \begin{cases} \frac{1}{3} & \text{si } \theta = 0 \\ \frac{2}{3}p(\theta) & \text{si } \theta \neq 0. \end{cases}$$

De este modo, la verosimilitud valuada en las hipótesis es

$$\begin{aligned} p(\mathbf{x} | \theta = 0) &\propto \exp(-\lambda t) \frac{(\lambda t)^{N_t}}{N_t!} \\ p(\mathbf{x} | \theta \neq 0) &\propto \int_0^\infty p(\mathbf{x} | \theta) p(\theta | \tau) d\theta \\ &= \int_0^\infty p(\theta | \mathbf{x}) K(\tau) d\theta \\ &= K(\tau). \end{aligned}$$

donde  $K(\tau)$  es la constante normalizadora de  $p(\theta | \mathbf{x})$ .

Por lo tanto las probabilidades a posteriori asignadas a las hipótesis son

$$p(\theta = 0 | \mathbf{x}) = \frac{\exp(-\lambda t) \left(\frac{\lambda t}{N_t}\right)^{N_t}}{\exp(-\lambda t) \left(\frac{\lambda t}{N_t}\right)^{N_t} + 2K(\tau)}$$

y

$$p(\theta > 0 | \mathbf{x}) = 1 - p(\theta = 0 | \mathbf{x}).$$

Se considera la siguiente función de pérdida

	$\theta = 0$	$\theta > 0$
$d_0$	0	1
$d_1$	1	0

entonces será tomada la decisión correspondiente a la alternativa con mayor probabilidad a posteriori.

■



Para todos los planes de muestreo discutido en esta tesis, se observó que la función de verosimilitud construida a partir de observaciones de un proceso Poisson estacionario tiene la misma forma funcional, salvo ciertas discrepancias en los exponentes. Estas diferencias están dadas por el tamaño de muestra. Esto explica porqué para el enfoque de máxima verosimilitud se obtiene siempre la misma expresión para el estimador puntual de la intensidad. Aquí es el tipo de muestreo lo que determina la distribución de dicho estimador.

También los estimadores resultan similares si se adopta el enfoque bayesiano, ya que toda inferencia se basa en la función de verosimilitud y en la distribución *a priori*; particularmente cuando se adoptan distribuciones *a priori* conjugadas. Sin embargo, las diferencias entre las funciones de verosimilitud pueden originar distribuciones de Jeffreys de formas distintas (Sección 5.2).

Para los planes de muestreo basados en observar el número de ocurrencias del proceso en una colección finita de intervalos (de longitud fija o aleatoria), el desempeño del estimador de máxima verosimilitud en términos de error cuadrático medio está determinado por dos factores: el primero es el tamaño de muestra, es decir, el número de subintervalos considerados en el muestreo. El segundo es el cociente entre la intensidad del proceso entre la longitud de los tiempos de observación (caso determinístico), o bien entre el parámetro de la distribución exponencial (Sección 5.2.2). Un valor pequeño para este cociente implica un aumento en la probabilidad de registrar ocurrencias en los tiempos de observación.

Los algoritmos propuestos para mejorar el estimador máximo verosímil en la Sección 4.5.3 cumplen con su objetivo de forma perceptible. Esto se afirma en base al estudio de simulación presentado en la misma sección. El estimador corregido usando el Algoritmo 4.5.3 resultó ser poco sensible a cambios en el valor de  $\phi_1$  (el parámetro de la distribución de los tiempos de observación para la primera etapa). Para distintos valores de los parámetros, se observó que los valores simulados presentaban poca variabilidad (repeticiones del mismo experimento), y una velocidad de convergencia (simulando un solo experimento, aumentando el tamaño de muestra) tal que la diferencia en valor absoluto entre el estimador y el verdadero valor de la intensidad es prácticamente despreciable (tomando como

referencia la intensidad misma) a partir de  $n = 50$ . Esto sólo tiene relevancia práctica si la primera etapa del algoritmo puede llevarse a cabo, es decir, si es factible tomar un número arbitrariamente grande de tiempos de observación con parámetro  $\phi_1$  antes de registrar la primera ocurrencia. Para el estimador corregido usando el Algoritmo 4.5.5, los resultados son prácticamente análogos si se tiene que  $\lambda \gg \phi_1$ . Conforme se hace aumentar  $\lambda$ , el estimador corregido usando el Algoritmo 4.5.5 se comporta de un modo más semejante al estimador máximo verosímil.

El hecho de que las distribuciones involucradas en los planes de muestreo pertenecieran a la Familia Exponencial facilitó la demostración de ciertas propiedades de los estimadores, por ejemplo: la completez y el cálculo de densidades conjugadas. Se tuvo además que las funciones de verosimilitud que se analizaron cumplen con las condiciones de regularidad (diferenciabilidad) necesarias para emplear resultados asintóticos.

Para los procesos no estacionarios y compuestos, no siempre es claro como proceder; a menos de que la función de verosimilitud pueda expresarse como miembro de la Familia Exponencial.



# Apéndice A

## Probabilidad

### A.1 Espacios y medidas de probabilidad

El primer objeto que se necesita en la definición de espacio de probabilidad es un conjunto, el cual se denota con  $\Omega$ . El segundo objeto se presenta a continuación.

Una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  es una colección  $\mathfrak{F}$  de subconjuntos de  $\Omega$  tal que tiene a  $\emptyset$  como miembro y es cerrada bajo complementación y uniones numerables. Si se pide que la colección sólo sea cerrada para uniones finitas, lo que se obtiene es álgebra de subconjuntos de  $\Omega$ . Los conjuntos que son elementos de la  $\sigma$ -álgebra se les llama  $\mathfrak{F}$ -medibles, o solamente medibles si la  $\sigma$ -álgebra está especificada en el contexto. Una  $\sigma$ -álgebra  $\mathfrak{G}$  es menor que otra  $\sigma$ -álgebra  $\mathfrak{F}$ , si para todo  $A \in \mathfrak{G}$ , se tiene que  $A \in \mathfrak{F}$ ; también se dice que  $\mathfrak{G}$  es una sub- $\sigma$ -álgebra de  $\mathfrak{F}$ .

Una medida de probabilidad  $\mathbf{P}$  sobre una  $\sigma$ -álgebra  $\mathfrak{F}$  de subconjuntos de  $\Omega$  es una función que va de  $\mathfrak{F}$  al intervalo  $[0, 1]$  tal que  $\mathbf{P}(\Omega) = 1$  y

$$\Pr \left( \bigcup_{m=1}^{\infty} A_m \right) = \sum_{m=1}^{\infty} \mathbf{P}(A_m)$$

para cada sucesión de conjuntos ajenos por parejas  $\{A_m \in \mathfrak{F} : m \in \mathbf{N}\}$ . Debido a esta propiedad, se dice que  $\mathbf{P}$  es *contablemente aditiva*. Para  $A \in \mathfrak{F}$ ,  $\mathbf{P}(A)$  es la *probabilidad de A*.

Un *espacio de probabilidad* es una terna  $(\Omega, \mathfrak{F}, \mathbf{P})$ , donde  $\Omega$  es un conjunto,  $\mathfrak{F}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$ , y  $\mathbf{P}$  es una medida de probabilidad sobre  $\mathfrak{F}$ . Al conjunto  $\Omega$  se le llama *espacio muestral*, y a sus miembros se les llama *puntos muestrales*. A los miembros de  $\mathfrak{F}$  se les denomina *eventos*.

Un *espacio medible* es un par  $(\Omega, \mathfrak{F})$ , donde  $\Omega$  es un conjunto y,  $\mathfrak{F}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$ . Sean  $(\Omega, \mathfrak{F})$  y  $(\Psi, \mathfrak{G})$  espacios medibles. Una función medible de  $(\Omega, \mathfrak{F})$  a  $(\Psi, \mathfrak{G})$  es una función  $X : \Omega \rightarrow \Psi$  tal que  $X^{-1}(B) \in \mathfrak{F}$  para todo  $B \in \mathfrak{G}$ . Cuando una medida de probabilidad  $\mathbf{P}$  se encuentra asociada al espacio medible  $(\Omega, \mathfrak{F})$ , de modo que  $(\Omega, \mathfrak{F}, \mathbf{P})$  es un espacio de probabilidad, a  $X$  se le llama *variable aleatoria* de  $(\Omega, \mathfrak{F}, \mathbf{P})$  a  $(\Psi, \mathfrak{G})$ .

Esta terminología tiende a acortarse. Por ejemplo, si las  $\sigma$ -álgebras  $\mathfrak{F}$  y  $\mathfrak{G}$  y la medida de probabilidad  $\mathbf{P}$  se sobreentienden del contexto, se dice que  $X$  es una variable aleatoria de  $\Omega$  a  $\Psi$ , o simplemente que  $X$  es una variable aleatoria que toma valores en  $\Psi$ .

Un espacio medible de interés es  $(\mathbf{R}^k, \mathfrak{B}(\mathbf{R}^k))$ , donde  $\mathfrak{B}(\mathbf{R}^k)$  es la mínima  $\sigma$ -álgebra que contiene a los conjuntos abiertos de  $\mathbf{R}^k$ . a dicha  $\sigma$ -álgebra se le llama la  $\sigma$ -álgebra de Borel de  $\mathbf{R}^k$ . Para estos casos se dirá que  $X$  toma valores en  $\mathbf{R}^k$ .

La medida de Lebesgue de  $\mathbf{R}^k$ , es la medida definida sobre  $(\mathbf{R}^k, \mathfrak{B}(\mathbf{R}^k))$ , tal que, a todo  $k$ -cubo

$$C = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]$$

con  $a_i < b_i$  asigna el valor

$$\mu(C) = \prod_{j=1}^k (b_j - a_j).$$

Sea  $(\Psi, \mathfrak{G})$  un espacio medible. La medida de contar  $\tau$  sobre este espacio es una medida dada por

$$\tau(A) \begin{cases} \#(A) & \text{si } A \text{ es finito} \\ \infty & \text{en caso contrario} \end{cases}$$

para todo  $A \in \mathfrak{G}$ .

Se dice que una medida  $\mu$  es  $\sigma$ -finita si existe una partición  $A_1, A_2, \dots$  de  $\Omega$ , tal que  $\mu(A_j) < \infty$ , para todo  $j$ .

Sea  $X$  una función  $X: \Omega \rightarrow \mathbf{R}$ . La mínima  $\sigma$ -álgebra de  $\Omega$  que hace que  $X$  sea una variable aleatoria, se le conoce como la  $\sigma$ -álgebra generada por  $X$ .

Sea  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio de probabilidad. Una proposición  $\mathcal{H}$  se cumple  $\mathbf{P}$ -casi seguramente si existe  $A \in \mathfrak{F}$  tal que  $\mathbf{P}(A) = 0$ , y

$$\{\omega \in \Omega : \mathcal{H} \text{ no se cumple}\} \subseteq A.$$

Este concepto también se define para espacios de medida.

## A.2 Lemas de clases monótonas

Sea  $\mathcal{C}$  una clase de subconjuntos de cierto conjunto  $\Omega$ . La familia  $\mathcal{C}$  es una clase monótona si:

- i) Si  $A_1, A_2, \dots \in \mathcal{C}$ , con  $A_i \subseteq A_j$  si  $i < j$ , y  $A = \bigcup_{i=1}^{\infty} A_i$ , entonces  $A \in \mathcal{C}$ .
- ii) Si  $A_1, A_2, \dots \in \mathcal{C}$ , con  $A_i \supseteq A_j$  si  $i < j$ , y  $A = \bigcap_{i=1}^{\infty} A_i$ , entonces  $A \in \mathcal{C}$ .

El siguiente resultado relaciona este concepto con el de  $\sigma$ -álgebra.

**Lema de Clases Monótonas.** Sea  $\mathfrak{F}_0$  un álgebra de conjuntos de  $\Omega$ , y  $\mathcal{C}$  una familia de subconjuntos de  $\Omega$  tal que es una clase monótona. Si  $\mathcal{C} \supset \mathfrak{F}_0$ , entonces  $\mathcal{C}$  incluye a la mínima  $\sigma$ -álgebra que contiene a  $\mathfrak{F}_0$ .

Sea  $\mathcal{D}$  una clase de subconjuntos de  $\Omega$ . Se dice que  $\mathcal{D}$  es un sistema de Dynkin ( $\mathcal{D}$ -sistema) si se cumplen las siguientes condiciones:

- i)  $\Omega \in \mathcal{D}$ .
- ii) Si  $A, B \in \mathcal{D}$ ,  $B \subset A$ , entonces  $A - B \in \mathcal{D}$ . Es decir,  $\mathcal{D}$  es cerrado bajo diferencias.
- iii) Si  $A_1, A_2, \dots \in \mathcal{D}$ , con  $A_i \subset A_j$  si  $i < j$ , y  $A = \bigcup_{i=1}^{\infty} A_i$ , entonces  $A \in \mathcal{D}$ .

Se cuenta con el siguiente resultado:

**Lema de Dynkin.** Sea  $\mathcal{J}$  una clase de subconjuntos de  $\Omega$ , de modo que  $\mathcal{J}$  es cerrado bajo intersecciones finitas. Si  $\mathcal{D}$  es un  $D$ -sistema y  $\mathcal{D} \supset \mathcal{J}$ , entonces  $\mathcal{D}$  incluye a la mínima  $\sigma$ -álgebra que contiene a  $\mathcal{J}$ .

### A.3 Teorema de la Probabilidad Total

Sea  $(\Omega, \mathfrak{F}, \mathcal{P})$  un espacio de probabilidad y sean  $B_1, B_2, \dots, B_n$  elementos de  $\mathfrak{F}$  tales que conforman una partición de  $\Omega$ . Supóngase además que  $\Pr\{B_i\} > 0$  para  $i \in \{1, 2, \dots, n\}$ , entonces para todo  $A \in \mathfrak{F}$  se tiene que

$$\Pr\{A\} = \sum_{j=1}^n \Pr\{A \mid B_j\} \Pr\{B_j\}.$$

### A.4 Teoría de integración

Sea  $g$  una función Borel-medible, definida sobre  $(\Psi, \mathcal{G})$ . Se dice que  $g$  es *simple*, si existen  $A_1, A_2, \dots, A_k$  en  $\mathcal{G}$  tales que forman una partición de  $\Psi$ , y  $x \in A_i$  implica  $g(x) = y_i$ , donde  $y_r \neq y_j$  si  $r \neq j$ . Se considera ahora una función simple  $g$ , definida sobre un espacio de medida  $(\Psi, \mathcal{G}, \mu)$ . La *integral de Lebesgue* de  $g$  respecto a  $\mu$  se define como

$$\int g d\mu = \sum_{i=1}^k y_i \mu(A_i)$$

si esta suma es finita; de no ser así, se dice que la integral *no existe*.

Se puede demostrar que toda función Borel-medible no negativa  $g$  sobre  $(\Psi, \mathcal{G}, \mu)$  puede expresarse como el límite puntual ( $\mu$ -casi seguramente) de una sucesión monótona creciente de funciones simples no negativas  $\{h_i : i \in \mathbf{N}\}$ . La integral de Lebesgue de  $g$  respecto a  $\mu$  se define como

$$\int g d\mu = \sup_{i \in \mathbf{N}} \left\{ \int h_i d\mu \right\},$$

Si este supremo es finito. Se dice entonces que  $g$  es  $\mu$ -integrable.

Sea  $g$  una función Borel-medible. Se tiene que  $g = g^+ - g^-$ , donde

$$g^+ = \max(0, g) \quad \text{y} \quad g^- = \max(0, -g).$$

Si  $g^+$  y  $g^-$  son  $\mu$ -integrables, la integral de  $g$  respecto a  $\mu$  se define como:

$$\int g d\mu = \int g^+ d\mu - \int g^- d\mu.$$

La Integral de Lebesgue posee las siguientes propiedades:

- i) Si  $\int g d\mu$  existe y  $c \in \mathbf{R}$ , entonces  $\int cgd\mu$  existe, y es igual a  $c \int g d\mu$ .
- ii) Si  $g(\omega) \geq h(\omega)$  para todo  $\omega \in \Psi$ , entonces  $\int g d\mu \geq \int h d\mu$  en el sentido de que si  $\int h d\mu$  existe y es mayor a  $-\infty$ , entonces  $\int g d\mu$  existe y  $\int g d\mu \geq \int h d\mu$ ; si  $\int g d\mu$  existe y es menor a  $+\infty$ , entonces  $\int h d\mu$  existe y  $\int g d\mu \geq \int h d\mu$ . Por tanto, si ambas integrales existen  $\int g d\mu \geq \int h d\mu$ , sin importar si sólo una, o ninguna sea finita.
- iii) Si  $\int h d\mu$  existe, entonces  $|\int h d\mu| \leq \int |h| d\mu$ .
- iv) Si  $\int h d\mu$  existe, entonces  $\int_A h d\mu := \int I_A h d\mu$  también existe, para todo  $A \in \mathcal{G}$ .

Un aspecto muy útil de la Integral de Lebesgue es, que bajo ciertas condiciones, es posible 'intercambiar' el símbolo de límite con el de integral. A continuación se enuncian los teoremas básicos de convergencia de integrales.

**Teorema de Convergencia Monótona.** Sea  $\{h_i : i \in \mathbf{N}\}$  una sucesión creciente de funciones Borel medibles no negativas, y sea  $h(\omega) = \lim_{n \rightarrow \infty} h_n(\omega)$ ,  $\omega \in \Psi$ . Entonces  $\int h_n d\mu \rightarrow \int h d\mu$ .

**Lema de Fatou.** Sean  $\{f_i : i \in \mathbf{N}\}$  y  $f$  funciones Borel medibles. Si  $f_n \geq f$  para toda  $n$ , donde  $\int f d\mu > -\infty$ , entonces

$$\liminf_{n \rightarrow \infty} \int f_n d\mu \geq \int \left( \liminf_{n \rightarrow \infty} f_n \right) d\mu.$$

**Teorema de Convergencia Dominada.** Si  $\{f_i : i \in \mathbf{N}\}$  y  $g$  son Borel medibles,  $|f_n| \leq g$  para toda  $n$ , donde  $g$  es  $\mu$ -integrable, y  $f_n(\omega) \rightarrow f(\omega)$ , para toda  $\omega \in \Psi \setminus A$ , con  $A \in \mathcal{G}$  y  $\mu(A) = 0$ , entonces  $f$  es  $\mu$ -integrable y  $\int f_n d\mu = \int f d\mu$ .

Sea  $(\Psi, \mathcal{G}, \mu)$  un espacio de medida, y sea  $p$  un número real con  $p \geq 1$ . El conjunto de las funciones Borel-medibles  $f$  tales que

$$\int |f|^p d\mu < \infty$$

se denota con  $\mathcal{L}^p(\Psi, \mathcal{G}, \mu)$ . Este conjunto tiene muchas propiedades importantes. En particular, se trata de un espacio vectorial normado y completo; donde

$$\|f\|_p = \left( \int |f|^p d\mu \right)^{\frac{1}{p}}$$

define a la norma.

### A.5 Función de Distribución

Sea  $X$  una variable aleatoria que toma valores en  $\mathbf{R}$ . La función de distribución de  $X$ , se denota con  $F_X$  y está dada por

$$F_X(x) = \text{Pr}\{X \leq x\} \quad \text{para todo } x \in \mathbf{R}.$$

Toda función de distribución  $F$  es *creciente* ( $a < b$  implica que  $F(a) < F(b)$ ) y *continua por la derecha* ( $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$ ).

Una *medida de Lebesgue-Stieltjes* sobre  $\mathbb{R}$  es una medida  $\mu$  definida sobre  $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$  tal que  $\mu(I) < \infty$  para todo intervalo acotado  $I$ . Existe una correspondencia uno a uno entre medidas de Lebesgue-Stieltjes y funciones de distribución.

Sea  $X$  una variable aleatoria. Una variable aleatoria  $X$  es llamada *discreta*, si existen  $\{a_i \in \mathbb{R} : i \in \mathbb{N}\}$  tales que  $\Pr\{X = a_i\} = q_i > 0$  y  $\sum_{i=1}^{\infty} q_i = 1$ . Si  $X$  es tal que  $F_X$  es continua en todo  $\mathbb{R}$ , se dice entonces que  $X$  es una variable aleatoria *continua*.

### A.6 Independencia

Sea  $\{A_1, A_2, \dots, A_n\}$  una familia de eventos. Se dice que dichos eventos son *independientes*, si para toda subfamilia  $\{A_{k_1}, A_{k_2}, \dots, A_{k_m}\}$ , ( $m \leq n$ ), se tiene que

$$\Pr\{A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_m}\} = \prod_{i=1}^m \Pr\{A_{k_i}\}.$$

Sea  $\{X_1, X_2, \dots, X_n\}$  una familia de variables aleatorias, dichas variables son *independientes*, si para toda subfamilia  $\{X_{k_1}, X_{k_2}, \dots, X_{k_m}\}$  se cumple que

$$\Pr\{X_{k_1} \leq x_1, X_{k_2} \leq x_2, \dots, X_{k_m} \leq x_m\} = \prod_{i=1}^m \Pr\{X_{k_i} \leq x_i\},$$

para todo  $(x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ .

### A.7 Funciones de densidad y de masa

**Teorema de Radon-Nikodým.** Sean  $\mathbf{P}_0$  y  $\mathbf{P}$  medidas  $\sigma$ -finitas sobre  $(\Omega, \mathfrak{F})$ , con  $\mathbf{P}$  absolutamente continua respecto a  $\mathbf{P}_0$ ; es decir  $\mathbf{P}_0(E) = 0$  implica  $\mathbf{P}(E) = 0$  para todo  $E \in \mathfrak{F}$ . También se dice que  $\mathbf{P}$  está dominada por  $\mathbf{P}_0$ . Entonces existe  $f \in \mathcal{L}^1(\Omega, \mathfrak{F}, \mathbf{P}_0)$  tal que

$$\mathbf{P}(E) = \int_E f d\mathbf{P}_0$$

A dicha  $f$  se le conoce como la *derivada de Radon-Nikodým de  $\mathbf{P}$  respecto a  $\mathbf{P}_0$* .

Una variable aleatoria  $Y$  tiene una *función de densidad de probabilidad*, si existe una función Borel-medible,  $f_Y : \mathbb{R} \rightarrow [0, \infty)$  tal que

$$\Pr(Y \in B) = \int_B f_Y(u) du \quad B \in \mathfrak{B}(\mathbb{R}).$$

Esto ocurre en particular cuando  $Y$  es continua. En términos del Teorema de Radon-Nikodým, la medida de Lebesgue-Stieltjes asociada a la variable aleatoria  $Y$ , tiene densidad  $f_Y$  relativa a la medida de Lebesgue.

La función de densidad es útil para calcular probabilidades. Se tiene que

$$\Pr(X \in (a, b)) = \int_a^b f_X(t) dt, \quad \text{para todo } x \in \mathbf{R}.$$

para variables aleatorias continuas. En el caso discreto

$$p_X(x) = \Pr\{X = x\} \quad x \in \mathbf{R},$$

aquí a la densidad se le conoce también como la *función de masa* de  $X$ .

## A.8 Media y Varianza

Si  $X$  es una variable aleatoria definida en  $(\Omega, \mathfrak{F}, \mathbf{P})$ , la *esperanza de  $X$*  se define como

$$E[X] = \int X d\mathbf{P}$$

bajo el supuesto de que la integral existe. Entonces  $E[X]$  es la integral de una función Borel medible  $X$  respecto a una medida de probabilidad  $\mathbf{P}$ , por lo tanto los resultados de Teoría de Integración pueden aplicarse.

Como casos particulares, se tiene que

$$E[X] = \sum_{i=1}^{\infty} x_i p_X(x_i)$$

si  $X$  discreta con puntos de masa  $x_1, x_2, \dots, x_j, \dots$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

si  $X$  es continua con función de densidad  $f_X$ .

## A.9 Probabilidad conjunta y marginal

La *ley de probabilidad conjunta* de dos variables aleatorias  $Y$  y  $Z$  se define como

$$\Pr_{Y,Z}(\Gamma) := \Pr\{(Y, Z) \in \Gamma\},$$

la cual es una función  $\Pr_{Y,Z} : \Gamma = \mathfrak{B}(\mathbf{R}) \times \mathfrak{B}(\mathbf{R}) \rightarrow [0, 1]$ . Equivalentemente, la *función de distribución conjunta* de  $Y$  y  $Z$  es

$$F_{Y,Z}(y, z) = \Pr\{Y \leq y, Z \leq z\}.$$

$Y$  y  $Z$  poseen una *función de densidad conjunta*,  $f_{Y,Z}$ , si

$$\begin{aligned} \Pr \{(Y, Z) \in \Gamma\} &= \int_{\Gamma} f_{Y,Z}(u) \, du \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} 1_{\Gamma}(u, v) f_{Y,Z}(u, v) \, du \, dv. \end{aligned}$$

donde  $1_{\Gamma}(u, v)$  es la función indicadora apropiada. Si  $Y, Z$  pertenecen a  $\mathcal{L}^2$ , su covarianza es

$$\begin{aligned} \text{cov}[Y, Z] &= E[(Y - E[Y])(Z - E[Z])] \\ &= E[YZ] - E[Y]E[Z]. \end{aligned}$$

Si dos variables aleatorias  $Y$  y  $Z$ , son independientes

$$E[YZ] = E[Y]E[Z]$$

y

$$\text{var}[Y + Z] = \text{var}[Y] + \text{var}[Z].$$

Sin embargo, el recíproco no se cumple.

Las *densidades marginales* de una función de densidad conjunta  $f_{Y,Z}$  están dadas por

$$\begin{aligned} f_Y(y) &= \int_{\mathbf{R}} f_{Y,Z}(y, v) \, dv \\ f_Z(z) &= \int_{\mathbf{R}} f_{Y,Z}(u, z) \, du. \end{aligned}$$

Una distribución marginal para variables aleatorias discretas es obtenida de modo similar, sólo que ahora se efectúa una suma. Si  $Y$  y  $Z$  son independientes

$$F_{Y,Z}(y, z) = F_Y(y)F_Z(z),$$

y, si ambas densidades existen, se tiene que

$$f_{Y,Z}(y, z) = f_Y(y)f_Z(z).$$

## A.10 Convoluciones

Para dos variables aleatorias independientes,  $Y$  y  $Z$ , la distribución de su suma,  $S = Y + Z$ , está dada por la *convolución*

$$\begin{aligned} f_{Y+Z}(y+z) &= \int_{\mathbf{R}} f_Y(s-u) f_Z(u) \, du \\ &= \int_{\mathbf{R}} f_Y(u) f_Z(s-u) \, du. \end{aligned}$$

La convolución es conmutativa y asociativa.

### A.11 Esperanza y probabilidad condicional

Sea  $X$  una variable aleatoria definida en  $(\Omega, \mathfrak{F}, \mathbf{P})$ , con  $E(|X|) < \infty$  y sea  $\mathfrak{G}$  una sub- $\sigma$ -álgebra de  $\mathfrak{F}$ . Se dice que  $Y = E[X | \mathfrak{G}]$  es la *esperanza condicional de  $X$  dado  $\mathfrak{G}$*  si:

- i)  $Y$  es  $\mathfrak{G}$ -medible y  $E(|Y|) < \infty$ .
- ii)  $\int_B X d\mathbf{P} = \int_B Y d\mathbf{P}$  para todo  $B \in \mathfrak{G}$ .

La esperanza condicional posee las siguientes propiedades.

- i) Si  $X \equiv c$  entonces  $E[X | \mathfrak{G}] = c$ ,
- ii)  $X \geq 0$  implica  $E[X | \mathfrak{G}] \geq 0$ ,
- iii) El operador esperanza es lineal, es decir

$$E[aX_1 + bX_2 | \mathfrak{G}] = aE[X_1 | \mathfrak{G}] + bE[X_2 | \mathfrak{G}],$$

- iv)  $X_1 \leq X_2$  entonces  $E[X_1 | \mathfrak{G}] \leq E[X_2 | \mathfrak{G}]$ ,
- v) Para una sucesión monótona  $0 \leq X_n \leq X_{n+1} \dots$   $n \in \{1, 2, \dots\}$  se tiene que

$$E\left[\lim_{n \rightarrow \infty} X_n | \mathfrak{G}\right] = \lim_{n \rightarrow \infty} E[X_n | \mathfrak{G}],$$

donde estas afirmaciones se cumplen con  $\mathbf{P}$ -casi seguramente.

La *varianza condicional* de  $Y$  dado  $X = x$  se define como

$$\text{var}[Y | X = x] = E[Y^2 | X = x] - (E[Y | X = x])^2.$$

Sea  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio de probabilidad y sean  $A$  y  $B$  elementos de  $\mathfrak{F}$ , con  $\Pr\{B\} > 0$ . La *probabilidad condicional* de  $A$  dado  $B$  se denota con  $\Pr\{A | B\}$ , y se define como

$$\Pr\{A | B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}.$$

Sean  $X$  y  $Y$  variables aleatorias discretas con función de masa conjunta  $p_{X,Y}(x, y)$ . La *función de masa condicional* de  $X$  dado  $Y$ , se denota con  $p_{X|Y}(\cdot | \cdot)$ , y está dada por

$$p_{X|Y}(x, y) = \frac{p_{X,Y}(x, y)}{p_Y(y)},$$

si  $p_Y(y) > 0$ , donde  $p_Y$  es la función de masa marginal de  $Y$ , y queda indefinida si  $p_Y(y) = 0$ .

Estas ideas puede generalizarse usando el concepto de esperanza condicional.

Sea  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio de probabilidad, y  $\mathfrak{G}$  una sub- $\sigma$ -álgebra de  $\mathfrak{F}$ . Se toma  $A \in \mathfrak{F}$  fijo. Existe una función  $\mathfrak{G}$ -medible  $\Pr(A | \mathfrak{G})$ , llamada la *probabilidad condicional de  $A$  dada  $\mathfrak{G}$* , tal que

$$\Pr(C \cap A) = \int_C \Pr(A | \mathfrak{G}) d\mathbf{P}.$$



Esta función es única  $\mathbf{P}$ -casi seguramente, y de hecho

$$\Pr(A \mid \mathfrak{G}) = E\{1_A \mid \mathfrak{G}\}.$$

El definir a la probabilidad condicional de esta forma permite, entre otras cosas, calcular probabilidades del tipo

$$\Pr(A \mid Y = y)$$

donde  $A \in \mathfrak{F}$  y  $Y$  es una variable aleatoria continua definida en  $(\Omega, \mathfrak{F}, \mathbf{P})$ . Un caso particular de interés es el de las *funciones de densidad condicional*.

Sean  $X$  y  $Y$  variables aleatorias continuas con función de densidad conjunta  $f_{X,Y}(x, y)$ . La *función de densidad condicional* de  $X$  dado  $Y$ , se denota con  $f_{X|Y}(\cdot \mid \cdot)$ , y está dada por

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

si  $f_Y(y) > 0$ , donde  $f_Y$  es la función de densidad marginal de  $Y$ , y queda indefinida si  $f_Y(y) = 0$ . Se cuenta con las siguientes identidades:

Sea  $(X, Y)$  un vector aleatorio dos-dimensional, y  $g(\cdot, \cdot)$  una función medible de dos variables. La *esperanza condicional* de  $g(X, Y)$  dado  $\{X = x\}$ , se denota con  $E[g(X, Y) \mid X = x]$ , y viene dada por

$$E[g(X, Y) \mid X = x] = \int_{-\infty}^{\infty} g(x, y) f_{Y|X}(y, x) dy,$$

si  $(X, Y)$  son conjuntamente continuas; y

$$E[g(X, Y) \mid X = x] = \sum g(x, y_j) p_{Y|X}(y_j, x),$$

si  $(X, Y)$  son conjuntamente discretas, donde la suma se toma sobre todos los posibles valores de  $Y$ .

## A.12 La Regla de Bayes

Sea  $(\Omega, \mathfrak{F}, \mathbf{P})$  un espacio de probabilidad y sean  $B_1, B_2, \dots, B_n$  elementos de  $\mathfrak{F}$  tales que conforman una partición de  $\Omega$ , y sea  $A \in \mathfrak{F}$ . Se tiene entonces que

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i) \Pr(B_i)}{\sum_{j=1}^n \Pr(A \mid B_j) \Pr(B_j)},$$

donde  $i \in \{1, 2, \dots\}$ . A esta identidad se le conoce como la Regla de Bayes.

Esto puede generalizarse de la siguiente manera.

Sean  $Y$  y  $Z$  con función de densidad conjunta estrictamente positiva sobre  $\mathbf{R}^2$ , se tiene que

$$\begin{aligned} f_{Y|Z}(y, z) &= \frac{f_{Y,Z}(y, z)}{f_Z(z)} \\ &= \frac{f_{Z|Y}(z, y) f_Y(y)}{f_Z(z)}, \end{aligned}$$

entonces

$$f_{Y|Z}(y, z) \propto f_Y(y) f_{Z|Y}(z, y).$$

Donde la constante de normalización, la cual está en función de  $y$ , está determinada por

$$\int_{\mathbf{R}} f_{Y|Z}(u, z) du = 1.$$

## A.13 Teoría Límite

### A.13.1 Tipos de Convergencia

**Convergencia casi segura:**  $Y_n \rightarrow Y$  *casi seguramente*, casi donde sea, o con probabilidad 1, si

$$\Pr \left\{ \lim_{n \rightarrow \infty} Y_n = Y \right\} = 1.$$

**Convergencia en  $\mathcal{L}^r$ :**  $Y_n \rightarrow Y$  en  $\mathcal{L}^r$  si cada  $Y_n$  pertenece a  $\mathcal{L}^r$  y  $Y$  pertenece a  $\mathcal{L}^r$  y

$$\lim_{n \rightarrow \infty} E[|Y_n - Y|^r] = 0.$$

También es llamada *convergencia en  $r$ -media*.

**Convergencia en probabilidad:**  $Y_n \rightarrow Y$  en probabilidad si, para toda  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr(|Y_n - Y| > \varepsilon) = 0.$$

**Convergencia en distribución:** Esta puede expresarse de diversas formas

$$\lim_{n \rightarrow \infty} \Pr(Y_n \leq y) = \Pr(Y \leq y),$$

o equivalentemente

$$\lim_{n \rightarrow \infty} E[h(Y_n)] = E[h(Y)],$$

para cualquier función  $h(\cdot)$  continua, acotada y que tome valores en los reales.

Estos tipos de convergencia se relacionan de las siguientes formas:

$$\begin{aligned} (Y_n \rightarrow Y \text{ en } \mathcal{L}^r) &\Rightarrow (Y_n \rightarrow Y \text{ en } \mathcal{L}^s) \quad r > s \geq 1 \\ (Y_n \rightarrow Y \text{ en } \mathcal{L}^r) &\Rightarrow (Y_n \rightarrow Y \text{ en probabilidad}) \quad r \geq 1 \\ (Y_n \rightarrow Y \text{ casi seguramente}) &\Rightarrow (Y_n \rightarrow Y \text{ en probabilidad}) \\ (Y_n \rightarrow Y \text{ en probabilidad}) &\Rightarrow (Y_n \rightarrow Y \text{ en distribución}) \end{aligned}$$

y de ninguna otra más. Entonces, todos los otros tipos de convergencia implican convergencia en distribución.

A.13.2 Leyes Asintóticas

Los siguientes resultados son ampliamente conocidos y pueden consultarse en muchos textos de probabilidad elemental.

**Leyes de grandes números.** Sean  $Y_1, Y_2, \dots$  variables aleatorias independientes e idénticamente distribuidas con  $E\{|Y^k|\} < \infty$  y  $S_n = Y_1 + Y_2 + \dots + Y_n$ . Entonces, la Ley Fuerte de los Grandes Números es

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = E\{Y_k\} \quad \text{casi seguramente.}$$

Si  $E\{Y_k^2\} < \infty$ , la convergencia es en  $\mathcal{L}^2$ . Si la convergencia es sólo en probabilidad, se tiene la Ley Débil de los Grandes Números.

**Teorema central del límite.** Sean  $Y_1, Y_2, \dots$  variables aleatorias i.i.d. con  $E\{Y_k\} = 0$  y  $0 < \sigma^2 = \text{var}\{Y_k\} < \infty$ ,  $S_n = Y_1 + Y_2 + \dots + Y_n$ , y

$$S_n^* = \frac{S_n}{\sigma\sqrt{n}}.$$

Entonces, para  $y \in \mathbb{R}$ , conforme  $n \rightarrow \infty$ ,  $S_n^*$  converge en distribución a una  $N(0, 1)$ :

$$\lim_{n \rightarrow \infty} \Pr\{S_n^* \leq y\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left[-\frac{u^2}{2}\right] du.$$

**Teorema de Slutsky.** Sea  $\{X_n, Y_n\}$ ,  $n \in \{1, 2, \dots\}$  una sucesión de pares de variables aleatorias. Entonces:

- a) Si  $X_n \rightarrow X$  en distribución y  $Y_n \rightarrow 0$  en probabilidad  $\Rightarrow X_n Y_n \rightarrow 0$  en probabilidad.
- b) Si  $X_n \rightarrow X$  en distribución y  $Y_n \rightarrow c$  en probabilidad  $\Rightarrow X_n + Y_n \rightarrow X + c$ ,  $X_n Y_n \rightarrow cX$  y  $X_n/Y_n \rightarrow X/c$  en distribución.

A.14 Función Generatriz de Momentos

La función generatriz de momentos de una variable aleatoria  $Y$ , es la función  $m_Y : \mathbb{R} \rightarrow \mathbb{R}$  tal que

$$m_Y(u) := E\{e^{uY}\}$$

Es la transformada de Laplace de la medida de probabilidad y existe solamente si  $Y \in \mathcal{L}^r$  para toda  $r$ . La  $i$ -ésima derivada de  $m_Y(u)$  con respecto a  $u$  es

$$m_Y^{(i)}(u) = E\{Y^i e^{uY}\}$$

entonces

$$m_Y^{(i)}(0) = E\{Y^i\}$$

es decir, el  $i$ -ésimo momento de  $Y$ .

Si  $Y$  y  $Z$  son variables aleatorias independientes

$$m_{Y+Z}(u) = m_Y(u) m_Z(u),$$

es decir, es posible expresar la función generatriz de  $X+Y$  en términos de las funciones generatrices de  $X$  y  $Y$ .

## A.15 Función Característica

Sea  $\mu$  una medida finita sobre  $\mathfrak{B}(\mathbb{R})$ . La *función característica* de  $\mu$  es la función  $\varphi_Y : \mathbb{R} \rightarrow \mathbb{C}$  dada por

$$\varphi_Y(u) = \int_{\mathbb{R}} e^{iux} d\mu(x), \quad u \in \mathbb{R}.$$

Entonces  $h$  es la transformada de Fourier de  $\mu$ . Si  $F$  es una función de distribución correspondiente a  $\mu$ , esto puede expresarse como

$$\varphi_Y(u) = \int_{\mathbb{R}} e^{iux} dF(x),$$

y se dice que  $h$  es la función característica de  $F$  (o de  $Y$ , si  $Y$  es la variable aleatoria con función de distribución  $F$ ).

Las funciones características son herramientas útiles en el estudio de sumas de variables aleatorias independientes, debido al siguiente resultado.

**Teorema A.15.1** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes, y sea  $S_n = X_1 + X_2 + \dots + X_n$ . La función característica de  $S_n$  es el producto de las funciones características de las  $X_i$ .

El resultado anterior permite calcular la función característica de  $S_n$ , conociendo sólo la distribución individual de las  $X_j$ 's. De hecho, una vez que la función característica es conocida, la función de distribución está totalmente determinada.

**Teorema A.15.2** Sean  $P_1$  y  $P_2$  dos medidas de probabilidad definidas sobre  $\mathfrak{B}(\mathbb{R})$ . Si

$$\int_{\mathbb{R}} e^{iux} dP_1(x) = \int_{\mathbb{R}} e^{iux} dP_2(x) \quad \text{para toda } u \in \mathbb{R},$$

entonces  $P_1 \equiv P_2$ .

Las funciones características poseen las siguientes propiedades elementales.

**Teorema A.15.3** Sea  $\varphi$  la función característica de una variable aleatoria finita con distribución  $F$ . Entonces

i)  $|\varphi(u)| \leq \varphi(0)$ .

ii)  $h$  es continua en  $\mathbb{R}$ .

iii)  $\varphi(-u) = \bar{\varphi}(u)$ , el conjugado complejo de  $\varphi$ .

iv) Si  $\int_{\mathbf{R}} |x|^r dF(x) < \infty$  para algún entero positivo  $r$ , entonces la  $r$ -ésima derivada de  $\varphi$  existe y es continua en  $\mathbf{R}$ . Se tiene además

$$\varphi^{(r)}(u) = \int_{\mathbf{R}} (ix)^r e^{iux} dF(x).$$



## Apéndice B

### Inferencia Paramétrica

#### B.1 La Familia Exponencial

Muchos de los modelos comúnmente usados en estadística están basados en distribuciones de probabilidad que tienen ciertas características en común. Estas pueden agruparse en familias. Una familia de especial interés es la *Familia Exponencial*, la cual contiene a la mayoría de las distribuciones usuales en estadística, y posee propiedades que facilitan ciertas manipulaciones de tipo matemático.

Se requieren dos conceptos previos.

El *dominio*  $\Theta$  del vector de parámetros, es aquel conjunto de valores para el cual la constante normalizadora (se denota con  $a_0(\theta)$ ) de la función de densidad, es estrictamente positiva.

El *soporte*  $\mathcal{Y}$  de la distribución es el conjunto de valores observables  $y$  para los cuales la función de densidad es positiva para todos los valores posibles del parámetro.

Cualquier función de densidad que pueda expresarse como una función no negativa de la forma

$$f_{\theta}(y) = a_0(\theta) t_0(y) e^{\theta^T t(y)}$$

se dice que tiene la *Representación Exponencial*. Aquí  $\theta \in \Theta$  es el vector de parámetros  $k$ -dimensional que indexa a la distribución y  $y \in \mathcal{Y}$  es un valor fijo, u observado de una variable aleatoria  $Y$ , donde  $t(y)$  y  $t_0(y)$  son funciones tales que pueden calcularse a partir de los datos, se dice que dichas funciones son *estadísticas*. Finalmente  $a_0(\theta)$  es una constante normalizadora.

Claramente los miembros de la Familia Exponencial no siempre están expresados en términos del parámetro  $\theta$ . Cuando esto sí ocurre, se le llama la parametrización natural o canónica de la familia indexada por el parámetro y la estadística,  $t(y)$  es llamada la estadística natural o canónica. La familia, en sí misma, es llamada una *Familia Lineal Exponencial* si  $t(y) = y$ .

Si  $0 \in \Theta$  y  $a_0(0) = 1$ , entonces

$$\frac{dF_{\theta}}{dF_0} = a_{\theta}(\theta) e^{\theta^T t(y)}.$$

A esto se le llama la representación estándar de la familia exponencial. El dominio del parámetro es

$$\Theta = \left\{ a_0^{-1}(\theta) = \int t_0(u) e^{\theta T(u)} < \infty \right\},$$

mientras el soporte de la familia es

$$\mathcal{T} = \left\{ T : \frac{dP_\theta}{dP_0} > 0, \quad \text{para todo } \theta \in \Theta \right\}.$$

El dominio del parámetro es un subconjunto de la cerradura del casco convexo del soporte de la estadística canónica correspondiente. El dominio del parámetro canónico es convexo, y a  $\mathcal{P}$  se le llama *llena*. Si  $\mathcal{P}$  es llena y  $\Theta$  es abierto, se dice entonces que  $\mathcal{P}$  es *regular*. Cualquier familia llena con soporte finito es regular porque  $\Theta = \mathbb{R}^k$ .

Se dice que una familia de distribuciones de probabilidad es *completa* si para toda estadística  $T(y) \in \mathcal{L}^1$  tal que  $\int T(y) dP(y) = 0$  para todo  $\theta \in \Theta$ , se tiene que  $T(y) \equiv 0$  para todo  $\theta \in \Theta$ . Para la Familia Exponencial, si  $\text{int}\Theta \subseteq \Theta$ , es decir, si  $\Theta$  no tiene puntos frontera, la familia es completa.

## B.2 El modelo estadístico paramétrico

Dada la formulación de un problema, se propone uno o más conjuntos de posibles modelos que puedan ayudar a explicar el mecanismo que origina las observaciones. Esto representa cierto conocimiento teórico previo acerca del fenómeno que se está estudiando. Este paso requiere la colaboración entre el estadístico y los investigadores que formularon el problema. Una vez hecho esto, se observa una muestra de uno o más valores  $y$ , de la variable aleatoria  $Y$ .

Este conjunto de modelos pueden representarse como un modelo estadístico paramétrico  $(\mathcal{Y}, \mathcal{P}, \Psi)$ , donde

$\mathcal{Y}$  es el *espacio muestral*.

$\mathcal{P}$  es la familia de modelos  $\{P_\psi : \psi \in \Psi\}$ .

$\Psi$  es el *espacio paramétrico* que indexa estos modelos.

En muchas situaciones se consideran simplificaciones adicionales. Por ejemplo, el conjunto de posibles modelos puede contener una única forma funcional.

Cuando las observaciones han sido obtenidas, se procede a formular proposiciones respecto a que miembros de  $\mathcal{P}$  resultan ser los más razonables para describir el fenómeno en cuestión. Los problemas de inferencia tratados en esta tesis están dentro del siguiente caso: Dada la forma funcional del modelo, determinar un conjunto de valores plausibles para el parámetro.

## B.3 Verosimilitud

La verosimilitud de un modelo es la probabilidad de los datos observados dado dicho modelo. Se trata de una función cuyo argumento es cierto vector de parámetros. Dicha función es tal que provee una medida comparativa entre varias distribuciones respecto a que tan bien pueden predecir las



observaciones que ya tuvieron lugar. Es una medida que apoya de manera relativa a unos modelos sobre otros, a pesar del hecho de que, en realidad las observaciones no hayan sido generadas por alguno de ellos.

La idea de cantidades observables implica una unidad finita de medición. La probabilidad de observar un punto es siempre cero; entonces las probabilidades deben estar en función de dicha unidad.

En general, la probabilidad del vector de valores observados  $\mathbf{y}$ , de la variable aleatoria  $Y$ , está dado por la distribución de probabilidad conjunta

$$\Pr_{\psi}(y_{1,1} < Y_1 \leq y_{2,1}, \dots, y_{1,n} < Y_n \leq y_{2,n}).$$

Aquí del valor observado  $y_i$  de la variable aleatoria, sólo puede decirse que cae en el intervalo  $y_{1,i} < y_i \leq y_{2,i}$ , de una unidad de longitud. Si las observaciones son independientes, se tiene que

$$\begin{aligned} & \Pr_{\psi}(y_{1,1} < Y_1 \leq y_{2,1}, \dots, y_{1,n} < Y_n \leq y_{2,n}) \\ &= \prod_{i=1}^n \Pr_{\psi}(y_{1,i} < Y_i \leq y_{2,i}) \\ &= \prod_{i=1}^n [F_{Y_i}(y_{2,i}) - F_{Y_i}(y_{1,i})]. \end{aligned}$$

La función de verosimilitud para este conjunto de observaciones, es la función del parámetro definida por

$$L_{\psi} \propto \Pr_{\psi}(y_{1,1} < Y_1 \leq y_{2,1}, \dots, y_{1,n} < Y_n \leq y_{2,n}),$$

es decir, está definida salvo por una constante de proporcionalidad.

Para el caso de variables aleatorias independientes y con función de densidad  $f$ , la verosimilitud puede escribirse como

$$L_{\psi} \propto \prod_{i=1}^n \int_{y_{1,i}}^{y_{2,i}} f_{\psi}(u_i) du_i.$$

Debido a que esta función puede ser muy compleja desde el punto de vista analítico, es muy frecuente usar la aproximación

$$L_{\psi} \propto \prod_{i=1}^n f_{\psi}(u_i) \Delta_i.$$

A esto se le conoce como la *aproximación por densidad* de la función de verosimilitud.

Una función de verosimilitud es incapaz de proveer afirmaciones absolutas sobre la validez de los modelos candidatos, pero permite llevar a cabo comparaciones relativas entre ellos. Se considerará el caso más simple, con sólo dos modelos posibles. Para el conjunto de datos observado  $\mathbf{y}$ , las comparaciones sólo pueden realizarse tomando en cuenta la probabilidad de  $\mathbf{y}$  dado cada modelo. En esta comparación, cualquier evento irrelevante  $E$ , independiente de  $\mathbf{y}$ , no debe ser tomado en cuenta.

Debido a que dicho evento es independiente, la probabilidad conjunta es  $\Pr(E) \Pr(\mathbf{y}; \psi)$ . Para que dicha información irrelevante desaparezca, debe tomarse el cociente de las probabilidades entre ambos modelos. Dentro de este contexto, cualquier comparación entre dos modelos  $P_{\psi_1}$  y  $P_{\psi_2}$ , con distintos valores fijos para el mismo vector de parámetros, deben basarse en un cociente de verosimilitudes

$$\text{LR}(\psi_1, \psi_2; \mathbf{y}) = \frac{L_{\psi_1}(\mathbf{y})}{L_{\psi_2}(\mathbf{y})}$$

con  $\psi_1, \psi_2 \in \Psi$ .

El modelo más factible dentro de la familia paramétrica indexada por  $\psi$ , para los datos observados está dada por

$$\hat{\psi} = \arg \left[ \sup_{\psi \in \Psi} L_{\psi}(\mathbf{y}) \right]$$

esto es, aquel valor de  $\psi$  que hace a las observaciones más probables, para la familia de modelos dada. El vector de parámetros correspondiente, el cual se denota con  $\hat{\psi}$ , se le llama el *estimado máximo verosímil*.

Es muy común trabajar con el logaritmo de la función de verosimilitud. A esta función se le conoce como la log-verosimilitud, y se denota con  $l_{\psi}$ .

El vector gradiente de primeras derivadas parciales de la función log-verosimilitud con respecto a vector de parámetros, con componentes

$$U_{\psi_i}(\psi; \mathbf{y}) = \frac{\partial l_{\psi}(\mathbf{y})}{\partial \psi_i}$$

es llamado la función de *score* o *vector de score*.

El score es importante para el cálculo de estimadores máximo verosímiles. Entonces, las ecuaciones de score, también llamadas ecuaciones de verosimilitud, se definen como

$$U_{\hat{\psi}}(\mathbf{y}) = 0.$$

Su solución conduce (bajo ciertas condiciones de regularidad) a estimadores máximo verosímiles.

La función de verosimilitud es usada para comparar modelos. La curvatura del logaritmo de dicha función en un punto permite visualizar que tanto varía la plausibilidad de ese valor del parámetro respecto a valores muy cercanos. El negativo de la matriz Hessiana de la log-verosimilitud tiene elementos

$$J_{\psi_i, \psi_j} = - \frac{\partial^2 l_{\psi}(\mathbf{y})}{\partial \psi_i \partial \psi_j}.$$

A esta matriz se le llama *la información observada de Fisher*, y se denota con  $J_{\psi}(\mathbf{y})$ .

Al valor esperado de  $J$ ,

$$F(\psi) = E \{ J_{\psi}(\mathbf{y}) \}$$

se le llama *la información esperada de Fisher*.

## B.4 Suficiencia

Muchas veces resulta conveniente resumir la información contenida en la muestra  $\mathbf{Y}$  mediante una función de dicha muestra, una estadística  $T(\mathbf{Y})$ . Por simplicidad, se supondrá que el tamaño de muestra no es aleatorio, sino que, por el plan de muestreo, está fijo en  $n$ . Por razones prácticas, se busca que  $T(\mathbf{Y})$  sea de menor dimensión que  $\mathbf{Y}$ ; esto provee de cierto tipo de reducción de los datos.

Entonces, cualquier estadística  $T$  tal que la distribución condicional de  $\mathbf{Y}$  dado  $t$  es la misma para todo  $\{\mathcal{P}_\psi : \psi \in \Theta\}$ , sin importar que valor  $t$  haya sido observado, se dice que es *suficiente* para  $\psi$ , dada la familia paramétrica de modelos  $\{\mathcal{P}_\psi : \psi \in \Theta\}$ . Se dice que una estadística  $T$  es *suficiente minimal*, si es suficiente y cualquier estadística de menor dimensión resulta no ser suficiente.

Las estadísticas suficientes poseen ciertas características importantes.

- i) Si  $T$  es suficiente, también lo será toda transformación biyectiva de  $T$ .
- ii) Si  $T$  es suficiente minimal, se trata necesariamente de una función de cualquier otra estadística suficiente.
- iii) Si la estadística  $T$  es suficiente, entonces  $\Pr\{\mathbf{Y} \in A \mid T = t\}$ , con  $A \in \mathfrak{F}$ , no depende de  $\psi$ , con probabilidad 1.

El siguiente resultado permite encontrar estadísticas suficientes a partir de la función de verosimilitud.

**Lema de Factorización de Neyman-Fisher.** *Una condición necesaria y suficiente para que  $T$  sea una estadística suficiente para  $\psi$  en la familia  $\{\mathcal{P}_\psi : \psi \in \Theta\}$  es que la función de verosimilitud pueda expresarse como el producto*

$$L_\psi(\mathbf{y}) = G_\psi(T(\mathbf{y}))H(\mathbf{y})$$

para todo  $\psi \in \Theta$ , donde el primer factor depende de  $\mathbf{y}$  sólo a través de  $T$  y el segundo factor no depende de  $\psi$ .

## B.5 Pruebas de hipótesis

### B.5.1 El caso general

Supóngase que se tiene una familia parametrizada de modelos  $\{\mathcal{P}_\psi : \psi \in \Psi\}$  y que se ha formulado una hipótesis  $H$  acerca del modelo o del conjunto de modelos que han generado las observaciones. Tales hipótesis deben ser especificadas antes de observar los datos; debe ser claro que hipótesis formuladas a partir de los datos rara vez serán rechazadas.

Supóngase ahora que  $\{\mathcal{P}_\psi : \psi \in \Psi\}$  es el conjunto de todos los modelos posibles que se este considerando. Una *hipótesis nula*  $\Psi_0 \subset \Psi$  es un subconjunto de modelos de especial interés y el cual está bajo prueba; esto se denota con  $H_0$ . Como su nombre sugiere, usualmente es especificada haciendo algunos parámetros iguales a cero. La hipótesis alternativa, denotada con  $H_A$ , la cual es de hecho  $\Psi_A = \Psi \setminus \Psi_0$ , esto es, el conjunto complementario de modelos.

Una hipótesis  $i$  es llamada *simple* si especifica por completo al modelo, es decir que  $\Psi_i$  contiene un único miembro y que la probabilidad de cualquier evento posible puede ser calculado. De otra forma

se dice que la hipótesis es *compuesta*. Usualmente las hipótesis compuestas especifican el modelo completamente excepto por los valores de algunos de los parámetros desconocidos.

Una hipótesis  $\Psi_1$  es *cierta* si el modelo  $\eta$  correcto para describir el mecanismo que genera las observaciones está contenido en ella ( $\eta \in \Psi_1$ ).

El objetivo de una *prueba de hipótesis* consiste en definir si un conjunto de observaciones dado basta para rechazar la hipótesis nula; debido a que, la probabilidad de que dichas observaciones hayan tenido lugar dado cualquier elemento de dicha hipótesis nula sea muy baja. De esta forma, una prueba de hipótesis se encuentra definida por un conjunto de *regiones críticas*  $\mathcal{Y}_\alpha$ , en el espacio muestral tales que

- i)  $\Pr\{\mathbf{Y} \in \mathcal{Y}_\alpha\} = \alpha$  para todo  $\alpha$
- ii)  $\mathcal{Y}_{\alpha_1} \subset \mathcal{Y}_{\alpha_2}$  si  $\alpha_1 < \alpha_2$ .

Entonces, para alguna  $\alpha$ , con  $0 < \alpha < 1$ ,  $\mathcal{Y}_\alpha$  es el conjunto de puntos en el espacio muestral tales que, de ser observados, se tomarían como evidencia para contradecir  $H_0$  al *nivel de significancia*  $\alpha$ . Esto se hace mediante la construcción de una *estadística de prueba*  $T$ , de la siguiente forma

$$\{\mathbf{Y} \in \mathcal{Y}_\alpha\} \text{ si y sólo si } \{T(\mathbf{Y}) \in G_\alpha\},$$

donde  $G_\alpha$  es cierto subconjunto de  $\mathbb{R}^k$ . A  $\mathcal{Y}_\alpha$  se le llama una *región crítica* de tamaño  $\alpha$ . Comúnmente habrá muchas de estas regiones, ya que la región crítica dependerá de la estadística de prueba que se esté considerando.

Una prueba de hipótesis  $\Upsilon$  al nivel de significancia  $\alpha$  está compuesta por una pareja  $(\mathcal{Y}_\alpha, T)$ .

Si se rechaza  $H_0$  cuando es cierta, se dice que se ha cometido un *error tipo I*, si por otra parte, no se rechaza  $H_0$  cuando esta es falsa, se dice que se ha cometido un *error tipo II*. El *tamaño del error tipo I (II)* se define como la probabilidad de cometer un error de tipo I (II).

Sea  $\Upsilon$  una prueba para la hipótesis nula  $H_0$ . La *función potencia* de la prueba  $\Upsilon$ , denotada por  $\pi_\Upsilon$ , se define como

$$\begin{aligned} \pi_\Upsilon(\psi) &= \Pr_\psi\{\mathbf{Y} \in \mathcal{Y}_\alpha\} \\ &= \Pr_\psi\{T(\mathbf{Y}) \in G_\alpha\}, \end{aligned}$$

es decir,  $\pi_\Upsilon(\psi)$  es la probabilidad de que  $H_0$  sea rechazada cuando el verdadero modelo es  $\mathcal{P}_\psi$ .

Sea  $\Upsilon$  una prueba para la hipótesis  $H_0 : \psi \in \Psi_0$  donde  $\Psi_0 \subset \Psi$ . El *tamaño de la prueba*  $\Upsilon$  para  $H_0$  viene dado por  $\sup_{\psi \in \Psi_0} [\pi_\Upsilon(\psi)]$ .

Se dice que una prueba  $\Upsilon^*$  para  $H_0 : \psi = \psi_0$  contra  $H_1 : \psi = \psi_1$  es una *prueba uniformemente más potente* (PUMP) de tamaño  $\alpha$  ( $0 < \alpha < 1$ ) si cumple

- i)  $\pi_{\Upsilon^*}(\psi_0) = \alpha$
- ii)  $\pi_{\Upsilon^*}(\psi_1) \geq \pi_\Upsilon(\psi_1)$  para cualquier otra prueba tal que  $\pi_\Upsilon(\psi_0) \leq \alpha$ .

El siguiente resultado muestra como encontrar la PUMP cuando se tiene que la hipótesis nula y alternativa son simples.

**Lema de Neyman-Pearson.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución  $F_\psi$ , donde  $\psi$  pertenece a  $\{\psi_0, \psi_1\}$ , y  $\alpha \in (0, 1)$ . Aquí  $L_\psi, i \in \{0, 1\}$  denota a la verosimilitud para cada valor del parámetro. Sea  $k^*$  una constante positiva y  $\mathcal{Y}_\alpha$  un subconjunto del espacio muestral tal que

$$\Pr_{\psi_0}\{(X_1, X_2, \dots, X_n) \in \mathcal{Y}_\alpha\} = \alpha,$$

y

$$\lambda = \text{LR}(\psi_0, \psi_1; \mathbf{x}) = \frac{\mathbf{L}_{\psi_0}}{\mathbf{L}_{\psi_1}} \leq k^*$$

si y sólo si  $(x_1, x_2, \dots, x_n) \in \mathcal{Y}_\alpha$ .

Sea  $H_0 : \psi = \psi_0$  y  $H_1 : \psi = \psi_1$ . Entonces la prueba  $\Upsilon^*$  que corresponde a rechazar  $H_0$  si y sólo si  $(x_1, x_2, \dots, x_n) \in \mathcal{Y}_\alpha$  es la PUMP de tamaño  $\alpha$  para  $H_0 : \psi = \psi_0$  y  $H_1 : \psi = \psi_1$ .

### B.5.2 Teoría asintótica

Supóngase que la distribución de las variables aleatorias observables  $\{Y_1, Y_2, \dots, Y_n\}$  depende de cierto vector de parámetros desconocidos. En el caso más simple, esto se reduce a un parámetro  $\psi$  con valores en  $\mathbf{R}$  y el problema de probar la hipótesis  $H : \psi = \psi_0$  contra las alternativas  $K : \psi > \psi_0$ .

Supóngase que  $H$  es rechazada cuando cierta estadística de prueba  $T_n = T(Y_1, Y_2, \dots, Y_n)$  toma un valor muy grande. Esto se expresa como  $T_n \geq G_n$  donde  $G_n$  está determinada por

$$\Pr_{\psi_0} \{T_n \geq G_n\} = \alpha. \tag{B.1}$$

Aquí  $\alpha$  es el nivel de significancia predeterminado que controla la probabilidad de rechazar a la hipótesis cuando esta sea verdadera.

Se considera ahora el problema y obtener un valor aproximado para  $G_n$  cuando  $n$  es grande. Con este propósito se sustituye (B.1) por la condición más débil que representa un nivel asintótico  $\alpha$ , el cual satisface

$$\Pr_{\psi_0} \{T_n \geq G_n\} \rightarrow \alpha, \quad \text{cuando } n \rightarrow \infty. \tag{B.2}$$

En muchas ocasiones, se tiene

$$\sqrt{n} (T_n - \psi_0) \rightarrow \mathbf{N}(0, \sigma^2(\psi_0)) \quad \text{en distribución,}$$

de modo que

$$\frac{\sqrt{n}}{\sigma(\psi_0)} (T_n - \psi_0) \rightarrow \mathbf{N}(0, 1) \quad \text{en distribución.}$$

El lado derecho de la expresión (B.2) tenderá a  $\alpha$  si se cumple

$$\frac{\sqrt{n}}{\sigma(\psi_0)} (G_n - \psi_0) \rightarrow z_\alpha,$$

donde

$$1 - \Phi(z_\alpha) = \alpha.$$

Aquí  $\Phi$  es la función de distribución para una  $\mathbf{N}(0, 1)$ . La prueba que se deriva de la condición  $T_n \geq G_n$  tendrá entonces nivel asintótico  $\alpha$ , si

$$G_n = \psi_0 + \frac{\sigma(\psi_0)}{\sqrt{n}} z_\alpha + o\left(\frac{1}{\sqrt{n}}\right),$$

y en particular cuando

$$G_n = \psi_0 + \frac{\sigma(\psi_0) z_\alpha}{\sqrt{n}}$$

Para este valor crítico, la prueba puede frasearse de la siguiente forma

$$\frac{\sqrt{n}(T_n - \psi_0)}{\sigma(\psi_0)} \geq z_\alpha.$$

## B.6 Teoría de estimación insesgada

Un *estimador puntual* es un tipo especial de estadística  $T$ , la cual se espera tome valores cercanos al verdadero valor del parámetro desconocido  $\psi$ . Aquí es conveniente distinguir entre el estimador como función aleatoria  $T$  y su valor observado, la estimación puntual  $T(\mathbf{y})$ . A continuación se mencionan algunos atributos usados para calificar un estimador.

**Sesgo.** Supóngase que  $T$  es una estadística que va a ser usada como un estimador puntual para cierto parámetro  $\psi$ . Entonces el sesgo de  $T$  se define como  $E[T - \psi]$ . Si el sesgo es cero, se dice que el estimador es *insesgado*.

**Error cuadrático medio.** El error cuadrático medio de una estadística  $T$ , usada como un estimador de  $\psi$  es  $E[(T - \psi)^2]$ . Es deseable que dicho error sea pequeño. Un criterio para valuar estimadores puntuales consiste en preferir a aquellos cuyo error cuadrático medio sea menor. El restringirse a la clase de estimadores insesgados simplifica enormemente la tarea de minimizar esta esperanza.

**Consistencia.** Un estimador  $T$  es débilmente consistente si converge en probabilidad al verdadero valor del parámetro  $\psi$ . Es decir se toma una sucesión de estadísticas  $\{T_n : n \in \mathbf{N}\}$  con la misma forma funcional, donde el  $n$ -ésimo elemento es calculado a partir de una muestra de tamaño  $n$ . Se dice que un estimador es *fuertemente consistente* si la convergencia es casi segura.

**Normalidad asintótica.** Un estimador consistente  $T$  de  $\psi$  es asintóticamente normal si

$$\sqrt{n}(T_n - \psi) \rightarrow \mathbf{N}(0, \sigma^2(\psi)) \quad \text{en distribución.}$$

Los siguientes resultados tienen por objeto encontrar al estimador insesgado con menor error cuadrático medio.

**Cota inferior de Cramér-Rao.** Un estimador  $T$ , tiene varianza mínima si alcanza la Cota Inferior de Cramér-Rao:

$$\begin{aligned} \text{var}[T] &\geq \frac{\left[1 + \frac{\partial E[T - \psi]}{\partial \psi}\right]^2}{F(\psi)} \\ &= \frac{\left[\frac{\partial E[T]}{\partial \psi}\right]^2}{F(\psi)}. \end{aligned}$$

La igualdad se alcanza si y sólo si el estimador es una función lineal del score.

**Teorema de Rao-Blackwell.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución  $F_\psi$ , y sea  $S$  una estadística suficiente y completa para  $\psi$ . Sea  $T$  un estimador insesgado de  $\tau(\psi)$ . Se define  $T'$  como  $T' = E\{T \mid S\}$ . Entonces

i)  $T'$  es una estadística y es una función de  $S$ .

ii)  $E_\psi [T'] = \tau(\psi)$ .

iii)  $\text{var}_\psi [T'] \leq \text{var}_\psi [T]$  para toda  $\psi$ , y  $\text{var}_\psi [T'] < \text{var}_\psi [T]$  para algún  $\psi$ , a menos que

$$\Pr \{T = T'\} = 1.$$

**Teorema de Lehmann-Scheffé.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución  $F_\psi$ , y sea  $S$  una estadística suficiente y completa para  $\psi$ . Si  $T^*$  es una función de  $S$ , y además es un estimador insesgado de  $\tau(\psi)$ , entonces  $T^*$  es el estimador de  $\tau(\psi)$  con mínima varianza, esto para toda  $\psi \in \Theta$ .

### B.7 Resultado de normalidad asintótica para estimadores

El siguiente teorema garantiza que, bajo ciertas condiciones de regularidad, el estimador máximo verosímil es asintóticamente normal.

**Teorema de normalidad asintótica.** Sea  $X$  una variable aleatoria con densidad  $f_\psi$  con respecto a una medida  $\sigma$ -finita  $\mu$ , donde  $\psi$  pertenece a un intervalo no degenerado de la recta real. En el caso de que  $X$  sea discreta,  $f_\psi$  denota a su función de masa. Se hacen los siguientes supuestos.

**Supuesto 1.**-Las derivadas  $\partial \ln f_\psi / \partial \psi$ ,  $\partial^2 \ln f_\psi / \partial \psi^2$  y  $\partial^3 \ln f_\psi / \partial \psi^3$  existen, para casi toda  $x$  contenida en un intervalo  $A$  tal que el verdadero valor del parámetro  $\psi_0 \in A$ .

**Supuesto 2.**-En el punto  $\psi_0$

$$E_{\psi_0} \left[ \frac{\partial f_\psi / \partial \psi}{f_\psi} \right] = 0, \quad E_{\psi_0} \left[ \frac{\partial^2 f_\psi / \partial \psi^2}{f_\psi} \right] = 0$$

**Supuesto 3.**-Para toda  $\eta \in A$ ,

$$\left| \frac{\partial^3 \ln f_\psi}{\partial \psi^3} \right| < M(x), \quad E_\eta [M(x)] < K$$

donde  $K$  no depende de  $\psi$ .

Sean  $x_1, x_2, \dots, x_n$   $n$  observaciones independientes y sean

$$l(\psi; x) = \sum \ln f_\psi(x_i)$$

$$\frac{\partial l}{\partial \psi} = \sum \frac{\partial f_\psi(x_i) / \partial \psi}{f_\psi(x_i)}$$

Entonces:

Sea  $\{\hat{\psi}_n : n \in N\}$  una sucesión de raíces para la ecuación de verosimilitud. tal que es consistente para  $\psi$ , y los supuestos 1-3 verdaderos, se sigue que

$$\sqrt{n} \left[ (\hat{\psi}_n - \psi_0) F(\psi_0) - \frac{1}{n} \frac{\partial l}{\partial \psi} \Big|_{\psi=\psi_0} \right] \rightarrow 0 \quad \text{con probabilidad 1} \quad (\text{B.3})$$

y la distribución asintótica de  $\sqrt{n} (\hat{\psi}_n - \psi_0)$  es  $N(0, 1/F(\psi_0))$ .

## B.8 Intervalos de confianza

Un intervalo de confianza al  $1 - \alpha$  ( $\alpha \in (0, 1)$ ) para el parámetro  $\psi$  es una pareja de estadísticas  $(T_1, T_2)$  tal que

$$P_{T,\psi} (T_1 \leq \psi \leq T_2) = 1 - \alpha \quad \text{para todo } \psi \in \Psi.$$

Estas estadísticas no están determinadas de manera única para un  $\alpha$  fijo. Es común elegir el intervalo más angosto, el cual, generalmente está basado en la función de verosimilitud.

Una *cantidad pivotal* (Fisher, 1941) es una función de una estadística suficiente y de un parámetro  $\psi$ , de modo que su distribución no depende del parámetro. En general sólo existe para variables aleatorias continuas (a menos que el parámetro sea discreto) y debe ser una función monótona de  $\psi$  para casi toda observación  $y$ . Las cantidades pivotaes facilitan la construcción de intervalos de confianza.

Un *intervalo asintótico para  $\psi$  al nivel  $1 - \alpha$*  se define como las sucesiones de estadísticas

$$\{(T_{1,n}, T_{2,n} : n \in N)\}$$

tales que

$$P_{T,\psi} \{T_{1,n} \leq \psi \leq T_{2,n}\} \rightarrow 1 - \alpha \quad \text{cuando } n \rightarrow \infty,$$

para todo  $\psi$ .



# Apéndice C

## Sobre la Teoría Bayesiana

### C.1 Generalidades

La Teoría Bayesiana ofrece una metodología para abordar problemas de inferencia. Dicha teoría se basa en dos principios:

1.- Se interpreta a la probabilidad como una característica subjetiva. En el sentido de que la probabilidad de un evento está determinada por las creencias del investigador respecto a la propensión que tiene este de ocurrir.

2.- Los problemas de inferencia son tratados con las herramientas que proporciona la Teoría de las Decisiones.

El adoptar la idea de probabilidad subjetiva permite al investigador hacer proposiciones de tipo probabilístico acerca de cualquier parámetro desconocido  $\theta$ , es decir, la afirmación  $\text{Pr} \{ \theta \in A \} = \alpha$  ( $A$  pertenece a cierta  $\sigma$ -álgebra definida sobre el espacio parametral  $\Theta$ ,  $\alpha \in [0, 1]$ ) siempre tiene sentido en este esquema. De esta forma se puede proponer una densidad  $p(\theta)$  que sintetice la información disponible respecto a  $\theta$ . Debido a que la asignación de probabilidades depende del conocimiento del fenómeno, la llegada de nueva información que lleve a modificar las creencias sobre  $\theta$ , implicaría cambios en las probabilidades antes mencionadas. Dicha actualización se lleva a cabo mediante el Teorema de Bayes [ver Ash]

$$p(\theta | x) = \frac{p(\theta) p(x | \theta)}{\int_{\Theta} p(\theta) p(x | \theta) d\theta}$$

donde  $\Theta$  representa al conjunto de todos los valores de  $\theta$  que el investigador considera factibles. A  $p(\theta)$  se le conoce como densidad *a priori*, mientras que a  $p(\theta | x)$  se le llama densidad *a posteriori*. Estos términos son relativos al conjunto de datos  $x$  que se esté considerando.  $p(x | \theta)$  es la *función de verosimilitud* asociada al experimento.

Una vez obtenida la densidad *a posteriori*  $p(\theta | x)$ , el último paso del Método Bayesiano consiste en derivar de ella conclusiones razonables respecto a  $\theta$ . Es decir, cualquier afirmación que se haga de  $\theta$  debe basarse en características de  $p(\theta | x)$ .

Para que dicho proceso se realice de manera óptima, la Teoría Bayesiana trata a todo problema de inferencia respecto a  $\theta$  como un problema de decisión en presencia de incertidumbre [ver Bernardo y Smith]. Un problema de decisión en presencia de incertidumbre es una estructura de la forma  $(\mathcal{D}, E, u)$  donde  $\mathcal{D}$  es el conjunto de opciones, el cual debe ser exhaustivo, es decir, que agote todas las posibilidades que puedan, en principio, parecer razonables. A cada  $d \in \mathcal{D}$  le corresponde un escenario  $\{E_i, c_i \mid E_i, i \in \Lambda\}_d$  ( $\Lambda$  un conjunto de índices) donde las  $\{E_i, i \in \Lambda\}$  conforman un álgebra de conjuntos, los cuales se identifican con los posibles estados de la naturaleza,  $\{c_i \mid E_i\}_d$  corresponde a la consecuencia de haber tomado la opción  $d$ , dado que  $E_i$  es el verdadero estado. Sobre  $\{E_i, i \in \Lambda\}$  se define una medida de probabilidad, que de hecho será  $p(\theta)$ . Para poder comparar consecuencias y en base a ello formular decisiones, se define lo que se denomina una función de utilidad.  $u$  es una función de utilidad para el decisor particular. Si  $u$  es la utilidad,  $l = -u$  es una posible función de pérdida asociada al problema de decisión.

Las relaciones entre la probabilidad subjetiva y la Teoría de las decisiones quedan establecidas en los llamados *Axiomas de coherencia* [ver Bernardo]. Esto origina una teoría matemáticamente consistente.

## C.2 Estimación

Un criterio muy empleado en estadística para encontrar estimadores óptimos consiste en considerar la clase de estimadores insesgados para el parámetro, y luego buscar al elemento de dicha clase que posea mínima varianza. Esto se hace con el fin de minimizar el error cuadrático medio. El error cuadrático medio de un estimador es la esperanza de la diferencia al cuadrado del estimador respecto al verdadero valor del parámetro. De acuerdo con la Teoría de Estimación Insesgada, el promedio es tomado respecto a la distribución del estimador.

Para la inferencia Bayesiana,  $\theta$  es una variable aleatoria, por lo tanto tiene sentido promediar el error cuadrático con respecto a la densidad *a posteriori* de  $\theta$ . Se tiene que

$$\begin{aligned} E[(\tau - \theta)^2 \mid x] &= E[\tau^2 \mid x] - E[2\tau\theta \mid x] + E[\theta^2 \mid x] \\ &= \tau^2 - 2\tau E[\theta \mid x] + E[\theta^2 \mid x] \\ &= \{\tau - E[\theta \mid x]\}^2 + \text{var}[\theta \mid x]. \end{aligned}$$

Se sigue que el estimador  $\tau$  que minimiza el error cuadrático medio (respecto a  $x$ ) es  $\tau = E[\theta \mid x]$ .

No existe restricción alguna que imponga el uso del error cuadrático medio como función de distancia entre  $\theta$  y  $\tau$ . Puede usarse el error absoluto  $|\tau - \theta|$  o alguna otra función. Para este enfoque la elección de una estimación  $\tau$  es vista como una *decisión* acerca del error en el que se ha de incurrir. Dicha elección depende de la forma que se elija para penalizar las distintas magnitudes del error. El error cuadrático medio indica de forma implícita que los errores grandes son más graves de lo que serían si se tomara el error absoluto. Dada cualquier función  $l(\tau - \theta)$  de pérdida, es posible encontrar un estimador que minimice la *pérdida esperada posterior*  $E[l(\tau - \theta) \mid x]$ . Este será el estimador bayesiano.

### 3.3 Análisis conjugado

En general la integral  $\int_{\Theta} p(\theta) p(x|\theta) d\theta$  no es fácil de calcular, y por consiguiente  $p(\theta|x)$  tampoco o es. Por ejemplo, si  $X \sim N(\theta, \sigma^2)$  y  $\theta$  tiene distribución Cauchy con vector de parámetros  $(\alpha, \beta)$ , entonces  $p(\theta|x)$  sólo puede evaluarse numéricamente. Existe la preocupación de encontrar distribuciones *a priori* para las cuales la distribución *a posteriori* sea fácil de obtener. Estas son las llamadas *distribuciones conjugadas*.

**Familia Conjugada.** Sea  $\mathcal{F}$  una clase de funciones de densidad  $p(x|\theta)$  (indexada por  $\theta$ ). Una clase  $\mathcal{P}$  de distribuciones *a priori* se dice que es una familia conjugada para  $\mathcal{F}$ , si  $p(\theta|x)$  está en la clase de  $\mathcal{P}$  para toda observación  $x$  y  $p \in \mathcal{P}$ .

Para una clase de densidades  $\mathcal{F}$ , una familia de conjugadas puede encontrarse frecuentemente examinando las funciones de verosimilitud  $L_{\theta}(x) = p(x|\theta)$ , y eligiendo como familia conjugada la clase de distribuciones que tengan la misma forma funcional que dichas funciones de verosimilitud. A una densidad *a priori* obtenida de esta forma se le conoce como *a priori conjugada natural*.

El siguiente resultado permite obtener de manera sencilla, densidades *a priori* conjugadas para una gran variedad de densidades.

**Análisis conjugado para la Familia Exponencial.** Si  $X = (X_1, X_2, \dots, X_n)$  es una muestra aleatoria de una distribución que pertenece a la Familia Exponencial, de modo que

$$p(x|\theta) = \prod_{j=1}^n f(x_j) [g(\theta)] \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) \sum_{j=1}^n h_i(x_j) \right\}$$

entonces la familia conjugada para  $\theta \in \Theta \subset \mathbb{R}^k$  es de la forma

$$p(\theta|\tau) = [K(\tau)]^{-1} [g(\theta)]^{\tau_0} \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) \tau_i \right\}, \quad \theta \in \Theta$$

donde  $\tau$  es tal que  $K(\tau) = \int_{\Theta} [g(\theta)]^{\tau_0} \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) \tau_i \right\} d\theta < \infty$ .

La demostración de este resultado puede revisarse en [Bernardo y Smith].

### C.4 Distribuciones a priori no informativas

Algunos estadísticos se muestran escépticos respecto al uso de información *a priori*. Un argumento muy usado por ellos en su contra es que debido a su naturaleza subjetiva, puede alterar o minimizar evidencia importante contenida en los datos. Por lo tanto afirman que la información *a priori* no debe usarse en un contexto científico. El concepto de distribuciones *a priori* no informativas surge como un intento de reconciliar estos argumentos con el punto de vista Bayesiano.

En un principio, las distribuciones uniformes fueron propuestas para representar situaciones donde muy poca o ninguna información se encuentra disponible, o bien, aunque exista, se decida no emplearla. Entonces, una distribución  $p(\theta) \propto c$  con  $\theta \in A \subset \mathbb{R}$  indica que, con la información disponible,

ninguno de los elementos de  $A$  es preferible a los otros. El adoptar esta idea conduce a ciertas dificultades. La primera es que  $p(\theta)$  no es propiamente una densidad si  $A$  es no acotado, ya que  $\int_A p(\theta) d\theta = \infty$ , lo cual contradice las reglas básicas de la probabilidad. Además, si  $\phi = \phi(\theta)$  es una transformación uno a uno de  $\theta$  y si  $\theta$  se distribuye uniformemente, entonces por el Teorema de Cambio de Variable [ver Bartle], la densidad de  $\phi$  es

$$p(\phi) = p(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|,$$

la cual es constante solamente en el caso que  $\phi$  este definida por una transformación lineal. Sin embargo, los mismos supuestos que llevan a definir  $p(\theta) \propto c$ , conducen a afirmar que  $p(\phi) \propto c$ , lo cual contradice la última deducción.

Para realizar inferencias, se trabaja únicamente con la distribución *a posteriori*, por lo que se debe exigir que esta sea una densidad, aunque la *a priori* resulte no serlo. Bajo este criterio, el que  $\int_A p(\theta) d\theta = \infty$  pierde importancia frente al hecho de que efectivamente pueda analizarse  $p(\theta | x)$  donde  $x$  denota al conjunto de observaciones.

La clase de distribuciones *a priori* no informativas propuesta por Jeffreys es invariante (respecto a transformaciones uno a uno), pero en muchos casos conduce a densidades impropias.

**Densidad a priori de Jeffreys.** Se considera una observación  $X$  con función de densidad  $p(x | \theta)$ . La distribución *a priori* no informativa de Jeffreys tiene densidad dada por

$$p(\theta) \propto [I(\theta)]^{\frac{1}{2}}, \quad \theta \in \Theta$$

En el caso multivariado, la densidad esta dada por

$$p(\theta) \propto [\det I(\theta)]^{\frac{1}{2}}.$$

Este tipo de distribuciones se discuten con detalle en [Migón y Gamerman].

## C.5 Criterio de información

Se tiene el siguiente escenario: De acuerdo al conocimiento del investigador, los modelos razonables para describir el fenómeno pertenecen a cierta familia  $\mathcal{F}$ , que involucra varias formas funcionales de modo que la dimensión  $k$  de la característica de interés  $\theta$  puede variar entre miembros de esta familia. Entonces se elige  $f \in \mathcal{F}$  tal que maximiza

$$AIC = \text{Criterio de información de Akaike} = \ln p(\mathbf{y} | \hat{\theta}) - k.$$

Aquí  $\hat{\theta}$  denota al estimador máximo verosímil, y  $k$  es una penalización sobre la dimensión del parámetro.

Este criterio es discutido en [Leonard].

## 3.6 Estimación por intervalo

El resumir la distribución *a posteriori* mediante su media, moda u otra medida similar no siempre es suficiente para hacer inferencias sobre el parámetro de interés  $\theta$ . Una característica de la distribución *a posteriori* que resulta ser de mucha utilidad es la región donde se acumula la mayor parte de la probabilidad final, es decir donde se encuentran los valores más creíbles de  $\theta$ . Para hacer esto, se definen los conceptos de *intervalos creíbles* y de *regiones de mayor densidad posterior* (*highest posterior density regions*).

Los intervalos creíbles son obtenidos directamente de la distribución final. Con un intervalo creíble se hace referencia directa a la probabilidad posterior de que  $\theta$  se encuentre en cierto intervalo, condicionado a la información *a priori* y a los datos.

**Región Creíble.** Sea  $p(\theta | x)$  la densidad *a posteriori* y  $\Theta$  el espacio parametral sobre el cual se esté trabajando. Se dice que  $A$  es un intervalo creíble (puede hablarse también de región creíble) al  $1 - \gamma$  si

$$\int_A p(\theta | x) d\theta = 1 - \gamma.$$

Sin embargo, para una probabilidad de  $1 - \gamma$ , con  $\gamma \in [0, 1]$ , pueden existir muchos intervalos creíbles para una misma distribución posterior. Por lo cual es necesario considerar un subconjunto de  $\Theta$  'óptimo' en algún sentido.

Se busca un conjunto tal que:

- i) Para un nivel de probabilidad  $1 - \gamma$  dado, tenga menor longitud posible, y
- ii) Contenga valores del parámetro que sean más plausibles que cualquier valor que se encuentre afuera del mismo.

¿Es posible calcular regiones que posean ambas propiedades? Afortunadamente la respuesta es sí, y se puede demostrar [ver O'Hagan], que ambos atributos son equivalentes.

Los bayesianos llaman a dichos conjuntos *regiones de mayor densidad posterior*. Se caracterizan por estar delimitadas por curvas de nivel de la distribución final.

**Región de mayor densidad posterior.** Sea  $p(\theta | x)$  la densidad *a posteriori* y  $\Theta$  el espacio parametral. Se dice que  $A$  es una región de mayor densidad posterior al  $1 - \gamma$  si cumple:

- i)  $\int_A p(\theta | x) d\theta = 1 - \gamma$ .
- ii)  $p(\psi_1 | x) \geq p(\psi_2 | x)$  si  $\psi_1 \in A$  y  $\psi_2 \in \Theta \setminus A$ .



## Apéndice D

# Guía de instrucciones de S-PLUS

El paquete S-PLUS permite elaborar programas mediante un lenguaje estructurado, el cual es muy similar a Borland C. En dicho lenguaje las variables son vectores o matrices de caracteres o de números reales. S-PLUS cuenta además con un modo gráfico.

A continuación se presenta una lista de las instrucciones empleadas en los programas desarrollados en esta tesis. Después de cada definición se presenta la *sintaxis correspondiente*.

En este lenguaje, la  $i$ -ésima entrada de un vector  $x$  se denota con  $x[i]$ .

**abs.-** Tiene por argumento un vector de números reales, su salida es un vector cuyas componentes son los valores absolutos de las componentes del vector original.

*sintaxis*

`abs(X)`

*ejemplo*

`abs(c(-3.6, -7))`

regresa

`(3, 6, 7)`

**cbind.-** Tiene por argumentos un conjunto de vectores. Construye una matriz que tiene por columnas dichos vectores (respeta el orden).

*sintaxis*

`cbind(X1, X2, ..., Xk)`

*ejemplo*

```
cbind(c(-3, 6, -7), c(2, 2, 2))
```

```
regresa
```

$$\begin{pmatrix} -3 & 2 \\ 6 & 2 \\ -7 & 2 \end{pmatrix}$$

**cumsum.**- Su argumento es un vector, regresa un vector cuyas componentes son las sumas acumuladas de las componentes del argumento.

*sintaxis*

```
cumsum(X)
```

*ejemplo*

```
cumsum(c(-3, 6, -7))
```

```
regresa
```

```
(-3, 3, -4)
```

**for.**- Se trata de un control de ciclo, realiza un conjunto de instrucciones mientras cierto índice varía en un rango determinado.

*sintaxis*

```
for (índice in primer valor:último valor)
```

```
{instrucciones}
```

*ejemplo*

```
for (j in 1:4)
```

```
{x[j] < -j}
```

```
regresa
```

```
x = (1, 2, 3, 4)
```

**if.**- Lleva a cabo instrucciones si la proposición lógica que le sirve como argumento es verdadera.

*sintaxis*

```
if (proposición lógica)
```

```
{instrucciones}
```

*ejemplo*

```
y < -0
```

```
if (x == 4)
```

```
{y < -1}
```

```
regresa
```

$$y = \begin{cases} 0 & \text{si } x \neq 4 \\ 1 & \text{si } x = 4 \end{cases}$$



**length.-** Esta función tiene por argumento un vector y da como salida un entero que es la dimensión de dicho vector.

*sintaxis*

`length(X)`

*ejemplo*

```
length(c(-3, 6, -7))
regresa
3
```

**lines.-** Superpone una gráfica de tipo 'l' sobre una gráfica ya existente. Requiere la misma sintaxis que `plot`.

**pch.-** Argumento auxiliar de `plot`. Modifica el caracter desplegado en cada punto del gráfico.

**plot.-** Tiene por argumentos dos vectores  $(u, v)$  y crea una gráfica, cuyos puntos son de la forma  $(u[i], v[i])$ ,  $u$  y  $v$  deben ser de la misma dimensión

*sintaxis*

`plot(X, Y)`

**points.-** Superpone una gráfica de tipo 'p' sobre una gráfica ya existente. Requiere la misma sintaxis que `plot`.

**seq.-** Crea un vector de números reales o enteros, se pueden especificar los límites que se deseen para las componentes del vector, así como la longitud del mismo

*sintaxis*

`seq(...)`

*ejemplo*

```
seq(5)      regresa   (1, 2, 3, 4, 5)
seq(0, 1, .01) regresa (0, 01, 02, ..., 1)
seq(-3, 3, length = 100) regresa 100 valores en (-3, 3)
```

**sort.-** Tiene por argumento un vector y regresa un vector cuyas entradas son las del argumento en orden ascendente.

*sintaxis*

`sort(X)`

*ejemplo*

```
sort(c(-3, 6, -7))
regresa
(-7, -3, 6)
```

**rep.**- Se trata de una función con dos argumentos. Crea un vector donde todos los componentes tienen el mismo valor.

*sintaxis*

`rep(y, k)`

*ejemplo*

```
length(.01, 5)
regresa
(.01, .01, .01, .01, .01)
```

**rexp.**- Genera un vector aleatorio donde cada componente proviene de una distribución Exponencial.

*sintaxis*

`rexp(n, beta)`

**rgamma.**- Genera un vector aleatorio donde cada componente proviene de una distribución Gamma ( $\alpha, \beta$ ).

*sintaxis*

`rgamma(n, alfa, beta)`

**rpois.**- Genera un vector aleatorio donde cada componente proviene de una distribución Poisson.

*sintaxis*

`rpois(n, lambda)`

**runif.**- Genera un vector aleatorio donde cada componente proviene de una distribución Uniforme.

*sintaxis*

`runif(n, a, b)`

**type.**- Argumento auxiliar de plot. Se usa para especificar algunas modalidades para gráficos.

*sintaxis*

<i>instrucción</i>	<i>tipo de gráfico</i>
<code>type='p'</code>	puntos
<code>type='l'</code>	se interpola mediante una curva suave
<code>type='h'</code>	puntos con líneas al eje x

**while.**- Se trata de un control de ciclo, evalúa una proposición lógica y lleva a cabo ciertas instrucciones mientras dicha proposición sea verdadera.

*sintaxis*

```
while (proposición lógica)
{instrucciones}
```

*ejemplo*

```

i <- -1
y <- rep(0,5)
while(i <= 4)
  {y[i] <- -1
  i <- -i + 1}
regresa
y = { 0 si x ≠ 4
      1 si x = 4
    
```

**xlab, ylab.**- Estos argumentos auxiliares de `plot` permiten colocar rótulos en los ejes (X,Y respectivamente) de un gráfico.

**xlim, ylim.**- Estos argumentos auxiliares de `plot` permiten controlar los límites de los ejes (X,Y respectivamente) de un gráfico.



## Referencias

- [1] Basawa, I.V.; Prakasa Rao, B.L.S. (1980) *Statistical Inference for Stochastic Processes*. Academic Press.
- [2] Berger, J.O. (1980) *Statistical Decision Theory and Bayesian Analysis*. 2da. ed. Springer Verlag.
- [3] Bernardo, J.M ; Smith, A.F. (1994) *Bayesian Theory*. John Wiley & Sons.
- [4] Bratley, P; Fox, B.L.; Schrage, L. (1987) *A Guide to Simulation*. 2da. ed. Springer Verlag.
- [5] Brémaud, P. (1980) *An Introduction to Probabilistic Modeling*. Springer Verlag.
- [6] Casella, G.; Berger, R.L. (1990) *Statistical Inference*. Duxbury Press.
- [7] Cinlar, E. (1975) *Intoduction to Stochastic Processes*. Prentice Hall.
- [8] Feldman, D.; Fox, M (1991) *Probability: The Mathematics of Uncertainty*. Dekker.
- [9] Feller, W. (1971) *An Intoduction to Probability Theory and Its Applications*. Vol.2. 2da. ed. John Wiley & Sons.
- [10] Gelman, A.; Carlin, B.; Stern, H.S.; Rubin, B. (1995) *Bayesian Data Analysis*. Chapman & Hall.
- [11] Johnson, N.L.; Kotz, S.; Balakrishnan, N. (1994) *Continous Univariate Distributions*. John Wiley & Sons.
- [12] Karlin, S.; Taylor, H.M. (1994) *An Introduction to Stochastic Modeling*. ed. rev. Academic Press.
- [13] Karr, A. F. (1991) *Point Processes and Their Statistical Inference*. 2da. ed. Dekker.

- [14] Kendall, M.; Stuart, A. (1979) *The Advanced Theory of Statistics. vol.2: Inference and Relationship.* 4ta. ed. Charles Griffing & Company Limited.
- [15] Kingman, J.F.C. (1993) *Poisson Processes.* Oxford University Press.
- [16] Lindsey, J.K. (1996) *Parametric Statistical Inference.* Oxford University Press.
- [17] Mood, A.M.; Graybill, F.A.; Boes, D.C. (1974) *Introduction to the Theory of Statistics.* Mc. Graw Hill
- [18] O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics. vol.2B: Bayesian Inference.* Arnold.
- [19] Rao, C.R. (1973) *Linear Statistical Inference and Its Applications.* John Wiley & Sons
- [20] Ross, S.M. (1980) *Introduction to Probability Models.* 2da. ed. Academic Press.
- [21] Shao, J. (1998) *Mathematical Statistics.* Springer Verlag.
- [22] Snyder, D.L.; Miller, M.I. (1991) *Random Point Processes in Time and Space.* 2da. ed. Springer Verlag.
- [23] Widder, D.V. (1989) *Advanced Calculus.* 2da. ed. Dover Publications.