



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DE SUPERVIVENCIA PARA
DIABETES EN LA POBLACIÓN
MEXICANA,
UN ENFOQUE BAYESIANO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

OSVALDO ESPIN GARCIA

DIRECTOR DE TESIS:

DRA. RUTH SELENE FUENTES GARCIA



2009



Universidad Nacional
Autónoma de México



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.

Hoja de datos del jurado

1. Nombre del alumno

Espin
García
Osvaldo
22 07 65 23
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
301811974

2. Datos del tutor

Dra.
Ruth Selene
Fuentes
García

3. Datos del sinodal 1

Dr.
Raúl
Rueda
Díaz del Campo

4. Datos del sinodal 2

Mat.
Margarita Elvira
Chávez
Cano

5. Datos del sinodal 3

Act.
Jaime
Vázquez
Alamilla

6. Datos del sinodal 4

Act.
Carlos Vladimir
Rodríguez
Caballero

7. Datos del trabajo escrito

Análisis de supervivencia para diabetes en la población mexicana
Un enfoque bayesiano
72 p
2009

Agradecimientos

A Dios, por sobretodas las cosas.

A mi madre, por su siempre incondicional apoyo y amor.

A mi padre, por ser tal y como es.

A mi hermano, por ser mi gran amigo, que siempre estará ahí.

A mi tío Mario por siempre estar al pendiente y motivarme a seguir adelante en todo momento.

A mi familia, por creer en mí.

A mis amigos de la facultad, ustedes saben quienes son, por haber estado y seguir estando en mi vida, por las horas de estudio y de relaxo, por hacer inolvidable mi paso por la universidad.

A todos mis amigos, por confiar en mí, por compartir momentos e ideas, ser ustedes mismos y ser una parte muy importante en mi vida.

A todos aquellos que tuvieron, tienen y tendrán un lugar de mi vida, en los buenos, no tan buenos y malos momentos, por las enseñanzas y lecciones aprendidas, por ayudar a ser la persona que soy.

Un muy especial agradecimiento al Dr. Sergio López que sin su valioso apoyo hubiese sido muy difícil llevar a cabo este trabajo.

A la Dra. Ruth Fuentes Garcia, por su perseverancia y excelente dirección de este trabajo.

A mis sinodales, por sus valiosa aportación para complementar este trabajo.

A la Universidad, mi alma mater, por su grandeza, por sus profesores, por su gente y por brindarme la oportunidad de pertenecer a ella.

Índice general

Introducción	v
1. Un acercamiento al análisis de supervivencia	1
1.1. Definiciones	2
1.1.1. La función de supervivencia	3
1.1.2. La función de riesgo	3
1.2. Relaciones y equivalencias	4
1.3. Censura de datos	4
1.4. Modelos de supervivencia	7
1.4.1. Paramétricos	8
1.4.2. No paramétricos	11
1.4.3. Semi paramétricos	13
2. La aproximación bayesiana	15
2.1. El paradigma bayesiano	15
2.2. El principio de verosimilitud	17
2.3. La función <i>a priori</i>	19
2.3.1. Obtención de una <i>a priori</i>	19
2.3.2. Familias conjugadas	20
2.3.3. Informativas y “No informativas”	21
2.4. Cálculo de estimadores	22
2.4.1. MCMC	22

3. Un modelo de supervivencia con inferencia bayesiana para diabetes	25
3.1. El problema inicial	25
3.2. Primeras aproximaciones	28
3.3. Aproximación bajo un esquema bayesiano	31
3.4. Resultados	37
Apéndices	42
A. Desarrollo de resultados	43
B. Código del muestreo de Gibbs	45
C. Resultados de diagnóstico	51
Conclusiones	69
Referencias	71

Introducción

Pronosticar o dar aproximaciones a futuros eventos ha sido siempre una práctica frecuente para los seres humanos. En tiempos remotos estos pronósticos se realizaban mediante métodos “mágicos” o poco ortodoxos. Con el paso del tiempo y gracias a los avances teóricos y tecnológicos de la ciencia estas aproximaciones han ido cambiando hasta llegar a metodologías rigurosamente científicas y bien fundamentadas teóricamente.

Hoy en día una herramienta fundamental para el pronóstico y aproximación de eventos que aún no ocurren es la estadística, ya que esta se sirve de la información previamente colectada para proyectar tendencias y posibles comportamientos futuros. La estadística al considerarse una ciencia que da una interpretación mas que una explicación de los fenómenos naturales, puede tener diversos enfoques y asepciones por lo cual lo más adecuado es encontrar aquel enfoque que cubra las necesidades específicas de cada problema planteado.

Dentro de la estadística se encuentra el análisis de supervivencia, el cual surge como una herramienta para medir la probabilidad de ocurrencia en un cierto evento, esto mediante estimaciones de su distribución de probabilidad. Para tal fin existen principalmente dos aproximaciones estadísticas, la forma paramétrica, la cual busca darle a la distribución una forma funcional conocida y posteriormente encontrar los parámetros que ajusten a dicha distribución y la no paramétrica, en la cual no interesa la forma de la distribución sino que la aproximación dada se encuentre cercana a esta.

La aproximación que se presentará en esta tesis conjuga ambos enfoques, debido a las características propias del problema que se desea abordar y a su vez para evitar supuestos

que podrían limitar los alcances del análisis a realizar.

Un problema que se presenta dentro del análisis de supervivencia es el de encontrar los parámetros que ajusten al modelo paramétrico propuesto, para tal fin la estadística bayesiana constituye una herramienta eficaz para encontrar dichos parámetros utilizando, además de la información contenida en la muestra, información previa que puede considerarse en los modelos.

Dentro del siguiente trabajo se presenta un enfoque bayesiano del análisis de supervivencia el cual se ha venido desarrollando a últimas fechas gracias a los avances en materia de cómputo y simulación principalmente. La ventaja al utilizar inferencia estadística bayesiana radica principalmente en que se puede incorporar de manera natural al modelo información “extra” que ayuda al modelo siempre y cuando represente de manera adecuada los fenómenos observados.

Asimismo, en el presente trabajo se plantea una aplicación del análisis de supervivencia con inferencia bayesiana en el campo de la medicina, específicamente en la epidemiología realizando un análisis de la mortalidad de la población mexicana debido a la diabetes; la cual se ha convertido en un grave problema de salud pública y uno de los objetivos del presente trabajo es proveer de una herramienta para analizarlo desde otro ángulo.

El trabajo se divide de la siguiente forma:

En el capítulo 1 se presentan nociones básicas y definiciones del análisis de supervivencia, así como algunos de los modelos más utilizados en la metodología “clásica”.

En el capítulo 2 se describe en lo que consiste la aproximación bayesiana, conceptos básicos del esquema bayesiano y resultados que serán de utilidad durante el desarrollo del presente trabajo.

El capítulo 3 presenta el problema que se aborda en esta tesis, se dan algunas aproximaciones a la solución del mismo, mediante métodos “clásicos” para posteriormente implementar un modelo de supervivencia haciendo uso de herramientas bayesianas presentado a su vez los resultados, gráficas y demás análisis referentes a la implementación del modelo propuesto. Para finalizar se presentan los apéndices correspondientes y las conclusiones alcanzadas.

Un acercamiento al análisis de supervivencia

El análisis de supervivencia es una rama de la estadística que se especializa en estudiar tiempos de ocurrencia de un cierto evento (plenamente identificado), esto se realiza mediante el análisis de variables aleatorias que miden el tiempo transcurrido desde un punto inicial hasta el “evento” a estudiar.

En muchos casos este enfoque se plantea como un análisis de los tiempos de transición de un estado a otro, como lo es, por ejemplo, el paso de la vida (estado 1) a la muerte (estado 2); al realizar estas consideraciones lo importante es decidir en que se va a enfocar el estudio, ya que algunas veces hay mayor importancia en estudiar la transición, es decir la incidencia de un estado a otro y en otros casos es el estado, es decir la prevalencia que se está modelando.

El análisis de supervivencia se presenta de manera natural al estudiar fenómenos relacionados con la medicina, como lo son el tiempo de recuperación tras alguna cirugía, recaída de algún padecimiento, desarrollo de enfermedades, tiempo libre de tumores, tiempo a la muerte; aunque también es ampliamente utilizado en otros campos como lo son los seguros (en tiempos de reclamación), la sociología (duración del primer matrimonio), la industria (tiempo de vida de componentes) o la mercadotecnia (duración de una suscripción a una revista).

Para realizar un estudio de esta índole se debe contar con la información correspondiente. Dentro de los datos que considera este análisis o “datos de supervivencia” se pueden enlistar:

- Respuesta a tratamientos determinados.
- Características del paciente relacionados con la respuesta.
- Mediciones de tiempo de supervivencia.
- Desarrollo de alguna enfermedad.

El objetivo que se plantea en esta clase de análisis está en pronosticar la ocurrencia de una determinada respuesta, la supervivencia de un individuo o estimar la esperanza de vida; aunque últimamente también se ha utilizado para identificar el riesgo y los “factores de pronóstico” que están relacionados con la respuesta, supervivencia y desarrollo del evento.

Un problema frecuente en el análisis de supervivencia es el de medir los tiempos exactos de ocurrencia del “evento”, ya que observaciones o “datos censurados” pueden ocurrir una vez finalizado el estudio o que los individuos involucrados en el estudio se pierdan y con ello parte de la información a utilizar.

A continuación se presentan algunas definiciones y conceptos del análisis de supervivencia que son importantes para familiarizarse con el tema.

1.1. Definiciones

Considérese a T una variable aleatoria real no-negativa continuamente distribuida la cual representa el tiempo de supervivencia de un individuo en una población. T definido en $[0, \infty)$. Sea $f(t)$ la función de densidad de T y sea la función de distribución

$$F(t) = P(T \leq t) = \int_0^t f(u)du.$$

Esta función es la ampliamente conocida en el ambito de la probabilidad, la interpretación en el análisis de supervivencia recae principalmente en la función de densidad la cual se puede entender como el límite de la probabilidad de que un individuo falle en el intervalo $(t, \Delta t)$ por unidad de tiempo Δt .

1.1.1. La función de supervivencia

La probabilidad de que un individuo llegue con vida al tiempo t está dada por la función de supervivencia $S(t)$, la cual se define como

$$S(t) = P(T > t) = 1 - F(t), \quad t \in [0, \infty),$$

es decir

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du.$$

Esta función debe satisfacer cuatro condiciones:

1. $S(0) = 1$,
2. $S(t) < 1, \quad \forall t$,
3. $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$,
4. $S(t)$ es no creciente *i.e.* si $u < t$ entonces $S(u) \geq S(t)$.

Esta función es la base del análisis de supervivencia ya que ejemplifica perfectamente el concepto de supervivencia una vez que queremos modelar el hecho de que una variable aleatoria T exceda un determinado tiempo t .

1.1.2. La función de riesgo

También conocida como tasa instantánea de falla, fuerza de mortalidad (en ciencias actuariales), tasa de mortalidad condicional y tasa de falla a una edad determinada, se encuentra definida como

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} \\ &= \frac{f(t)}{1 - F(t)}. \end{aligned} \tag{1.1}$$

Esta función calcula la probabilidad de que la falla (evento) ocurra en un intervalo corto de tiempo, siempre y cuando el individuo haya llegado con vida al inicio del intervalo.

Con la función anterior se define la función de riesgo acumulada la cual está dada por

$$H(t) = \int_0^t h(x)dx.$$

1.2. Relaciones y equivalencias

A continuación se presentan algunas de las expresiones que ligán las definiciones de la sección anterior, esto con el fin de encontrar alternativas para una modelación más adecuada.

1. De la definición de función de riesgo y función de supervivencia se sigue que

$$\begin{aligned}h(t) &= \frac{f(t)}{1 - F(t)} \\ &= \frac{f(t)}{S(t)}.\end{aligned}\tag{1.2}$$

2. De la relación entre función de densidad y distribución

$$\begin{aligned}f(t) &= \frac{d}{dt}F(t) \\ &= \frac{d}{dt}(1 - S(t)) \\ &= -S'(t).\end{aligned}\tag{1.3}$$

3. Por 1.2 y 1.3

$$\begin{aligned}h(t) &= \frac{-S'(t)}{S(t)} \\ &= -\frac{d}{dt} \log S(t).\end{aligned}\tag{1.4}$$

4. Integrando la función de riesgo y dado que $S(0) = 1$, por 1.4 se tiene que

$$-\int_0^t h(x)dx = \log S(t) \text{ o } H(t) = -\log S(t),$$

por tanto

$$S(t) = e^{-\int_0^t h(x)dx} = e^{-H(t)}.\tag{1.5}$$

5. Por 1.5 y 1.2

$$f(t) = h(t)e^{-H(t)}.\tag{1.6}$$

1.3. Censura de datos

Algo característico del análisis de supervivencia y que le diferencia de otras herramientas estadísticas es la censura de datos, esta se refiere a que dentro del análisis de supervivencia

se consideran aquellas observaciones que por alguna razón contienen información incompleta, es decir, se incluye en el análisis toda la información de las observaciones, aún aquellas que pudieron perderse durante el estudio o cuyo “evento” sucede después de finalizado el estudio. La importancia de conocer sobre los datos censurados radica primordialmente en saber que es lo “faltante” y las implicaciones que esto traerá al análisis posterior.

Existen tres posibles casos de datos censurados: censurados por la derecha, censurados por la izquierda y censurados por intervalo.

- **Censurado por la derecha:** Se da cuando sabemos que el “evento” a estudiar ocurre en algún punto posterior al tiempo de duración del estudio (o tiempo de recopilación de datos); es decir, cuando se rebasa el periodo de observación y no es posible determinar la ocurrencia exacta.
- **Censurado por la izquierda:** Ocurre cuando la información correspondiente a una observación queda incompleta debido a que se pierde el seguimiento de la observación o en el caso médico el paciente “sale” del estudio y se vuelve imposible recuperar los datos para el análisis.
- **Censurado por intervalo:** Es similar a la censura por la izquierda solo que esta se diferencia en que no se puede establecer un punto determinado de “pérdida de la información” y esta se define que ocurre dentro de un intervalo.

Asimismo, se consideran tres tipos de censura.

- **Tipo I:** se fija un tiempo determinado para observar el desarrollo de las variables (duración del estudio) y sobre ese intervalo de tiempo se observan las censuras ocurridas. En la figura 1.1 se muestra un ejemplo gráfico.
- **Tipo II:** se determina una cantidad m de observaciones con las cuales concluirá el estudio, *i.e.* si se tienen 100 sujetos en observación se determina que el estudio finalizará con la ocurrencia del 30° “evento”. En la figura 1.2 se muestra un ejemplo gráfico.
- **Tipo III:** se considera el tiempo fijo para la observación del desarrollo de las variables pero estas no entran todas al mismo tiempo al estudio, *i.e.* el tiempo de supervivencia

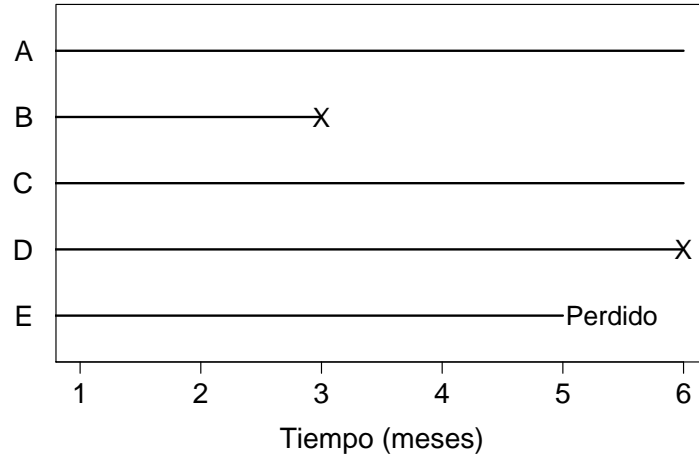


Figura 1.1: Gráfica tipo de censura I

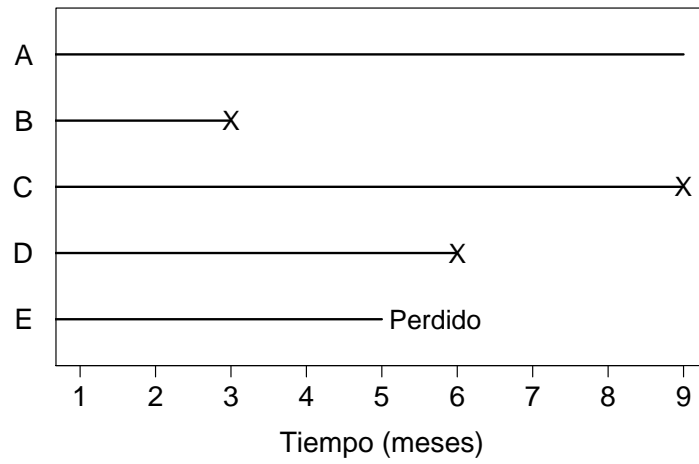


Figura 1.2: Gráfica tipo de censura II

se considera a partir de la entrada del sujeto al estudio y hasta la ocurrencia del “evento”. En la figura 1.3 se muestra un ejemplo gráfico.

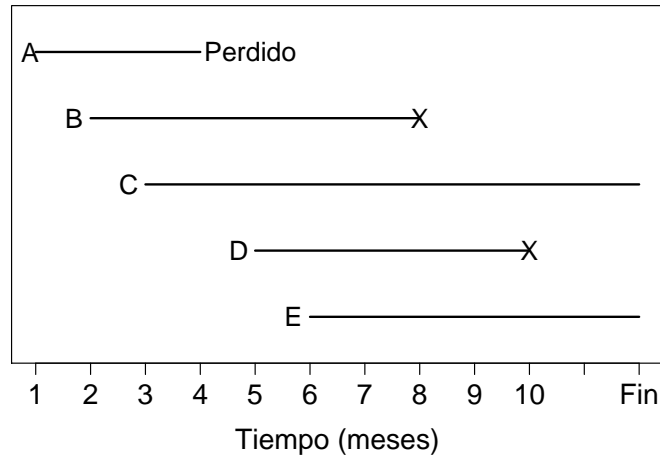


Figura 1.3: Gráfica tipo de censura III

A los *tipos I* y *II* se les conoce como “censura simple” y al *tipo III* como “censura progresiva” o “censura aleatoria” dentro de estos el tipo de datos censurados que presentan es el censurado por la derecha.

1.4. Modelos de supervivencia

Dado el espectro de este enfoque se han presentado diversas técnicas para abordar el problema del análisis de supervivencia, existen aproximaciones paramétricas y no paramétricas (aunque también existen conjunciones de ambas o semiparamétricas), cada uno tiene un planteamiento diferente para la resolución, de acuerdo a las características, necesidades y objetivos que se planteen. A continuación se presentan los métodos más utilizados.

1.4.1. Paramétricos

Dentro del análisis de supervivencia existen determinadas distribuciones que son ampliamente conocidas y aplicadas en este contexto, dentro de las principales se encuentran:

Distribución exponencial

Es la función más simple y la más importante de las distribuciones de supervivencia. Algunas las aplicaciones de esta función se encuentra al describir patrones de vida en sistemas electrónicos, como por ejemplo sistemas de generación de estados de cuenta bancarios, cajeros automáticos y componentes de radares. Es usualmente asociada a patrones de falla puramente aleatorios, también es famosa por la propiedad de “falta de memoria” con el único detalle que esta propiedad requiere que la edad de la persona (componente, animal) no afecte su supervivencia futura. Está caracterizada por una tasa de riesgo constante λ , como su único parámetro. Valores grandes de λ se asocian a riesgo alto y por tanto baja supervivencia y viceversa.

El tiempo de supervivencia T sigue una distribución exponencial con parámetro λ si su función de densidad está dada por

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \quad \lambda > 0, \quad (1.7)$$

por tanto las funciones de distribución acumulada, supervivencia y riesgo están dadas por

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0 \quad (1.8)$$

$$S(t) = e^{-\lambda t}, \quad t \geq 0 \quad (1.9)$$

$$h(t) = \lambda \quad (1.10)$$

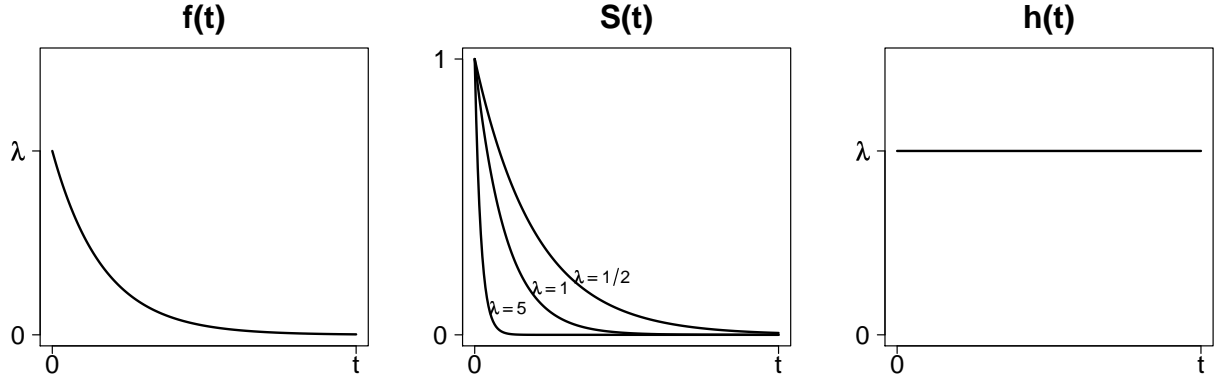


Figura 1.4: Gráficas para la distribución exponencial(λ)

Distribución weibull

Es una generalización de la distribución exponencial; se diferencia de la anterior en que no supone una tasa de riesgo constante, logrando con esto aumentar sus posibles aplicaciones. Sus primeras aplicaciones se dieron en rentabilidad, mortalidad humana, en pacientes con cáncer de pulmón y tiempos de mejora en cirugías mayores. Está caracterizada por dos parámetros: γ que determina la forma de la distribución y λ que determina su escala. Cuando $\gamma = 1$ se obtiene el caso exponencial. Cuando t crece, la tasa de riesgo aumenta cuando $\gamma > 1$ y viceversa.

Las funciones de densidad, distribución acumulada, supervivencia y riesgo están dadas a continuación

$$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}e^{(-\lambda t)^\gamma}, \quad t \geq 0, \quad \gamma, \lambda > 0 \quad (1.11)$$

$$F(t) = 1 - e^{(-\lambda t)^\gamma}, \quad (1.12)$$

$$S(t) = e^{(-\lambda t)^\gamma}, \quad (1.13)$$

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1} \quad (1.14)$$

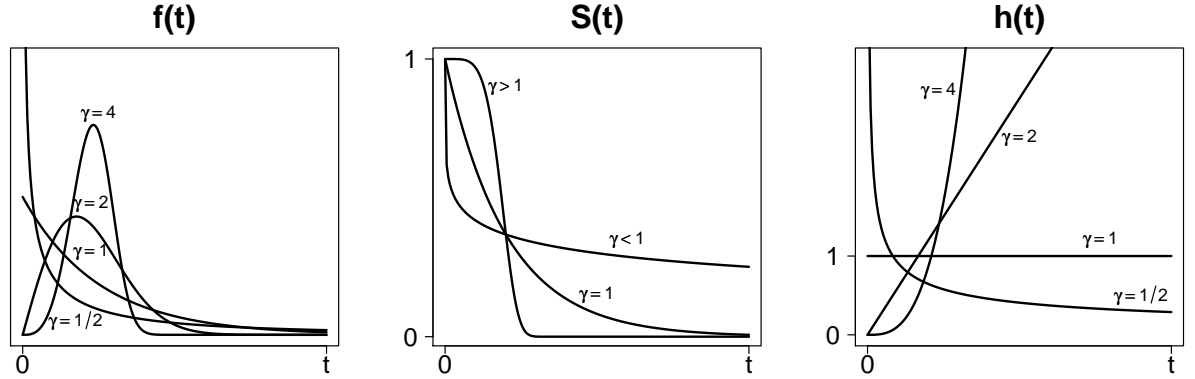


Figura 1.5: Gráficas para la distribución weibull($\gamma, \lambda = 1$)

Distribución gamma

Esta distribución incluye a las distribuciones exponencial y ji-cuadrada. Inicialmente fue utilizada para describir el tiempo útil de los vasos de cristal de cafeteras y para la duración de materiales, de igual forma se ha utilizado en la industria y supervivencia humana. Esta distribución surge de suponer que la falla o muerte ocurre en n periodos, o cuando n subfallas han sucedido, es decir, al tiempo T_1 , la primera subfalla ocurre, al tiempo T_2 la segunda subfalla ocurre y así sucesivamente. La falla total o muerte ocurre al final del n -ésimo periodo con la n -ésima subfalla. El tiempo de supervivencia T es entonces $T_1 + T_2 + \dots + T_n$, se supone que dichos tiempos son independientes y distribuidos exponencialmente con función de densidad $\lambda \exp(-\lambda t_i)$, $i = 1, \dots, n$. Lo cual quiere decir que las subfallas ocurren de forma independiente con tasa constante λ . La distribución de T es llamada *distribución de Erlang* y ha sido ampliamente usada en teoría de colas y procesos de vida. Si se generaliza la distribución de Erlang reemplazando el parámetro n por un parámetro real $\gamma > 0$ se obtiene la *distribución gamma*. La distribución gamma está caracterizada por dos parámetros γ y λ , los cuales determinan la forma y la escala respectivamente. Si $0 < \gamma < 1$ la tasa de riesgo disminuye monótonamente de infinito a λ conforme avanza el tiempo; si $\gamma > 1$ la tasa de riesgo crece monótonamente de 0 a λ cuando t avanza y si $\gamma = 1$ la tasa de riesgo se vuelve constante con tasa λ como en el caso exponencial.

Las funciones de densidad, distribución acumulada, supervivencia y riesgo son:

$$f(t) = \frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} e^{-\lambda t}, \quad t > 0, \quad \gamma, \lambda > 0 \quad (1.15)$$

$$F(t) = \int_0^t \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x} dx, \quad (1.16)$$

$$S(t) = \int_t^\infty \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x} dx, \quad (1.17)$$

$$h(t) = f(t)/S(t) \quad (1.18)$$

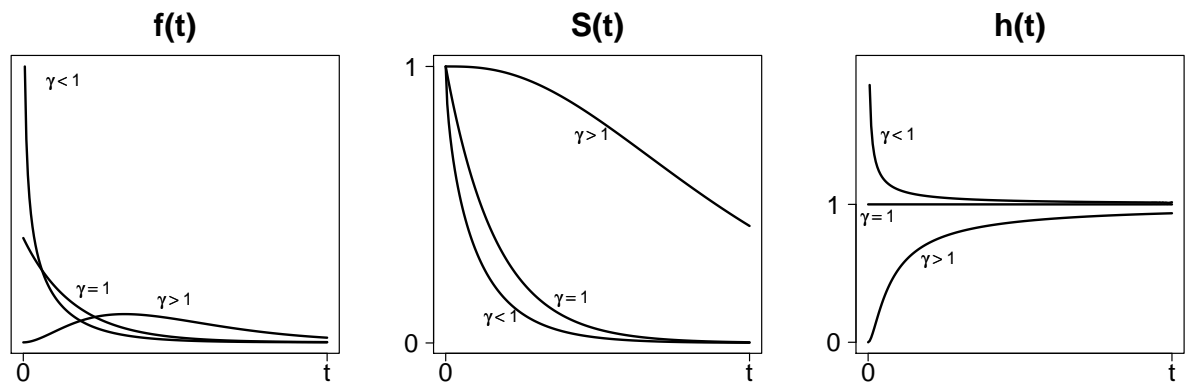


Figura 1.6: Gráficas para la distribución gamma($\gamma, \lambda = 1$)

1.4.2. No paramétricos

A diferencia de los paramétricos estos modelos no especifican distribución alguna, en esta sección el modelo que se verá más detalladamente será el Kaplan-Meier.

Kaplan-Meier

Este es uno de los métodos no paramétricos de estimación más frecuentemente utilizados (si no el más utilizado) en el ámbito de la supervivencia, debido a que su implementación es relativamente sencilla y se llega a buenos ajustes para la función de supervivencia, sin embargo debido a que lo que realiza este método es aproximar una función discreta a la función de supervivencia (la cual es continua) no es posible estimar directamente la función de riesgo, que en muchos casos es de interés analizar. Aunque esta se puede aproximar mediante otros métodos.

Cabe señalar que este método únicamente considera el número de elementos que se encuentran con vida a un determinado tiempo t y por tanto únicamente tiempos de supervivencia reportados, es decir que en ningún momento considera información contenida en covariables, lo cual podría ser una limitante para un análisis más a fondo.

Estimación de Kaplan-Meier

Considérense T_1, T_2, \dots, T_n tiempos de observación, independientes y sus correspondientes indicadores de ocurrencia (muerte) D_1, D_2, \dots, D_n . Para cada t se define el nivel de riesgo como $R(t) = \sum_i 1 \{T_i \leq t\}$, la cual es una función escalonada, decreciente, continua por la izquierda y que se puede interpretar como el número de personas bajo observación al tiempo t . Se particiona el tiempo en intervalos de la forma $0 = x_0 < x_1 < \dots < x_m = \infty$, por conveniencia, se supone que la censura solo ocurre en el extremo final de cada intervalo o en aquellos intervalos en los que no se haya presentado ocurrencia (muerte), asimismo se toman los intervalos como equidistantes (a excepción del último intervalo, el cual abarca hasta ∞), ahora se procede a parametrizar la distribución mediante las probabilidades de muerte condicionadas dentro de los intervalos, las cuales están dadas por

$$p_j = P(x_{j-1} < T \leq x_j | x_{j-1} < T).$$

Se supone que el individuo se encuentra con vida al inicio del intervalo; esta es una forma discretizada de la función de riesgo. Si se condiciona el número de muertes con respecto a $R(x_{j-1})$, el número de muertes sigue una distribución binomial con parámetro p_j ; esto es, para cada individuo que llegue con vida al inicio del intervalo, existe una variable aleatoria D_{ij} , tal que:

$$P(D_{ij} = 1) = p_j, \quad P(D_{ij} = 0) = 1 - p_j.$$

El estimador para p_j resulta de manera natural $D_{.j} = \sum_i D_{ij}$ dividido entre $R(x_{j-1})$, esto es, $\hat{p}_j = D_{.j}/R(x_{j-1})$. Entonces, para estimar la función de supervivencia en algún punto final del intervalo, por ejemplo x_q , debemos notar primero que $S(x_q) = \prod_{j=1}^q (1 - p_j)$. Ahora, de la expresión anterior y considerando los estimadores para p_j

$$\hat{S}(x_q) = \prod_{j=1}^q \left\{ 1 - \frac{D_{.j}}{R(x_{j-1})} \right\}$$

Para considerar esta una estimación no-paramétrica, se refina la partición de los intervalos de tal forma que no pueda ser definido en un punto q y supóngase que quiere estimarse el valor en un punto t cualquiera, en caso de que t no sea un extremo final de un intervalo se considera el extremo final inmediato anterior, si los intervalos son suficientemente finos, no habrá ocurrencias (muertes) entre estos dos puntos a menos que ocurra exactamente en t . Si se analizan detenidamente los intervalos se observa que no hay muchos con ocurrencias, por lo cual estos intervalos contribuyen a la función de supervivencia con factores de 1, por tanto estos pueden ser omitidos, entonces solo habrá t_1, t_2, \dots, t_q tiempos donde se hayan observado ocurrencias, donde cada uno de estos tiempos de muerte, t_r , $r = 1, 2, \dots, q$, se encuentra contenido en un intervalo, $j(r)$. Si se supone una distribución continua, únicamente podrá ocurrir una muerte en cada uno de estos tiempos. Entonces el estimador para la función de supervivencia queda expresado como:

$$\hat{S}(t) = \prod_{r=1}^q \left\{ 1 - \frac{1}{R(x_{j(r)-1})} \right\}.$$

Dado que cada tiempo de muerte contribuye a la función y la manera de asignar los intervalos está dada por $j(r)$ y este únicamente describe el nivel de riesgo justo antes del tiempo t_r , se puede cambiar el término $R(x_{j(r)-1})$ por $R(t_r)$. En la práctica puede suceder que existan muchas ocurrencias en el mismo punto y por lo tanto al tiempo t_r ocurren N_r muertes. Entonces la fórmula para el estimador de Kaplan-Meier está dada por:

$$\hat{S}(t) = \prod_{r=1}^q \left\{ 1 - \frac{N_r}{R(t_r)} \right\}. \quad (1.19)$$

Cabe notar que este modelo únicamente considera los tiempos de supervivencia y no así a las covariables que intervienen en el proceso, esto podría limitar las conclusiones o posibles interpretaciones en un modelo donde intervengan covariables. La importancia de este método es que da una aproximación muy precisa a la función de supervivencia y su análisis e interpretación se presenta de manera natural, lo cual para muchos casos es suficiente.

1.4.3. Semi paramétricos

Riesgos proporcionales

Los modelos de riesgos proporcionales son una alternativa semiparamétrica para modelar la función de riesgo basándose en la información que se tiene, a continuación se presenta una

descripción de esta aproximación.

Considérese que la función de riesgo se ve afectada principalmente por dos factores, el primero es como varía esta a través del tiempo y la segunda es referente al efecto de las covariables al describir como el riesgo se relaciona con diversos factores.

El *principio de riesgos proporcionales* supone que la función de riesgo puede ser expresada como la función de riesgo inicial multiplicada por el efecto de las covariables dentro del modelo, esto es:

$$h(t|x) = g(x)h_0(t),$$

donde $h_0(t)$ es la función de riesgo inicial y $g(x)$ una función que caracteriza el efecto de las covariables dentro del modelo.

En otras palabras, si se cambia de variable (observación), obtendremos una nueva función de riesgo la cual será proporcional a la función de riesgo inicial.

El modelo de riesgos proporcionales de Cox está dado por

$$h(t|x) = h_0(t)\exp\{G(x, \beta)\}$$

, donde $h_0(t)$ es la función de riesgo inicial y β es el vector de coeficientes; el segundo término se escribe en forma exponencial debido a que debe ser positivo. El modelo implica que la proporción de riesgo entre dos individuos es constante a través del tiempo; siempre y cuando las covariables no se alteren en el tiempo. Se supone a su vez que el efecto de las covariables es multiplicativo, dejando a la función de riesgo expresada como

$$h(t|x) = h_0(t)\exp(x'\beta)$$

donde $\eta = x'\beta$ es llamado el predictor lineal. Este modelo implica que la proporción de riesgo entre dos individuos depende de la diferencia entre sus predictores lineales en cualquier momento.

La aproximación bayesiana

El enfoque bayesiano surge como una de las opciones más frecuentes para realizar inferencia sobre un parámetro determinado esto mediante la función posterior y/o la función predictiva que son el objetivo principal de la estadística bayesiana. Muchas son las discusiones que se han presentado con respecto a este enfoque, aunque en muchos casos el análisis se realiza de manera natural y deja de utilizar resultados asintóticos tales como el teorema del límite central y la ley de los grandes números las cuales no se verifican en todas la muestras.

Por tal motivo se presenta como una opción viable para estudiar fenómenos de supervivencia cuando los estudios se realizan sobre grupos particularmente pequeños además de que se puede hacer uso de la información previa resultante de estudios o análisis anteriores, esto puede dotar de una mayor relevancia a la investigación en cuestión. No obstante siempre se deberán tener consideraciones al aplicar cualquier modelo sea bayesiano o no.

2.1. El paradigma bayesiano

Sea una serie de observaciones, x_1, \dots, x_n las cuales serán utilizadas para el análisis estadístico, dichas observaciones provienen de una función de distribución paramétrica, *i.e.* x_i ($i = 1, \dots, n$) tiene una distribución con densidad $f_i(x_i|\theta_i, x_1, \dots, x_{n-1})$, tal que el parámetro θ_i es desconocido y la función f_i conocida. Lo anterior puede ser expresado como:

$$x \sim f(x|\theta),$$

donde x es el vector de observaciones y θ es el conjunto de parámetros, $\theta_1, \dots, \theta_n$, el cual pertenece a un espacio paramétrico. Por *espacio paramétrico* se entiende el conjunto Θ de

todos los valores que puede tomar θ .

Este modelo representa de manera adecuada observaciones únicas, observaciones dependientes y observaciones independientes e idénticamente distribuidas (*i.i.d.*) de una distribución $f(x_i|\theta)$, en cuyo caso si $x = (x_1, \dots, x_n)$ son v.a.i.i.d. se sigue que:

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

Se define a $f(x|\theta)$ como un *modelo probabilístico paramétrico* si las observaciones de una variable aleatoria se distribuyen de acuerdo a $f(x|\theta)$, donde únicamente el parámetro θ es desconocido y pertenece a un espacio vectorial Θ de dimensión finita.

Una vez conocido el modelo se procede a hacer inferencia sobre el parámetro recurriendo al *Teorema de Bayes*. Sean A y B dos eventos tales que $P(B) \neq 0$, entonces $P(A|B)$ y $P(B|A)$ se relacionan mediante

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{P(B|A)P(A)}{P(B)}. \end{aligned}$$

Asimismo Thomas Bayes introdujo una versión continua de su resultado para un caso particular y posteriormente Laplace generalizó dicho resultado, esto es que dadas dos variables aleatorias x y y , con distribución condicional $f(x|y)$ y distribución marginal $g(y)$, la distribución de y dado x está dado por

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}.$$

Bayes y Laplace consideraron que la *incertidumbre* sobre el parámetro θ podía ser modelado mediante una distribución de *probabilidad* (π) sobre Θ llamada distribución *a priori*. Entonces la inferencia se basa sobre la distribución de θ condicionado a x , esto es $\pi(\theta|x)$, llamada distribución *a posteriori* o simplemente distribución posterior, la cual está definida como

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}.$$

Y dado que $\int_{\Theta} f(x|\theta)\pi(\theta)d\theta$ no depende de θ la expresión anterior puede escribirse como:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta), \tag{2.1}$$

es decir la función a priori multiplicada por la verosimilitud.

Es así como se define el *modelo estadístico paramétrico bayesiano* el cual está conformado por un *modelo estadístico paramétrico*, $f(x|\theta)$ y una distribución *a priori*, $\pi(\theta)$ sobre los parámetros.

Otro paso importante dentro del enfoque bayesiano radica en la facilidad de hacer predicciones, basándose en la función posterior, previamente definida en el modelo estadístico paramétrico bayesiano y la cual se presenta a continuación. Considérese que se desea realizar inferencia sobre observaciones desconocidas, la distribución de x la cual se presenta como datos observables pero desconocidos es

$$\begin{aligned} p(x) &= \int p(x, \theta) d\theta \\ &= \int \pi(\theta) f(x|\theta) d\theta. \end{aligned} \tag{2.2}$$

La cual es llamada *distribución predictiva a priori*; considérese ahora que han sido observados los datos x y se desea predecir a una observación \tilde{x} del mismo proceso, a la distribución de \tilde{x} se le conoce como *distribución predictiva posterior* y se define como

$$\begin{aligned} p(\tilde{x}|x) &= \int p(\tilde{x}, \theta|x) d\theta \\ &= \int p(\tilde{x}|\theta, x) p(\theta|x) d\theta. \end{aligned} \tag{2.3}$$

2.2. El principio de verosimilitud

Para dar entrada a este principio se retomará el concepto de función de verosimilitud.

Definición: *Función de verosimilitud* Sean $x = \{x_1, x_2, \dots, x_n\}$ observaciones de una población con función de densidad $f(x|\theta)$, la función de verosimilitud de θ denotada por $\ell(\theta|x)$, se define como la función de densidad conjunta:

$$\ell(\theta|x) = f_{X_1, \dots, X_n}(x_1, \dots, x_n|\theta).$$

El valor de $\ell(\theta|x)$ se denomina como *verosimilitud* de θ . Esta función se encuentra asociada a la probabilidad de x dado que se considera a θ como el valor correcto o “adecuado” del

parámetro, la cual cambia si se varía el valor de θ , es decir, se puede hablar de una mayor probabilidad o mayor “plausibilidad” en los datos dependiendo del valor de θ asociado.

Principio de verosimilitud: La información acerca de θ proporcionada por una serie de observaciones x se encuentra contenida en la función de verosimilitud $\ell(\theta|x)$. Más aún si x_1 y x_2 son dos series de observaciones tales que ambas dependen del mismo parámetro θ , entonces existe una constante c la cual satisface:

$$\ell_1(\theta|x_1) = c\ell_2(\theta|x_2),$$

para cada θ , esto es, ambas observaciones aportan la misma información sobre θ y deben conducir a inferencias idénticas.

Habrá que notar que el principio de verosimilitud es válido únicamente cuando:

1. La inferencia se realiza sobre el mismo parámetro θ y
2. θ contiene “toda” la información desconocida del modelo.

Es importante aclarar que el principio de verosimilitud no afirma que “toda” la información sobre θ está contenida en $\ell(\theta|x)$ sino toda la información “experimental”, en otras palabras, que únicamente contiene todo aquello que las observaciones aportan a θ sin considerar otra información relevante como conocimientos previos u otros factores adicionales.

A pesar de las implicaciones del principio de verosimilitud, este cuenta con algunas limitaciones inherentes a su misma definición. Toda vez que este principio no dice nada sobre como desarrollar un análisis estadístico ni sobre su implementación en dicho análisis. Esto es, el principio de verosimilitud no provee marco alguno para, por ejemplo, realizar predicciones acerca de futuros valores de x ni para incorporar información adicional existente acerca del modelo.

Dentro de los diversas esquemas que se han propuesto para utilizar la función de verosimilitud para llegar a conclusiones sobre θ , el enfoque bayesiano provee un marco apropiado para tal fin, como se verá en las siguientes secciones.

Definición: *Intercambiabilidad* Una sucesión infinita $x_1, x_2, \dots, x_n, \dots$ de variables aleatorias se dice que es *intercambiable* si para toda $n = 2, 3, \dots$, se cumple que:

$$x_1, \dots, x_n \stackrel{D}{=} x_{\pi(1)}, \dots, x_{\pi(n)} \quad \forall \pi \in S(n),$$

donde $S(n)$ es el grupo de permutaciones de $\{1, \dots, n\}$.

Observación: Si x_1, \dots, x_n, \dots son independientes e idénticamente distribuidas entonces son intercambiables, pero el regreso no es necesariamente cierto.

La importancia de la intercambiabilidad radica en que es un concepto más débil que el de variables independientes e idénticamente distribuidas y por tanto más asequible y que además formaliza la idea de que futuras muestras se comportarán como muestras pasadas, lo cual conduce a que la distribución de dichas variables depende únicamente del valor de estas y no del orden en que fueron obtenidas.

2.3. La función *a priori*

En esta sección se discutirán algunas de las principales cuestiones referentes a la función *a priori*, la cual es de suma importancia en el contexto bayesiano.

2.3.1. Obtención de una *a priori*

Un primer paso dentro de la inferencia bayesiana radica en la correcta interpretación de la información previa con la que se cuenta, es de crucial importancia dar un análisis adecuado a la información con la que se cuenta y la manera en que esta será introducida al modelo en cuestión, hay muchas discusiones acerca de la manera en que se procede a “traducir” la información previa para que pueda ser utilizada como una función *a priori* que refleje la intención de los datos previos. A continuación se presentan algunas consideraciones a tener para un adecuado manejo de la información *a priori* con la que se cuenta.

Algunas de las grandes discusiones acerca de la utilización del enfoque bayesiano para realizar inferencia radican en la implementación de una función *a priori*, estas radican pri-

mordialmente en la imposibilidad de encontrar “la distribución” *a priori*, este es en esencia el principal problema de la estadística bayesiana. Otro punto a considerar es la influencia que tiene la función *a priori* sobre la inferencia, ya que de influir demasiado traería ciertas consecuencias en los resultados, toda vez que dependen de la elección (subjetiva) de la distribución *a priori*, una manera de medir esta influencia en los resultados está en métodos de análisis de sensibilidad lo cual nos da una idea de cuanto puede afectar a la inferencia la utilización de una distribución u otra. Otra forma de abordar este inconveniente se encuentra en implementar modelos jerárquicos para aminorar la arbitrariedad en la elección de una función *a priori*, para más información sobre la implementación de modelos jerárquicos ver Gelman et al (2004 Cap 5).

2.3.2. Familias conjugadas

Dado que tanto $\pi(\theta|x)$ como $\pi(\theta)$ son funciones de distribución sobre θ el paso de una función a otra está dado por la información aportada por x y en este sentido $\pi(\theta|x)$ es una actualización de la función *a priori*. Y para fines prácticos se consideraba deseable que tanto $\pi(\theta|x)$ como $\pi(\theta)$ pertenecieran a la misma familia.

Definición: Sea $P = \{p(x|\theta)|\theta \in \Theta\}$ una familia paramétrica. Una clase (o colección) de distribuciones de probabilidad f es una *familia conjugada* para P si para todo $p(x|\theta) \in P$ y $\pi(\theta) \in f$ se cumple que $\pi(\theta|x) \in f$.

La anterior definición provee de una serie de familias paramétricas para las cuales se cumple lo anterior. Entre estas encontramos:

Fam. paramétrica	Fam. conjugada
Bernoulli(θ)	Beta($\theta \alpha, \beta$)
Poisson(λ)	Gamma($\lambda \alpha, \beta$)
Geométrica(θ)	Beta($\theta \alpha, \beta$)
Exponencial(λ)	Gamma($\lambda \alpha, \beta$)
Uniforme($0, \theta$)	Pareto($\theta \alpha, \beta$)
Normal(μ)	Normal($\mu \mu_0, \lambda_0$)
Normal(λ)	Gamma($\lambda \alpha, \beta$)
Normal(μ, λ)	Normal-Gamma($\mu, \lambda \mu_0, n_0, \alpha, \beta$)

Tabla 2.1: Familias conjugadas conocidas

Para la tabla 2.1 en la normal el parámetro λ es llamado *precisión* y se define como en inverso de la varianza.

2.3.3. Informativas y “No informativas”

Como ya se mencionó, uno de los cuestionamientos que han emergido dentro del marco bayesiano se encuentra en la correcta utilización de la información previa con la que se cuenta, en el sentido de que tanto se puede decir (o no decir) sobre el problema mediante una función a priori.

Para tal fin se ha buscado proponer funciones a priori que sean acordes con la información con la que se cuenta y es aquí donde surge un problema, ¿cómo proceder cuando no se tiene información previa?, es decir, una función a priori que no refleja información alguna sobre θ (o en su defecto, no se quiere o no se puede utilizar dicha información), es aquí donde la utilización de funciones *a priori no informativas* surge para darle respuesta a estos inconvenientes.

A menudo sucede que al implementar una función a priori no informativa de manera natural, esta resulta una *a priori impropia*, es decir el valor de la integral (suma) de la función de densidad (probabilidad) sobre todo el soporte no es finito. El problema a enfrentar entonces es el de determinar funciones a priori no informativas. Para tal fin hay diversas y muy variadas técnicas para determinar dicha distribución, una primera idea está en asig-

narle la misma probabilidad a cada posible parámetro, en el caso discreto esta idea resulta sencilla, con esto se pretende no favorecer de ninguna forma a algún valor del parámetro en particular, lo cual se logra mediante la utilización de una distribución uniforme sobre el intervalo, el inconveniente surge cuando lo anterior no puede llevarse a cabo.

Ahora la pregunta está en que sucede una vez que la función a priori es determinada pero el producto resultante (función posterior) no es una función conocida (es decir, no es conjugada), en la siguiente sección se presentará la forma de abordar este inconveniente.

2.4. Cálculo de estimadores

El objetivo final de la estadística bayesiana es encontrar la función posterior y así proveer de estimaciones para el modelo planteado, en la práctica para llevarlo a cabo se utilizan diversas técnicas para calcular los resultados. Como se ha visto en las secciones anteriores el principal problema que surge al desarrollar un esquema bayesiano radica en determinar la forma funcional de la función posterior, lo cual en muchas ocasiones no es posible, una de las técnicas que a últimas fechas ha sido ampliamente utilizada para sortear este inconveniente se encuentra en los métodos de simulación *Markov Chain Monte Carlo* (MCMC). A continuación se dará una breve introducción a estos métodos y en donde radica su importancia.

2.4.1. MCMC

Los métodos MCMC surgen de la idea original de la simulación Monte Carlo, la cual consiste en aproximar integrales mediante promedios de las funciones conocidas utilizando resultados asintóticos, pero a diferencia de estos, los métodos MCMC requieren un poder computacional mayor.

Estos métodos se usan con frecuencia para generar “aproximaciones” a una función de distribución objetivo las cuales se realizan mediante algoritmos para muestrear dicha función basándose en cadenas de Markov; cuya principal idea es que para producir aproximaciones aceptables a las integrales y/o a funcionales dependientes de una función de distribución es suficiente con generar una cadena de Markov ($\theta^{(m)}$) cuya distribución límite es la distribución objetivo de interés. Para estos métodos la convergencia es asegurada por la siguiente

proposición (Teorema ergódico).

Proposición: Sea $\theta^{(1)}, \theta^{(2)}, \dots$ una *Cadena de Markov* homogénea, irreducible y aperiódica, con *espacio de estados* Θ y *distribución límite* $p(\theta|x)$. Entonces si $t \rightarrow \infty$,

(i) $\theta^{(t)} \xrightarrow{D} \theta$, donde $\theta \sim p(\theta|x)$;

(ii) $\frac{1}{t} \sum_{i=1}^t g(\theta^{(i)}) \rightarrow E(g(\theta)|x)$.

Muestreo de Gibbs

Como se mencionó anteriormente lo que se desea es simular una cadena de Markov $(\theta^{(1)}, \theta^{(2)}, \dots)$ cuya distribución límite está dada por $p(\theta|x)$ la cual es la distribución posterior del modelo estadístico bayesiano ya presentado. El muestreo de Gibbs es un algoritmo, cuya particularidad radica en que la simulación la realiza mediante funciones univariadas, que son más fáciles de simular que las distribuciones conjuntas las cuales suelen tener formas más complejas.

Considérese a $\theta = (\theta_1, \dots, \theta_k)$ una partición del vector θ , tal que $\theta_i \in \mathbb{R}^{d_i}$ y $\sum_{i=1}^k d_i = d$. Asimismo

$$\begin{aligned} & p(\theta_1|\theta_2, \dots, \theta_k, x) \\ & \quad \vdots \\ & p(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k, x) \quad (i = 2, \dots, k-1) \\ & \quad \vdots \\ & p(\theta_k|\theta_1, \dots, \theta_{k-1}, x) \end{aligned}$$

son llamadas *condicionales completas* (o *totales*) las cuales para cada $i \in (1, \dots, k)$ cumplen lo siguiente:

$$p(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k, x) \propto p(\theta|x),$$

vista como función de θ_i . Hay que notar que estas características son las que ayudan a encontrar la función posterior, para tal fin, las funciones que deben ser fáciles de simular deben ser estas densidades condicionales completas, lo cual sucede en muchos casos, por esta razón el muestreo de Gibbs se ha convertido en un algoritmo cuya implementación es relativamente sencilla.

A continuación se muestra el algoritmo, cuya idea principal es simular una cadena de Markov en la que el paso $t + 1$ se obtiene a partir del paso t :

1. Generar un valor inicial de $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ paso $t = 0$
2. Generar una observación de $\theta_1^{(t+1)}$ a partir de $p(\theta_1 | \theta_2^{(t)}, \dots, \theta_k^{(t)}, x)$
 Generar una observación de $\theta_2^{(t+1)}$ a partir de $p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, x)$
 \vdots
 Generar una observación de $\theta_k^{(t+1)}$ a partir de $p(\theta_k | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, x)$

De esta forma, la sucesión $\theta^{(1)}, \theta^{(2)}, \dots$ es una realización de una cadena de Markov, con distribución límite $p(\theta|x)$ y kernel de transición

$$p(\theta^{(t+1)} | \theta^{(t)}) = \prod_{i=1}^k p(\theta_i^{(t+1)} | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_k^{(t)}, x).$$

Un modelo de supervivencia con inferencia bayesiana para diabetes

En este capítulo se presenta un modelo de supervivencia para la diabetes haciendo uso de herramientas bayesianas. La inquietud de implementar un modelo con estas características surge de la necesidad de pronosticar o dar una explicación a futuro utilizando datos actuales y así conocer el impacto de la mortalidad por diabetes en la población mexicana esto para poder cuantificar y prever posibles afectaciones derivadas de esto en el sector salud principalmente.

3.1. El problema inicial

Se tiene una base de datos que contiene el número de muertes por diabetes ocurridos de 1979 a 1998 desglosados por edad de ocurrencia. Adicionalmente se cuenta con la supervivencia general de la población comprendida en los mismos años y desglosada por edad de igual manera. También se cuenta con los nacidos vivos para los años comprendidos entre 1893 y 1998. Con esta información se construye una base de datos que contendrá las muertes por diabetes para una cohorte determinada así como el número de vivos a edad x para cada cohorte. Con base en la información que se cuenta, se desea encontrar las variaciones que ha sufrido la función de supervivencia a través del tiempo, las posibles tendencias de esta y

una vez determinado lo anterior, pronosticar comportamientos.

Denotemos con $M_j^d(x)$ al número de muertes por diabetes a edad x de la j -ésima cohorte y con $L_j(x)$ al número de vivos a edad x de la j -ésima cohorte. El horizonte temporal que se maneja es de 100 años, medidos de uno en uno, esto debido a que es la forma más práctica en que se pueden hacer las mediciones y captaciones correspondientes. Se supone que los vivos se encuentran a inicio del año en cuestión y los muertos son registrados al final del año.

De esta forma los datos quedan acomodados en el siguiente arreglo

Año\Cohorte	1	2	...	j	...	87	...	106
0						$M_{87}^d(0)$...	$M_{106}^d(0)$
1					...	$M_{87}^d(1)$...	
⋮				
x			...	$M_j^d(x)$...			
⋮					
86	$M_1^d(86)$	$M_2^d(86)$...					
⋮	⋮	...						
105	$M_1^d(105)$							

Un arreglo similar se presenta para el número de vivos a edad x , con la única diferencia de que para este arreglo el número de vivos a edad 0 es conocido para todas las cohortes.

Dadas las características de la información (la ventana de observación y el desglose implementados), se presenta incompleta para ciertos periodos de tiempo por cohorte, la parte triangular superior del arreglo hace referencia a datos pasados y la parte inferior del arreglo presentado, cuyos datos no son observados aún. Esto comprende un problema para una adecuada implementación de algún método. Para subsanar lo anterior se procede a hacer algunos supuestos sobre el comportamiento (pasado) tanto del número de vivos a edad x como del número de muertes por diabetes observadas a edad x .

Estimación del número de vivos

Se desea encontrar el número de vivos para el triángulo superior de la matriz. Para esto se procede como sigue, se supone que el porcentaje de decrementos se mantiene sin

variaciones considerables durante periodos de tres años, entonces, la estimación procede como sigue: $L_j(x) = L_j(x + 1) * promedio(\frac{L_{j+1}(x)}{L_{j+1}(x+1)}, \frac{L_{j+2}(x)}{L_{j+2}(x+1)}, \frac{L_{j+3}(x)}{L_{j+3}(x+1)})$, tomando el número entero debido a que esta estimación representa el número de vivos a edad x . Para llenar completamente todo el triángulo superior se debe aplicar dicha estimación diagonalmente, es decir, no es posible estimar $L_j(x)$ y después $L_j(x - 1)$ ya que no se cuenta con la información de $L_{j+1}(x - 1)$ la cual es necesaria para la estimación correspondiente. Así, el llenado de la información faltante se sigue de la siguiente manera: $L_1(x), L_2(x - 1), L_3(x - 2), \dots, L_j(x - j + 1), \dots$, siempre que dicho valor sea consistente con los índices de edad y número de cohortes.

Estimación del número de muertes por diabetes

De manera análoga se realiza la estimación para el número de muertes por diabetes, únicamente que lo que se supone constante por periodos de tres años es la tasa de muerte por diabetes. Así que para encontrar el número de muertes por diabetes correspondiente se multiplica el número de nacidos vivos en la j -ésima cohorte por la tasa de muerte respectiva, la cual es calculada como sigue: $M_j(x) = M_j(x + 1) * promedio(\frac{M_{j+1}(x)}{M_{j+1}(x+1)}, \frac{M_{j+2}(x)}{M_{j+2}(x+1)}, \frac{M_{j+3}(x)}{M_{j+3}(x+1)})$ donde $M_j(x)$ es la proporción de muertos por diabetes a edad x de la j -ésima cohorte, que es obtenida mediante $M_j^d(x)/L_j(0)$ para aquellos valores donde esta información exista. Una vez que se obtienen todas las estimaciones correspondientes, el número de muertes por diabetes se obtiene al multiplicar $L_j(0)$ por su respectiva tasa de muerte.

Al realizar esto con las ciento cinco cohortes correspondientes a los años comprendidos entre 1893 y 1998 se presentan ciertas perturbaciones en la función de supervivencia (deja de ser monótona decreciente), debiéndose principalmente a errores propios de los censos de los que se obtuvo la información. Estos errores se presentan principalmente en los años cuya terminación es 5 o 0. Este fenómeno se ha estudiado en demografía y se debe principalmente a la tendencia de redondear hacia esas cifras la edad declarada, lo que causa diferencias entre la estimación de la supervivencia de la cohorte inicial y la supervivencia observada años más tarde. Para aminorar este impacto en la función de supervivencia se agrupan las cohortes de forma quinquenal. Las siguientes secciones abordan el problema y se implementan diversas aproximaciones para la modelación de la función de supervivencia.

3.2. Primeras aproximaciones

Un primer acercamiento a la función de supervivencia de forma heurística es considerar directamente los eventos a analizar, sea SD_x el evento de sobrevivir a la diabetes a edad x y D_x el evento de morir por diabetes a edad x , la relación entre estos se define mediante

$$\{SD_x\} \equiv \{D_1^c\} \cap \{D_2^c\} \cap \dots \cap \{D_{x-1}^c\}.$$

De manera similar se define $\{D_x\}$ en términos de dos eventos. Sean $\{A_{x-1}\}$ el evento de llegar con vida a edad $x-1$ y $\{B_x\}$ el evento de morir por diabetes entre las edades $x-1$ y x . Por tanto

$$\{D_x\} \equiv \{A_{x-1}\} \cap \{B_x\}.$$

Tomando probabilidades

$$\begin{aligned} P(D_x) &= P(A_{x-1} \cap B_x) \\ &= P(B_x|A_{x-1})P(A_{x-1}). \end{aligned}$$

De este modo los estimadores de la probabilidad de morir por diabetes a edad x en la j -ésima cohorte en términos de los datos con los que se cuenta, se expresa de la siguiente manera

$$\begin{aligned} P(D_x^j) &= P(B_x^j|A_{x-1}^j)P(A_{x-1}^j) \\ &= \frac{M_j^d(x)}{L_j(x-1)} \frac{L_j(x-1)}{L_j(0)}, \quad \forall x \in [1, \omega], \end{aligned} \quad (3.1)$$

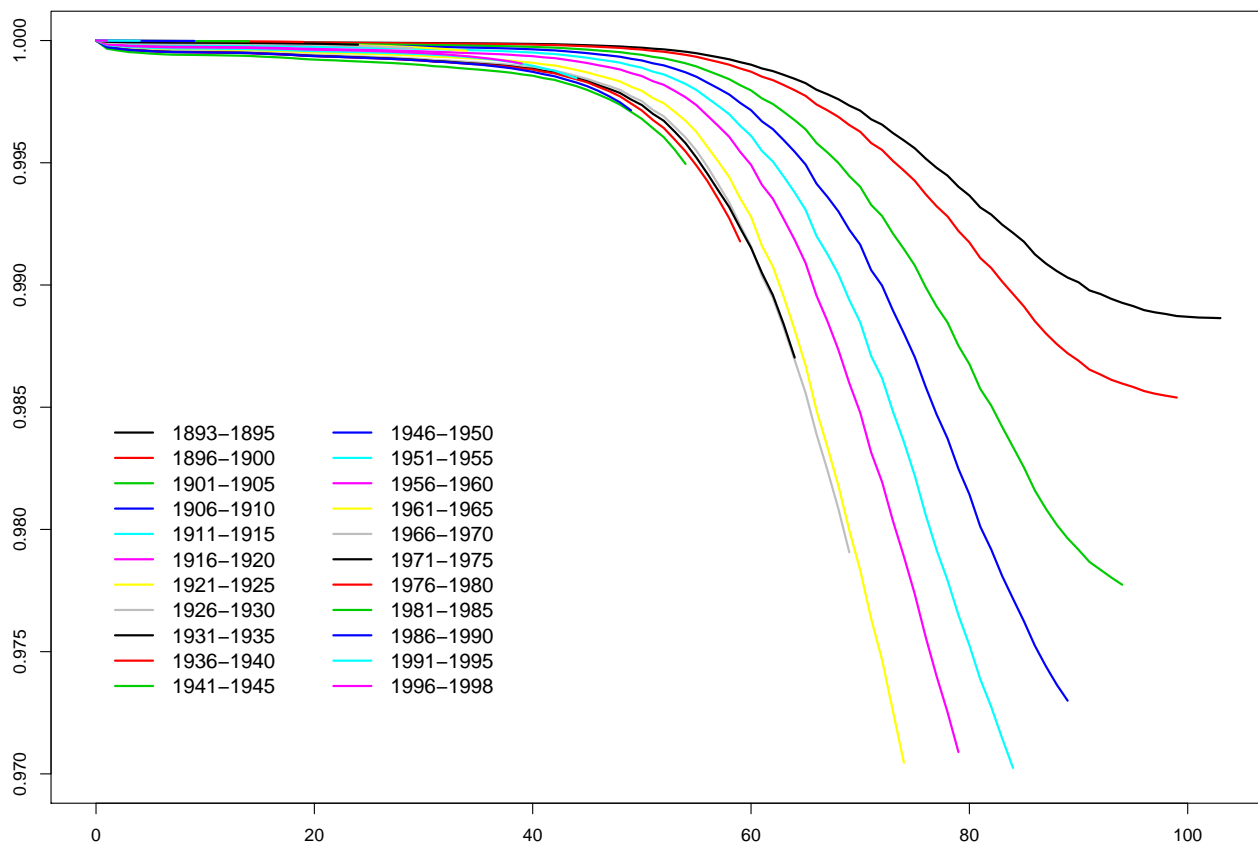
donde ω es la edad máxima de muerte observada.

Así, la función de supervivencia a edad x para la j -ésima cohorte:

$$\begin{aligned} P(SD_x^j) &= P(D_1^{j^c} \cap D_2^{j^c} \cap \dots \cap D_{x-1}^{j^c}) \\ &= \prod_{t=1}^{x-1} P(D_t^{j^c}) \\ &= \prod_{t=1}^{x-1} \{1 - P(D_t^j)\} \\ &= \prod_{t=1}^{x-1} \left\{1 - \frac{M_j^d(t)}{L_j(t-1)} \frac{L_j(t-1)}{L_j(0)}\right\}. \end{aligned} \quad (3.2)$$

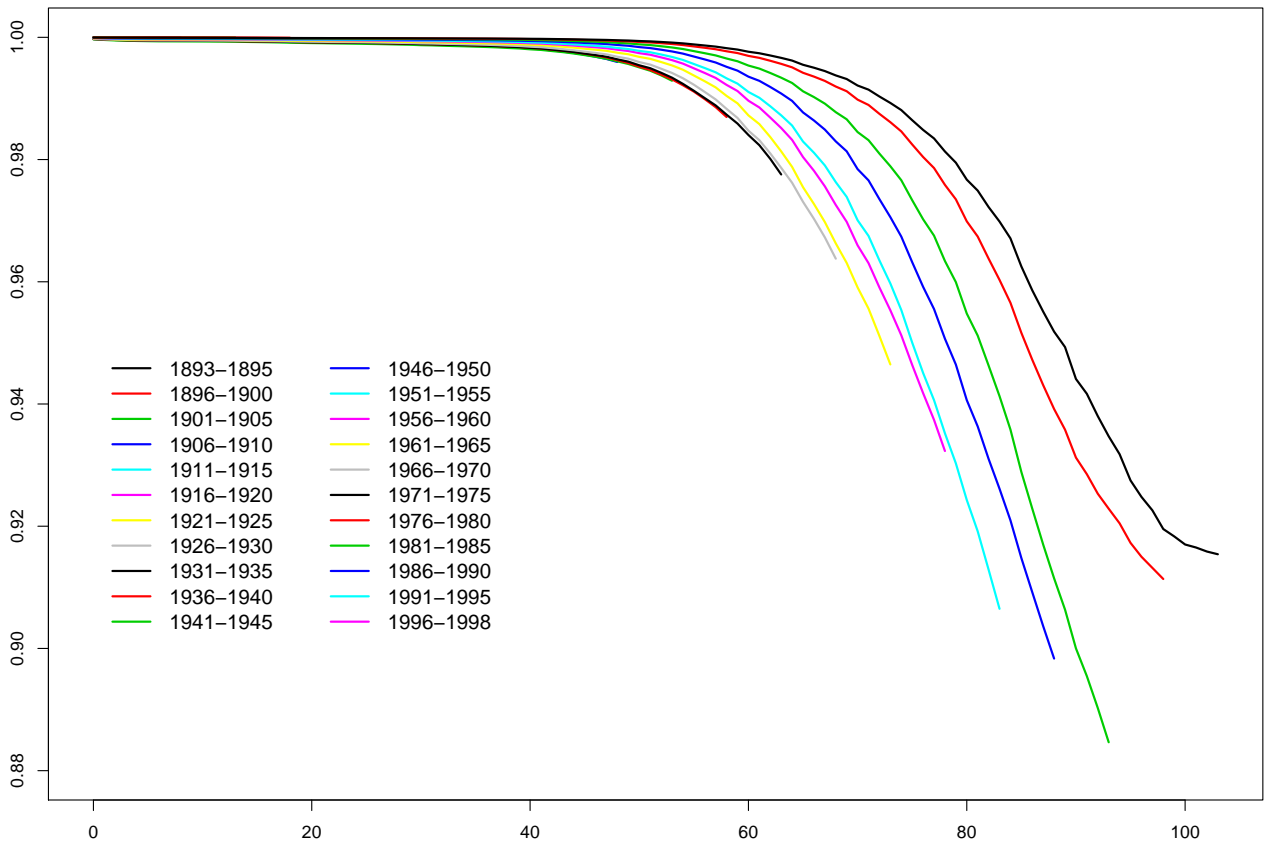
Los curvas resultantes de este enfoque se muestran en la figura (3.1).

Figura 3.1: Estimación de la función de supervivencia aproximación condicionada



Otra aproximación más frecuentemente utilizada está dada por el estimador Kaplan-Meier, el cual da una forma gráfica a la curva de supervivencia y de manera empírica muestra el comportamiento de la supervivencia de las diferentes cohortes. De primera instancia se utilizará este estimador, el cual es uno de los más utilizados para los estudios de esta naturaleza. Como ya se presentó en el capítulo I, este estimador está dado por: $\hat{S}_{K-M}^j(t) = \prod_{t=1}^n \left\{ 1 - \frac{M_j^d(t)}{L_j(t)} \right\}$ donde j es el número de quinquenio correspondiente. Las curvas resultantes se muestran en la figura (3.2).

Figura 3.2: Estimación de la función de supervivencia, estimador Kaplan-Meier



En ambos gráficos se alcanza a observar que las curvas tienden a decrecer más rápidamente conforme se presentan los quinquenios, en el primer enfoque este decremento es más marcado a pesar de que la supervivencia total no se ve afectada tanto como en el segundo enfoque, cuyas curvas se encuentran más cercanas.

Como se observa en las gráficas de ambos enfoques a pesar de que alcanzan diferentes valores, se muestran comportamientos similares con respecto al incremento en el riesgo de morir por diabetes conforme los quinquenios van avanzando, pero debido al enfoque utilizado no es posible cuantificar exactamente de cuanto son estos incrementos.

Tras realizar estos primeros esbozos se generan cuestionamientos con relación a la relevancia que estos análisis presentan, esto después de utilizar algunos supuestos como el

comportamiento previo de las muertes por diabetes y la estimación utilizada, lo cual puede traer consigo serias desaveniencias en cuanto a su implementación, de igual forma el poder predictivo de estos métodos es limitado o nulo debido a la naturaleza misma de su enfoque al ser este de mayor carácter descriptivo.

3.3. Aproximación bajo un esquema bayesiano

Nieto y Walker (2001), publican un trabajo sobre modelación de tasas de riesgo utilizando un enfoque bayesiano con procesos de markov beta y gamma, que se enfocan en la generalización del proceso beta en tiempo discreto con incrementos independientes que propuso Hjort (1990) y en generalizar el proceso gamma independiente para modelar tasas de riesgo constantes en intervalos discretos propuesto por Walker y Mallick (1997); la principal aportación de este modelo se basa en la incorporación de procesos a priori con incrementos Markovianos, debido a que no siempre es apropiada una estructura de incrementos independientes, es decir que no es realista suponer que dos tasas de riesgo consecutivas carecen de relación entre ellas. Este modelo cambia esta perspectiva al incorporar elementos de dependencia entre tasas mediante procesos de Markov, lo cual enriquece los modelos propuestos.

El interés en este modelo radica en que para la función de supervivencia no es necesaria una distribución en específico sino que trabaja con la expresión general de esta; a su vez utiliza intervalos de tiempo discretos (o discretizados) y esto se ajusta bien a lo que el problema original presenta, debido a las características de la información con que se cuenta; de igual manera, el modelo es flexible en cuanto a la incorporación de mayor información (cualidad en los enfoques bayesianos) lo cual será de gran ayuda para evitar supuestos excesivos en el comportamiento de los datos.

Procesos Markov Beta y Gamma para modelar tasas de riesgo

Antes de iniciar se introduce la notación que será utilizada. Sean $ga(\alpha, \beta)$ una densidad gamma con media α/β ; $po(\lambda)$ una función de probabilidad poisson con media λ ; $po-ga(\alpha, \beta, c)$ una densidad poisson-gamma con media $c\alpha/\beta$; $be(\alpha, \beta)$ una densidad beta con media $\alpha/(\alpha + \beta)$; $bi(c, \pi)$ una función de probabilidad binomial con $0 < \pi < 1$ y media

$c\pi$; $bi-be(\alpha, \beta, c)$ una densidad binomial-beta con media $c\alpha/(\alpha + \beta)$; $\Gamma(\cdot)$ la función gamma y T_1, \dots, T_n tiempos de falla independientes -posiblemente con censura por la derecha incluida-.

A continuación se presentan ambos enfoques, el proceso gamma en (1) y el proceso beta en (2).

(1) La idea consiste en la partición del eje temporal en intervalos, así $0 = \tau_0 < \tau_2 < \dots$, y se define λ_k a la tasa de riesgo en el intervalo $(\tau_{k-1}, \tau_k]$. Aquí Walker y Mallick (1997) propusieron una densidad inicial de tal forma que se toma $\{\lambda_k\}$ como variables aleatorias independientes con distribución gamma; para el siguiente trabajo se propone de manera alternativa un proceso Markov como a priori que generaliza la idea de variables gamma independientes y a su vez mantiene a la a priori gamma como conjugable. Sobre algunas otras aproximaciones mediante procesos de Markov, ver Gamerman (1991), Gray (1994) y Arjas y Gasbarra (1994).

(2) La idea consiste en tiempos de falla discretos, los cuales ocurren en el conjunto $\{\tau_1, \tau_2, \dots\}$ y la tasa de riesgo al tiempo τ_k está dada por π_k ; esto es $P(T = \tau_k | T \geq \tau_k) = \pi_k$. Hjort (1990) tomó a priori independientes y con distribución beta para $\{\pi_k\}$ que resulta conjugable y el estimador bayes generaliza al estimador Nelson-Aalen; para el siguiente trabajo se propone un proceso de Markov a priori para modelar las $\{\pi_k\}$ lo cual generaliza el modelo de procesos beta independientes propuesto por Hjort.

Procesos de Markov beta y gamma

Para ambas aproximaciones. Sea θ_k la tasa de riesgo (λ_k o π_k). Para tener una interpretación del modelo de Markov, es de importancia asegurar que:

$$E[\theta_{k+1} | \theta_k] = a + b\theta_k,$$

donde (a, b) son fijos y dependen de k . No se asegura que marginalmente θ_k sea beta o gamma sino cuando $\{\theta_k\}$ son independientes es cuando estas se vuelven beta o gamma.

Esta aproximación sirve para introducir un proceso latente $\{u_k\}$ tal que $\{\theta_k\}$ provenga de:

$$\theta_1 \rightarrow u_1 \rightarrow \theta_2 \rightarrow u_2 \rightarrow \dots$$

donde se asignan las densidades condicionales $f(u|\theta)$ y $f(\theta|u)$ basadas en la densidad conjunta $f(u, \theta)$.

(1) Se comienza con $\lambda_1 \sim ga(\alpha_1, \beta_1)$ y se toman $u_k|\lambda_k \sim po(c_k \lambda_k)$ y $\lambda_{k+1}|u_k \sim ga(\alpha_{k+1} + u_k, \beta_{k+1} + c_k)$ y así sucesivamente. Estas actualizaciones surgen de la densidad conjunta,

$$f(u, \lambda) = ga(\lambda|\alpha, \beta)po(u|c\lambda)$$

el cual es parecido a una actualización de un muestreo de Gibbs, la diferencia es que los parámetros (α, β, c) están cambiando en cada iteración y por tanto la cadena no es estacionaria y marginalmente $\{\lambda_k\}$ no son gammas. De cualquier forma

$$E[\lambda_{k+1}|u_k] = \frac{\alpha_{k+1} + u_k}{\beta_{k+1} + c_k}$$

entonces

$$E[\lambda_{k+1}|\lambda_k] = \frac{\alpha_{k+1} + c_k \lambda_k}{\beta_{k+1} + c_k}.$$

Si $c_k = 0$ entonces $P(u_k = 0) = 1$ y de aquí $\{\lambda_k\}$ son variables gamma independientes. También es sencillo mostrar que

$$Var[\lambda_{k+1}|\lambda_k] = \frac{\alpha_{k+1} + 2c_k \lambda_k}{(\beta_{k+1} + c_k)^2},$$

por tanto si α_{k+1} y β_{k+1} son pequeños, lo cual es lo más recomendable a menos de que se cuente con información a priori bien sustentada, entonces

$$E[\lambda_{k+1}|\lambda_k] \approx \lambda_k$$

y

$$Var[\lambda_{k+1}|\lambda_k] \approx 2\lambda_k/c_k$$

para $c_k > 0$. En el caso que se tome $\alpha_k = \alpha_1$ y $\beta_k = \beta_1$ constantes para toda k entonces el proceso $\{u_k\}$ es poisson-gamma con marginales $u_k \sim po-ga(\alpha_1, \beta_1, c_k)$. Más aún el proceso $\{\lambda_k\}$ es estrictamente estacionario y marginalmente $\lambda_k \sim ga(\alpha_1, \beta_1)$.

(2) Se comienza con $\pi_1 \sim be(\alpha_1, \beta_1)$ y se toman $u_k|\pi_k \sim bi(c_k, \pi_k)$ y $\pi_{k+1}|u_k \sim be(\alpha_{k+1} + u_k, \beta_{k+1} + c_k - u_k)$ y así sucesivamente. Esto surge de la conjugada de una binomial-beta con densidad conjunta

$$f(u, \pi) = be(\pi|\alpha, \beta)bi(u|c, \pi).$$

Así,

$$E[\pi_{k+1}|u_k] = \frac{\alpha_{k+1} + u_k}{\alpha_{k+1} + \beta_{k+1} + c_k}$$

entonces

$$E[\pi_{k+1}|\pi_k] = \frac{\alpha_{k+1} + c_k\pi_k}{\alpha_{k+1} + \beta_{k+1} + c_k}.$$

Como en el proceso gamma, si se toma $c_k = 0$ entonces $P(u_k = 0) = 1$ y entonces $\{\pi_k\}$ se vuelven variables beta independientes. Para α_{k+1} y β_{k+1} pequeños

$$E[\pi_{k+1}|\pi_k] \approx \pi_k$$

y

$$Var[\pi_{k+1}|\pi_k] \approx \frac{2\pi_k(1 - \pi_k)}{1 + c_k}$$

para $c_k > 0$. Si se toman $\alpha_k = \alpha_1$ y $\beta_k = \beta_1$ constantes para toda k , el proceso $\{u_k\}$ es un proceso binomial-beta con marginales $u_k \sim bi-be(\alpha_1, \beta_1, c_k)$. Y más aún el proceso $\{\pi_k\}$ se vuelve estrictamente estacionario y marginalmente $\pi_k \sim be(\alpha_1, \beta_1)$.

Análisis posterior

Dadas $\{u_k\}$, $\{\theta_k\}$ son independientes y debido a eso es posible actualizarlas utilizando los datos disponibles y $\{u_k\}$. Esto es, se tiene $f(\theta|u, \text{datos})$, donde $u = (u_1, u_2, \dots)$, $\theta = (\theta_1, \theta_2, \dots)$ y $f(\theta_k|u_{k-1}, u_k) \propto f(u_k|\theta_k)f(\theta_k|u_{k-1})$. También es posible obtener

$$f(u|\theta, \text{datos}) = \prod_k f(u_k|\theta_k, \theta_{k+1})$$

donde $f(u_k|\theta_k, \theta_{k+1}) \propto f(\theta_{k+1}|u_k)f(u_k|\theta_k)$. A continuación se presentan las particularidades de ambos procesos.

(1) Sea T una variable aleatoria continua con función de distribución $F(t) = P(T \leq t)$ definida en $[0, \infty)$. Se particiona el eje temporal en intervalos, sean $0 = \tau_0 < \tau_1 < \tau_2 < \dots$,

y sea λ_k la tasa de riesgo en el intervalo $(\tau_{k-1}, \tau_k]$ entonces la función de riesgo se encuentra dada por

$$\lambda(t) = \sum_{k=1}^{\infty} \lambda_k I_{(\tau_{k-1}, \tau_k]}(t).$$

Por tanto, la función de distribución y la función de densidad dadas $\{\lambda_k\}$, son

$$F(t|\{\lambda_k\}) = 1 - \exp\{-\Lambda(t)\}, \quad (3.3)$$

$$f(t|\{\lambda_k\}) = \lambda(t) \exp\{-\Lambda(t)\}, \quad (3.4)$$

donde $\Lambda(t) = \int_0^t \lambda(s) ds$. También se tiene que

$$f(\lambda_k | u_{k-1}, u_k) = ga(\alpha_k + u_{k-1} + u_k, \beta_k + c_{k-1} + c_k).$$

Entonces, dada una muestra T_1, T_2, \dots, T_n de $f(\cdot|\{\lambda_k\})$, se sigue que

$$f(\lambda_k | u_{k-1}, u_k, \text{datos}) = ga(\alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} + c_k + m_k), \quad (3.5)$$

donde n_k es el número de observaciones no censuradas en el intervalo $(\tau_{k-1}, \tau_k]$ y $m_k = \sum_i r_{ki}$; además

$$r_{ki} = \begin{cases} \tau_k - \tau_{k-1} & t_i > \tau_k \\ t_i - \tau_{k-1} & t_i \in (\tau_{k-1}, \tau_k] \\ 0 & \text{e.o.c.} \end{cases}$$

Adicionalmente se tiene que

$$P(u_k = u | \lambda_k, \lambda_{k+1}) \propto \frac{\{c_k(\beta_{k+1} + c_k)\lambda_k\lambda_{k+1}\}^{u_k}}{\Gamma(\alpha_{k+1} + u_k)\Gamma(u_k + 1)}, \quad (3.6)$$

con $u = 0, 1, 2, \dots$. De aquí es posible implementar un muestreo de Gibbs, dadas las condicionales totales anteriores con el fin de obtener análisis posteriores.

(2) Sea T una variable aleatoria discreta con valores en el conjunto $\{\tau_1, \tau_2, \dots\}$ con función de densidad $f(\tau_k) = P(T = \tau_k)$. Sea π_k la tasa de riesgo en el punto τ_k entonces la función de distribución y la función de densidad dadas $\{\pi_k\}$ están dadas por

$$F(\tau_j|\{\pi_k\}) = 1 - \prod_{k=1}^j (1 - \pi_k), \quad (3.7)$$

$$f(\tau_j|\{\pi_k\}) = \pi_j \prod_{k=1}^{j-1} (1 - \pi_k). \quad (3.8)$$

Por otro lado, la densidad condicional de π_k es

$$f(\pi_k|u_{k-1}, u_k) = be(\alpha_k + u_{k-1} + u_k, \beta_k + c_{k-1} - u_{k-1} + c_k - u_k).$$

Entonces, dada una muestra T_1, T_2, \dots, T_n de $f(\cdot|\{\pi_k\})$ se sigue que

$$f(\pi_k|u_{k-1}, u_k, \text{datos}) = be(\alpha_k + u_{k-1} + u_k + n_k, \beta_k - u_{k-1} + c_{k-1} - u_k + c_k + m_k), \quad (3.9)$$

donde n_k es el número de fallas en τ_k , $m_k = \sum_i r_{ki}$ y

$$r_{ki} = \begin{cases} 1 & t_i > \tau_k \\ 0 & \text{e.o.c.} \end{cases}$$

es decir, m_k denota al número de individuos sin fallas en τ_k . De forma análoga al caso anterior se tiene que

$$P(u_k = u|\pi_k, \pi_{k+1}) \propto \frac{(\pi_{k+1}\pi_k)^{u_k} [(1 - \pi_{k+1})(1 - \pi_k)]^{-u_k}}{\Gamma(\alpha_{k+1} + u_k)\Gamma(\beta_{k+1} + c_k - u_k)\Gamma(u_k + 1)\Gamma(c_k - u_k + 1)}, \quad (3.10)$$

con $u = 0, 1, \dots, c_k$. Al igual que con el proceso anterior, la implementación de un muestreo de Gibbs es viable utilizando las condicionales totales mostradas.

Ahora bien, este modelo es presentado para datos dentro de una misma población de seguimiento, en el caso que se presenta en este trabajo se aborda la supervivencia para varias poblaciones a seguir, las cuales están determinadas por los quinquenios en que fueron agrupados los datos. Aquí es donde se deben tener ciertas consideraciones con respecto a la dimensión que se le adiciona a los datos y el impacto que esto genera en el modelo a utilizar.

Es así que, a partir de los modelos presentados se realizan las consideraciones necesarias para adecuarlo al problema planteado originalmente, primero se toman los intervalos para el proceso gamma como $\tau_0 = 0$ y $\tau_k = \tau_{k-1} + h$, con $h = 1$ y para el proceso beta se considera el conjunto $\{\tau_0, \tau_1, \dots\}$ como $\{0, 1, \dots\}$.

Un problema también se presenta al determinar la forma en que será utilizada la información de los datos en referencia a su carácter predictivo, es decir, el como afecta la información de quinquenios anteriores en quinquenios posteriores. Para efectos de este trabajo la influencia de los quinquenios pasados estará dada por los parámetros α y β del quinquenio siguiente. Esto es para definir los parámetros de inicio donde no se cuenta con

información (parte inferior de la matriz de datos) se consideran los datos posteriores del quinquenio inmediato anterior. Así, la muestra generada por el muestreo de Gibbs en el quinquenio j para el año k será utilizada para el año k pero del quinquenio $j + 1$ siempre y cuando en la entrada $a_{(k,j+1)}$ de la matriz de datos no se cuente con información registrada.

3.4. Resultados

Proceso Gamma

El análisis se enfoca a la mortalidad en las edades posteriores a los 20 años, esto debido a que para edades inferiores es muy baja y para efectos de cualquier estudio esta no es realmente significativa. Tras realizar el muestreo de Gibbs para los 22 quinquenios y edades posteriores a 20 años utilizando $\alpha = 0.1$, $\beta = 0.1$, $c_k = 10$ constantes, un número de iteraciones de 5000 y un periodo de calentamiento de 3000 se obtienen los siguientes promedios ergódicos. Cabe señalar que por razones de espacio no se incluyen todas las gráficas sino los resultados para algunos quinquenios y años para observar el comportamiento general de las cadenas simuladas (gráficas C.1 a C.8).

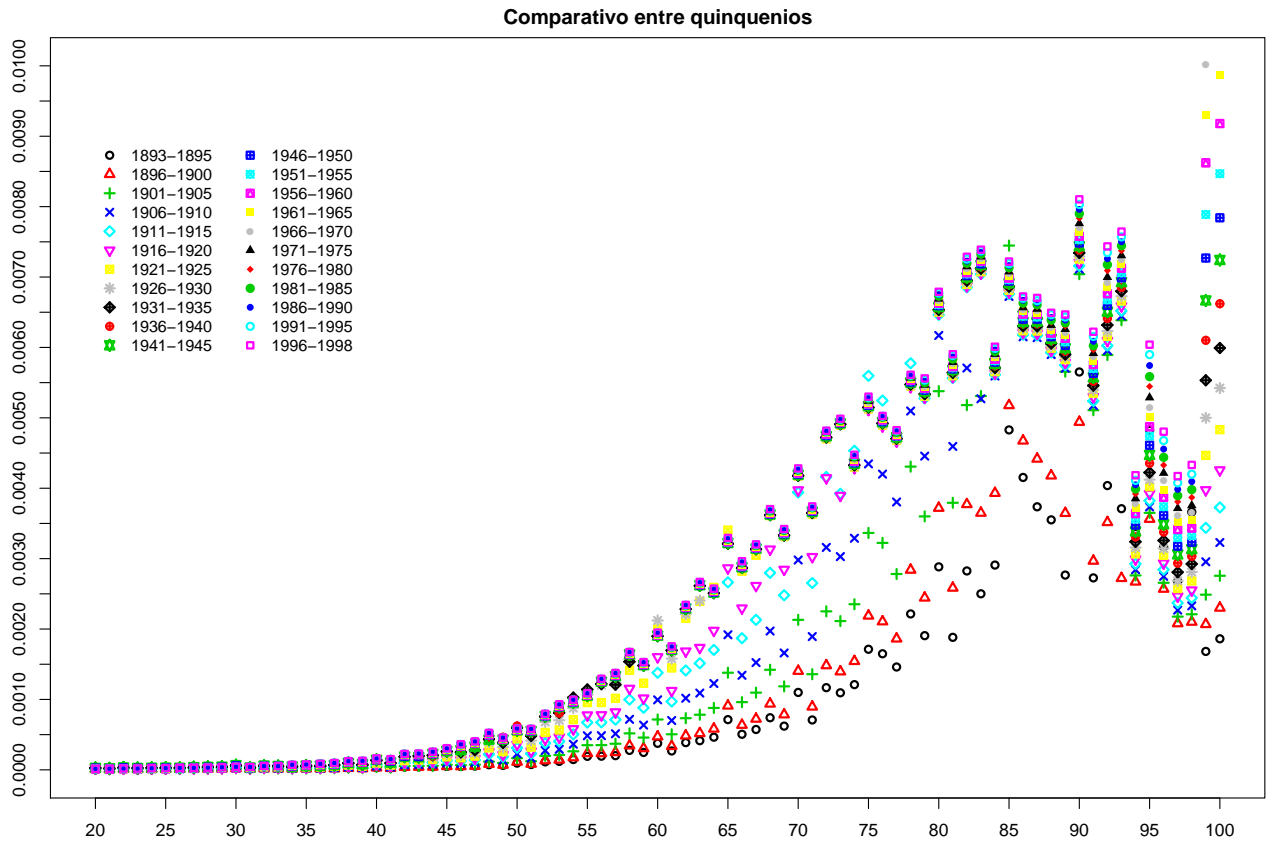
Proceso Beta

Para este proceso se utilizó $\alpha = 1$, $\beta = 1$, $c_k = 10$ constantes para todas las edades y quinquenios, un número de iteraciones de 5000 y un periodo de calentamiento de 3000 donde se obtienen los siguientes promedios ergódicos e histogramas. De nuevo se incluyen algunas las gráficas (C.9 a C.16) para mostrar el comportamiento para este caso.

Asimismo se realizan las comparaciones referentes a las tasas de riesgo para todos los quinquenios, los resultados se presentan en las gráficas 3.3 y 3.4.

Una vez que se obtienen todos los resultados se observan principalmente dos cosas: la primera es referente al comportamiento en sí de las tasas de riesgo, las cuales como se preveía van incrementando conforme los quinquenios avanzan, siendo de esta forma consistentes con lo que presentaban los anteriores enfoques. La segunda observación es en relación a que el enfoque bayesiano completa lo que las anteriores aproximaciones dejaron truncado, esto

Figura 3.3: Comparativo de tasas de riesgo modelo poisson-gamma todos los quinquenios



mediante una metodología bien sustentada teóricamente. Esto es importante porque hace posible realizar una cuantificación plausible del riesgo de muerte por diabetes en el futuro.

Adicionalmente se analiza la capacidad predictiva de los modelos, esto mediante la distribución predictiva posterior (Capítulo II 2.3). Se hizo una simulación de la distribución predictiva posterior, considerando la posterior de cada uno de los modelos para el quinquenio 1 y se comparo con los datos reales del quinquenio 2. Se decidió que estos quinquennios podían proporcionar una mejor óptica de la capacidad predictiva ya que son los quinquennios que mayor información real poseen. Los comparativos se observan en las gráficas 3.5 y 3.6.

Figura 3.4: Comparativo de tasas de riesgo modelo binomial-beta todos los quinquenios

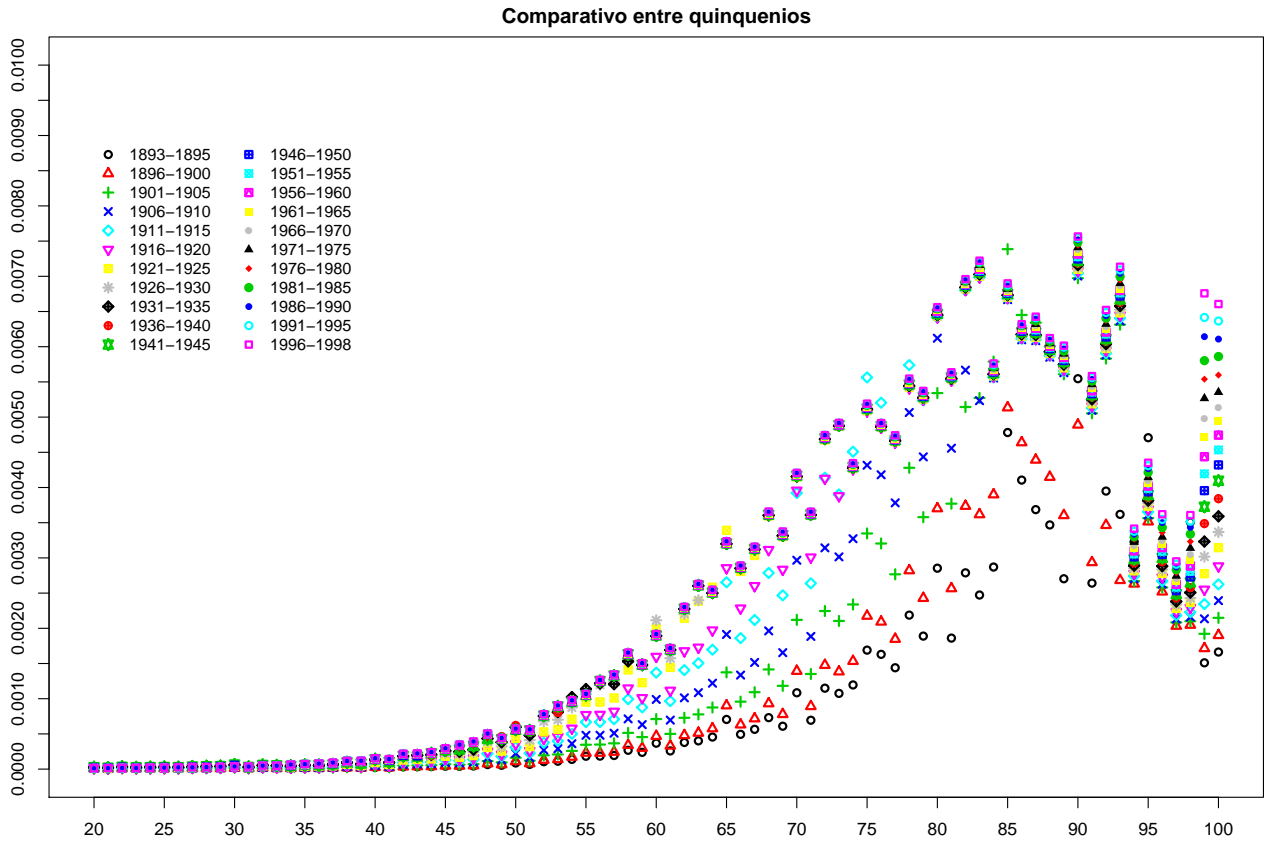


Figura 3.5: Predictiva posterior modelo poisson-gamma contra muertes reales

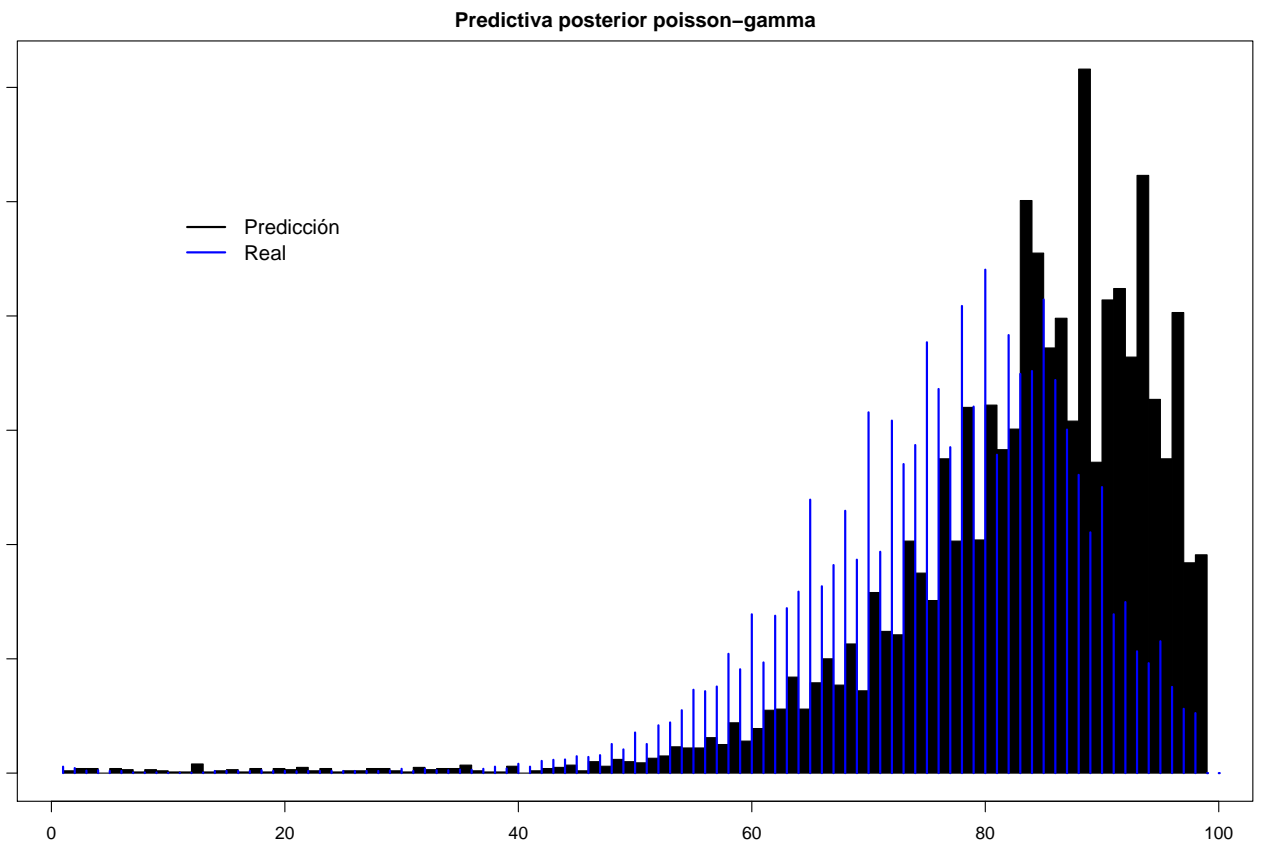
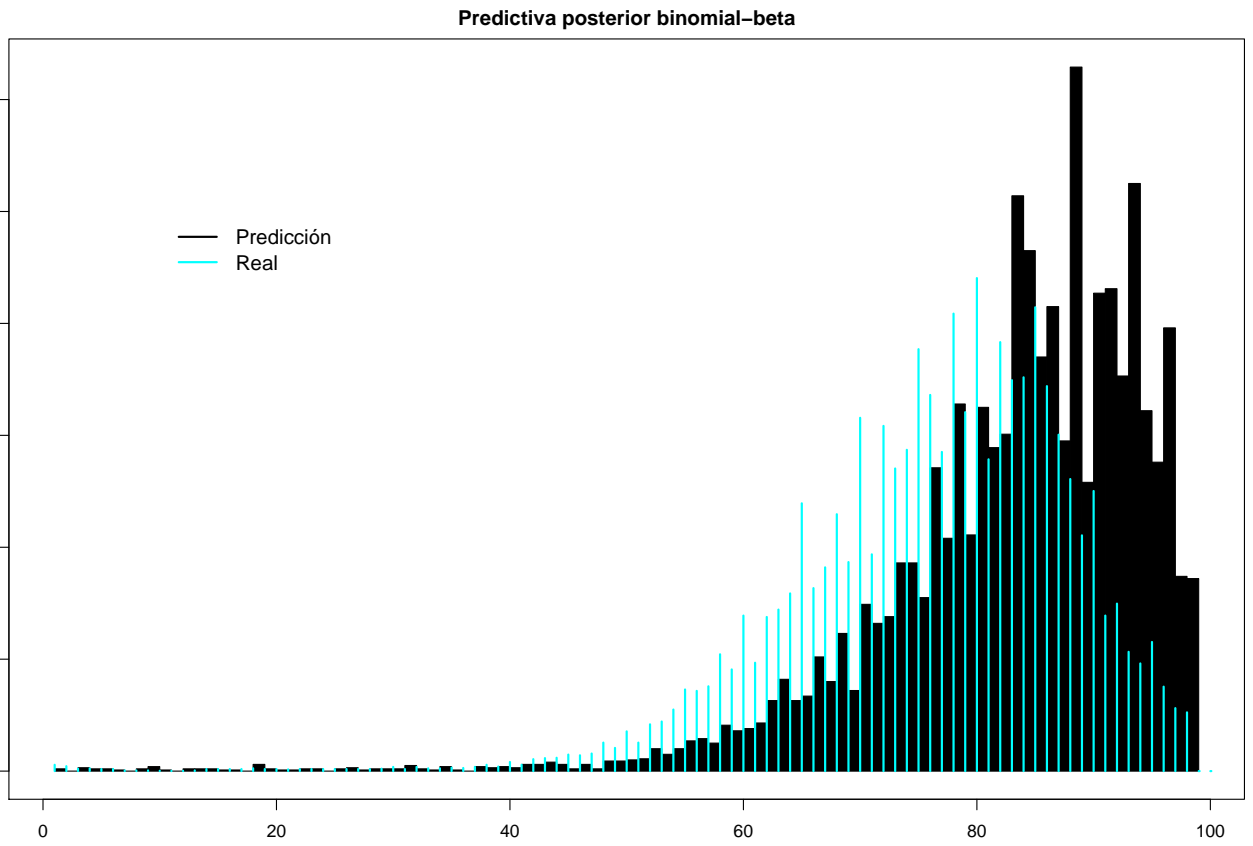


Figura 3.6: Predictiva posterior modelo binomial-beta contra muertes reales



Desarrollo de resultados

Resultado 3.5

$$\begin{aligned}
f(\lambda_k | u_{k-1}, u_k, \text{datos}) &= \frac{f(\lambda_k, u_{k-1}, u_k, \text{datos})}{f(u_{k-1}, u_k, \text{datos})} \\
&= \frac{f(\text{datos} | \lambda_k, u_{k-1}, u_k) f(\lambda_k, u_{k-1}, u_k)}{f(u_{k-1}, u_k, \text{datos})} \\
&= \frac{f(\text{datos} | \lambda_k) f(\lambda_k | u_{k-1}, u_k) f(u_{k-1}, u_k)}{f(u_{k-1}, u_k) f(\text{datos})} \\
&= \frac{f(\text{datos} | \lambda_k) f(\lambda_k | u_{k-1}, u_k)}{f(\text{datos})} \\
&\propto f(\text{datos} | \lambda_k) f(\lambda_k | u_{k-1}, u_k) \\
&= \lambda_k^{n_k} e^{-\lambda_k \sum_1^{n_k} t_i} \times \\
&\quad \frac{(\beta_k + c_{k-1} + c_k)^{\alpha_k + u_{k-1} + u_k} e^{-(\beta_k + c_{k-1} + c_k)\lambda_k} \lambda_k^{\alpha_k + u_{k-1} + u_k - 1}}{\Gamma(\alpha_k + u_{k-1} + u_k)} \\
&= \lambda_k^{\alpha_k + u_{k-1} + u_k + n_k - 1} e^{-\lambda_k (\beta_k + c_{k-1} + c_k + \sum_1^{n_k} t_i)} \times \\
&\quad \frac{(\beta_k + c_{k-1} + c_k)^{\alpha_k + u_{k-1} + u_k}}{\Gamma(\alpha_k + u_{k-1} + u_k)} \\
&= ga(\alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} + c_k + m_k).
\end{aligned}$$

Resultado 3.6

$$\begin{aligned}
P(u_k = u | \lambda_k, \lambda_{k+1}) &\propto ga(\alpha_{k+1} + u_k, \beta_{k+1} + c_k) po(c_k \lambda_k) \\
&= \frac{(\beta_{k+1} + c_k)^{\alpha_{k+1} + u_k} e^{-\lambda_{k+1} (\beta_{k+1} + c_k)} \lambda_{k+1}^{\alpha_{k+1} + u_k - 1} e^{-c_k \lambda_k (c_k \lambda_k)^{u_k}}}{\Gamma(\alpha_{k+1} + u_k) \Gamma(u_k + 1)} \\
&= \frac{\{c_k (\beta_{k+1} + c_k) \lambda_k \lambda_{k+1}\}^{u_k} e^{-(\lambda_{k+1} (\beta_{k+1} + c_k) + c_k \lambda_k)} \{(\beta_{k+1} + c_k) \lambda_{k+1}\}^{\alpha_k}}{\Gamma(\alpha_{k+1} + u_k) \Gamma(u_k + 1) \lambda_{k+1}} \\
&\propto \frac{\{c_k (\beta_{k+1} + c_k) \lambda_k \lambda_{k+1}\}^{u_k}}{\Gamma(\alpha_{k+1} + u_k) \Gamma(u_k + 1)}.
\end{aligned}$$

Resultado 3.9

$$\begin{aligned}
f(\pi_k|u_{k-1}, u_k, \text{datos}) &= \frac{f(\pi_k, u_{k-1}, u_k, \text{datos})}{f(u_{k-1}, u_k, \text{datos})} \\
&= \frac{f(\text{datos}|\pi_k, u_{k-1}, u_k)f(\pi_k, u_{k-1}, u_k)}{f(u_{k-1}, u_k, \text{datos})} \\
&= \frac{f(\text{datos}|\pi_k)f(\pi_k|u_{k-1}, u_k)f(u_{k-1}, u_k)}{f(u_{k-1}, u_k)f(\text{datos})} \\
&= \frac{f(\text{datos}|\pi_k)f(\pi_k|u_{k-1}, u_k)}{f(\text{datos})} \\
&\propto f(\text{datos}|\pi_k)f(\pi_k|u_{k-1}, u_k) \\
&= \pi_k^{n_k}(1 - \pi_k)^{m_k} \times \\
&\frac{\Gamma(\alpha_k\beta_k + c_{k-1} + c_k)\pi_k^{\alpha_k+u_{k-1}+u_k-1}(1 - \pi_k)^{\beta_k-u_{k-1}+c_{k-1}-u_k+c_k-1}}{\Gamma(\alpha_k + u_{k-1} + u_k)\Gamma(\beta_k - u_{k-1} + c_{k-1} - u_k + c_k)} \\
&= \frac{\Gamma(\alpha_k + \beta_k + c_{k-1} + c_k)}{\Gamma(\alpha_k + u_{k-1} + u_k)\Gamma(\beta_k - u_{k-1} + c_{k-1} - u_k + c_k)} \times \\
&\pi_k^{\alpha_k+u_{k-1}+u_k+n_k-1}(1 - \pi_k)^{\beta_k-u_{k-1}+c_{k-1}-u_k+c_k+m_k-1} \\
&= be(\alpha_k + u_{k-1} + u_k + n_k, \beta_k - u_{k-1} + c_{k-1} - u_k + c_k + m_k).
\end{aligned}$$

Resultado 3.10

$$\begin{aligned}
P(u_k = u|\pi_k, \pi_{k+1}) &\propto be(\alpha_{k+1} + u_k, \beta_{k+1} + c_k - u_k)bi(c_k, \pi_k) \\
&= \frac{\Gamma(\alpha_{k+1} + \beta_{k+1} + c_k)\pi_{k+1}^{\alpha_{k+1}+u_k-1}(1 - \pi_{k+1})^{\beta_{k+1}+c_k-u_k-1}}{\Gamma(\alpha_{k+1} + u_k)\Gamma(\beta_{k+1} + c_k - u_k)} \times \\
&\frac{\Gamma(c_k + 1)\pi_k^{u_k}(1 - \pi_k)^{c_k-u_k}}{\Gamma(u_k + 1)\Gamma(c_k - u_k + 1)} \\
&= \Gamma(\alpha_{k+1} + \beta_{k+1} + c_k)\Gamma(c_k + 1)\pi_{k+1}^{\alpha_{k+1}}(1 - \pi_{k+1})^{\beta_{k+1}+c_k-1} \times \\
&\frac{(\pi_{k+1}\pi_k)^{u_k} [(1 - \pi_{k+1})(1 - \pi_k)]^{-u_k}}{\Gamma(\alpha_{k+1} + u_k)\Gamma(\beta_{k+1} + c_k - u_k)\Gamma(u_k + 1)\Gamma(c_k - u_k + 1)} \\
&\propto \frac{(\pi_{k+1}\pi_k)^{u_k} [(1 - \pi_{k+1})(1 - \pi_k)]^{-u_k}}{\Gamma(\alpha_{k+1} + u_k)\Gamma(\beta_{k+1} + c_k - u_k)\Gamma(u_k + 1)\Gamma(c_k - u_k + 1)}.
\end{aligned}$$

Código del muestreo de Gibbs

El siguiente código fue implementado en el paquete estadístico R para realizar el muestreo de Gibbs dadas las condicionales totales que se presentaron en el modelo poisson-gamma y binomial-beta.

```
#Modelo Poisson-Gamma
a1<-.1
b1<-.1
c_k<-10
N<-5100
B<-3100
du_pg<-function(x,a1,be,ck,lk,lk1)
#función de densidad de la condicional total de u del modelo poisson-gamma
{
co=double(1)
for (y in 0:(2*ck)) co<-co+exp(y*(log(ck)+log(ck+be)+log(lk)+log(lk1)));
-lgamma(y+1)-lgamma(a1+y))
res<-exp(x*(log(ck)+log(ck+be)+log(lk)+log(lk1))-lgamma(x+1);
-lgamma(a1+x))/co
res
}
setwd("") #determina el directorio donde se encuentran las bases de datos
set.seed(147) #fija una semilla para la generación de aleatorios
#lee los datos con los que se cuenta
```

```

Md<-read.csv("Md(x)_quinq.csv",header=TRUE,sep = ",")
S<-read.csv("L(x)_quinq.csv",header=TRUE,sep = ",")
#esta parte llena la matriz de alfas y betas
am<-read.csv("Alpha_quinq.csv",header=TRUE,sep = ",")
bm<-read.csv("Beta_quinq.csv",header=TRUE,sep = ",")
#asignando espacio y valores a las variables a utilizar
qui<-22 #número de quinquenios
anios<-102 #número de años
theta<-c(rep(0,N*qui*anios))
dim(theta)<-c(anios,qui,N)
u<-c(rep(0,N*qui*anios))
dim(u)<-c(anios,qui,N)
ck<-c(rep(c_k,anios))
#llena la matriz de alfas y betas con las iniciales dadas
for (i in 1:qui)
{
  for (j in 1:anios)
  {
    if(am[j,i+1]!=-1) am[j,i+1]=a1
    if(bm[j,i+1]!=-1) bm[j,i+1]=b1
  }
}
#Aqui comienza el muestreo de Gibbs para el modelo poisson-gamma
for (j in 1:qui)
{
  for (k in 1:anios)
  {
    if(am[k,j+1]==-1)
    {
      #Estima los parámetros alfa y beta de una distribución gamma con
      #base en la información previa cuando no se cuenta con datos para
      #realizar la simulación.
      am[k,j+1]=a1
    }
  }
}

```

```

Md[k,j+1]=median(theta[k,j-1,(B+1):N])^2/(mad(theta[k,j-1,(B+1):N])^2)
bm[k,j+1]=b1
S[k,j+1]=median(theta[k,j-1,(B+1):N])/mad(theta[k,j-1,(B+1):N])^2
}
}
theta[1,j,1]=rgamma(1,am[1,j+1],bm[1,j+1])
for (k in 2:anios)
{
u[k-1,j,1]=rpois(1,ck[k-1]*theta[k-1,j,1])
theta[k,j,1]=rgamma(1,(am[k,j+1]+u[k-1,j,1]),(bm[k,j+1]+ck[k-1]))
}
u[anios,j,1]=rpois(1,ck[anios]*theta[anios,j,1])
for (i in 2:N)
{
theta[1,j,i]=rgamma(1,am[1,j+1]+u[1,j,i-1]+Md[1,j+1],bm[1,j+1]+ck[1];
+S[1,j+1])
for (k in 2:anios)
{
theta[k,j,i]=rgamma(1,am[k,j+1]+u[k-1,j,i-1]+u[k,j,i-1]+Md[k,j+1];
,bm[k,j+1]+ck[k-1]+ck[k]+S[k,j+1])
}
for (k in 1:(anios-1))
{
x=rpois(20,ck[k])
u[k,j,i]=sample(x,1,prob=du_pg(x,am[k+1,j+1],bm[k+1,j+1],ck[k];
,theta[k,j,i],theta[k+1,j,i]))
}
x=rpois(20,ck[anios])
u[anios,j,i]=sample(x,1,prob=du_pg(x,am[anios,j+1],bm[anios,j+1],ck[anios];
,theta[anios,j,i],theta[anios,j,i]))
}
}

```

```

#Modelo Binomial-Beta
a1<-.1
b1<-.1
c_k<-10
N<-5100
B<-3100
du_bb<-function(x,al,be,ck,lk,lk1)
#función de densidad de la condicional total de u del modelo binomial-beta
{
co=0
for (y in 0:ck) co<-co+exp(y*(log(lk)+log(lk1)-log(1-lk)-log(1-lk1)));
-lgamma(y+1)-lgamma(ck-y+1)-lgamma(al+y)-lgamma(be+ck-y))
res<-exp(x*(log(lk)+log(lk1)-log(1-lk)-log(1-lk1))-lgamma(x+1);
-lgamma(ck-x+1)-lgamma(al+x)-lgamma(be+ck-x))/co
res
}
setwd("") #determina el directorio donde se encuentran las bases de datos
set.seed(147) #fija una semilla para la generación de aleatorios
#lee los datos con los que se cuenta
Md<-read.csv("Md(x)_quinq.csv",header=TRUE,sep = ",")
S<-read.csv("L(x)_quinq.csv",header=TRUE,sep = ",")
#esta parte llena la matriz de alfas y betas
am<-read.csv("Alpha_quinq.csv",header=TRUE,sep = ",")
bm<-read.csv("Beta_quinq.csv",header=TRUE,sep = ",")
#asignando espacio y valores a las variables a utilizar
qui<-22 #número de quinquenios
anios<-102 #número de años
theta<-c(rep(0,N*qui*anios))
dim(theta)<-c(anios,qui,N)
u<-c(rep(0,N*qui*anios))
dim(u)<-c(anios,qui,N)

```



```

ck<-c(rep(c_k,anios))
#llena la matriz de alfas y betas con las iniciales dadas
for (i in 1:qui)
{
  for (j in 1:anios)
  {
    if(am[j,i+1]!=-1) am[j,i+1]=a1
    if(bm[j,i+1]!=-1) bm[j,i+1]=b1
  }
}
#Aqui comienza el muestreo de Gibbs para el modelo binomial-beta
for (j in 1:qui)
{
  for (k in 1:anios)
  {
    if(am[k,j+1]==-1)
    {
      #Estima los parámetros alfas y betas de una distribución beta con
      #base en la información previa cuando no se cuenta con datos para
      #realizar la simulación.
      am[k,j+1]=a1
      Md[k,j+1]=median(theta[k,j-1,(B+1):N])*(median(theta[k,j-1,(B+1):N]));
      -mad(theta[k,j-1,(B+1):N])^2/mad(theta[k,j-1,(B+1):N])^2
      bm[k,j+1]=b1
      S[k,j+1]=(1-median(theta[k,j-1,(B+1):N]))*(median(theta[k,j-1,(B+1):N]));
      -mad(theta[k,j-1,(B+1):N])^2/mad(theta[k,j-1,(B+1):N])^2
    }
  }
theta[1,j,1]=rbeta(1,am[1,j+1],bm[1,j+1])
for (k in 2:anios)
{
  u[k-1,j,1]=rbinom(1,ck[k-1],theta[k-1,j,1])
  theta[k,j,1]=rbeta(1,am[k,j+1]+u[k-1,j,1],bm[k,j+1]+ck[k-1]-u[k-1,j,1])
}

```

```

    }
u[anios,j,1]=rbinom(1,ck[anios],theta[anios,j,1])
for (i in 2:N)
  {
theta[1,j,i]=rbeta(1,am[1,j+1]+(u[1,j,i-1]+Md[1,j+1])/F1,bm[1,j+1]+(ck[1];
-u[1,j,i-1]+S[1,j+1])/F1)
for (k in 2:anios)
  {
theta[k,j,i]=rbeta(1,am[k,j+1]+u[k-1,j,i-1]+u[k,j,i-1]+Md[k,j+1];
,bm[k,j+1]+ck[k-1]+ck[k]-u[k-1,j,i-1]-u[k,j,i-1]+S[k,j+1])
  }
for (k in 1:(anios-1))
  {
x=rbinom(20,ck[k],.5)
u[k,j,i]=sample(x,1,prob=du_bb(x,am[k+1,j+1],bm[k+1,j+1],ck[k];
,theta[k,j,i],theta[k+1,j,i]))
  }
x=rbinom(20,ck[anios],.5)
u[anios,j,i]=sample(x,1,prob=du_bb(x,am[anios,j+1],bm[anios,j+1],ck[anios];
,theta[anios,j,i],theta[anios,j,i]))
  }
}

```

Apéndice C

Resultados de diagnóstico

A continuación se presentan algunas de los resultados generados por el algoritmo para el modelo poisson-gamma y binomial beta; histogramas y promedios ergódicos para diferentes edades y quinquenios. Debido a la cantidad de años y quinquenios simulados es imposible mostrarlos todos en este apéndice.

Figura C.1: Promedios ergódicos para la tasa de riesgo, modelo poisson-gamma, edad 20

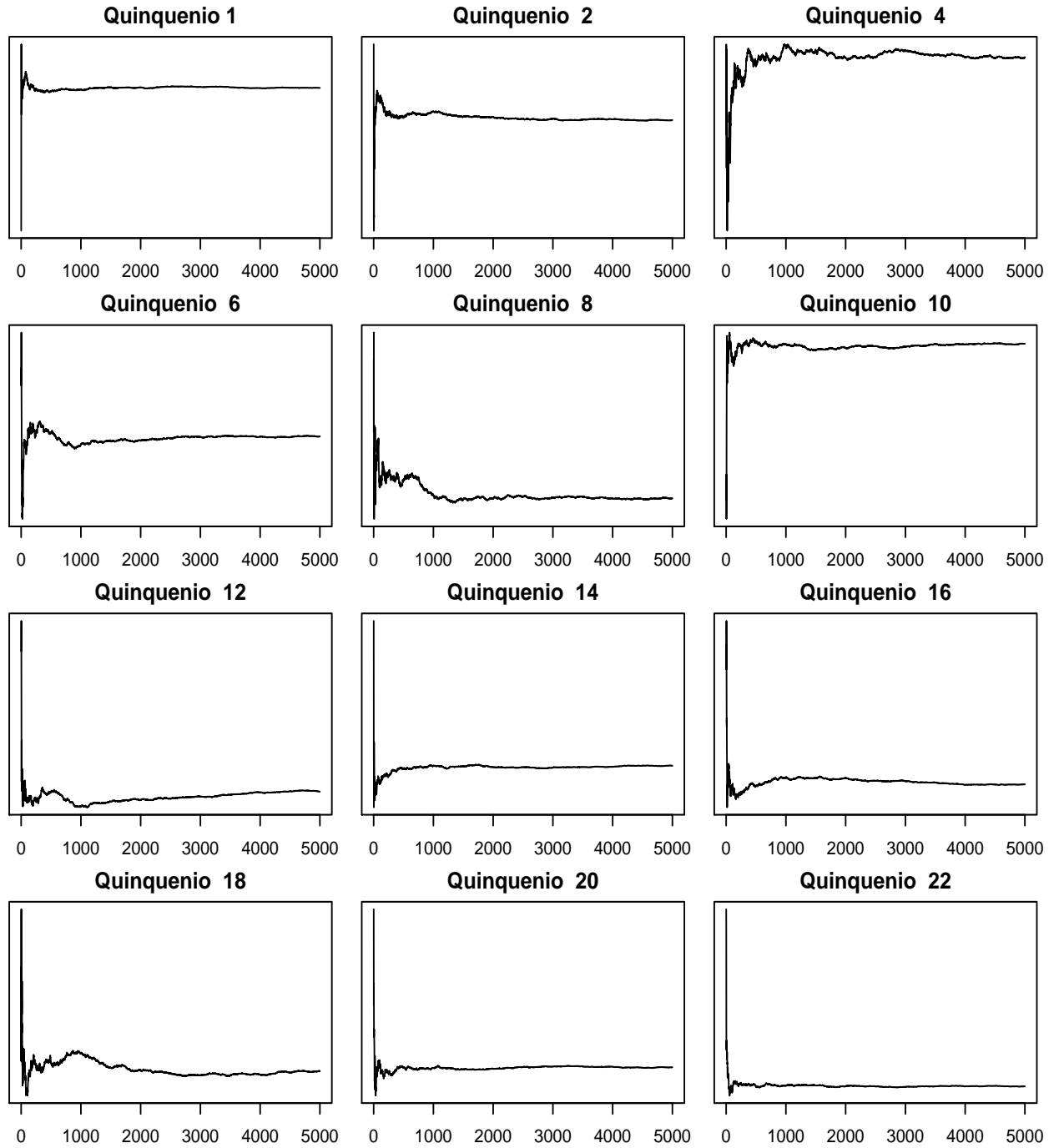


Figura C.2: Promedios ergódicos para la tasa de riesgo, modelo poisson-gamma, edad 40

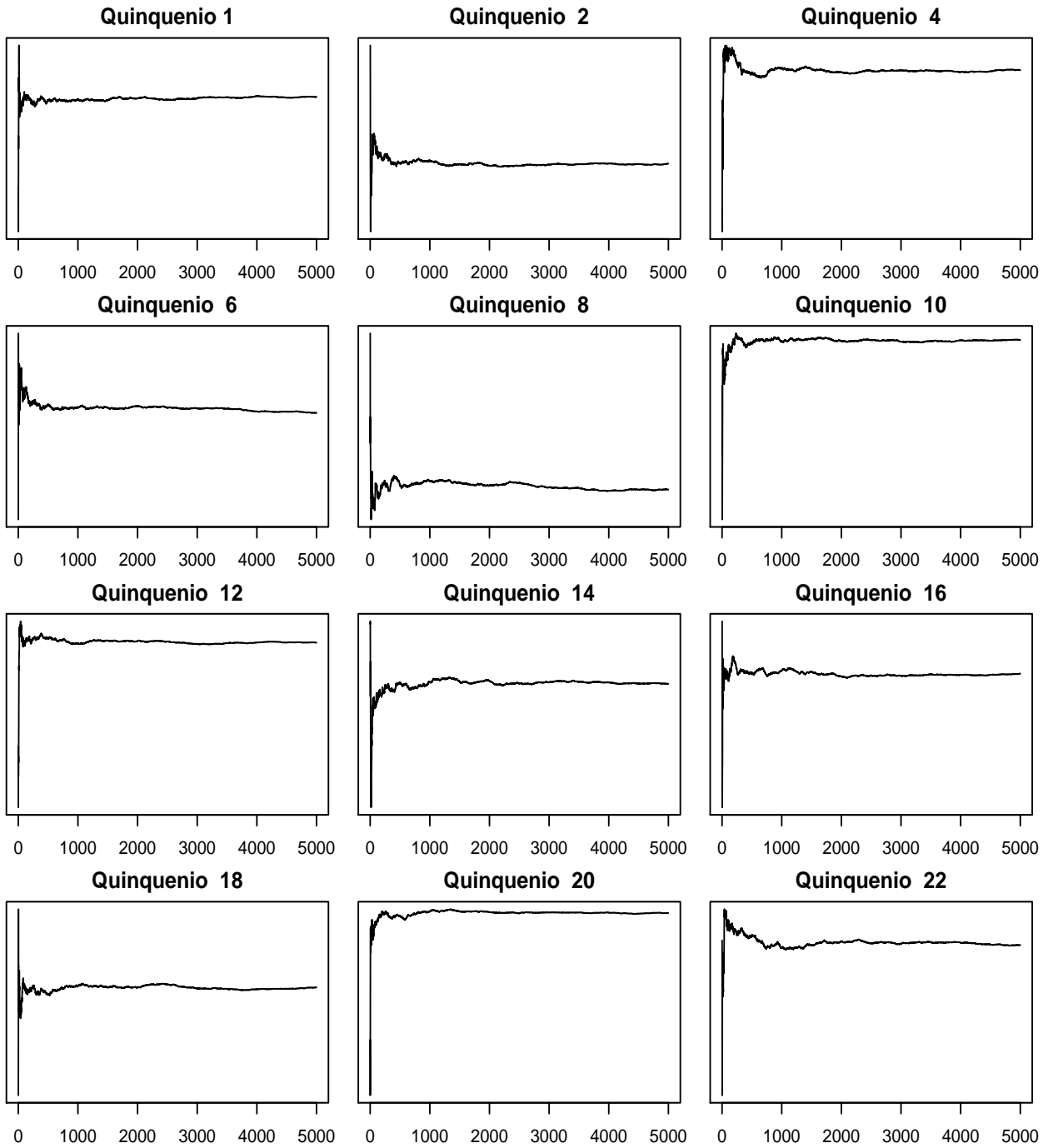


Figura C.3: Promedios ergódicos para la tasa de riesgo, modelo poisson-gamma, edad 60

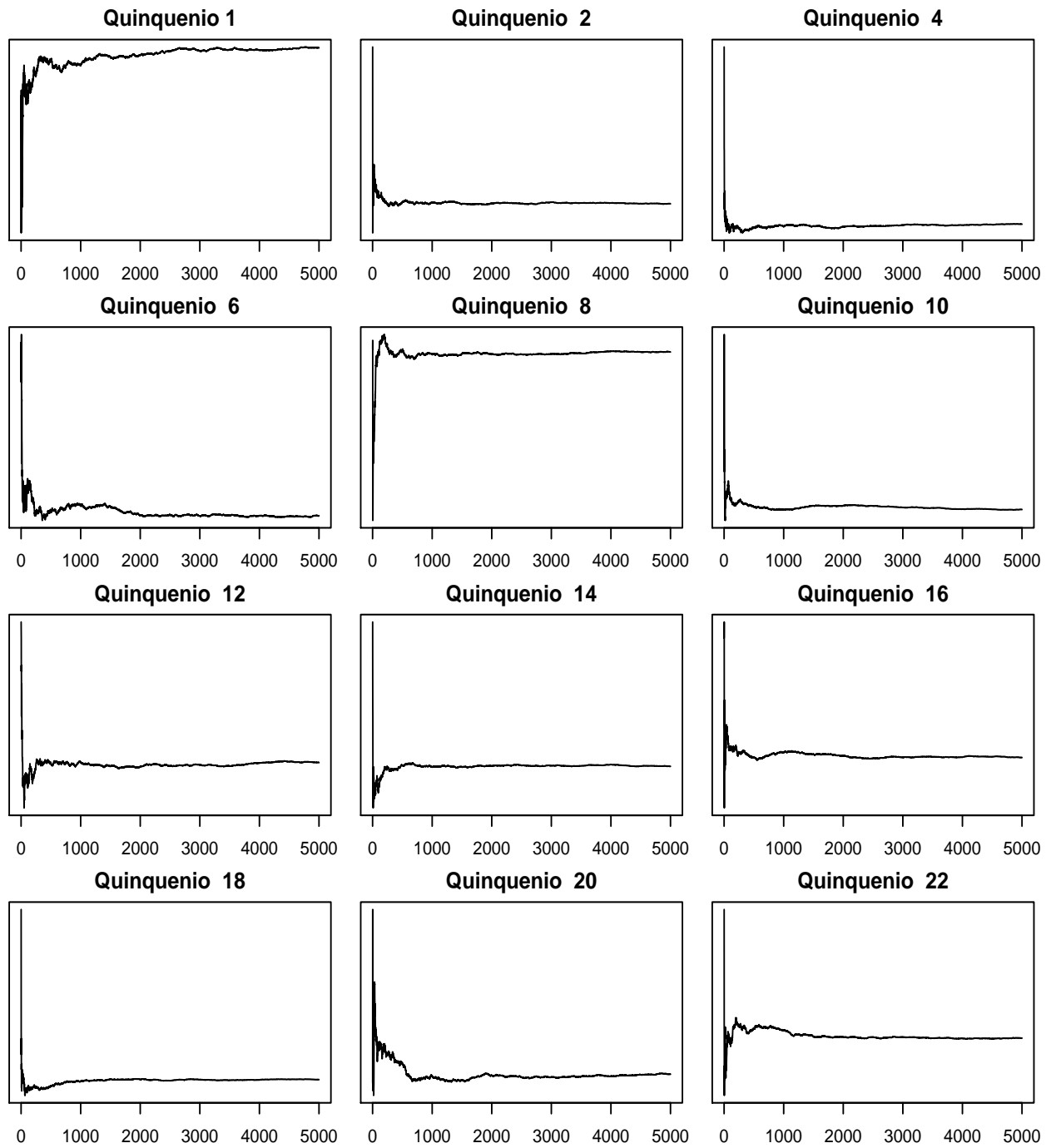


Figura C.4: Promedios ergódicos para la tasa de riesgo, modelo poisson-gamma, edad 80

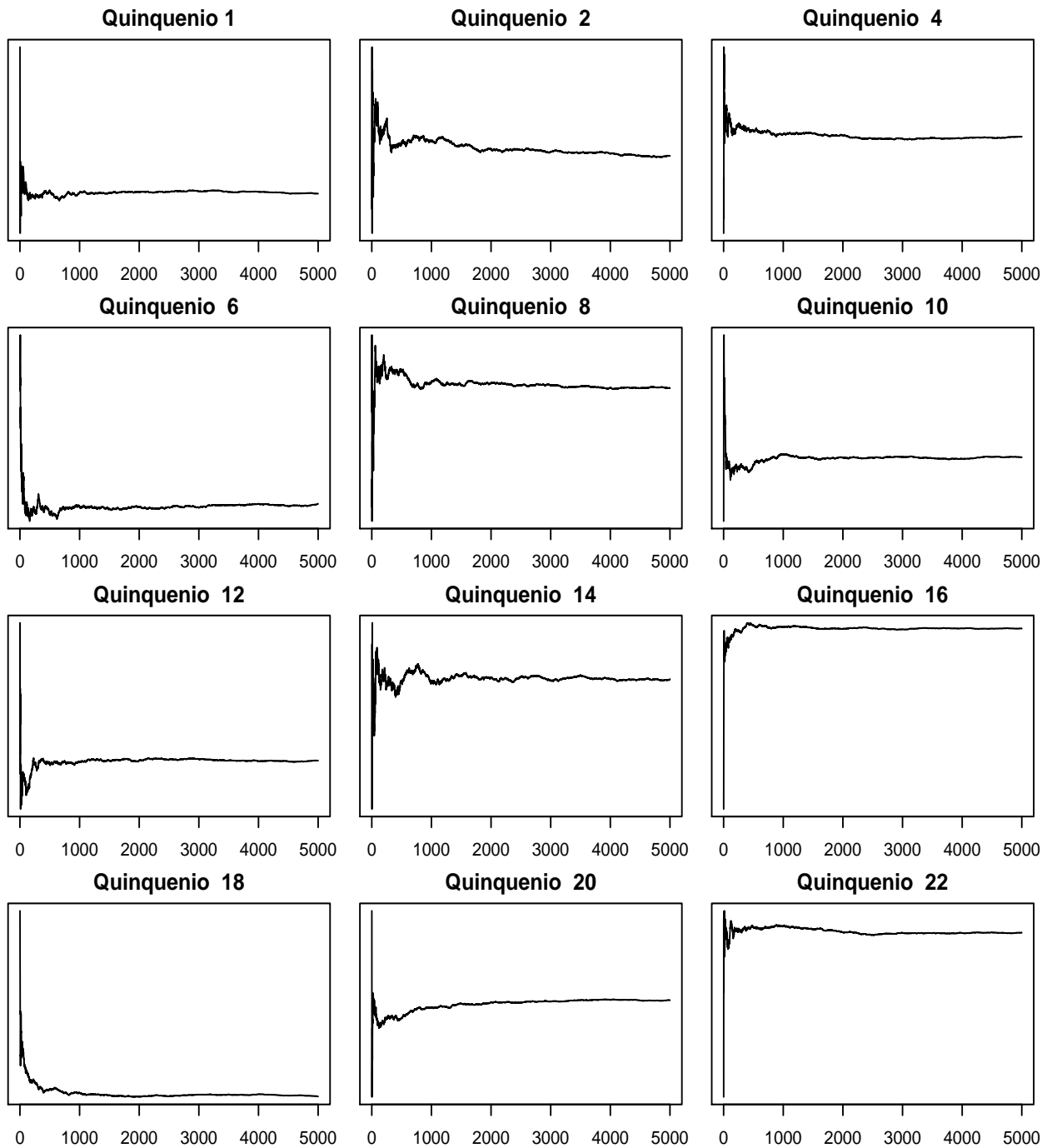


Figura C.5: Histogramas para la tasa de riesgo, modelo poisson-gamma, edad 20

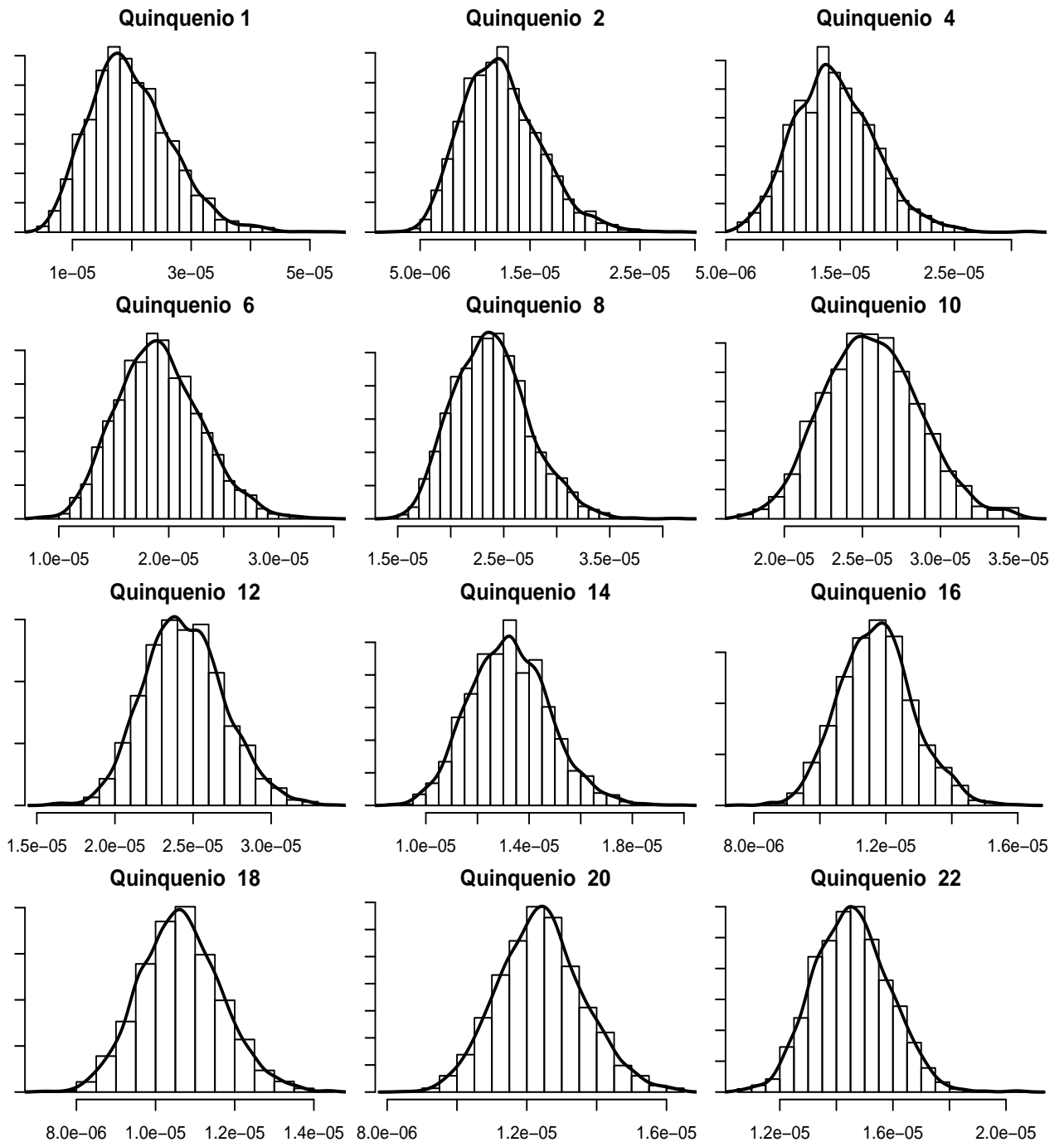


Figura C.6: Histogramas para la tasa de riesgo, modelo poisson-gamma, edad 40

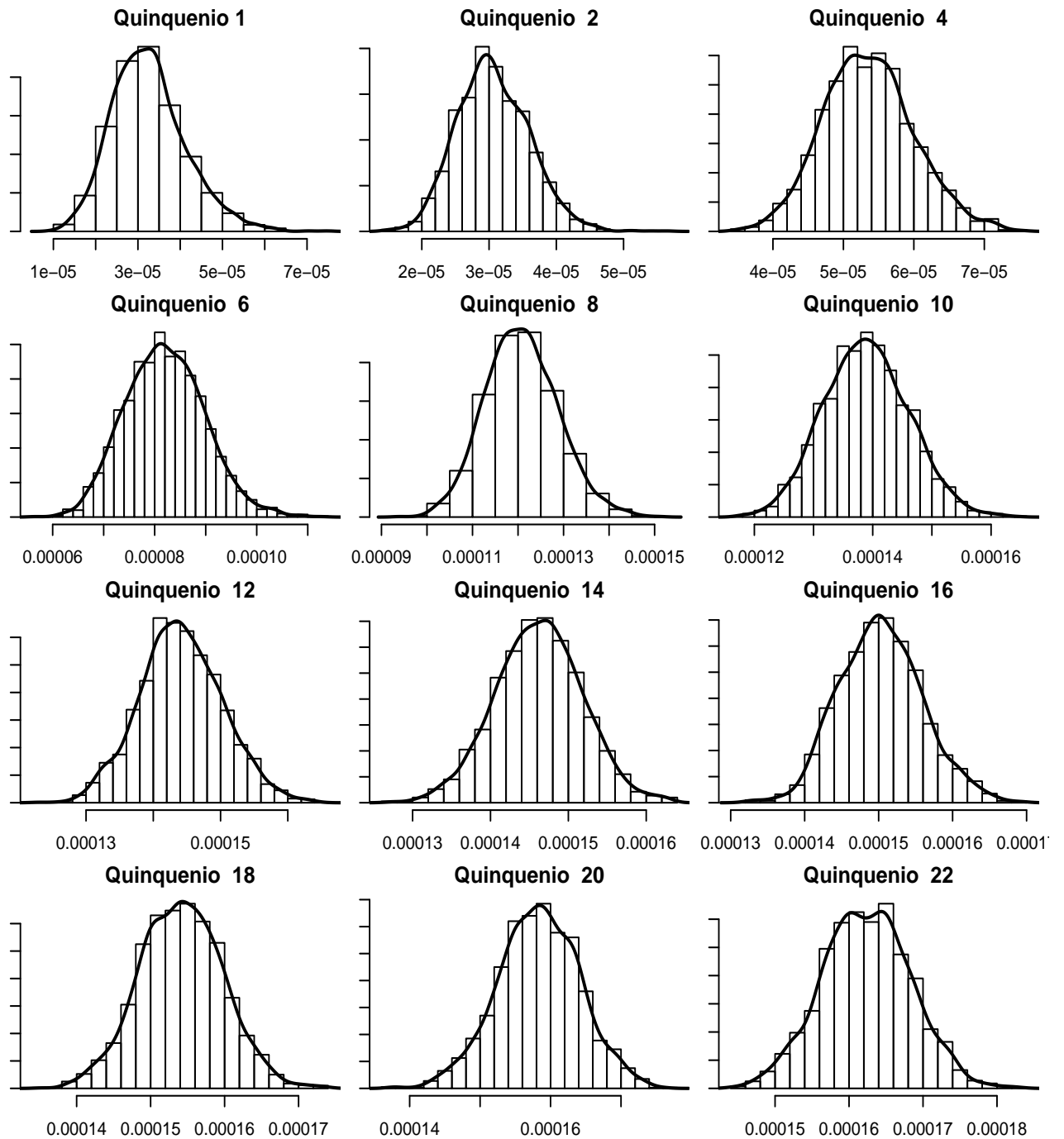


Figura C.7: Histogramas para la tasa de riesgo, modelo poisson-gamma, edad 60

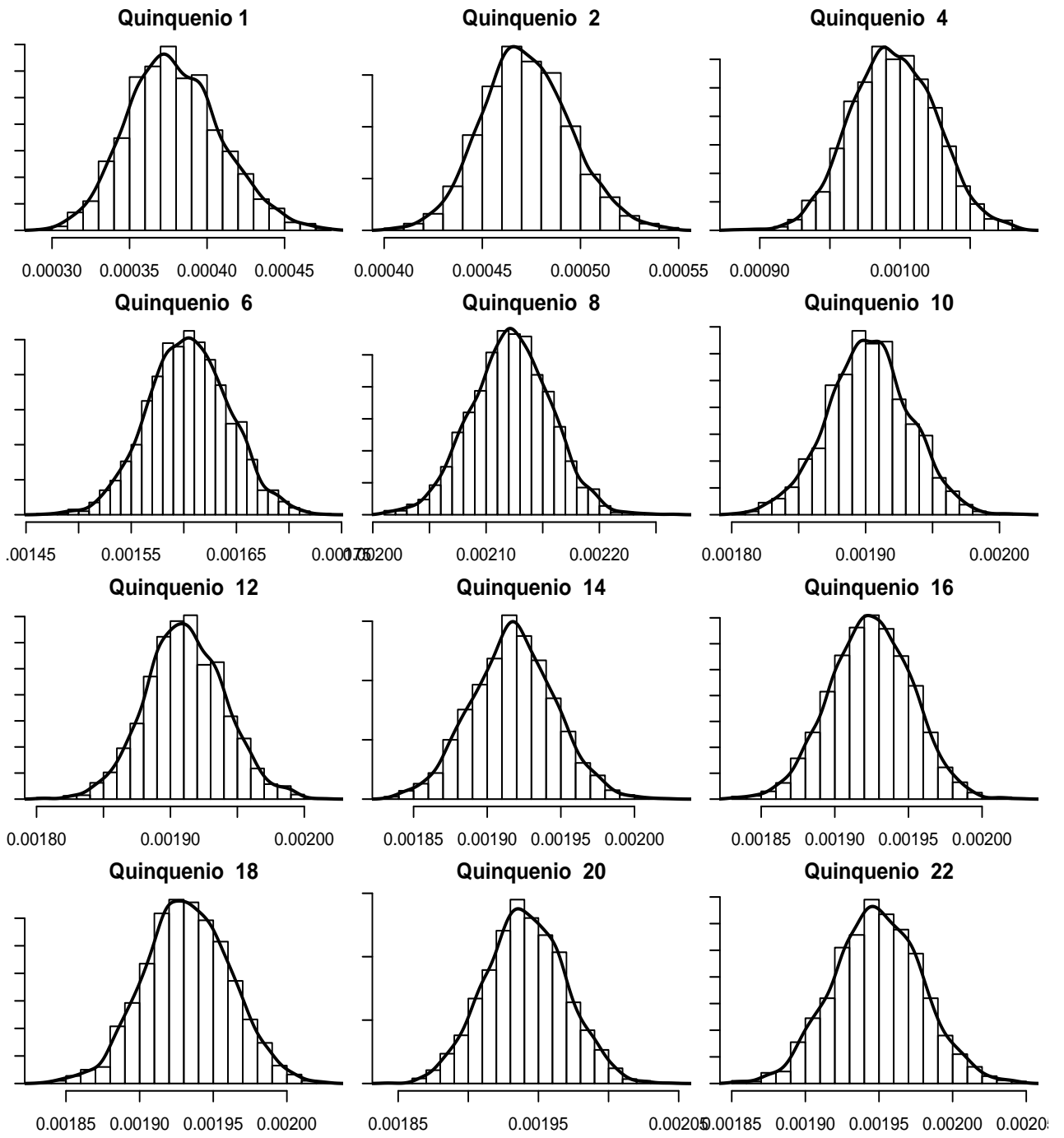


Figura C.8: Histogramas para la tasa de riesgo, modelo poisson-gamma, edad 80

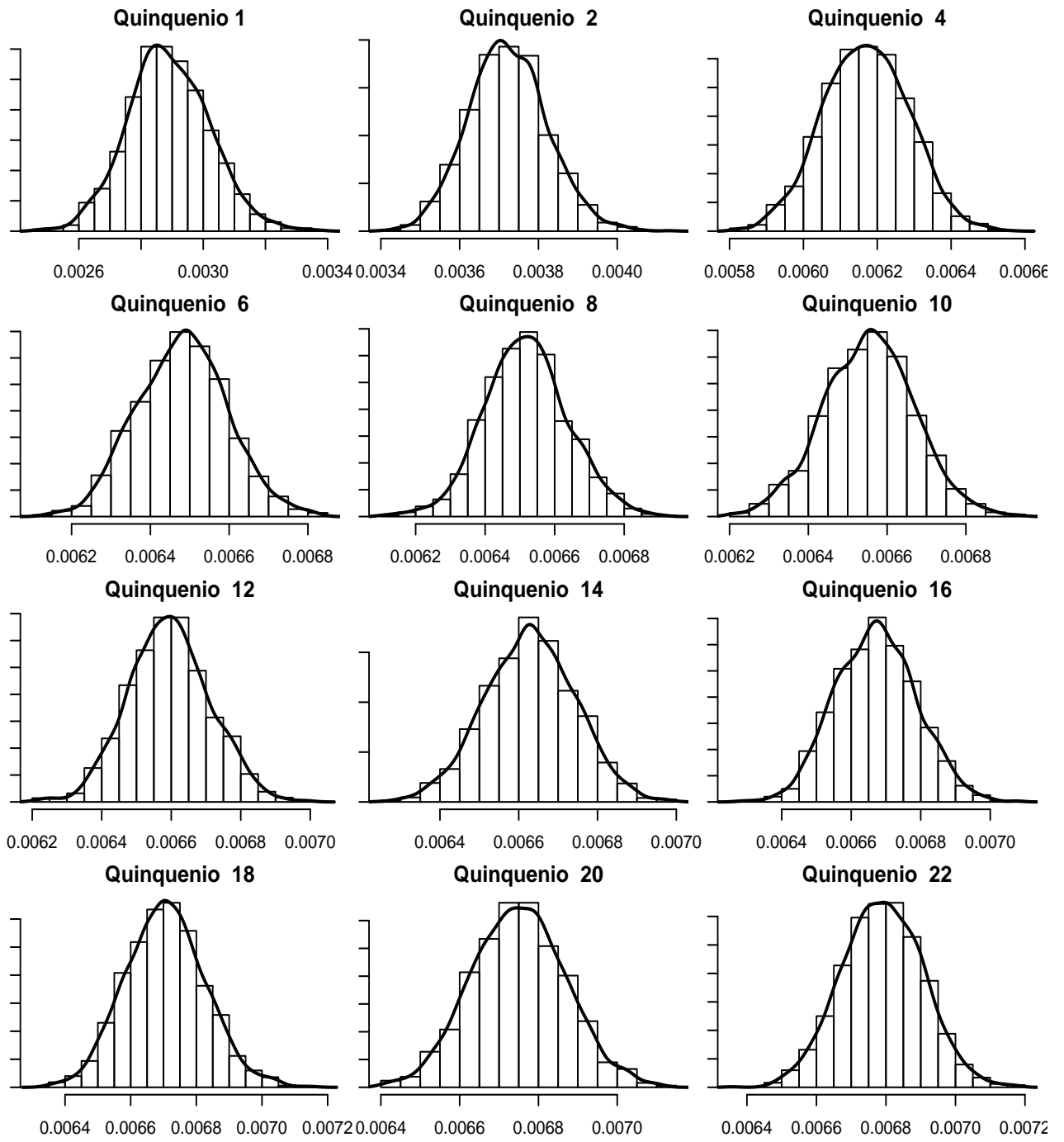


Figura C.9: Promedios ergódicos para la tasa de riesgo, modelo binomial-beta, edad 20

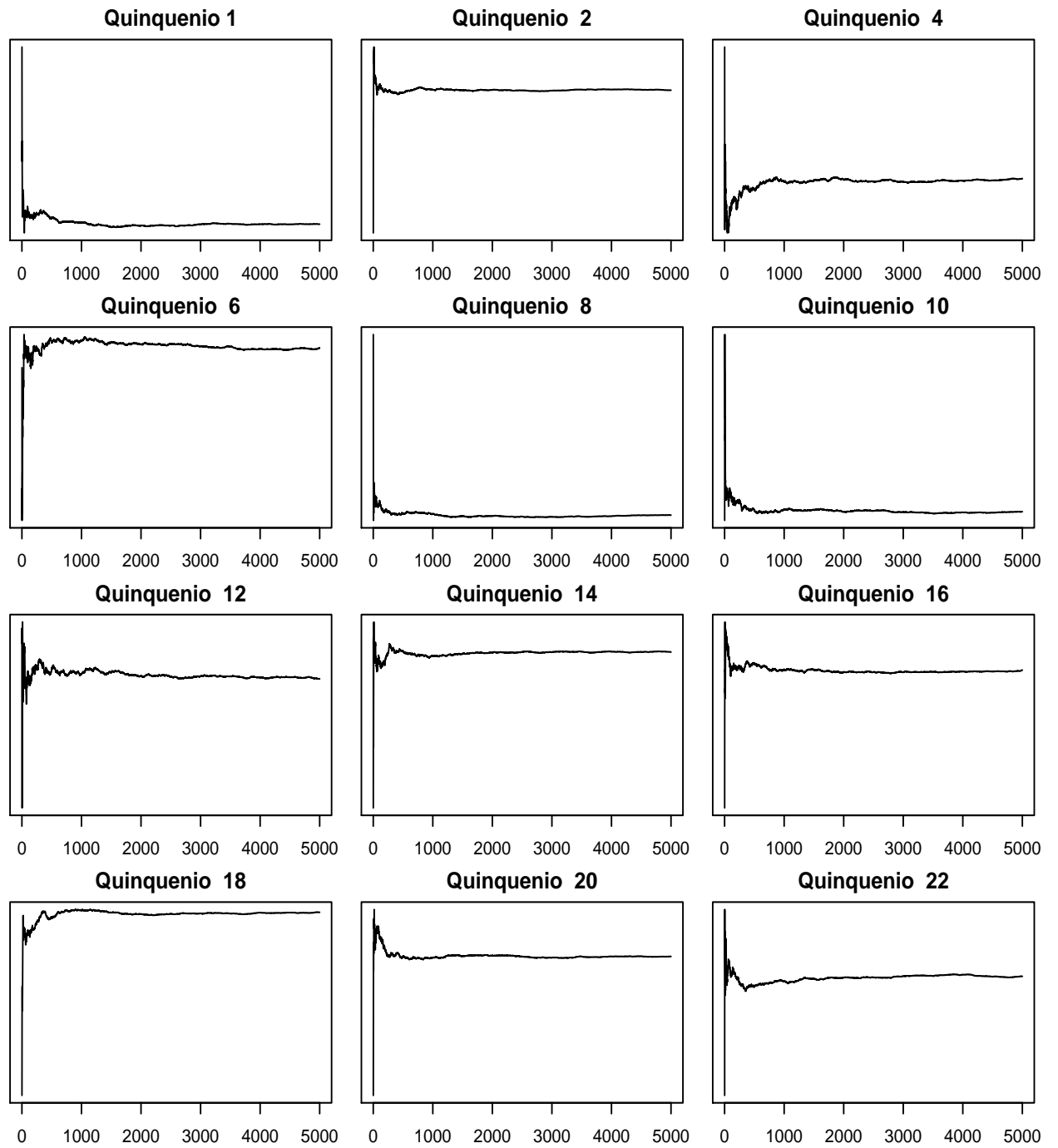


Figura C.10: Promedios ergódicos para la tasa de riesgo, modelo binomial-beta, edad 40

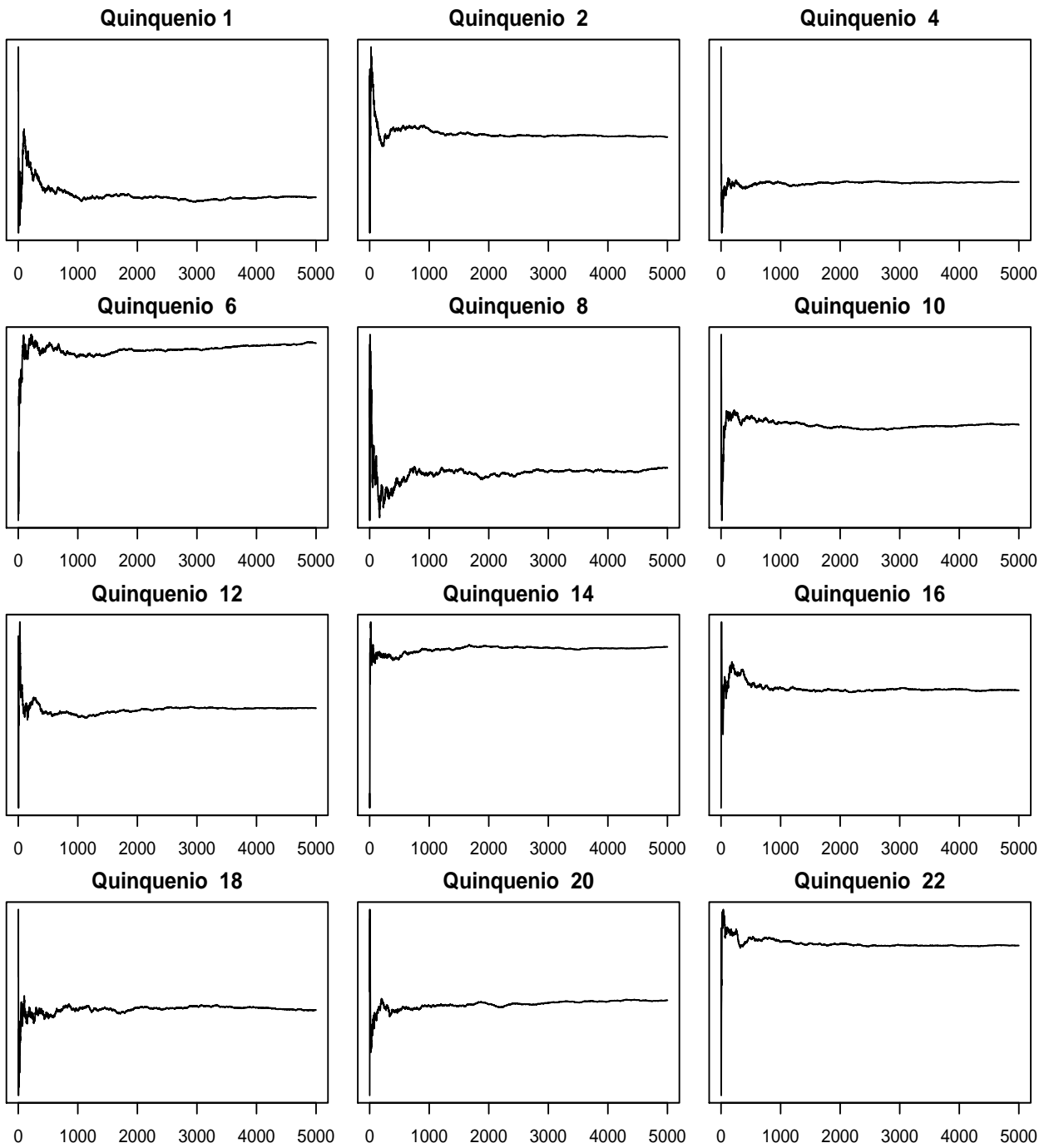


Figura C.11: Promedios ergódicos para la tasa de riesgo, modelo binomial-beta, edad 60

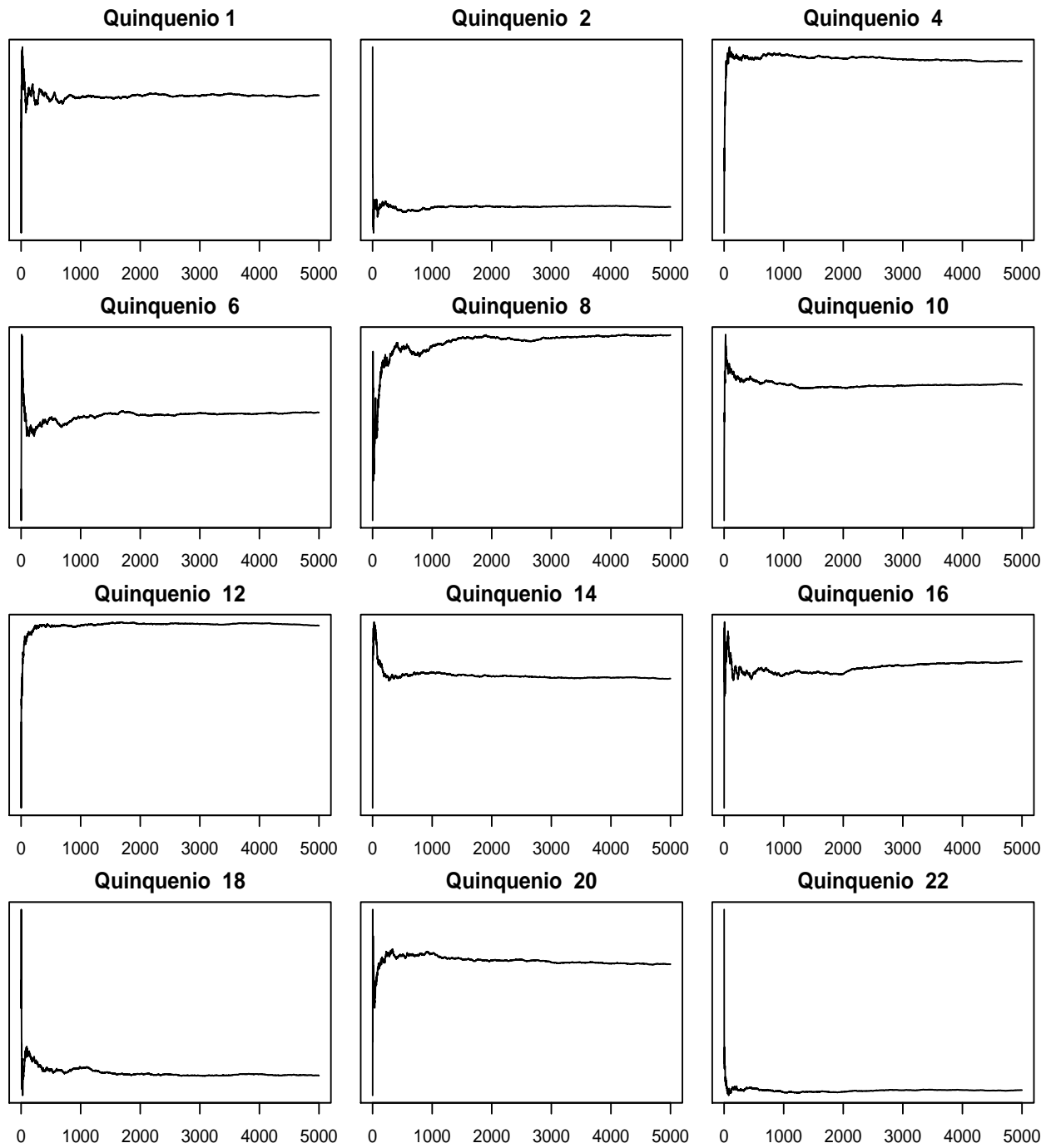


Figura C.12: Promedios ergódicos para la tasa de riesgo, modelo binomial-beta, edad 80

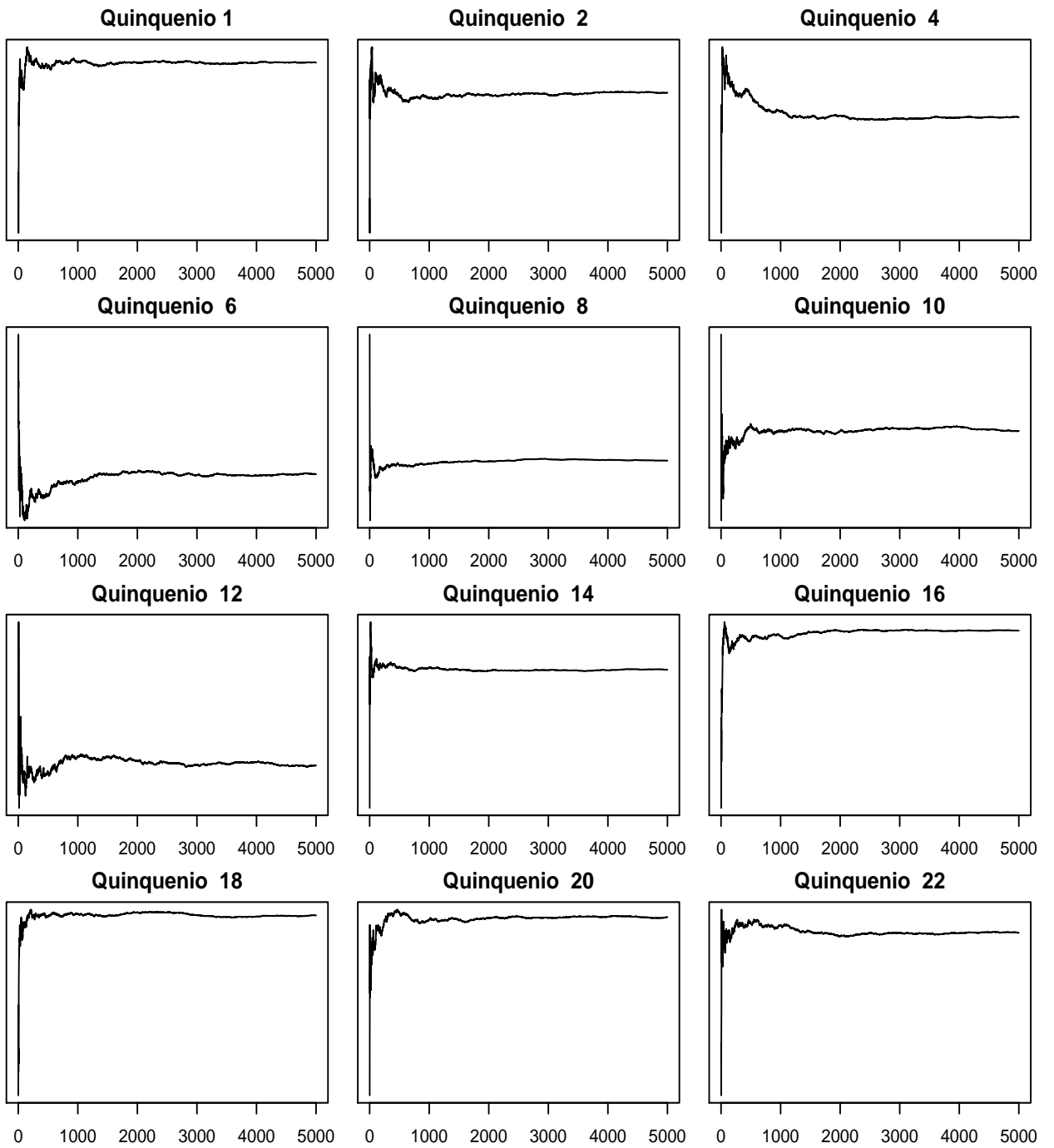


Figura C.13: Histogramas para la tasa de riesgo, modelo binomial-beta, edad 20

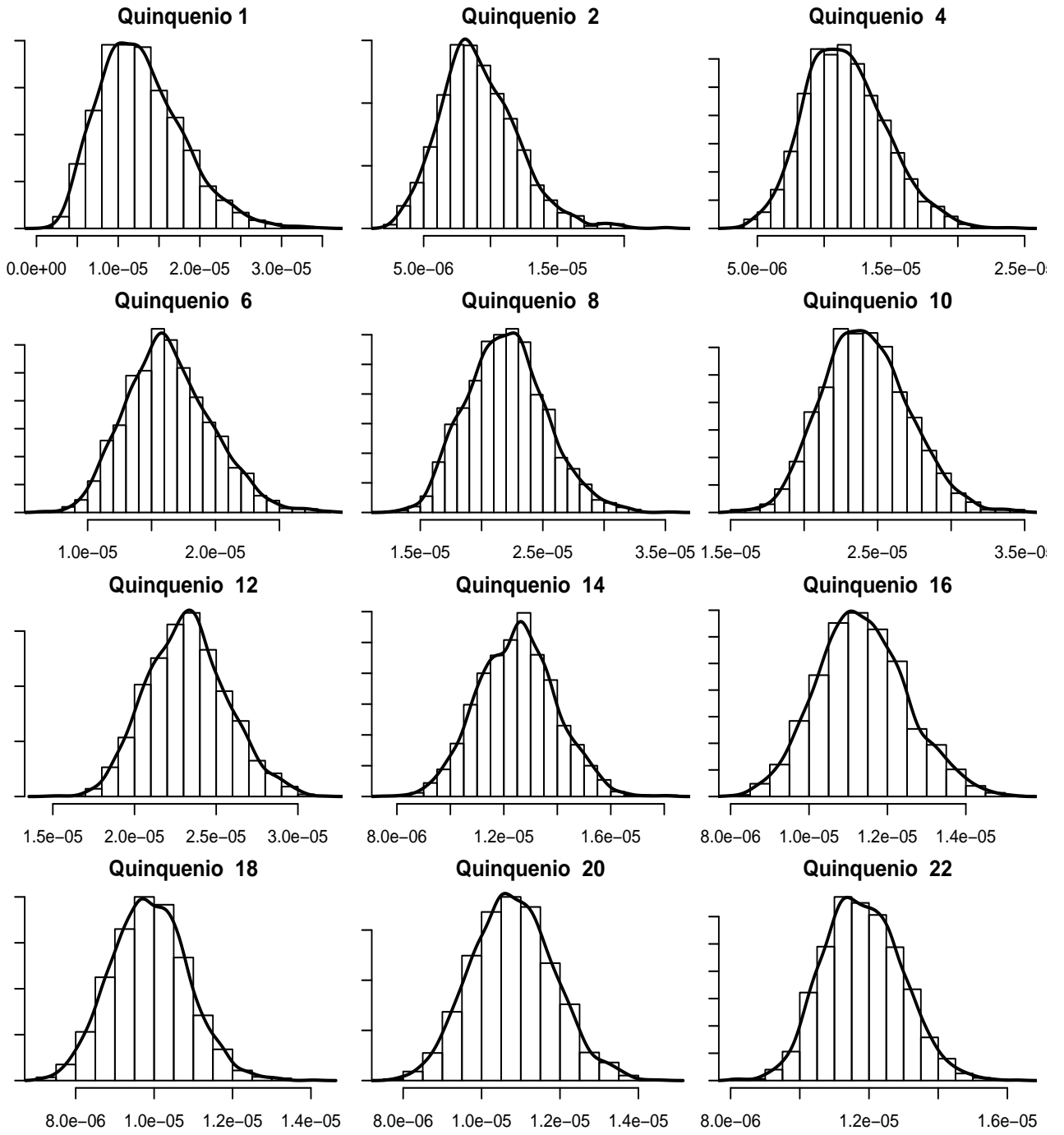


Figura C.14: Histogramas para la tasa de riesgo, modelo binomial-beta, edad 40

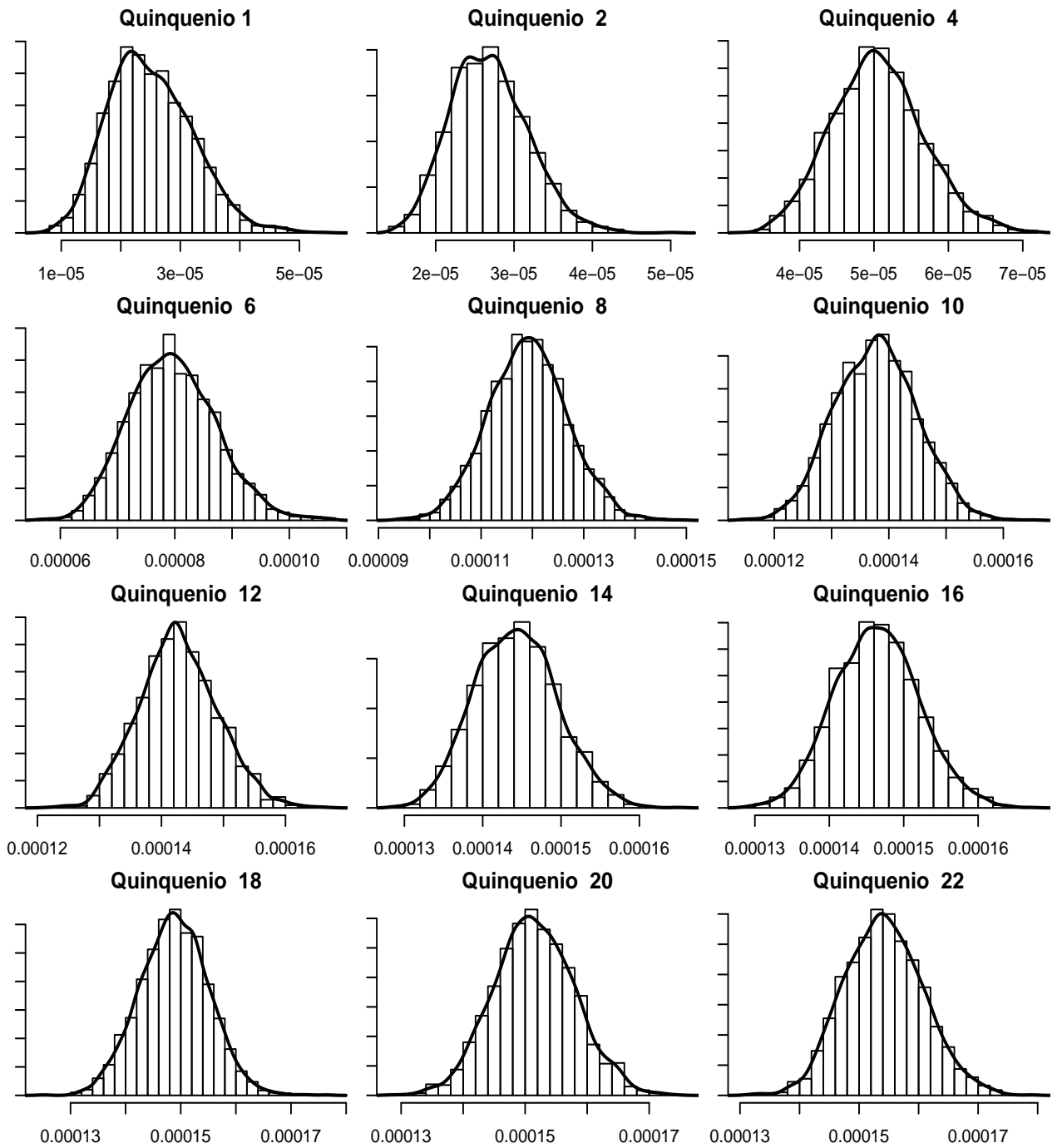


Figura C.15: Histogramas para la tasa de riesgo, modelo binomial-beta, edad 60

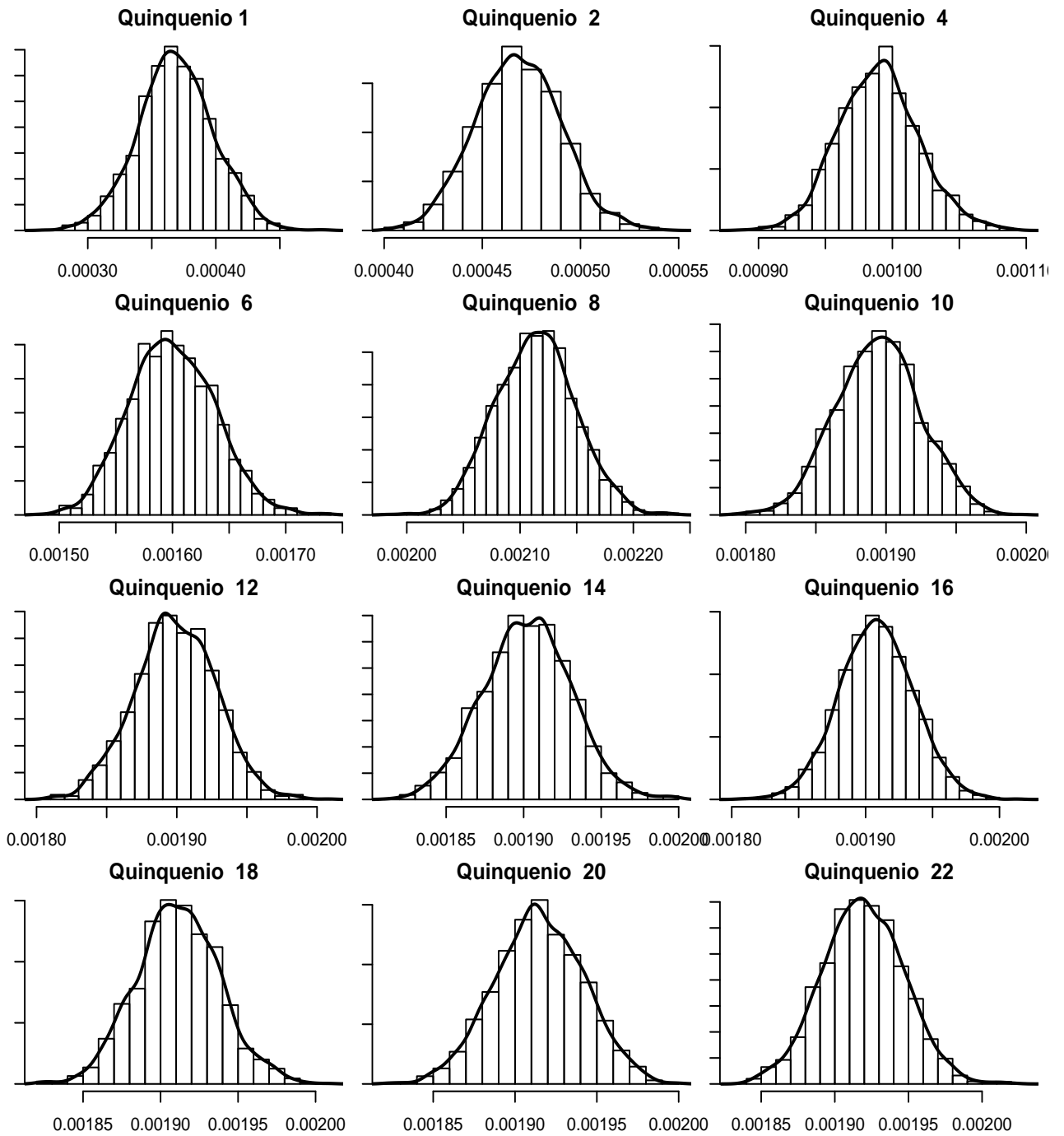
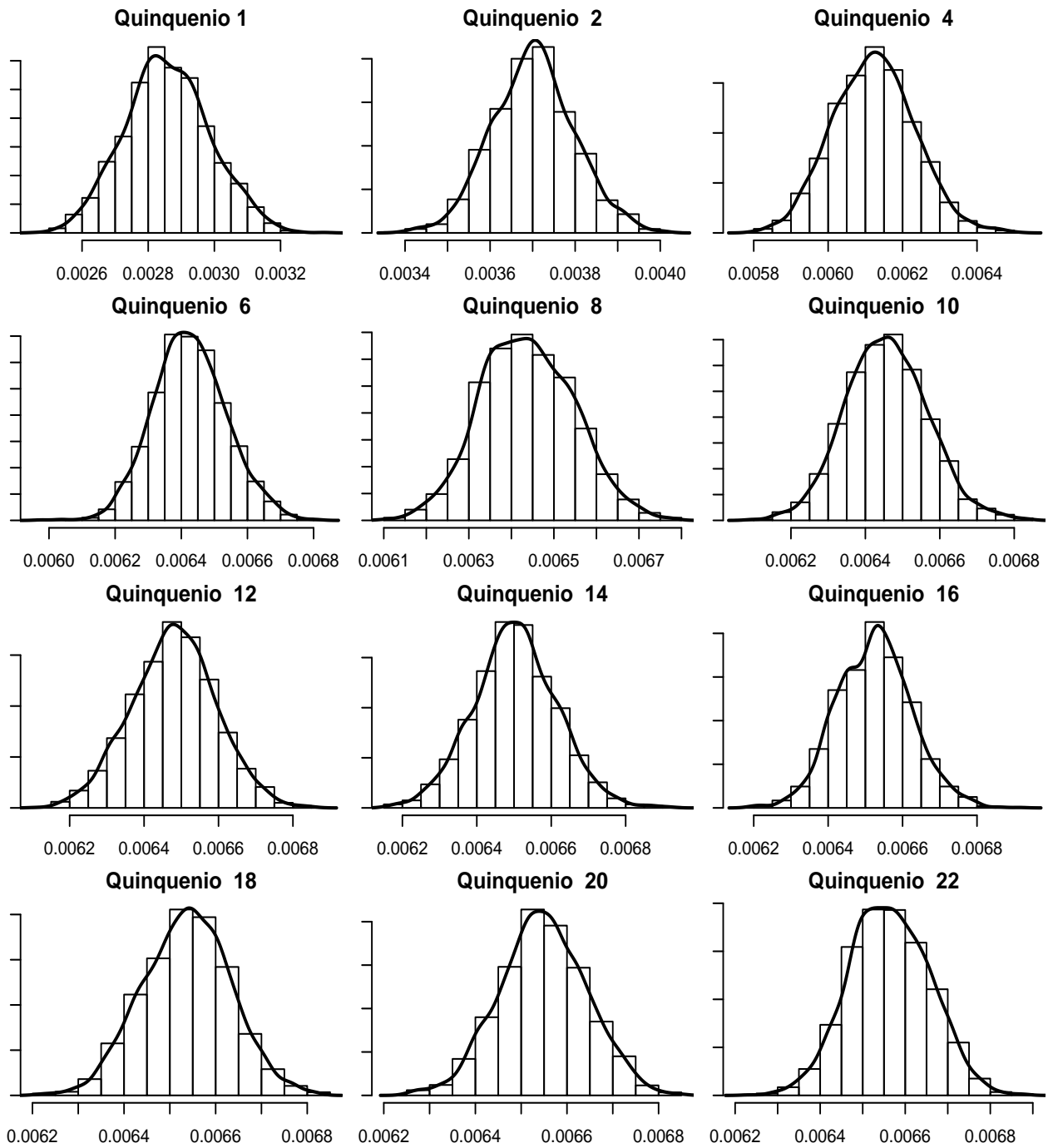


Figura C.16: Histogramas para la tasa de riesgo, modelo binomial-beta, edad 80



Conclusiones

El análisis de supervivencia es una rama de la estadística que se enfoca en el estudio de los tiempos de ocurrencia de eventos considerando a estos como variables aleatorias y mediante datos observados poder realizar inferencias sobre las probabilidades de dichas ocurrencias y analizar las mismas mediante funciones características de esta rama como lo son las funciones de supervivencia y de riesgo las cuales pueden representarse mediante funciones conocidas en el ámbito de la probabilidad como lo son la función Weibull o la exponencial o incluso mediante modelos no paramétricos los cuales solo dan aproximaciones numéricas o estimaciones a estas funciones como lo puede ser el modelo Kaplan-Meier.

El incorporar un enfoque bayesiano fue de suma importancia para el fin que se persigue durante el trabajo, ya que utiliza información previa para hacer inferencia sobre los parámetros de interés en los quinquenios posteriores y se complementa bien al utilizar los datos con los que se cuenta pudiendo decidir dependiendo del caso en donde se puede o no dar mayor peso a los datos, cuando estos existen o a la información a priori, en los casos en que no se cuenta con datos para realizar la inferencia, siendo así más o menos informativos dependiendo de si se tienen o no datos, proveyendo así de alternativas para la modelación del problema planteado. Otro punto importante del modelo utilizado radica en la dependencia de las tasas de riesgo via los procesos beta y gamma, lo cual al generar una dependencia entre la tasa de riesgo del año k con el inmediato anterior $k - 1$ aporta una estructura que hace factible las proyecciones realizadas.

Tras realizar un análisis de supervivencia de diabetes de la población mexicana surgen diversas interrogantes sobre el aumento de esta con el paso del tiempo, es importante poder

cuantificar estos incrementos para así poder prever el impacto que se podría presentar para el sector salud primordialmente, el enfoque presentado aporta para este fin una metodología que puede ser implementada sistemáticamente, esto debido al esquema bayesiano primordialmente ya que una principal cualidad de este esquema es precisamente la posibilidad de actualizar datos de manera sencilla. Otro punto importante del modelo propuesto radica en el papel que juega la dependencia via procesos de markov de las tasas de riesgo, el cual es esencial para el pronóstico de las futuras tasas.

Con respecto a la implementación de la rutina en el paquete estadístico R, la principal dificultad que se presentó se dio al muestrear de la condicional total de la variable latente u lo cual fue solucionado completando la constante con la que la función de probabilidad sumaba uno. En cuanto a la eficiencia del programa y debido a que se corrieron alrededor de 5000 iteraciones del muestreo de gibbs para todas las edades(20 a 100) de los 22 quinquenios, el tiempo que se tardó en completar la rutina fue alrededor de 5 horas, lo cual no presenta mayores inconvenientes.

Cabe señalar que dentro de cualquier pronóstico existe un momento en que este ya no es plausible o sus resultados se alejan de la realidad, esto se debe a que conforme se va avanzando en el tiempo los datos dejan de ser observados y por tanto se vuelven a hacer las proyecciones con datos estimados generando márgenes de error en cada ocasión que el proceso se repite. Este trabajo no es la excepción y la prueba de esto se da al observar las tasas de riesgo finales para todos los quinquenios, donde se observa que llega un cierto quinquenio a partir del cual todas las tasas de riesgo para los quinquenios posteriores son muy parecidas y esto no tendría que ser necesariamente cierto lo que al parecer sucede es que ya no es posible hacer mayor pronóstico con la información con la que se cuenta y es por eso que se presentan datos similares a partir de ese quinquenio.

Referencias

- [1] Arjas, E.; Gasbarra, D. (1994) *Nonparametric bayesian inference from right censored survival data using the Gibbs sampler* Statist. Sinica Vol. 4.

- [2] Erdely Ruiz, Arturo *Apuntes de estadística bayesiana* México, 2004.

- [3] Gamerman, Dani (1991) *Dynamic bayesian models for survival data* Appl. Statist. Vol. 40.

- [4] Gamerman, Dani *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference* Chapman and Hall, London, 1997.

- [5] Gelman, Andrew; et al. *Bayesian data analysis* Chapman and Hall, Florida, 2004.

- [6] Gilks, W. R.; Richardson, S.; Spiegelhalter, D.J. *Markov Chain Monte Carlo in practice* Chapman and Hall, London, 1996.

- [7] Gray, R. J. (1994) *A bayesian analysis of institutional effects in multicenter cancer clinical trial* Biometrics Vol. 50.

- [8] Gutiérrez Peña, Eduardo, *Métodos computacionales en la inferencia bayesiana* en línea, <http://www.dpye.iimas.unam.mx/eduardo/MCB>, 2003.
- [9] Hjort, N. L. (1990) *Nonparametric bayes estimators based on beta processes in models for life history data* Ann. Statist. Vol. 18.
- [10] Hosmer, David W.; Lemeshow, Stanley *Applied survival analysis: regression modeling of time to event data* J. Wiley, New York, 1999.
- [11] Ibrahim, Joseph G.; Chen, Ming-Hui; Sinha, Debajyoti *Bayesian survival analysis* Springer, New York, 2001.
- [12] Lee, Elisa T.; Wentu Wang, John *Statistical methods for survival data analysis* J. Wiley, New York, 2003.
- [13] Nieto Barajas, Luis E.; Walker, Stephen G. (2001) *Markov beta and gamma processes for modelling hazard rates* Scandinavian Journal of Statistics Vol. 29.
- [14] Robert, Christian P. *The bayesian choice* Springer, New York, 1994.
- [15] Walker, Stephen G.; Mallick, B. K. (1997) *Hierarchical generalized linear models with bayesian nonparametric mixing* J. Roy. Statist. Soc. Ser. Vol. B 59.
- [16] Walsh, B. *Markov Chain Monte Carlo and Gibbs sampling* en línea, <http://www.stat.columbia.edu/liam/teaching/neurostat-spr07/papers/mcmc/mcmc-gibbs-intro.pdf>, 2004.