



"El saber de mis hijos
hará mi grandeza"

UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Posgrado en Matemáticas

Inferencia estadística en proceso semi-markoviano
Weibull

T E S I S

Que para obtener el grado de:

Maestro en Ciencias (Matemáticas)

Presenta:

Francisco Figueroa Ramírez

Directores de Tesis: Dr. José Arturo Montoya Laos
Dr. Oscar Vega Amaya

Hermosillo, Sonora, México, Agosto de 2016.

SINODALES

Dra. Gudelia Figueroa Preciado
Universidad de Sonora

Dr. Carlos Pacheco González
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Dr. José Arturo Montoya Laos
Universidad de Sonora

Dr. Oscar Vega Amaya
Universidad de Sonora



"El Saber de Mis Hijos
Hará Mi Grandeza"



UNIVERSIDAD DE SONORA

ACTA DE EXAMEN DE GRADO

En la ciudad de Hermosillo, Sonora, siendo las 17:00 horas del día 26 de agosto de 2016, se reunieron en el Auditorio del Departamento de Matemáticas de la Universidad de Sonora, los integrantes del jurado:

DR. CARLOS PACHECO GONZÁLEZ
DR. OSCAR VEGA AMAYA
DR. JOSÉ ARTURO MONTOYA LAOS
DRA. GUDELIA FIGUEROA PRECIADO

bajo la presidencia del primero y fungiendo como secretario el último, para realizar el examen de grado del programa de Maestro en Ciencias Matemáticas, a:

FRANCISCO FIGUEROA RAMIREZ

quien de acuerdo a la opción de examen de grado presentó un trabajo de investigación titulado

"Inferencia Estadística en Proceso Semi-markoviano Weibull"

Francisco Figueroa R

**FRANCISCO FIGUEROA
RAMIREZ**

208202458

El jurado, después de debatir entre sí reservada y libremente, emitió el siguiente dictamen:

APROBADO POR UNANIMIDAD

y para constancia se levantó la presente acta.

Acta: 40

Foja: 40

Libro: I

**DR. JESÚS ADOLFO
MINJAREZ SOSA,**
Coordinador del Programa de
Maestro en Ciencias
Matemáticas de la Universidad
de Sonora, hace constar que las
firmas que anteceden
corresponden al jurado que
intervino en el examen de
grado.

Hermosillo, Sonora, a 26 de
agosto de 2016

Adolfo Minjarez Sosa

**DR. JESÚS ADOLFO
MINJAREZ SOSA**

Coordinador de programa

Carlos Pacheco Gonzalez
**DR. CARLOS PACHECO
GONZÁLEZ**
Presidente

Oscar Vega Amaya
DR. OSCAR VEGA AMAYA
Sinodal

Gudelia Figueroa Preciado
**DRA. GUDELIA FIGUEROA
PRECIADO**
Secretario

Jose Arturo Montoya Laos
**DR. JOSÉ ARTURO MONTOYA
LAOS**
Sinodal



DIVISION DE CIENCIAS
EXACTAS Y NATURALES.
COORDINACIÓN
POSGRADO EN MATEMÁTICAS

Agradecimientos

Gracias a Dios por permitirme tener y disfrutar a mi familia, gracias a toda mi familia, en especial a mi madre (Cecilia Ramírez Córdova) y hermanos (Cecilia Figueroa Ramírez y Oscar Figueroa Ramírez), por estar a mi lado en cada momento de mi vida. Gracias Dios, por ser guía y darme la oportunidad de cumplir mis metas.

Me gustaría que estas líneas sirvieran para expresar mi más profundo y sincero agradecimiento a los doctores José Arturo Montoya Laos y Oscar Vega Amaya, directores de esta investigación, por la orientación, el seguimiento y la supervisión continúa de la misma, pero sobre todo por la motivación y el apoyo recibido a lo largo de estos años.

Muchas gracias a mis sinodales Dra. Gudelia Figueroa Preciado y Dr. Carlos Pacheco González, por haber destinado parte de su tiempo a la revisión de este trabajo, gracias por sus observaciones, correcciones y aprobación.

A todos los profesores que formaron parte de la maestría, por ofrecer sus conocimientos para mejorar profesionalmente.

A mi novia (Flor Carolina Silva Moreno) por apoyarme, motivarme y hacerme feliz.

A todos ellos, muchas gracias.

Francisco Figueroa Ramírez
Agosto de 2016

Dedicatoria

Este trabajo va dedicado a personas muy especiales, que siempre han estado a mi lado apoyándome y dando lo mejor de si para que siga progresando personal y profesionalmente.

A ***Francisco Figueroa Castro***, mi querido y santo Padre, que desde donde esta, siempre me da fuerzas para cumplir todas mis metas. Te quiero mucho Papá.

Francisco Figueroa Ramírez
Agosto de 2016

Índice general

1. Introducción	1
2. Procesos semi-markovianos	3
2.1. Simulación de un proceso semi-markoviano Weibull	7
3. Inferencia estadística	11
3.1. Teoría de verosimilitud	11
3.1.1. La función de verosimilitud	11
3.1.2. Estimador de máxima verosimilitud (emv)	14
3.2. Métodos de valuación de modelos	16
3.2.1. Métodos heurísticos	16
3.2.2. Prueba de hipótesis	20
3.2.3. Pruebas de significacia: <i>p-valor</i>	22
3.3. Teoría bootstrap	23
3.3.1. El método bootstrap	23
4. Estimación de parámetros del modelo Weibull y tamaño de muestra	29
4.1. Estimación de máxima verosimilitud de los parámetros del modelo Weibull	29
4.2. Función de verosimilitud Weibull	30
4.3. Tamaño de muestra esperado	31
5. Inferencias en un proceso semi-markoviano Weibull: Datos de terremotos	37
6. Conclusiones	43
7. Apéndice A	45
Bibliografía	47

Capítulo 1

Introducción

Los modelos semi-markovianos proporcionan un marco matemático muy flexible para el modelado de fenómenos aleatorios que evolucionan en el tiempo. Como extensiones de los procesos de Markov y de renovación, los procesos semi-markovianos tienen una potencial riqueza en diversas aplicaciones reales. Por ejemplo, en finanzas se pueden utilizar para el modelado de la clasificación de crédito y fiabilidad. En ciencias de la salud se pueden utilizar para el pronóstico y evolución de enfermedades. En sismología, los procesos semi-markovianos han sido utilizados para el modelado de ocurrencias de tipos de terremotos.

En la literatura de Probabilidad y Estadística, el modelo semi-markoviano con distribución de tiempo de espera Weibull ha sido ampliamente estudiado. Esto puede deberse a la flexibilidad del modelo Weibull para el modelado de tiempos de espera y a que la distribución exponencial es un caso particular de la distribución Weibull, que corresponde a un proceso markoviano en tiempo continuo.

Hacer inferencia estadística sobre los parámetros involucrados en un modelo semi-markoviano con distribución de tiempo de espera Weibull o sobre parámetros de interés que son funciones de los parámetros del modelo, es un problema relevante que requiere de un conocimiento profundo del fenómeno aleatorio bajo estudio, del marco teórico del proceso, de diferentes conceptos y metodologías de inferencia estadística, de los problemas de estimación asociados con la distribución Weibull, entre otros.

En esta tesis se aborda el problema de inferencia estadística en procesos semi-markovianos con distribución de tiempo de espera Weibull. A continuación se describe de manera general los aspectos desarrollados en cada capítulo.

En el Capítulo 1 se expone la construcción del proceso semi-markoviano, el cual involucra una cadena de Markov y un proceso de renovación. Se presentan las demostraciones de algunas de sus propiedades usando conceptos y resultados básicos de probabilidad.

En el Capítulo 2 se desarrollan conceptos relevantes del enfoque de verosimilitud y la teoría Bootstrap, utilizados aquí para estimar los parámetros involucrados en el proceso semi-markoviano bajo estudio y para estimar características de interés

(parámetros que son funciones de los parámetros del modelo) a través de intervalos de confianza, respectivamente. Algunos de estos conceptos son: La función de verosimilitud, estimador de máxima verosimilitud, el método bootstrap, el algoritmo bootstrap y los intervalos de confianza bootstrap. Además, se presentan métodos de valuación de modelos como: Métodos heurísticos, prueba de hipótesis y pruebas de significancia.

En el Capítulo 3 se hace una revisión histórica de los problemas de estimación asociados con la distribución Weibull para el caso de tamaños de muestra pequeños. Se muestra que la forma de la función de verosimilitud es muy informativa ya que permite identificar no solo dificultades estadísticas en la estimación sino también problemas computacionales de optimización. Además, se hace un estudio para explorar, a través de simulaciones y de forma teórica, el comportamiento de la matriz de frecuencias de transición del proceso. Esto con el objetivo de mostrar el número esperado de transiciones y anticipar posibles dificultades de estimación asociadas a un tamaño de muestra pequeño.

En el Capítulo 4 se hace inferencia estadística sobre los parámetros del proceso semi-markoviano con distribución de tiempo de espera Weibull, descrito en el el Capítulo 1, con base en datos reales de terremotos de gran magnitud que ocurrieron en el siglo XX y se localizaron en el llamado “Norte de Anatolia Zona de Falla” en Turquía. En particular, se hace estimación puntual y valuación de modelos usando las herramientas estadísticas del Capítulo 2. Además, se calculan estimadores puntuales e intervalos de confianza bootstrap para una probabilidad de transición que es función de parámetros involucrados en el modelo.

En el Capítulo 5 se presentan las conclusiones generales de esta tesis.

Las simulaciones, cálculos de funciones de verosimilitud, optimizaciones y las figuras mostradas en esta tesis se realizaron con la ayuda del lenguaje de programación R versión 3.0.2 instalado en una computadora personal Intel Core Duo.

Capítulo 2

Procesos semi-markovianos

Los procesos semi-markovianos proporcionan una estructura muy flexible para el modelado de fenómenos que evolucionan en el tiempo ya que combinan la estructura probabilística de una cadena de Markov y de un proceso de renovación. En otras palabras, los procesos semi-markovianos modelan sistemas cuyas transiciones entre estados ocurren de acuerdo a una cadena de Markov y los tiempos en que ocurren dichas transiciones son aleatorios y forman un proceso de renovación.

Para fijar ideas, supóngase que en el tiempo $\tau_0 = 0$ el sistema se encuentra en el estado inicial $X_0 = x_0$, donde permanece un tiempo aleatorio $\tau_1 \geq 0$. En este instante, el sistema se mueve a un nuevo estado $X_1 = x_1$, donde permanece un tiempo aleatorio $\tau_2 \geq 0$, y así sucesivamente. A los tiempos τ_n , $n \in \mathbb{N}$, se les llama *tiempos de permanencia* y a los tiempos

$$T_{n+1} := T_n + \tau_{n+1}, \quad n \in \mathbb{N}, \quad T_0 := 0,$$

se les llama *tiempos de salto* del sistema. Naturalmente, los tiempos de salto T_n , $n \in \mathbb{N}$, son no-decrecientes puesto que los tiempos de permanencia son no-negativos. Al proceso X_n , $n \in \mathbb{N}_0$, se le llama *proceso de estados* y al conjunto \mathcal{H} donde toma valores se le llama *espacio de estados* del sistema.

Definición 1. *Un proceso de renovación Markoviano (X_n, τ_{n+1}) , $n \in \mathbb{N}_0$, con espacio de estados discreto \mathcal{H} es un proceso estóastico tal que*

$$\begin{aligned} P(X_{n+1} = y, \tau_{n+1} \leq t | X_0 = x_0, \tau_1 = t_1, \dots, \tau_n = t_n, X_n = x) \\ = P(X_{n+1} = y, \tau_{n+1} \leq t | X_n = x), \end{aligned}$$

para toda colección de estados $x_1, x_2, \dots, x, y \in \mathcal{H}$, números reales no-negativos t_1, t_2, \dots, t_n , y todo número natural $n \in \mathbb{N}$.

En todo lo que resta de este trabajo considérese un proceso de renovación markoviano (X_n, τ_{n+1}) , $n \in \mathbb{N}_0$, con espacio de estados numerable \mathcal{H} . Supóngase además que dicho proceso es homogéneo, es decir,

$$P(X_{n+1} = y, \tau_{n+1} \leq t | X_n = x) = P(X_{k+1} = y, \tau_{k+1} \leq t | X_k = x)$$

para todo $x, y \in \mathcal{H}$ y $n, k \in \mathbb{N}_0$.

Definición 2. *A la función de probabilidades*

$$Q(y, t | x) := P(X_{n+1} = y, \tau_{n+1} \leq t | X_n = x), \quad x, y \in \mathcal{H}, n \in \mathbb{N}_0, t \in \mathbb{R},$$

se le llama kernel semi-markoviano.

Para ilustrar los conceptos anteriores, considérese el fenómeno de ocurrencia de sismos. Naturalmente, las variables de interés son los tiempos en que se presentan los sismos y su intensidad medida con la escala de Richter. Supóngase que los sismos se clasifican en h tipos diferentes x_1, x_2, \dots, x_h dependiendo de su intensidad. En este caso, T_n representa el instante en que se presenta el n -ésimo sismo, X_n indica su tipo de acuerdo a la clasificación $\mathcal{H} = \{x_1, x_2, \dots, x_h\}$ y τ_n es el tiempo transcurrido entre el $(n-1)$ -ésimo y el n -ésimo sismo. Entonces, $Q(y, t | x)$ es la probabilidad de que el próximo sismo sea del tipo $y \in \mathcal{H}$ y que éste se presente antes de que transcurran t unidades de tiempo desde la ocurrencia del anterior bajo el supuesto de que este fue del tipo $x \in \mathcal{H}$.

El kernel semi-markoviano tiene las siguientes propiedades:

- $Q(y, t | x) \geq 0$ para todo $x, y \in \mathcal{H}$ y $t \geq 0$.
- $\sum_{y \in \mathcal{H}} Q(y, t | x) \leq 1$ para todo $x, y \in \mathcal{H}$ y $t \geq 0$.
- $Q(y, t | x)$ es no-decreciente en la variable t para todo $x, y \in \mathcal{H}$. En consecuencia, la función

$$\theta_{xy} := \lim_{t \rightarrow \infty} Q(y, t | x), \quad x, y \in \mathcal{H},$$

está bien definida. Además,

$$Q(y, t | x) \leq \theta_{xy} \quad \forall x, y \in \mathcal{H}, t \in \mathbb{R}. \quad (2.1)$$

- Nótese que

$$\theta_{xy} = P(X_{n+1} = y | X_n = x) \quad \forall x, y \in \mathcal{H}, n \in \mathbb{N}_0.$$

Por lo tanto, el proceso de estados $\{X_n\}$ es una cadena de Markov con probabilidad de transición θ ...

Defínase la función

$$H(t | y) := P(\tau_{n+1} \leq t | X_n = y), \quad t \in \mathbb{R}, y \in \mathcal{H},$$

y obsérvese que

$$H(t | y) = \sum_{x \in \mathcal{H}} Q(x, t | y) \quad \forall x, y \in \mathcal{H}, t \in \mathbb{R}.$$

En otras palabras, $H(\cdot|y)$ es la distribución del tiempo de permanencia del sistema en el estado $y \in \mathcal{H}$. Entonces, el tiempo medio de permanencia en y es

$$\mu(y) := \int_0^\infty [1 - H(t|y)] dt.$$

Además, si $\theta_{xy} > 0$, por (2.1) existe una única función $F(\cdot|y, x)$ tal que

$$Q(y, \cdot|x) = F(\cdot|y, x) \theta_{xy}. \quad (2.2)$$

Por otra parte, si $\theta_{xy} = 0$, se toma $F(\cdot|y, x) := 0$. En resumen,

$$F(t|y, x) = \begin{cases} \frac{Q(y, t|x)}{\theta_{xy}} & \text{si } \theta_{xy} > 0, \\ 0 & \text{si } \theta_{xy} = 0. \end{cases} \quad (2.3)$$

Proposición 1. Sean $x, y \in \mathcal{H}$ estados arbitrarios del proceso de renovación markoviano. Entonces,

$$F(t|y, x) = P(\tau_{n+1} \leq t | X_n = x, X_{n+1} = y) \quad (2.4)$$

para todo $t \geq 0$ y $n \in \mathbb{N}_0$. Además,

$$H(t|x) = \sum_{y \in \mathcal{H}} F(t|y, x) \theta_{xy} \quad (2.5)$$

para todo $x \in \mathcal{H}$ y $t \geq 0$.

Demostración. De la definición de $H(t|y)$ se obtiene el resultado (2.5) de la siguiente manera:

$$\begin{aligned} H(t|x) &= \sum_{y \in \mathcal{H}} Q(y, t|x) \\ &= \sum_{\{y: Q(y, t|x) > 0\}} F(t|y, x) \theta_{xy} \\ &= \sum_{y \in \mathcal{H}} F(t|y, x) \theta_{xy}. \end{aligned}$$

Para demostrar (2.4), supóngase que $\theta_{xy} > 0$, entonces

$$F(t|y, x) = \frac{Q(y, t|x)}{\theta_{xy}}.$$

Por la definición de probabilidad condicional se sigue que

$$\begin{aligned} F(t|y, x) &= \frac{Q(y, t|x)}{\theta_{xy}} \\ &= \frac{P(X_{n+1} = y, \tau_{n+1} \leq t | X_n = x)}{P(X_{n+1} = y | X_n = x)} \\ &= \frac{P(X_{n+1} = y, \tau_{n+1} \leq t, X_n = x)}{P(X_n = x)} \bigg/ \frac{P(X_{n+1} = y, X_n = x)}{P(X_n = x)} \\ &= \frac{P(X_{n+1} = y, \tau_{n+1} \leq t, X_n = x)}{P(X_{n+1} = y, X_n = x)} \\ &= P(\tau_{n+1} \leq t | X_n = x, X_{n+1} = y), \end{aligned}$$

lo cual prueba (2.4).

■

Teorema 1. Para cada $n \in \mathbb{N}$, las variables $\tau_1, \tau_2, \dots, \tau_n$ son condicionalmente independientes dadas las variables X_0, X_1, \dots, X_n . Además,

$$\begin{aligned} P(\tau_1 \leq t_1, \dots, \tau_n \leq t_n | X_0 = x_0, \dots, X_n = x_n) \\ = F(t_1 | x_0, x_1) F(t_2 | x_1, x_2) \cdots F(t_n | x_{n-1}, x_n) \end{aligned}$$

para todo $x_0, x_1, \dots, x_n \in \mathcal{H}$ y números reales no-negativos t_1, t_2, \dots, t_n .

Demostración. Se realizará la demostración sólo para el caso $n = 2$, puesto que exhibe el esquema para la demostración en el caso general. Usando las propiedades y definiciones de la probabilidad condicional se pueden verificar la siguiente secuencia de igualdades:

$$\begin{aligned} P(\tau_1 \leq t_1, \tau_2 \leq t_2 | X_0 = x_0, X_1 = x_1, X_2 = x_2) \\ &= \frac{P(\tau_1 \leq t_1, \tau_2 \leq t_2, X_0 = x_0, X_1 = x_1, X_2 = x_2)}{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)} \\ &= \frac{P(X_2 = x_2, \tau_2 \leq t_2 | X_0 = x_0, \tau_1 \leq t_1, X_1 = x_1) P(X_0 = x_0, \tau_1 \leq t_1, X_1 = x_1)}{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)} \\ &= \frac{P(X_2 = x_2, \tau_2 \leq t_2 | X_1 = x_1) P(X_0 = x_0, \tau_1 \leq t_1, X_1 = x_1)}{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)} \\ &= \frac{P(X_2 = x_2, \tau_2 \leq t_2 | X_1 = x_1) P(X_1 = x_1, \tau_1 \leq t_1 | X_0 = x_0) P(X_0 = x_0)}{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)} \\ &= \frac{Q(x_2, t_2 | x_1) Q(x_1, t_1 | x_0) P(X_0 = x_0)}{P(X_0 = x_0) \theta_{x_0 x_1} \theta_{x_1 x_2}} \\ &= F(t_2 | x_1, x_2) F(t_1 | x_0, x_1). \end{aligned}$$

■

Para cada par de estados $x, y \in \mathcal{H}$, denotese por τ_{xy} a la variable aleatoria con función de distribución $F(t | x, y)$. Obsérvese que τ_{xy} representa el tiempo que tardó en ocurrir la transición del estado x al estado y , bajo el supuesto de que dicha transición efectivamente ya ocurrió. Entonces, para cada $x, y \in \mathcal{H}$, la esperanza condicional

$$\begin{aligned} \mu(y | x) &:= E[\tau_{xy} | X_0 = x, X_1 = y] \\ &= \int_0^\infty [1 - F(t | y, x)] dt, \end{aligned}$$

representa el tiempo medio que toma una transición de x a y .

Definición 3. Para cada $t \geq 0$, defínase

$$N(t) := \sup \{n \geq 1 : T_n \leq t\}, \quad t \geq 0.$$

Nótese que $N(t)$ registra el número de transiciones en el intervalo de tiempo $(0, t]$. Para los procesos de renovación markovianos, de forma análoga a lo que ocurre en teoría clásica de renovación, también se tiene que

$$T_n \leq t \Leftrightarrow N(t) \geq n. \quad (2.6)$$

La evolución del proceso de renovación markoviano (X_n, τ_{n+1}) , $n \in \mathbb{N}_0$, también puede registrarse por medio del proceso en tiempo continuo

$$Z(t) = X_{N(t)}, \quad t \geq 0.$$

La relación (2.6) implica que

$$N(t) = n \Leftrightarrow T_n \leq t < T_{n+1}.$$

Entonces,

$$Z(t) = X_n \quad \text{si} \quad T_n \leq t < T_{n+1}.$$

Al proceso $Z(t)$, $t \geq 0$, se le llama *proceso semi-markoviano* y al proceso en tiempo discreto X_n , $n \in \mathbb{N}$, se le llama *cadena inmersa* en el proceso semi-markoviano.

Nótese que, a priori, no se puede garantizar que el proceso semi-markoviano $Z(t)$, $t \geq 0$, está definido sobre todo el semi-eje de los reales positivos. Esto se debe a la posibilidad de que la sucesión de los tiempos de salto T_n , $n \in \mathbb{N}$, sea acotada, permitiendo que en ciertos intervalos de tiempo ocurran una infinidad de transiciones.

Observación 1. *La siguiente condición garantiza que la variable $N(t)$ es finita con probabilidad uno para cada $t \geq 0$. Existen constantes $\alpha, \beta > 0$ tal que*

$$1 - H(\alpha | x) > \beta \quad \forall x \in \mathcal{H}.$$

En otras palabras esta condición evita que el proceso de renovación markoviano experimente demasiadas transiciones en intervalos cortos de tiempo. Su demostración se puede consultar en Ross (1970, p. 88).

2.1. Simulación de un proceso semi-markoviano Weibull

Sea $Z(t)$, $t \geq 0$, un proceso semi-markoviano con espacio de estados $\mathcal{H} = \{1, 2, 3\}$. Puesto que

$$Q(y, t | x) = F(t | x, y) \theta_{xy} \quad \forall x, y \in \mathcal{H}, t \geq 0,$$

la evolución del proceso semi-markoviano queda completamente especificado si se proporcionan las funciones de distribución $F(\cdot | x, y)$ y las probabilidades de transición θ_{xy} para cada $x, y \in \mathcal{H}$. Supóngase entonces que

$$F(t | x, y) = \int_0^t f(s | x, y) ds, \quad t \geq 0, x, y \in \mathcal{H},$$

donde

$$f(s|x, y) := \left(\frac{\alpha_{xy}}{\beta_{xy}} \right) s^{\alpha_{xy}-1} \exp(-s^{\alpha_{xy}}/\beta_{xy}), \quad s \geq 0, \quad (2.7)$$

y los parámetros α_{xy} , β_{xy} , $x, y \in \mathcal{H}$, corresponden a las entradas de las matrices

$$A = \begin{bmatrix} 1.20 & 4.42 & 0.67 \\ 0.47 & 3.12 & 1.47 \\ 0.88 & 0.95 & 0.41 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 985.87 & 1743475000000 & 46.94 \\ 13.90 & 1297043000 & 14162.69 \\ 161.29 & 1690.608 & 15.25 \end{bmatrix}.$$

Por otra parte, las probabilidades de transición θ_{xy} , $x, y \in \mathcal{H}$, corresponden a las entradas de la matriz

$$\Theta = \begin{bmatrix} 0.50 & 0.16 & 0.34 \\ 0.29 & 0.19 & 0.52 \\ 0.51 & 0.06 & 0.43 \end{bmatrix}. \quad (2.8)$$

A continuación se describe el procedimiento de simulación. Para hacer esto, tómese $Z(0) = z_0 = 1$ y $\tau_0 = 0$.

- (a) Primero se simulará la transición del sistema del estado 1 a un estado y . Nótese que esto corresponde a simular una variable aleatoria con distribución $(\theta_{11}, \theta_{12}, \theta_{13}) = (0.50, 0.16, 0.34)$. Esta simulación puede hacerse de varias maneras, por ejemplo, simulando una muestra de un elemento de una variable multinomial con parámetros $(0.50, 0.16, 0.34)$.
- (b) Una vez simulada la transición y suponiendo que fue hacia el estado $X_1 = y$ se procede a simular el tiempo $\tau_{z_0 y}$ que tardó en ocurrir dicha transición. De acuerdo a (2.7), esto corresponde a simular una variable aleatoria Weibull con parámetros α_{1y} y β_{1y} . Esta simulación puede realizarse simulando una observación p de una uniforme en el intervalo $(0, 1)$ y evaluando la inversa de la función de distribución Weibull en p . Nótese que $\tau_1 = \tau_{z_0 y}$.
- (c) Se repiten los pasos (a) y (b) pero ahora considerando como estado inicial a $Z(\tau_1) = X_1 = y$.

La Figura (2.1) muestra los resultados de una simulación de las primeras 1000 transiciones del proceso semi-markoviano especificado por (2.7) y (2.8).

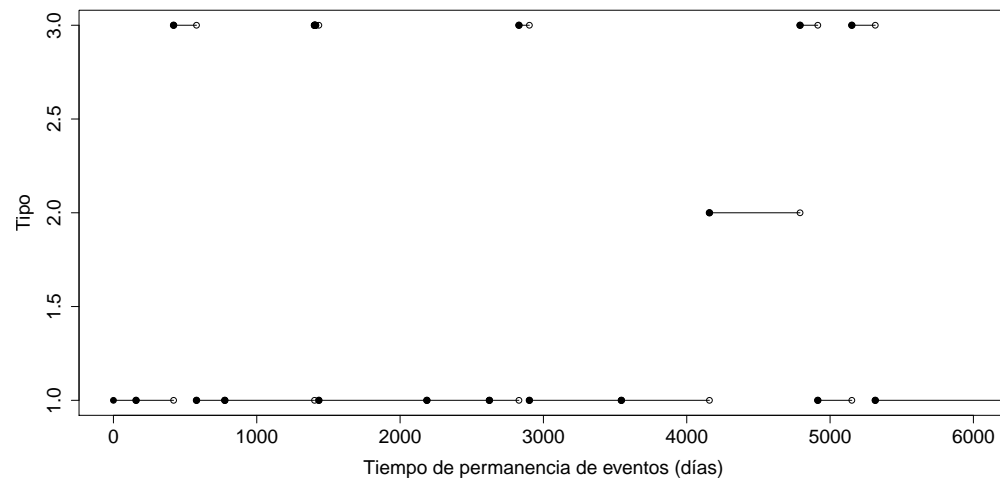


Figura 2.1: Simulación de un proceso semi-markoviano.

Capítulo 3

Inferencia estadística

En este capítulo se presentan los siguientes conceptos básicos del enfoque de verosimilitud y la teoría Bootstrap como: La función de verosimilitud, estimador de máxima verosimilitud, el método bootstrap, el algoritmo bootstrap y los intervalos de confianza bootstrap. También se presentan métodos de valuación de modelos, entre ellos métodos heurísticos, prueba de hipótesis y pruebas de significancia. Todos estos conceptos teóricos estadísticos serán usados más adelante para hacer inferencia en los procesos semi-markovianos aplicados a datos reales y simulados.

3.1. Teoría de verosimilitud

El enfoque de verosimilitud es ampliamente usado por el uso objetivo que hace de la información estadística para la estimación de parámetros, así como porque sus resultados son estadísticamente eficientes. A continuación se describen conceptos fundamentales de esta teoría que se utilizarán posteriormente.

3.1.1. La función de verosimilitud

Supóngase que se ha formulado un modelo de probabilidad para un fenómeno aleatorio de interés y que dicho modelo involucra un vector de parámetros desconocidos $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \Lambda \subset \mathbb{R}^m$ donde Λ es la región de valores posibles que puede tomar λ . Además, supóngase que el resultado de observar el fenómeno aleatorio de interés es un conjunto de datos que puede considerarse como un evento observado E_{obs} en el espacio muestral del modelo de probabilidad. La probabilidad del evento observado E_{obs} se puede determinar del modelo, y en general, será una función del vector de parámetros desconocidos λ , $P(E_{obs}; \lambda)$. La *función de verosimilitud* o simplemente *verosimilitud* de λ se define como la probabilidad del evento observado E_{obs} pero vista como función de λ , esto es,

$$L(\lambda; E_{obs}) = P(E_{obs}; \lambda). \quad (3.1)$$

La verosimilitud en (3.1) se definió para el caso general, ya que un evento observado E_{obs} se puede asociar a un vector de variables aleatorias continuas o discretas $X = (X_1, X_2, \dots, X_n)$, que pueden ser independientes e idénticamente distribuidas o dependientes con alguna función de probabilidad conjunta paramétrica.

Definición 4. Sea $X = (X_1, X_2, \dots, X_n)$ un vector de variables aleatorias discretas con función de probabilidad conjunta $P(x; \lambda) : \mathbb{R}^n \rightarrow [0, 1]$ y que $x = (x_1, x_2, \dots, x_n)$ es un valor observado de X . La función de verosimilitud $L(\lambda; x) : \Lambda \rightarrow [0, 1]$ se define como

$$L(\lambda; x) = P(x; \lambda) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \lambda).$$

Cuando $X = (X_1, X_2, \dots, X_n)$ es un vector de variables aleatorias discretas independientes e idénticamente distribuidas con función de probabilidad $P(x'; \lambda)$ y $x = (x_1, x_2, \dots, x_n)$ es una muestra observada de X , entonces

$$\begin{aligned} P(X = x; \lambda) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \lambda) \\ &= \prod_{i=1}^n P(X_i = x_i; \lambda) \\ &= \prod_{i=1}^n P(x_i; \lambda). \end{aligned}$$

Por lo tanto, en este caso la función de verosimilitud del parametro λ es

$$L(\lambda; x) = \prod_{i=1}^n P(x_i; \lambda). \quad (3.2)$$

Cuando se tienen variables aleatorias continuas independientes e idénticamente distribuidas con función de densidad $f(x; \lambda)$ y un instrumento de medición que registra las observaciones con una precisión $\pm\epsilon/2$, la verosimilitud se define como en (3.2) escribiendo $P(x_i; \lambda)$ como

$$P(x_i; \lambda) = P\left(x_i - \frac{\epsilon}{2} \leq X_i \leq x_i + \frac{\epsilon}{2}\right) = \int_{x_i - \frac{\epsilon}{2}}^{x_i + \frac{\epsilon}{2}} f(x; \lambda) dx.$$

Para más detalles consulte Barnard y Sprott (1983).

Entonces, la función de verosimilitud de λ para la muestra observada se suele expresar como

$$L(\lambda; x) = \prod_{i=1}^n f(x_i; \lambda). \quad (3.3)$$

Cabe mencionar aquí que (3.3) se obtiene por una aproximación a la verosimilitud de λ a través del teorema de valor medio para integrales, y despreciando constantes de proporcionalidad.

La función de verosimilitud juega un papel fundamental en la inferencia estadística. Su rol principal es inferir qué valores del vector de parámetros $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ son razonables para la función de probabilidad $P(x; \lambda)$ que haya sido elegida para

el fenómeno aleatorio de interés a partir de un evento observado (datos). Esto es particularmente relevante después de un experimento, cuando ya fueron observados los datos.

A continuación se ilustra el concepto de función de verosimilitud con el proceso semi-markoviano introducido en la Sección 1.1. En este caso, los parámetros son las entradas de la matriz de transición $\Theta = (\theta_{xy})$ y de las matrices $A = (\alpha_{xy})$ y $B = (\beta_{xy})$ que especifican la densidad condicional $f(\cdot | x, y)$ donde $x, y \in \mathcal{H} = \{1, 2, 3\}$.

Ejemplo 1. Sea $Z(t)$, $t \geq 0$, el proceso semi-markoviano descrito en la Sección 1.1. Supóngase que el evento observado E_{obs} consiste en los estados que el sistema visitó en las primeras n -transiciones y los tiempos en los cuales dichas transiciones ocurrieron; es decir,

$$E_{obs} = \{T_0 = t_0, Z(T_0) = z_0, \dots, T_n = t_n, Z(T_n) = z_n\}.$$

Tomando $t_0 = 0$, este evento puede re-escribirse equivalentemente como

$$E_{obs} = \{\tau_0 = 0, Z(T_0) = z_0, \tau_1 = t_1, Z(T_1) = z_1, \dots, \tau_n = t_n - t_{n-1}, Z(T_n) = z_n\},$$

y este a su vez como

$$E_{obs} = \{Z_0 = z_0, Z(T_1) = z_1, \tau_{z_0 z_1} = t_1, Z(T_2) = z_2, \tau_{z_1 z_2} = t_2 - t_1, \dots, Z(T_n) = z_n, \tau_{z_{n-1} z_n} = t_n - t_{n-1}\}.$$

Para cada par de estados $x, y \in \mathcal{H} = \{1, 2, 3\}$, denótese por n_{xy} el número de transiciones que ocurrieron del estado x al estado y . Entonces, se tendrán n_{xy} observaciones de la variable τ_{xy} que se denotan como

$$\tau_{xy1}, \tau_{xy2}, \dots, \tau_{xyn_{xy}}.$$

Observe que debe cumplirse que $n = \sum_{x,y \in \mathcal{H}} n_{xy}$. Entonces, la función de verosimilitud es

$$\begin{aligned} L(\Theta(\theta_{xy}), A(\alpha_{xy}), B(\beta_{xy}); E_{obs}) &= \prod_{x=1}^3 \prod_{y=1}^3 \theta_{xy}^{n_{xy}} \prod_{k=1}^{n_{xy}} f(\tau_{xyk}; \alpha_{xy}, \beta_{xy}) \\ &= \prod_{x=1}^3 \prod_{y=1}^3 \theta_{xy}^{n_{xy}} \prod_{k=1}^{n_{xy}} \left(\frac{\alpha_{xy}}{\beta_{xy}} \right) \tau_{xyk}^{\alpha_{xy}-1} \exp(-\tau_{xyk}/\beta_{xy}) \\ &= \prod_{x=1}^3 \prod_{y=1}^3 \theta_{xy}^{n_{xy}} \left(\frac{\alpha_{xy}}{\beta_{xy}} \right)^{n_{xy}} \left(\prod_{k=1}^{n_{xy}} \tau_{xyk}^{\alpha_{xy}-1} \right) \exp\left(-\frac{1}{\beta_{xy}} \sum_{k=1}^{n_{xy}} \tau_{xyk}\right), \end{aligned} \quad (3.4)$$

donde $\sum_{y=1}^3 \theta_{xy} = 1$ para toda $x \in \mathcal{H} = \{1, 2, 3\}$.

3.1.2. Estimador de máxima verosimilitud (emv)

La función de verosimilitud $L(\lambda; x)$ proporciona la plausibilidad de diferentes valores posibles de λ . Así, un estimador puntual razonable de λ es aquel valor de λ que maximiza $L(\lambda; x)$.

Definición 5. *El estimador de máxima verosimilitud de λ es cualquier valor $\hat{\lambda} \in \Lambda$ que cumple que*

$$L(\hat{\lambda}; x) = \sup_{\lambda \in \Lambda} L(\lambda; x).$$

Nótese que $L(\lambda; x) = P(x; \lambda) \leq 1$ debido a que $P(x; \lambda)$ es una probabilidad como función de x para cada λ . Así, el supremo de $L(\lambda; x)$ en $[0, 1]$ existe y es finito. Sin embargo, puede ocurrir que el emv no exista, o si existe, puede que no sea único. De aquí en adelante se supondrá que existe un único valor $\hat{\lambda} \in \Lambda$ que maximiza a la función de verosimilitud $L(\lambda; x)$.

El emv del parámetro λ es el valor más plausible de λ . Es decir, el emv $\hat{\lambda}$ es el valor de λ que explica mejor a la muestra observada en el sentido de que maximiza su probabilidad bajo el modelo de probabilidad propuesto para el fenómeno aleatorio de interés.

Para encontrar el emv de λ es necesario maximizar la función de verosimilitud $L(\lambda; x)$. Sin embargo, usualmente $L(\lambda; x)$ es el producto de varios términos y resulta matemáticamente conveniente trabajar con el logaritmo natural de la función de verosimilitud.

Definición 6. *La función log-verosimilitud se define como*

$$l(\lambda; x) = \ln [L(\lambda; x)].$$

El valor $\hat{\lambda}$ que maximiza $L(\lambda; x)$ también maximiza $l(\lambda; x)$ puesto que el logaritmo natural es una transformación monótona creciente.

A continuación se muestra el uso de la log-verosimilitud junto con un procedimiento típico de cálculo para obtener el emv de los parámetros θ_{xy} , α_{xy} y β_{xy} para cada $x, y \in \mathcal{H}$ del modelo semi-markoviano descrito en el Ejemplo 1. Además, se calcula el valor numérico de estos estimadores para los datos simulados del capítulo anterior.

Ejemplo 2. *La función log-verosimilitud correspondiente a la verosimilitud del Ejemplo 1 es*

$$\begin{aligned} l(\theta_{xy}, \alpha_{xy}, \beta_{xy}; \tau_{xyk}) &= \sum_{x=1}^3 \sum_{y=1}^3 n_{xy} [\log(\theta_{xy}) + \log(\alpha_{xy}) - \log(\beta_{xy})] \\ &\quad + (\alpha_{xy} - 1) \sum_{k=1}^{n_{xy}} \log(\tau_{xyk}) - \frac{1}{\beta_{xy}} \sum_{k=1}^{n_{xy}} (\tau_{xyk})^{\alpha_{xy}}, \quad (3.5) \end{aligned}$$

donde $\sum_{y=1}^3 \theta_{xy} = 1$ para toda $x \in \mathcal{H} = \{1, 2, 3\}$.

Aplicando el método de multiplicadores de Lagrange se tiene la siguiente función lagrangiana

$$\begin{aligned} h(\theta_{xy}, \alpha_{xy}, \beta_{xy}; \tau_{xyk}, \lambda_x) &= \sum_{x=1}^3 \sum_{y=1}^3 n_{xy} [\log(\theta_{xy}) + \log(\alpha_{xy}) - \log(\beta_{xy})] \\ &\quad + (\alpha_{xy} - 1) \sum_{k=1}^{n_{xy}} \log(\tau_{xyk}) - \frac{1}{\beta_{xy}} \sum_{k=1}^{n_{xy}} (\tau_{xyk})^{\alpha_{xy}} \\ &\quad - \sum_{x=1}^3 \lambda_x \sum_{y=1}^3 (\theta_{xy} - 1), \end{aligned} \quad (3.6)$$

donde $\sum_{y=1}^3 \theta_{xy} = 1$ para toda $x \in \mathcal{H} = \{1, 2, 3\}$. Para obtener los estimadores de máxima verosimilitud, se obtienen las derivadas parciales de (3.6) con respecto a θ_{xy} , α_{xy} y β_{xy} , se igualan a cero y se resuelven las siguientes ecuaciones resultantes:

$$\frac{\partial(h(\theta_{xy}, \alpha_{xy}, \beta_{xy}; \tau_{xyk}, \lambda_x))}{\partial \theta_{xy}} = \frac{n_{xy}}{\theta_{xy}} - \lambda_y = 0, \quad (3.7)$$

$$\begin{aligned} \frac{\partial(h(\theta_{xy}, \alpha_{xy}, \beta_{xy}; \tau_{xyk}, \lambda_x))}{\partial \alpha_{xy}} &= \frac{n_{xy}}{\alpha_{xy}} + \sum_{k=1}^{n_{xy}} \log(\tau_{xyk}) - \frac{1}{\beta_{xy}} \sum_{k=1}^{n_{xy}} (\tau_{xyk})^{\alpha_{xy}} \log(\tau_{xyk}) \\ &= 0, \end{aligned} \quad (3.8)$$

$$\frac{\partial(h(\theta_{xy}, \alpha_{xy}, \beta_{xy}; \tau_{xyk}, \lambda_x))}{\partial \beta_{xy}} = -\frac{n_{xy}}{\beta_{xy}} + \frac{1}{(\beta_{xy})^2} \sum_{k=1}^{n_{xy}} (\tau_{xyk})^{\alpha_{xy}} = 0. \quad (3.9)$$

De la ecuación (3.7) se obtiene que $\theta_{xy} = n_{xy}/\lambda_x$. Puesto que $\sum_{y=1}^3 \theta_{xy} = 1$ para toda $x \in \mathcal{H} = \{1, 2, 3\}$, entonces, el emv de la matriz de transición Θ es simplemente la proporciones muestrales; esto es,

$$\hat{\Theta} = \left\{ \hat{\theta}_{xy} \right\}, \text{ donde } \hat{\theta}_{xy} = n_{xy} / \sum_{y=1}^3 n_{xy}, \quad x, y \in \mathcal{H}. \quad (3.10)$$

Para más detalles consulte Ross (1980, pág. 364).

Para obtener el emv de α_{xy} y β_{xy} se procederá resolviendo la ecuación (3.9) para β_{xy} ; así, el emv restringido de β_{xy} para cada valor fijo de los otros parámetros es

$$\beta_{xy}(\alpha_{xy}) = \frac{\sum_{k=1}^{n_{xy}} \tau_{xyk}^{\alpha_{xy}}}{n_{xy}}. \quad (3.11)$$

Ahora, reemplazando (3.11) en (3.8) se obtiene la siguiente ecuación

$$\frac{1}{n_{xy}} \sum_{k=1}^{n_{xy}} \log(\tau_{xyk}) = \frac{\sum_{k=1}^{n_{xy}} (\tau_{xyk})^{\alpha_{xy}} \log(\tau_{xyk})}{\sum_{k=1}^{n_{xy}} \tau_{xyk}^{\alpha_{xy}}} - \frac{1}{\alpha_{xy}}. \quad (3.12)$$

El emv de α_{xy} se obtiene resolviendo numéricamente (3.12); es decir, $\hat{\alpha}_{xy}$ hace cero la ecuación (3.12). Luego, el emv de β_{xy} se obtiene reemplazando $\hat{\alpha}_{xy}$ en (3.11);

esto es,

$$\hat{\beta}_{xy} = \frac{\sum_{k=1}^{n_{xy}} \tau_{xyk}^{\hat{\alpha}_{xy}}}{n_{xy}}. \quad (3.13)$$

A continuación se presentan las estimaciones puntuales de los parámetros utilizando los datos simulados del capítulo anterior, Sección 1.1. Esta información será utilizada más adelante para ejemplificar herramientas estadísticas de valoración de modelos.

Los parámetros estimados $\hat{\alpha}_{xy}$, $\hat{\beta}_{xy}$, $x, y \in \mathcal{H} = \{1, 2, 3\}$, corresponden a las entradas de las matrices

$$\hat{A} = \begin{bmatrix} 1.18 & 4.83 & 0.75 \\ 0.31 & 4.42 & 1.57 \\ 0.90 & 0.90 & 0.42 \end{bmatrix} \quad y \quad \hat{B} = \begin{bmatrix} 868.40 & 26765430000000 & 85.99 \\ 4.40 & 5867402000000 & 30318.98 \\ 191.89 & 1164.91 & 16.65 \end{bmatrix}.$$

Por otra parte, las probabilidades de transición estimadas $\hat{\theta}_{xy}$, $x, y \in \mathcal{H} \{1, 2, 3\}$, corresponden a las entradas de la matriz

$$\hat{\Theta} = \begin{bmatrix} 0.46 & 0.17 & 0.36 \\ 0.26 & 0.19 & 0.54 \\ 0.51 & 0.06 & 0.42 \end{bmatrix}.$$

En la siguiente sección se presentarán algunos métodos estadísticos para valorar el ajuste de modelos a datos. Además, algunas pruebas que ayuden a discernir, con base en los datos, posibles modelos planteados para el fenómeno bajo estudio.

3.2. Métodos de valuación de modelos

Las siguientes técnicas estadísticas son muy útiles para valorar al modelo que será utilizado para hacer las inferencias.

3.2.1. Métodos heurísticos

En esta sección se presentan métodos poco rigurosos pero sencillos, de fácil uso e interpretación para valorar modelos estadísticos.

Histogramas y densidades ajustadas

Una técnica sencilla para verificar si la elección de un modelo es adecuada consiste en superponer en la misma gráfica el histograma de los datos y la densidad estimada del modelo (el “mejor” candidato bajo los datos). Es importante mencionar que si el número de datos de la muestra es pequeño, por el ejemplo $n = 5$, el concepto de histograma debe de tomarse con mucha reserva, es decir, no se puede otorgar la misma significancia para $n = 100$ o $n = 500$. Se debe tener mucho cuidado con la

elección del número de clases o intervalos, porque si se consideran pocas, el histograma tendrá un aspecto demasiado burdo para ser informativo, y si se consideran muchas, entonces el histograma tendrá un aspecto demasiado rugoso como para recoger aspectos relevantes.

Nótese que el histograma se aproxima a la densidad de la variable aleatoria X , cualquiera que esta sea, sin hacer supuestos acerca de su naturaleza.

Supóngase que X tiene una densidad $f(x; \lambda)$ y $\hat{\lambda}$ es el emv de λ . Sea $\hat{f}(x; \lambda) = f(x; \hat{\lambda})$ la *densidad ajustada o estimada* con los datos. Si en efecto X tiene densidad $f(x; \lambda)$, entonces la densidad ajustada \hat{f} converge a f . Pero si en realidad, la densidad X es una, y para la estimación del parámetro vía el enfoque de verosimilitud se considero una alternativa diferente, entonces no necesariamente ocurrirá que la densidad ajustada se parezca a la densidad verdadera de X . Así, cuando la densidad estimada sea muy parecida a la forma del histograma, entonces esto se considera un ajuste adecuado del modelo a los datos puesto que el histograma se aproxima al modelo verdadero. En contraste, cuando la densidad estimada y la forma del histograma no presentan una semejanza razonablemente buena, se concluirá que el ajuste del modelo no es el adecuado para los datos.

En la Figura 3.1 se muestra la densidad y densidad ajustada al histograma de frecuencias relativa de acuerdo a los datos simulados del proceso semi-markoviano $Z(t)$ descrito en la Sección 1.2. Las gráficas de las transiciones $(1 \rightarrow 2)$, $(2 \rightarrow 1)$, $(2 \rightarrow 2)$, $(2 \rightarrow 3)$ y $(3 \rightarrow 2)$ fueron omitidas en la Figura 3.1 debido a los pequeños tamaños de muestra y a las diferencias aparentemente grandes entre los parámetros que se utilizaron para simular y sus respectivos parámetros estimados. Sin embargo, a pesar de estas discrepancias entre los parámetros y sus estimaciones, las densidades verdaderas y las estimadas son parecidas y se ajustan razonablemente bien al histograma, en particular cuando la muestra no es pequeña; véase Figura (3.2). Todos los parámetros fueron estimados mediante el método de máxima verosimilitud.

El gráfico cuantil-cuantil

En estadística, una *gráfica cuantil-cuantil* es un método visual que permite verificar si un modelo estadístico es adecuado para la muestra observada. En lo que sigue, supóngase que se propone que la muestra observada x_1, x_2, \dots, x_n bajo estudio proviene de una sucesión de v.a.i.i.d X_1, X_2, \dots, X_n con función de distribución $F(x; \lambda)$, de la cual se desconoce λ . El objetivo es valorar que efectivamente $F(x; \lambda)$ describe bien a los datos. La gráfica cuantil-cuantil se construye siguiendo los siguientes pasos:

1. Ordenar las observaciones de menor a mayor,

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Estas observaciones ordenadas son los llamados *cuantiles empíricos*.

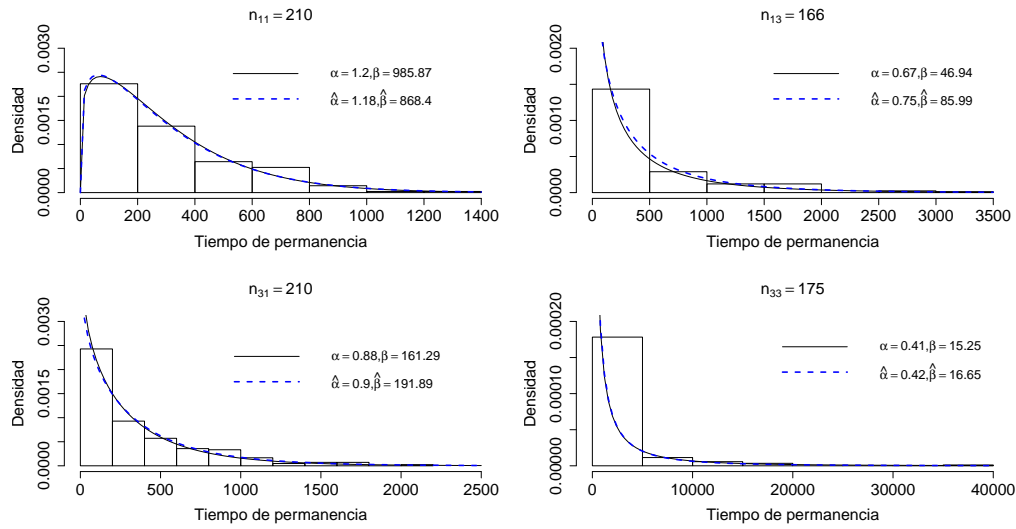


Figura 3.1: Histograma con la densidad y densidad ajustada del modelo Weibull.

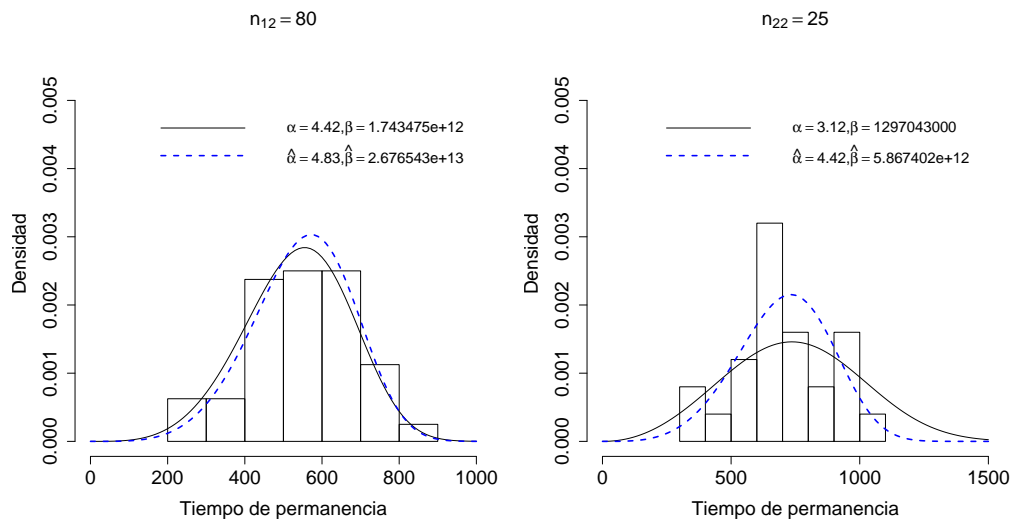


Figura 3.2: Histograma con la densidad y densidad ajustada del modelo Weibull.

- Determinar los valores

$$p_i = \frac{i - 0,5}{n}, \quad i = 1, 2, \dots, n.$$

Denotese por $Q_E(p)$ al cuantil empírico de orden p ($0 < p < 1$) de las observaciones. Calcule

$$x_i = Q_E(p_i), \quad i = 1, 2, \dots, n.$$

- Determinar los cuantiles de orden p_i , $i = 1, 2, \dots, n$, de la distribución teórica representada por la función de distribución $F(x; \lambda)$, es decir,

$$Q_T(p_i; \lambda) = F^{-1}(p_i; \lambda), \quad i = 1, 2, \dots, n.$$

Estos valores son llamados cuantiles teóricos. Nótese que el valor de $Q_T(p_i; \lambda)$ es desconocido puesto que el valor de λ no se conoce. Sin embargo, es posible aproximar $Q_T(p_i; \lambda)$ con

$$\hat{Q}_T(p_i) = Q_T(p_i; \hat{\lambda}) = F^{-1}(p_i; \hat{\lambda}), \quad i = 1, 2, \dots, n,$$

donde $\hat{\lambda}$ es un estimador puntual para λ . Se puede tomar a $\hat{\lambda}$ como el emv de λ .

- Representar en un plano cartesiano al conjunto de puntos $(\hat{Q}_T(p_i), Q_E(p_i))$, $i = 1, 2, \dots, n$ y la recta $y = x$. Si estos puntos están muy cerca y serpenteando alrededor de la recta, es evidencia de que el modelo es adecuado. En contraste, cualquier patrón que dé evidencia de que los puntos se alejan de la recta $y = x$, se interpretará como un mal ajuste del modelo a los datos.

La Figura 3.3 corresponde a la gráfica cuantil-cuantil bajo el modelo Weibull. Es decir, los parámetros del modelo fueron reemplazados por sus correspondientes emv. Se puede observar que los puntos serpentean alrededor de la recta $y = x$.

Nube cuantil-cuantil

Esta técnica gráfica muestra si existe evidencia para rechazar o no rechazar al modelo elegido para los datos. El proceso de construcción de este gráfico se describe a continuación:

- Primero elija un modelo adecuado para los datos y estimen sus parámetros. En esta etapa se pueden usar los emv.
- Use la densidad estimada para simular M muestras de tamaño n . Se recomienda tomar $M = 10,000$ y n igual al tamaño de la muestra observada.
- Estime los cuantiles teóricos con la muestra observada y calcule los empíricos para cada muestra simulada. Nótese que los cuantiles estimados son siempre los mismos para cada muestra simulada.

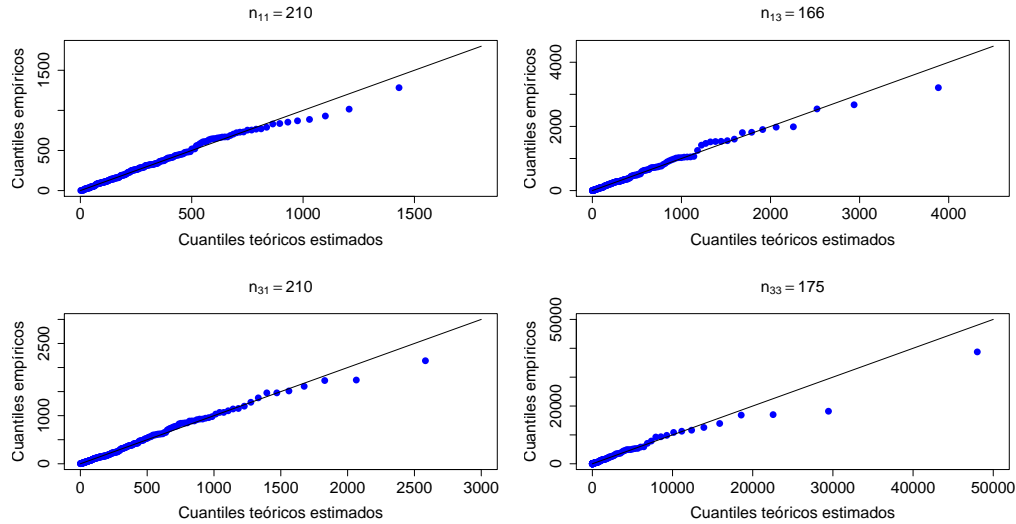


Figura 3.3: Gráfica cuantil-cuantil para el modelo Weibull.

4. Se coloca en una misma figura la gráfica cuantil-cuantil para cada muestra simulada y se superponen junto con la de la muestra observada. Cuando la nube de puntos captura completamente a la recta $y = x$, entonces se considera que no hay evidencia para rechazar el modelo supuesto para los datos.

La Figura 3.4 muestra la nube cuantil-cuantil para el modelo Weibull. Se puede observar la variabilidad de los cuantiles empíricos con respecto a los cuantiles teóricos determinados bajo el modelo Weibull. Se observa una alta variabilidad para cuantiles altos. Nótese que la recta $y = x$ se encuentra totalmente contenida en la nube de puntos; esto indica que el modelo evaluado puede ser adecuado para los datos.

Además de herramientas visuales para la valuación de modelos existen también herramientas teóricas. A continuación se presenta la prueba de hipótesis y la prueba de significancia p-valor.

3.2.2. Prueba de hipótesis

Una *prueba de hipótesis* es un examen de evidencia en forma de datos que da lugar a una de dos resoluciones posibles: rechazar H_0 a favor de H_1 , o no rechazar H_0 . Bajo este planteamiento, hay dos tipos de errores que se pueden cometer. El *error de Tipo I* se comete cuando se resuelve rechazar H_0 a favor de H_1 siendo que H_0 es “cierta”, y el *error de Tipo II* se comete cuando se resuelve no rechazar H_0 cuando H_0 es “falsa”. Usualmente, en la teoría estadística se consideran las hipótesis de tal forma que el *error de Tipo I* es más grave que el *error Tipo II*.

En general una prueba de hipótesis consta de dos ingredientes:

1. Una estadística $T = T(X_1, X_2, \dots, X_n)$ llamada la estadística de prueba.

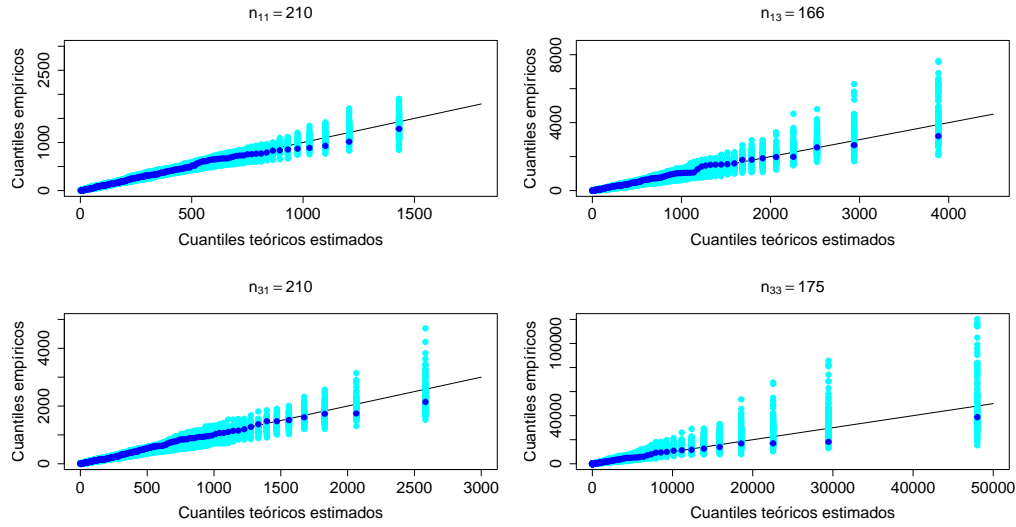


Figura 3.4: Nube cuantil-cuantil para el modelo Weibull.

2. Un subconjunto C de valores posibles de T llamado la región crítica de la prueba.

El criterio a utilizar consiste en rechazar H_0 si y sólo si $T \in C$. Cuando $T \in C$ se dice que la prueba es significativa y cuando $T \notin C$ se dice que la prueba es no significativa. La región crítica no depende de la muestra x_1, x_2, \dots, x_n , lo que quiere decir que aún antes de tomar la muestra, la región crítica ya ha sido especificada. Los datos intervienen para tomar o no la resolución de rechazar H_0 , lo cual se realiza con la región crítica, al comparar el valor de T con el conjunto C . Nótese que una prueba de hipótesis es de nivel α , con $0 \leq \alpha \leq 1$, si

$$\sup_{\lambda \in \Lambda} P_{\lambda}(T \in C) \leq \alpha.$$

Por tal razón, muchas veces se denota la región crítica C con el subíndice α para indicar que se trata de una prueba de nivel α .

En esta tesis se utiliza la estadística de prueba de la razón de verosimilitud definida como

$$D = -2 \log \left[\frac{L(1, \hat{\beta}, \hat{\theta}; x)}{L(\hat{\alpha}, \hat{\beta}, \hat{\theta}; x)} \right], \quad (3.14)$$

donde $L(1, \hat{\beta}, \hat{\theta}; x)$ es la verosimilitud basada en el proceso semi-markoviano con la distribución entre tiempos de saltos es exponencial y $L(\hat{\alpha}, \hat{\beta}, \hat{\theta}; x)$ es la verosimilitud basada en el proceso semi-markoviano con la distribución entre tiempos de saltos como Weibull. Esta estadística de prueba se distribuye como una χ^2 con grados de libertad igual a la diferencia de parámetros entre los modelos (Casella y Berger 1990). Así, la región crítica queda determinada por el cuantil $1 - \alpha$ de una distribución χ^2

con los grados de libertad señalados.

A continuación se ilustra esta prueba de hipótesis con el proceso semi-markoviano simulado en el Capítulo I. En este caso se tienen 9 grados de libertad y se elegirá una región crítica de nivel $\alpha = 0.05$. El valor de la estadística de prueba es

$$D_0 = -2 \log \left[\frac{L(1, \hat{\beta}, \hat{\theta}; x)}{L(\hat{\alpha}, \hat{\beta}, \hat{\theta}; x)} \right] = 714.481, \quad (3.15)$$

y la región crítica o área de rechazo de nivel $\alpha = 0.05$ es $C_\alpha = (16.92, \infty)$. Así, se rechaza la hipótesis nula H_0 . Es decir, se rechaza el modelo exponencial con una confianza del 95 %.

La teoría estadística sobre la que se basa el discurso para pruebas de hipótesis, recibe la denominación de teoría de Neyman-Pearson. Nótese que bajo este enfoque, dada una muestra observada y un nivel α el resultado de la prueba es binario, en el sentido de concluir “rechazar H_0 ” o “no rechazar H_0 ”. Esto puede ser criticable; por ejemplo, si la región crítica de una prueba de hipótesis fuera el conjunto $C_\alpha = (16.92, \infty)$ y datos en dos situaciones diferentes dieran lugar a valores de la estadística de prueba $D_1 = 17.45$ y $D_2 = 250.03$. La actitud de Neyman-Pearson diría simplemente, en ambos casos, “rechazar H_0 ”, siendo que es intuitivamente claro que ambas situaciones son diferentes en alguna cualidad. En el segundo caso, se rechaza con mayor fuerza que en el primero, y al decir sólo “rechazar H_0 ” no involucramos esta fuerza de la evidencia en contra de H_0 . Un enfoque que considera la evidencia en contra de una hipótesis es descrito a continuación.

3.2.3. Pruebas de significancia: *p-valor*

El concepto de *p-valor* tiene por objetivo cuantificar la fuerza con la que se rechaza una hipótesis nula H_0 . Se describe a través de una probabilidad. Tiene la interpretación de ser la probabilidad de haber observado un valor “más extremo” de una estadística de prueba $T(X)$ que ya se observó $t(x)$, o bien, la probabilidad de haber rechazado H_0 sólo por azar. La definición matemática del *p-valor* (Mood et al. 1985) es

$$p\text{-valor} = P(x) = \sup_{\lambda \in \Lambda} P_\lambda [T(X) \geq t(x)].$$

De esta forma, un *p-valor* “grande” denota que la evidencia en contra de H_0 es débil y un *p-valor* “chico” denota que los datos contienen mucha evidencia en contra de H_0 . En este sentido de *p-valores*, se puede no hablar de pruebas de hipótesis, sino de *pruebas de significancia*, donde la cuantificación del concepto abstracto de significancia es el *p-valor*.

En este trabajo de tesis para calcular el *p-valor* se considerará la estadística de prueba (3.14). Así, para nuestro caso de estudio simulado se tiene que la estadística

de prueba se distribuye como una χ^2 con 9 grados de libertad y el p-valor se puede calcular a través de la siguiente expresión:

$$p\text{-valor} = P[\chi_9^2 > D_0] = 1 - P[\chi_9^2 \leq D_0], \quad (3.16)$$

donde D_0 es el valor de la estadística (3.14) bajo los datos observados. Nótese que de acuerdo a (3.15) al rechazar H_0 se espera que el p – *valor* sea menor que $\alpha = 0.05$. Sin embargo, se procederá a calcularlo para fines de ejemplificar esta metodología.

Usando la estadística dada en (3.15) y la expresión (3.16), se obtiene el siguiente resultado

$$\begin{aligned} p\text{-valor} &= P[\chi_9^2 > D_0] \\ &= 1 - P[\chi_9^2 \leq 714.481] \\ &= 0.0000. \end{aligned} \quad (3.17)$$

Así, como se había previsto, el p-valor encontrado refleja evidencia contundente en contra del modelo exponencial.

A continuación se ilustrará el uso del modelo semi-markoviano para el modelado de terremotos, clasificados en tres tipos de intensidades, ocurridos en Turquía. Además se aplicarán los métodos de inferencia descritos en este capítulo para hacer estimaciones de cantidades relevantes e informativas.

3.3. Teoría bootstrap

Bootstrap es un método estadístico de remuestreo y de uso intensivo del computador, propuesto por Efron (1979, 1982), que se utiliza frecuentemente para hacer inferencia sobre una característica de interés de una función de distribución cuando no se conoce la distribución muestral del estimador de ésta característica. Por ejemplo, en esta tesis una característica de interés es la probabilidad de transición en el tiempo t , la cual depende de parámetros involucrados en el proceso semi-markoviano. En este caso, el emv de una probabilidad de transición al tiempo t no tiene una distribución conocida. En este capítulo se definirán conceptos relevantes de la teoría bootstrap que serán retomados en secciones posteriores para hacer inferencias sobre dicha probabilidad de transición; en particular, para construir intervalos de confianza para esta cantidad.

3.3.1. El método bootstrap

Supóngase que se tiene una muestra $x = (x_1, x_2, \dots, x_n)$ de variables aleatorias $X = (X_1, X_2, \dots, X_n)$ con función de distribución conjunta F y se desea hacer inferencias acerca de una característica de F , llámese $\sigma(F) \in \mathbb{R}^k$. Con esta muestra se puede obtener una estimación \hat{F}_n de F ya sea paramétrica o no paramétrica. Por una estimación paramétrica de F se entenderá una función de distribución que

es un miembro de una familia paramétrica de distribuciones y cuyos parámetros son estimados de alguna manera, por ejemplo, por máxima verosimilitud, usando la muestra que se observó. Se entenderá por una estimación no paramétrica de F a una función de distribución que aproxima a F y que se basa únicamente en la muestra con la que se cuenta; no es necesario estimar ningún parámetro para su construcción.

Nótese que $\sigma(\cdot)$ es un funcional cuyo dominio es la familia de distribuciones $\{G\}$, donde G es una función de distribución y cuyo contradominio es \mathbb{R}^k . Por tanto, se puede hablar de $\sigma(\hat{F}_n)$, la característica de interés pero correspondiente a la distribución \hat{F}_n . Dependiendo de si \hat{F}_n es una estimación paramétrica o no paramétrica de F , a $\sigma(\hat{F}_n)$ se le llama *estimación bootstrap paramétrica* o *estimación bootstrap no paramétrica* de la característica $\sigma(F)$, respectivamente.

No siempre es posible encontrar una expresión cerrada para $\sigma(\cdot)$ o ésta puede ser muy compleja. En esos casos se recurre a aproximar $\sigma(\hat{F}_n)$ mediante simulaciones. Antes de decir como se hace dicha aproximación, se definirá un nuevo concepto que facilita la explicación de los procedimientos utilizados en la teoría de bootstrap para hacer estimaciones.

Definición 7 (Muestra bootstrap). *Supóngase que se tiene una muestra $x = (x_1, x_2, \dots, x_n)$ de variables aleatorias $X = (X_1, \dots, X_n)$ con distribución conjunta F y \hat{F}_n es una estimación paramétrica o no paramétrica de F obtenida con esta muestra. Una muestra bootstrap de tamaño l es una muestra de l variables aleatorias con distribución \hat{F}_n .*

Comúnmente una letra mayúscula seguida de “*”, esto es X^* , denotará una variable aleatoria con distribución \hat{F}_n , y la minúscula correspondiente seguida de “*”, esto es x^* , una observación de esta variable aleatoria, a menos que se indique otra cosa. El símbolo “*” después de cualquier expresión que dependa de la distribución F (por ejemplo una esperanza, varianza, etc), será utilizado para indicar que dicha expresión es con respecto a \hat{F}_n en lugar de F .

Algoritmo bootstrap

Primero se piensa en un estimador fuertemente consistente de la estimación bootstrap $\sigma(\hat{F}_n)$ de la característica de interés (véase Apéndice A), llamémosle T . Posteriormente se hacen los siguientes pasos:

1. Simular M muestras bootstrap independientes de tamaño l :

$$\begin{aligned} x_1^* &= \{x_{(1,1)}^*, x_{(1,2)}^*, \dots, x_{(1,l)}^*\} \\ x_2^* &= \{x_{(2,1)}^*, x_{(2,2)}^*, \dots, x_{(2,l)}^*\} \\ &\vdots \\ x_M^* &= \{x_{(M,1)}^*, x_{(M,2)}^*, \dots, x_{(M,l)}^*\} \end{aligned}$$

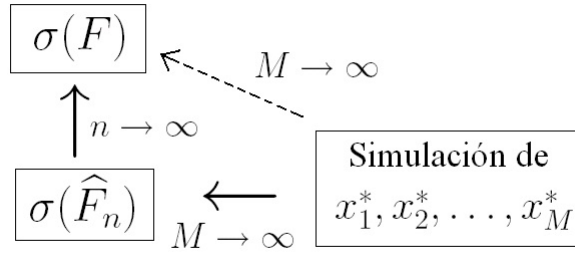


Figura 3.5: Estimación Bootstrap

2. Se calculan observaciones auxiliares m_i^* , para $i = 1, 2, \dots, M$, una por cada muestra bootstrap, que son necesarias para obtener una estimación basada en T de $\sigma(\hat{F}_n)$.
3. Estimar $\sigma(F)$ mediante

$$t_M = T(m_1^*, m_2^*, \dots, m_M^*).$$

En este algoritmo, M es un número muy grande y no se tiene una regla general para su elección pero típicamente se toma mayor o igual a 1,000, mayormente del orden de 10,000 en varias aplicaciones. El valor de l se toma dependiendo en la forma del estimador consistente de $\sigma(\hat{F}_n)$, algunas veces conviene tomarlo igual al tamaño de la muestra observada.

En caso en que se tenga una expresión de $\sigma(G)$ en función de G en la cual no sea muy difícil evaluar \hat{F}_n , es preferible estimar $\sigma(F)$ simplemente evaluando \hat{F}_n en $\sigma(\cdot)$ que utilizar el algoritmo bootstrap para obtener una aproximación a ella, ya que de esta forma se obtiene una estimación bootstrap $\sigma(\hat{F}_n)$ de la característica $\sigma(F)$ de F debido a la consistencia fuerte del estimador T , pero a t_M se le llamará también una *estimación bootstrap* de $\sigma(F)$, la cual será paramétrica o no paramétrica dependiendo de si \hat{F}_n es una estimación paramétrica o no paramétrica de F .

La metodología bootstrap es adecuada dependiendo de qué tan buena sea la aproximación de \hat{F}_n a F y qué tan suave sea el funcional $\sigma(\cdot)$ en F . En la Figura 3.5, la aproximación horizontal depende de la consistencia fuerte del estimador de $\sigma(\hat{F}_n)$ basado en las M muestras bootstrap independientes. La aproximación vertical depende de la continuidad del funcional $\sigma(\cdot)$ en F y de la aproximidad de \hat{F}_n a F en algún sentido cuando $n \rightarrow \infty$. Algunas veces se piensa que cuando $M \rightarrow \infty$ la estimación bootstrap se está acercando a $\sigma(F)$, la flecha punteada en la Figura 3.5 se refiere a esta creencia errónea. A lo que realmente se está acercando dicha estimación es a $\sigma(\hat{F}_n)$. Pero $\sigma(\hat{F}_n)$ podría no ser muy cercana a $\sigma(F)$ si $\sigma(\cdot)$ no es continua en F o si \hat{F}_n está lejos de F .

Intervalos de confianza Bootstrap

La técnica bootstrap, si bien permite encontrar una estimación puntual, su mayor utilidad se centra en generar intervalos de confianza. Existen varios métodos de intervalos de confianza bootstrap. En esta tesis se utiliza el método percentil, que se describe a continuación.

El método percentil es simple; luego de tener las estimaciones bootstrap,

$$m_1^*, m_2^*, \dots, m_M^*,$$

un intervalo de confianza tipo percentil para t_m del $(1 - \alpha)100\%$ está dado por los cuantiles $(Q_T(p_1), Q_T(p_2))$ de las M estimaciones bootstrap, con

$$p_1 = \alpha/2 \quad \text{y} \quad p_2 = 1 - \alpha/2.$$

Para ejemplificar el método percentil, a continuación se estimaran los intervalos bootstrap para una probabilidad de transición al tiempo t , considerando los datos simulados en la Sección 1.1 y los parámetros estimados en el Ejemplo 2.

Ejemplo 3. *Supóngase que se tienen datos de la distribución $Q(y, t | x)$. Una vez estimados los parámetros que caracterizan la distribución, en este caso por máxima verosimilitud, se procede a estimar la probabilidad de transición al tiempo t , calculada como*

$$Q_{xy} = \int_0^t f(s | x, y) ds,$$

donde Q_{xy} es la probabilidad de que el próximo evento sea del tipo $y \in \mathcal{H}$ y que éste se presente en un tiempo aleatorio bajo el supuesto que este fue del tipo $x \in \mathcal{H}$ y $f(s | x, y)$ es la densidad Weibull dada en la ecuación (2.7). Si $\hat{\theta}_{xy}$, $\hat{\alpha}_{xy}$ y $\hat{\beta}_{xy}$ son los estimadores basados en la muestra X_1, X_2, \dots, X_n , entonces el estimador de Q_{xy} será \hat{Q}_{xy} y se puede calcular de la ecuación

$$\hat{Q}_{xy} = \int_0^t f(s | x, y) ds = 1 - \exp \left\{ - \left(\frac{t^{\hat{\alpha}_{xy}}}{\hat{\beta}_{xy}} \right) \right\}.$$

Así, no es posible determinar que la distribución de este estimador sea miembro de una determinada familia paramétrica y poder obtener inferencias sobre Q_{xy} . Aquí es donde se recurre a la teoría bootstrap para obtener inferencias sobre esta cantidad de interés; en particular, intervalos de confianza sobre Q_{xy} . A continuación se describe el procedimiento para hacer estimación por intervalos sobre la probabilidad de transición Q_{31} ($3 \rightarrow 1$). Para hacer esto, tómese los parámetros estimados en el Ejemplo 2.

- (a) Primero se simulará un proceso semi-markoviano como en la Sección 1.1, utilizando los parámetros estimados en el Ejemplo 2, la cual será una muestra bootstrap de tamaño 500.

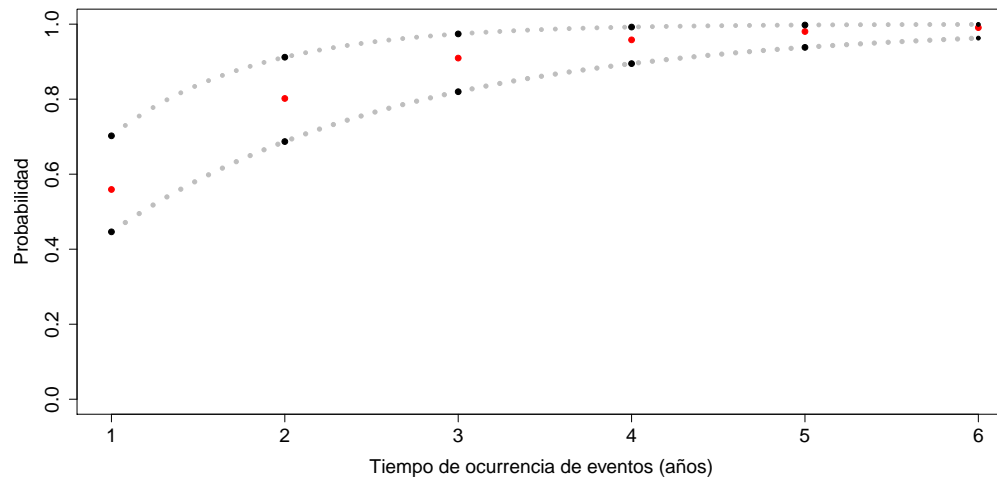


Figura 3.6: Intervalos bootstrap para Q_{31} .

- (b) Una vez simulada la muestra bootstrap, se estimaran los parámetros como en el Ejemplo 2 y será la muestra de parámetros bootstrap.
- (c) Con la muestra de parámetros bootstrap, se procede a calcular la probabilidad de interés.
- (d) Se repiten los pasos (a) – (c) 1000 veces y se procede a calcular el intervalo bootstrap para la probabilidad bootstrap acumulada con un $\alpha = 0.025$.

En la Tabla 3.1 se presentan los resultados de la simulación durante los próximos seis años; así como el intervalo bootstrap y el valor de la verdadera probabilidad de transición Q_{31} . La Figura 3.6 muestra los intervalos bootstrap de forma continua en el tiempo, donde se puede ver como cubre a la verdadera probabilidad de transición.

Tiempo	Q_{31}	límite inferior	límite superior
1	0.5592	0.4463	0.7023
2	0.8019	0.6870	0.9119
3	0.9095	0.8199	0.9740
4	0.9581	0.8948	0.9923
5	0.9803	0.9381	0.9977
6	0.9907	0.9628	0.9993

Cuadro 3.1: Estimaciones puntuales.

Capítulo 4

Estimación de parámetros del modelo Weibull y tamaño de muestra

En el capítulo anterior se mostró que la estimación de los parámetros del modelo, la matriz Θ de probabilidades de transición y los parámetros asociados a los modelos Weibull, se hace de manera independiente. Para datos observados del proceso, el parámetro Θ se estima a partir de las frecuencias de transición observadas (n_{xy}) y los parámetros asociados a los modelos Weibull se estiman con los tiempos de permanencia ($\tau_{xy1}, \tau_{xy2}, \dots, \tau_{xyn_{xy}}$). La estimación puntual de θ_{xy} se puede obtener fácilmente para todo tamaño de muestra $n_{xy} > 0$ con base en la fórmula matemática explícita dada en (3.10). En contraste, la estimación por máxima verosimilitud de los dos parámetros α_{xy} y β_{xy} de cada modelo Weibull requiere un proceso de optimización de la función de verosimilitud o log-verosimilitud correspondiente a este modelo, que depende del tamaño de muestra observado n_{xy} . Más aún, es importante enfatizar que las propiedades estadísticas de los estimadores de máxima verosimilitud $\hat{\alpha}_{xy}$ y $\hat{\beta}_{xy}$ también dependen de n_{xy} .

En este capítulo se hace una revisión histórica de la estimación por máxima verosimilitud de los parámetros de la distribución Weibull utilizada en el modelo semi-markoviano bajo estudio. Además, para el caso de una muestra pequeña, se identifican posibles dificultades computacionales y estadísticas con base en la forma de la función de verosimilitud Weibull. Por último, se hace un estudio donde se explora el comportamiento de la matriz de frecuencias de transición obtenida con base en escenarios del proceso semi-markoviano. Esto con el objetivo de mostrar el número esperado de transiciones y anticipar posibles dificultades de estimación asociadas a un tamaño de muestra pequeño.

4.1. Estimación de máxima verosimilitud de los parámetros del modelo Weibull

La distribución Weibull es utilizada para modelar una gran variedad de fenómenos aleatorios en diferentes campos de la ciencia: por ejemplo, en confiabilidad, biología, sismología, entre otros. En la literatura estadística, el problema de hacer inferencia sobre los parámetros del modelo Weibull, con base en una muestra (completa o con censura), ha sido discutido por Lehman (1962), Cohen (1965), Thoman et al. (1969), Rockette et al. (1974), entre otros. En particular, el problema de estimación

ha sido abordado desde el enfoque de máxima verosimilitud, método de momentos, mínimos cuadrados, estimación insesgada, etcétera. Además, se ha comparado la eficiencia de los estimadores obtenidos por estos métodos con base en el criterio del error cuadrático medio y el tamaño de muestra.

Cuando se tienen tamaños de muestra pequeños, el estimador de máxima verosimilitud suele tener un sesgo muy grande (con respecto al de otros estimadores); sin embargo, disminuye cuando el tamaño de muestra crece (Cohen, 1965; Thoman et al. 1969). En particular, Thoman et al. (1969) mostró una muy baja eficiencia del emv para muestras de tamaño $n = 5$, en comparación con el estimador obtenido por Menon (1963).

En la siguiente sección se exhibe el problema de estimación de los parámetros del modelo Weibull, cuando se tiene un tamaño de muestra pequeño; pero desde un punto de vista diferente al del sesgo. Se mostrará que la forma de la verosimilitud, para muestras pequeñas, puede conducir no solo a dificultades estadísticas (relacionadas con el sesgo) sino también a problemas computacionales de optimización.

4.2. Función de verosimilitud Weibull

Sean X_1, \dots, X_n variables aleatorias independientes idénticamente distribuidas con función de densidad Weibull:

$$f(x; \alpha, \beta) = \left(\frac{\alpha}{\beta}\right) x^{\alpha-1} \exp(-x^\alpha/\beta), \quad (4.1)$$

donde $x \geq 0$, $\alpha > 0$ y $\beta > 0$.

La función de verosimilitud (aproximación continua) de α y β para una muestra observada x_1, \dots, x_n es

$$L(\alpha, \beta; x_1, \dots, x_n) = \left(\frac{\alpha}{\beta}\right)^n \left(\prod_{i=1}^n x_i^{\alpha-1}\right) \exp\left(-\sum_{i=1}^n x_i^\alpha/\beta\right). \quad (4.2)$$

A continuación se aplicará el enfoque de verosimilitud, descrito en el Capítulo 2, para hacer inferencia sobre los parámetros α y β con base en un juego de datos simulados.

En la Figura 4.1 se muestra la gráfica de la función de verosimilitud relativa de los parámetros α y β de la distribución Weibull, con base en los $n = 5$ datos mostrados en la Tabla 4.1. Estos datos fueron simulados de una distribución Weibull con parámetros $\alpha = 0.95$ y $\beta = 1690.608$. Los valores de los parámetros fueron seleccionados de acuerdo a las entradas de las matrices de A y B de la Sección 1.1. Es importante recordar que estas matrices tienen en sus entradas estimaciones de los parámetros de un modelo Weibull, obtenidas con base en datos reales de terremotos; véase Alvarez (2005). Más aún, los valores seleccionados de $\alpha = 0.95$ y $\beta = 1690.608$ corresponden

a estimaciones hechas con un tamaño de muestra relativamente pequeño ($n = 2$) de tiempos de permanencia. En la Figura 4.1 se observa que la función de verosimilitud relativa de α y β crece rápidamente hasta alcanzar su máximo y luego decrece lentamente manteniéndose casi plana. Esta forma de la función de verosimilitud relativa puede ocasionar problemas de optimización. Además, si se quisiera dar intervalos de verosimilitud-confianza, estos fueran muy anchos o podría ocurrir que el límite superior fuera infinito. En la Figura 4.2 se muestran los contornos correspondientes a la función de verosimilitud relativa de α y β . Se marcó con un asterisco (*) la ubicación del emv de los parámetros α y β , $\hat{\alpha} = 0.85$ y $\hat{\beta} = 990.725$, y con un signo más (+) el valor de los parámetros con el que fueron simulados los datos.

Aquí se ha mostrado que tamaños de muestra pequeños pueden conducir a dificultades a la hora de querer estimar los parámetros de un modelo Weibull. Por esta razón, en el contexto de esta tesis, es deseable tener tamaños de muestra relativamente grandes o moderados para poder hacer inferencias razonables y útiles. Con el objetivo de explorar el comportamiento de la matriz de frecuencias del proceso semi-markoviano bajo estudio, en la siguiente sección se calcula el número esperado de transiciones (tamaño de muestra) que deben de ocurrir del estado x al estado y , para $x, y \in \mathcal{H}$.

1	6455.06
2	5245.19
3	2197.70
4	19.77
5	3887.93

Cuadro 4.1: Datos simulados de una distribución Weibull de parámetros $\alpha = 0.95$ y $\beta = 1690.608$.

4.3. Tamaño de muestra esperado

Las definiciones de los procesos estocásticos de teoría de cadenas de Markov que a continuación se presentan, determinan una manera para calcular teóricamente el número esperado de transiciones; es decir, las entradas esperadas de la matriz de frecuencias. Todo esto para determinar el tamaño de muestra esperado para poder tener estimaciones razonablemente buenas de los parámetros.

Sea $1_{[X_m=x, X_{m+1}=y]}$ la función indicadora definida por

$$1_{[X_m=x, X_{m+1}=y]} = \begin{cases} 1, & \text{si } X_{m+1} = y, \\ 0, & \text{si } X_{m+1} \neq y; \end{cases}$$

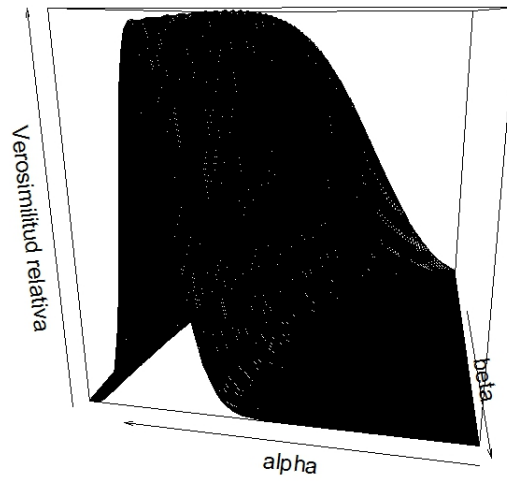


Figura 4.1: Función de verosimilitud relativa de α y β para los datos simulados de la Tabla 4.1.

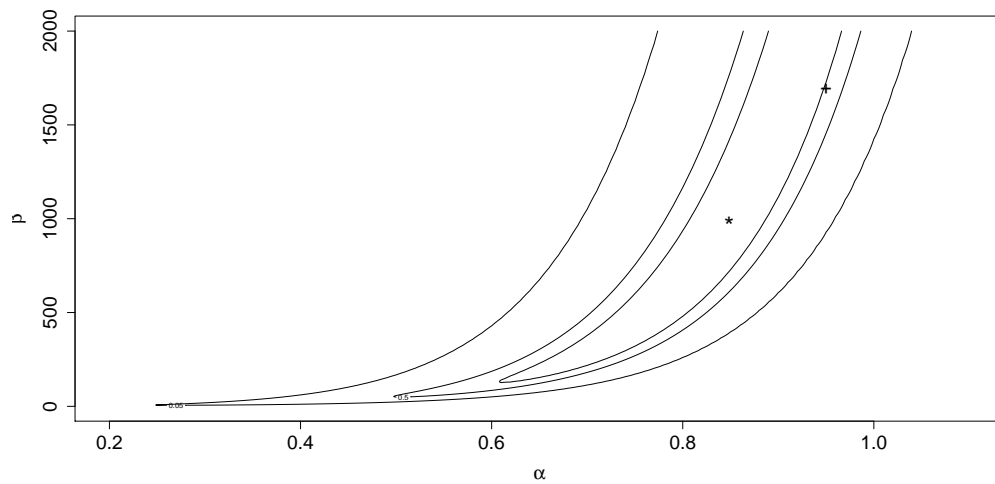


Figura 4.2: Contornos de la función de verosimilitud relativa de α y β para los datos simulados de la Tabla 4.1.

esto es, $1_{[X_m=x, X_{m+1}=y]} = 1$ si la cadena está en el estado y en un salto desde el estado x , y $1_{[X_m=x, X_{m+1}=y]} = 0$ en otro caso. Entonces

$$n_{xy} = \sum_{m=0}^n 1_{[X_m=x, X_{m+1}=y]}, \quad (4.3)$$

donde n es el número total de transiciones.

Por otro lado, sea $E_z(\cdot)$ el valor esperado de una variable aleatoria en terminos de una cadena de Markov que inicia en el estado z . Por ejemplo,

$$E_z(1_{[X_m=x, X_{m+1}=y]}) = P_z(X_m = x, X_{m+1} = y). \quad (4.4)$$

Así, el número esperado de visitas del estado x al estado y , de acuerdo a la Ecuación (4.3) y (4.4), es

$$\begin{aligned} E_z(n_{xy}) &= E_z\left(\sum_{m=0}^n 1_{[X_m=x, X_{m+1}=y]}\right) \\ &= \sum_{m=0}^n E_z(1_{[X_m=x, X_{m+1}=y]}) \\ &= \sum_{m=0}^n P_z(X_m = x, X_{m+1} = y). \end{aligned} \quad (4.5)$$

Ahora, por probabilidad condicional se tiene lo siguiente:

$$\begin{aligned} E_z(n_{xy}) &= \sum_{m=0}^n P(X_{m+1} = y | X_m = x) P_z(X_m = x) \\ &= \sum_{m=0}^n \theta_{xy} P_z(X_m = x) \\ &= \theta_{xy} \sum_{m=0}^n P^m(z, x). \end{aligned} \quad (4.6)$$

Por último, si X_m , $m \geq 0$, es una cadena de Markov irreducible y recurrente positiva con distribución estacionaria $\mu(\cdot)$, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^n P^m(x, y) = \mu(y). \quad (4.7)$$

Por lo tanto, si n es suficientemente grande entonces el número esperado de visitas del estado x al estado y es aproximadamente $n\theta_{xy}\mu(x)$,

$$E_z(n_{xy}) \approx n\theta_{xy}\mu(x). \quad (4.8)$$

Para más detalles véase Hoel, Port y Stone (1972, pág. 72).

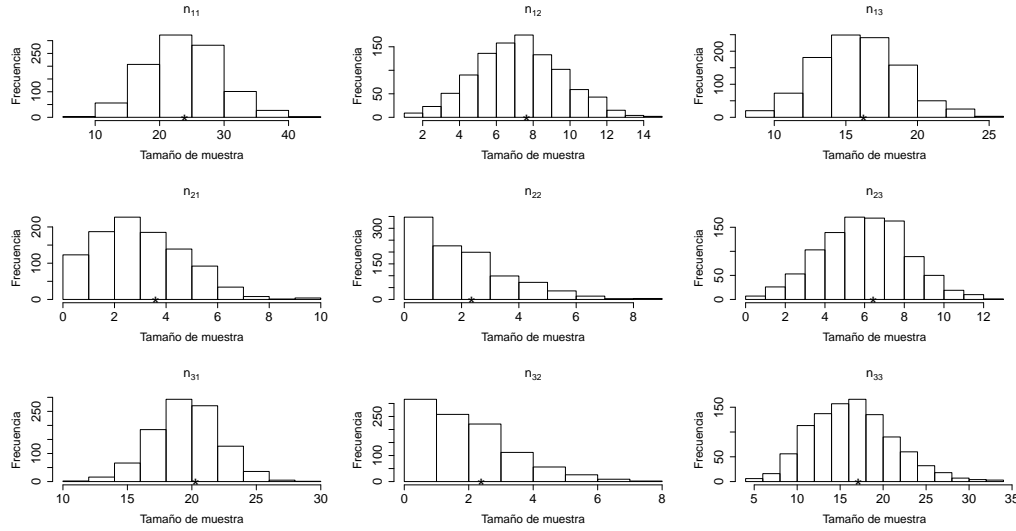


Figura 4.3: Histogramas para el tamaño de muestra para un proceso semi-markoviano de 100 transiciones.

En la Tabla 4.2 se muestra el número esperado de visitas para cada entrada de la matriz de frecuencias, calculado de acuerdo a la Ecuación (4.8), para un proceso semi-markoviano con matriz de transición dada en la Sección 1.1 del Capítulo 1. En la primera fila de ésta tabla se presentan los resultados obtenidos para un escenario de 100 transiciones; mientras que en la segunda fila se consideraron 500 transiciones. Adicionalmente, se hizo una simulación que consistió en realizar 1000 repeticiones del proceso semi-markoviano (con parámetros dados en la Sección 1.1 del Capítulo 1) y se contó el número de veces que el proceso saltó del estado x al estado y en una transición, para $x, y \in \mathcal{H}$. Las Figuras 4.3 y 4.4 muestran los histogramas para cada entrada de la matriz de frecuencia de acuerdo al estudio de simulación. En cada histograma se marcó con un asterisco (*) el valor teórico obtenido de acuerdo a la Ecuación (4.8). Los resultados muestran que 100 transiciones del proceso bajo estudio producen tamaños de muestra relativamente pequeños que pueden ser inapropiados para hacer inferencia sobre los parámetros del modelo. En contraste, 500 transiciones del proceso producen un número esperado de transiciones que parecen ser mucho más adecuados para hacer inferencia estadística.

n	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	n_{31}	n_{32}	n_{33}
100	24	8	16	3	2	6	20	2	17
500	119	38	81	18	12	32	101	12	85

Cuadro 4.2: Entradas esperadas de la matriz de frecuencia de un proceso semi-markoviano.

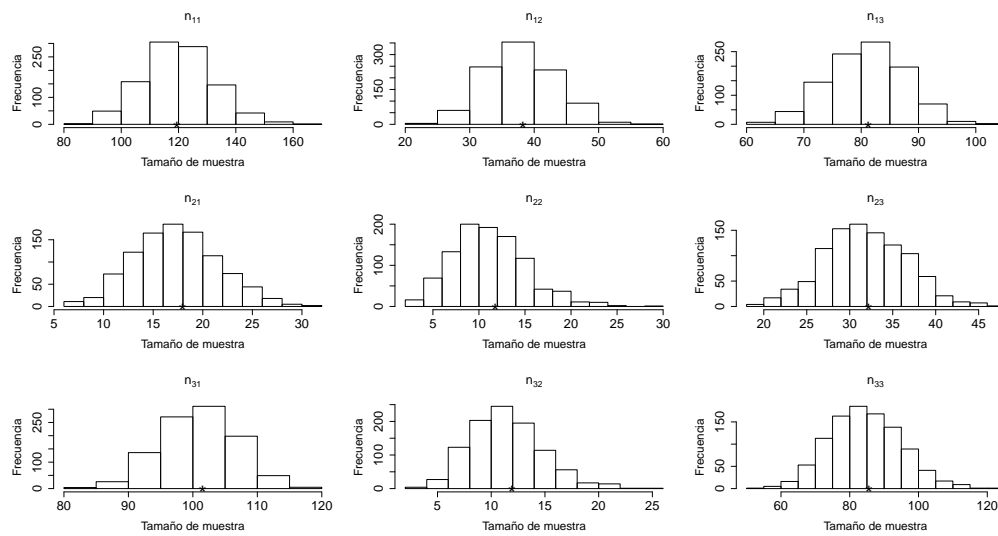


Figura 4.4: Histogramas para el tamaño de muestra para un proceso semi-markoviano de 500 transiciones.

Capítulo 5

Inferencias en un proceso semi-markoviano Weibull: Datos de terremotos

En este capítulo se hará inferencia estadística sobre los parámetros del proceso semi-markoviano, descrito en el Capítulo 1, con base en datos reales de terremotos de gran magnitud que ocurrieron en el siglo XX y se localizaron en el llamado “Norte de Anatolia Zona de Falla” en Turquía. En particular, se hará estimación puntual de los parámetros y se usarán los métodos de valuación de modelos, descritos en en Capítulo 2, para valorar el ajuste del modelo semi-markoviano a los datos. Además, se calcularán estimadores puntuales e intervalos de confianza bootstrap para una probabilidad de transición (Q_{31}), considerando diferentes tiempos. Cabe mencionar que los datos que serán utilizados aquí fueron presentados y analizados por Álvarez (2005). El autor también utiliza un modelo semi-markoviano con distribución Weibull para los tiempos entre eventos; sólo que considera datos censurados en el proceso de inferencia. También es importante mencionar que Álvarez (2005) presenta estimaciones puntuales para Q_{31} ; pero no calcula intervalos de confianza para este parámetro de interés.

Los datos considerados para este caso de estudio son: La frecuencia de ocurrencia de tres tipos o categorías de terremotos y los tiempos de permanencia en cada categoría. Los terremotos fueron clasificados de acuerdo a su magnitud en la escala de Richter usando la Tabla 5.1. En la Figura 5.1 se muestra una gráfica descriptiva del comportamiento de estos datos. La matriz de frecuencias esta dada por

$$N = \begin{bmatrix} 18 & 6 & 12 \\ 3 & 2 & 5 \\ 15 & 2 & 12 \end{bmatrix}.$$

Tipo	Categoría
Tipo 1	[5.5, 6]
Tipo 2	(6, 6.5]
Tipo 3	> 6.5

Cuadro 5.1: Categorización de la escala de Richter.

Con base en la matriz de frecuencias observadas, se tiene que el estimador de máxima verosimilitud de la matriz de probabilidades de transición es

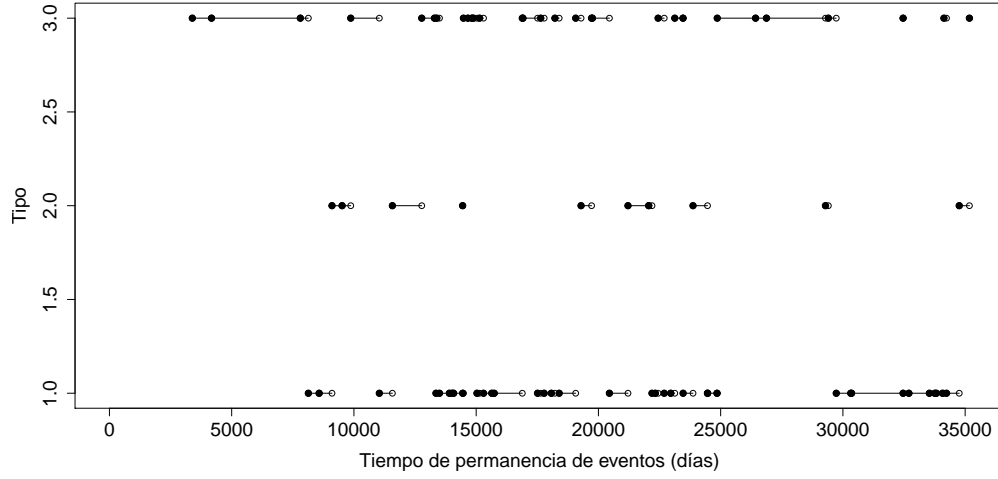


Figura 5.1: Tiempos de permanencia de tres tipos de terremotos, clasificados de acuerdo a su magnitud en la escala de Richter usando la Tabla 5.1, que ocurrieron en el siglo XX y se localizaron en el “Norte de Anatolia Zona de Falla” en Turquía.

$$\hat{\Theta} = \begin{bmatrix} 0.50 & 0.17 & 0.33 \\ 0.30 & 0.20 & 0.50 \\ 0.52 & 0.07 & 0.41 \end{bmatrix}.$$

Ahora, antes de realizar las estimaciones de los parámetros α_{xy} y β_{xy} de cada modelo Weibull, es importante observar que la frecuencia con la que ocurrieron las transiciones $(1 \rightarrow 2)$, $(2 \rightarrow 1)$, $(2 \rightarrow 2)$, $(2 \rightarrow 3)$ y $(3 \rightarrow 2)$ es muy pequeña. Como se mostró en la Sección 3.3; un tamaño de muestra pequeño puede contener poca información sobre los parámetros desconocidos del modelo y esto puede conducir a una serie de problemas estadísticos y computacionales. En vista de todo esto y en la espera de inferencias razonables, de aquí en adelante sólo se estimarán modelos Weibull para las transiciones $(1 \rightarrow 1)$, $(1 \rightarrow 3)$, $(3 \rightarrow 1)$ y $(3 \rightarrow 3)$. Los resultados de las estimaciones puntuales por máxima verosimilitud para α_{xy} y β_{xy} son:

$$\hat{A} = \begin{bmatrix} 1.18 & - & 0.68 \\ - & - & - \\ 0.85 & - & 0.70 \end{bmatrix} \quad \text{y} \quad \hat{B} = \begin{bmatrix} 774.33 & - & 46.20 \\ - & - & - \\ 115.17 & - & 101.67 \end{bmatrix}.$$

La Figura 5.2 muestra la densidad ajustada Weibull al histograma de frecuencia relativa de los datos. Se observa que el modelo se ajusta razonablemente bien al histograma. Las Figuras 5.3 y 5.4 corresponden a la gráfica cuantil-cuantil y nube cuantil-cuantil bajo el mejor modelo Weibull, respectivamente. En la Figura 5.3 se observa que los puntos serpentean alrededor de la recta $y = x$; es decir, se muestra un ajuste razonablemente bueno del modelo a los datos. En la Figura 5.4 se observa que la recta $y = x$ se encuentra totalmente contenida en la nube de puntos; es decir,

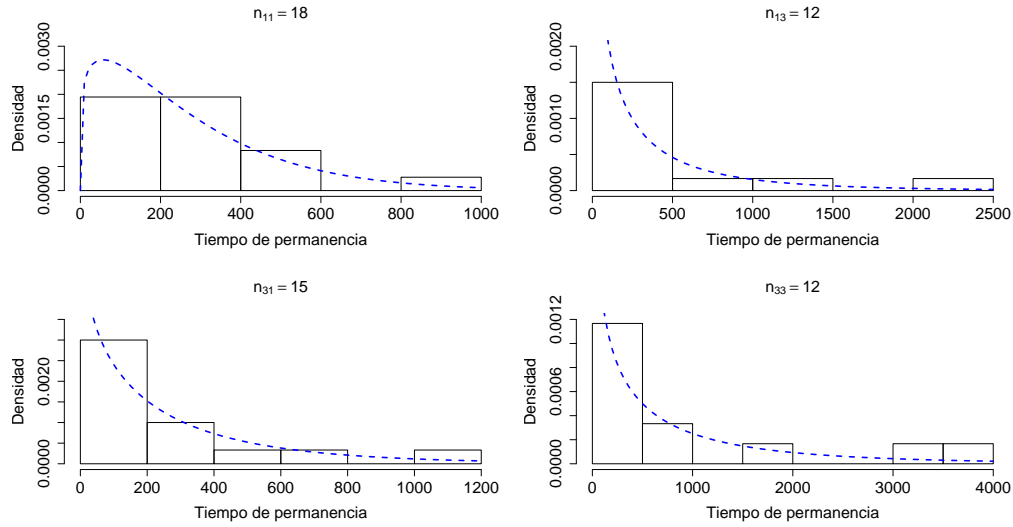


Figura 5.2: Histograma con la densidad ajustada del modelo Weibull: Datos terremotos.

esto indica que el modelo puede ser adecuado para los datos.

A continuación se hizo estimación por intervalos bootstrap sobre la probabilidad de transición Q_{31} ($3 \rightarrow 1$) con los datos reales, ya que es la cantidad de interés; el procedimiento y el valor de la probabilidad de cobertura de los intervalos (95 %) son los mismos que se utilizaron en el Ejemplo 3 de la Subsección 2.3.1. En la Tabla 5.2 se presentan los resultados de la estimación para los próximos 6 años. La Figura 5.5 muestra esta información de forma gráfica, junto con otros intervalos bootstrap para Q_{31} , del mismo nivel de confianza, calculados para una partición más fina del intervalo de tiempo.

Tiempo	Q_{31}	límite inferior	límite superior
1	0.5592	0.4660	0.7566
2	0.8019	0.7063	0.9410
3	0.9095	0.8348	0.9856
4	0.9581	0.9054	0.9965
5	0.9803	0.9448	0.9991
6	0.9907	0.9672	0.9998

Cuadro 5.2: Estimaciones puntuales.

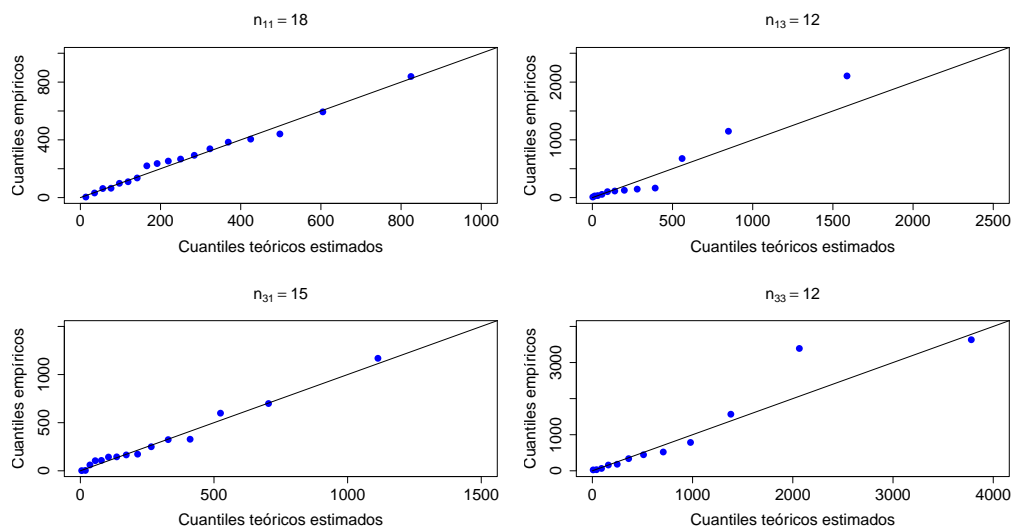


Figura 5.3: Gráfica cuantil-cuantil para el modelo Weibull: Datos terremotos.

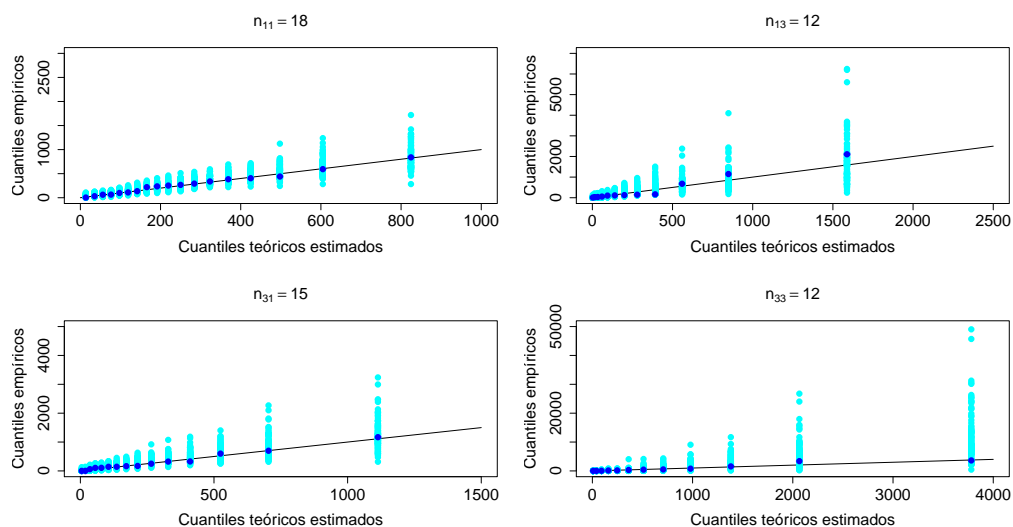


Figura 5.4: Nube cuantil-cuantil para el modelo Weibull: Datos terremotos.

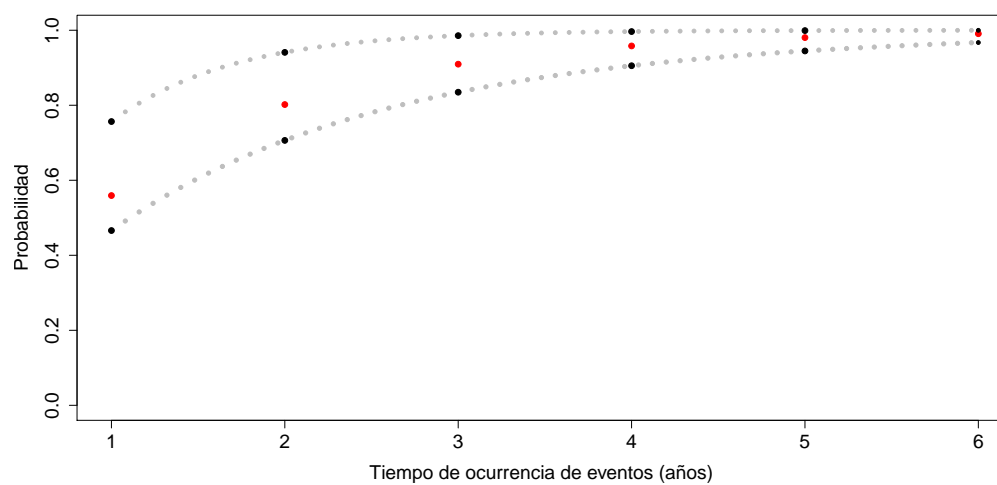


Figura 5.5: Intervalos bootstrap para Q_{31} .

Capítulo 6

Conclusiones

En este trabajo de tesis se abordó el problema de hacer inferencia estadística en procesos semi-markovianos con tiempo de espera Weibull. En particular, se mostró el uso del método de verosimilitud para hacer estimación puntual sobre los parámetros involucrados en este modelo, la matriz de transición y los parámetros asociados a las distribuciones Weibull, y sobre cantidades de interés que son funciones de los parámetros del modelo, probabilidades de transición en el tiempo. Además, se usó el método bootstrap para hacer estimación por intervalos sobre dichas probabilidades de transición. Estas metodologías estadísticas, junto con diferentes herramientas estadísticas de valoración de modelos, fueron ilustradas tanto con datos simulados como con datos reales de ocurrencias de tipos de terremotos.

Para el caso de estimación de parámetros de las distribuciones Weibull, consideradas en el modelo semi-markoviano bajo estudio, se mostró que la función de verosimilitud puede tener una forma plana para muestras pequeñas. Este comportamiento anómalo de la verosimilitud en presencia de muestras pequeñas resulta ser muy informativo y advierte sobre posibles problemas estadísticos en la estimación y sobre dificultades computacionales a la hora de querer encontrar los estimadores. Cabe mencionar que las herramientas presentadas en el Capítulo 3 permiten explorar el comportamiento del número esperado de transiciones de un estado a otro (tamaños de muestra) para escenarios del proceso semi-markoviano Weibull. Esto puede ayudar a anticipar posibles dificultades asociadas a un tamaño de muestra pequeño.

La definición de los estados involucrados en los procesos semi-markovianos Weibull puede afectar drásticamente los resultados de la inferencia estadística. Por ejemplo, en el caso de estudio de terremotos desarrollado en el Capítulo 4, la frecuencia (tamaño de muestra) para las transiciones $(1 \rightarrow 2)$, $(2 \rightarrow 1)$, $(2 \rightarrow 2)$, $(2 \rightarrow 3)$ y $(3 \rightarrow 2)$ fue pequeña y conduce a dificultades de estimación. Nótese que se trata de transiciones que llegan o salen del estado 2, categoría de terremoto intermedia.

Capítulo 7

Apéndice A

El siguiente concepto es muy relevante, ya que juega un papel importante en la teoría del bootstrap para garantizar que de alguna manera las estimaciones bootstrap converjan al parámetro o característica de una distribución y por tanto esta metodología de estimación sea de interés, es el siguiente

Definición 8. *Un estimador $T_n = T_n(X_1, X_2, \dots, X_n)$ de λ es consistente si la sucesión $\{T_n, n \geq 1\}$ converge a λ en algún sentido apropiado como los siguientes:*

- *T_n es un estimador fuertemente consistente de λ si $\{T_n, n \geq 1\}$ converge casi seguramente (c.s.) a λ cuando $n \rightarrow \infty$, es decir, si $\{T_n, n \geq 1\}$ converge a λ excepto en un conjunto de medida cero.*
- *T_n es un estimador débilmente consistente de λ si $\{T_n, n \geq 1\}$ converge en probabilidad a λ cuando $n \rightarrow \infty$, es decir, si*

$$\lim_{n \rightarrow \infty} P(|T_n - \lambda| > \epsilon) = 0 \quad \forall \epsilon > 0.$$

Si con alguna forma de convergencia γ se tiene que

$$\hat{F}_n \xrightarrow{\gamma} F, \text{ cuando } n \rightarrow \infty,$$

y $\sigma(\cdot)$ es continua en F , en el sentido de que para toda sucesión de distribuciones $\{G_n, n \geq 1\}$ que tiende con la convergencia γ a F cuando $n \rightarrow \infty$, se cumple que $\{\sigma(G_n), n \geq 1\}$ tiende con alguna forma de convergencia ω a $\sigma(F)$ cuando $n \rightarrow \infty$, entonces se tendrá que la convergencia ω

$$\sigma(\hat{F}_n) \xrightarrow{\omega} \sigma(F), \text{ cuando } n \rightarrow \infty. \quad (7.1)$$

En el caso en que \hat{F}_n sea la distribución empírica, el Teorema de Glivenko-Cantelli, cuya demostración se puede consultar en Resnick (2001, p.224), establece que

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{\text{c.s.}} 0, \text{ cuando } n \rightarrow \infty.$$

Por lo tanto, \hat{F}_n converge uniformemente en \mathbb{R} a F cuando $n \rightarrow \infty$, con probabilidad 1. Sin embargo en la práctica muchas veces no se tiene forma de verificar si el funcional $\sigma(\cdot)$ es efectivamente continuo para garantizar que (7.1) sea cierto.

Si se satisfacen las suposiciones anteriores de continuidad del funcional $\sigma(\cdot)$ junto con la consistencia fuerte del estimador de $\sigma(\hat{F})$, se puede tener la tranquilidad de que el método bootstrap proporcionará estimaciones adecuadas.

Bibliografía

- [1] ÁLVAREZ, E. (2005). *Estimation in Stationary Markov Renewal Processes, with Application to Earthquake Forecasting in Turkey*. Springer Science & Business Media.
- [2] BILLINGSLEY, P. (1979). *Probability and Measure*. New York: Jhon Wiley & Sons.
- [3] CASELLA, G. y BERGER, R.L. (2002). *Statistical Inference*. California: Duxbury.
- [4] COHEN, A.C. (1965). *Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples*. Technometrics, Vol. 7, No. 4, pp. 579-588.
- [5] EFRON, B. y TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- [6] HOEL, P.G., PORT, S.C. y STONE, C.J. (1971). *Introduction to Probability Theory*. Boston: Houghton Mifflin Company.
- [7] HOEL, P.G., PORT, S.C. y STONE, C.J. (1972). *Introduction to Stochastic Processes*. Boston: Houghton Mifflin Company.
- [8] LEHMAN, E.H. (1962). *Shapes, Moments and Estimators of the Weibull Distribution*. Transactions of IEEE-Reliability 11, 32-38.
- [9] MONTROYA, J.A. (2008). *La Verosimilitud Perfil en la Inferencia Estadística*. Tesis de Doctorado en Ciencias con Especialidad en Probabilidad y Estadística, CIMAT, Guanajuato, Gto., México.
- [10] ROCKETTE, H., ANTLE, C.E. y KLIMKO, L.A. (1974). *Maximum Likelihood Estimation with the Weibull Model*. Journal of the American Statistical Association, Vol. 69, No. 345, pp. 246-249.
- [11] ROSS, S.M. (1996). *Stochastic Processes*. New York: Jhon Wiley & Sons.
- [12] SERFLING, R. (1980). *Aproximation Theorems of Mathematical Statistics*. New York: Jhon Wiley & Sons.
- [13] SPROTT, D.A. (2000). *Statistical Inference in Science*. New York: Springer-Verlag.
- [14] THOMAN, D.R., BAIN, L.J. y ANTLE, C.E. (1969). *Inferences on the Parameters of the Weibull Distribution*. Technometrics, Vol. 11, No. 3, pp. 445-460.