

00324



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

UN MÉTODO BAYESIANO PARA GRADUACIÓN DE TABLAS DE MORTALIDAD.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
M A T E M Á T I C A
P R E S E N T A :
MARÍA CRISTINA AGUILAR BUSTOS

TESIS CON
FALLA DE ORIGEN

DIRECTOR DE TESIS:
DR. EDUARDO ARTURO GUTIÉRREZ PEÑA



FACULTAD DE CIENCIAS
UNAM

2003

Autorizo a la Dirección General de Bibliotecas de UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional

NOMBRE: María Cristina Aguilar Bustos

FECHA: 07 - Julio - 2003

FIRMA: [Firma manuscrita]



FACULTAD DE CIENCIAS
SECCION ESCOLAR



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.

PAGINACION DISCONTINUA



SECRETARÍA DE EDUCACIÓN PÚBLICA
 INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

DRA. MARÍA DE LOURDES ESTEVA PERALTA
 Jefa de la División de Estudios Profesionales de la
 Facultad de Ciencias
 Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

"Un Método Bayesiano para la Graduación de Tablas de Mortalidad"

realizado por **AGUILAR BUSTOS MARÍA CRISTINA**
 con número de cuenta **09010193-5**, quién cubrió los créditos de la carrera de **MATEMÁTICAS**.
 Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
 Propietario

DR. EDUARDO ARTURO GUTIERREZ PEÑA

[Handwritten signature]

Propietario

DR. ALBERTO CONTRERAS CRISTAN

[Handwritten signature]

Propietario

MAT. MARGARITA ELVIRA CHAVEZ CANO

[Handwritten signature]

Suplente

DR. RAUL RUEDA DIAZ DEL CAMPO

[Handwritten signature]

Suplente

M. en C. SALVADOR ZAMORA MUÑOZ

[Handwritten signature]
 zamora / 11/03/03 J.S.

Consejo Departamental de Matemáticas



[Handwritten signature]

M. en C. JOSE ANTONIO GOMEZ ORTEGA

SECRETARÍA DE EDUCACIÓN PÚBLICA
 CONSEJO DEPARTAMENTAL
 DE
 MATEMÁTICAS

**TESIS CON
 FALLA DE ORIGEN**

Agradecimientos

Durante el desarrollo de la presente tesis, he recibido el apoyo de numerosas personas a las que quiero expresar mi gratitud:

A Dios

Por estar a mi lado a lo largo de esta vida, siendo mi soporte y guiando mis pasos, permitiéndome existir y estar aquí, siempre llena de alegría, además por darme la oportunidad de tener personas que me motivan y me dan aliento en el camino.

A mi familia

A mi padre, por su amor, y sabiduría para guiarme en el camino de la vida.

A mi madre, por su cariño, comprensión y apoyo sin condiciones ni medida.

A mi hermana, por su ejemplo y por el apoyo que me ha brindado, para mi formación, por fomentar en mí el deseo de saber, de conocer lo novedoso y abrirme las puertas al mundo ante mi curiosidad insaciable.

A mi hermano, por su ejemplo y comprensión, su infinita ayuda moral y material, por alentarme a seguir adelante. Por enseñarme todo el valor y toda la fuerza en un solo abrazo.

A "Pinto" por ser la luz que acompaña mi existir.

A mi director de tesis

Dr. Eduardo Arturo Gutiérrez Peña, por su paciencia y sabiduría al dirigir esta tesis, por su gran calidad humana y por el apoyo que me brindó, quien no sólo fue el director de esta tesis, sino que este trabajo no hubiera sido posible sin las importantes observaciones, y sus acertadas y apreciables críticas; además del tiempo aún más de lo requerido.

A mis sinodales

Mat. Margarita Chávez Cano, Dr. Alberto Contreras Cristán, Dr. Raúl Rueda Díaz del Campo y M. en C. Salvador Zamora Muñoz, quienes me brindaron su apoyo, tiempo, trabajo y experiencia sin ningún tipo de interés, y por medio de las discusiones y preguntas, me hicieron crecer en conocimiento además por su invaluable evaluación de este trabajo.

A la Comisión Nacional de Seguros y Fianzas

De manera especial al apoyo del Dr. Manuel Mendoza Ramírez Por brindarnos la documentación necesaria, basada en investigaciones previas de la Experiencia Mexicana en el sector asegurador.

A la UNAM

Institución que me brindó la oportunidad de formarme profesionalmente a lo largo de muchos años y de la que siempre estaré orgullosa de pertenecer.

A Juan Manuel

Por existir y estar conmigo, apoyándome incondicionalmente. Porque dentro de una de sus sonrisas eternas aprendí que la vida esta llena de satisfacciones no materiales.

A Javier Páez

Por guiar mis conocimientos hacia nuevos horizontes. Además por su sabio y oportuno consejo por su paciencia y gran amistad.

A Giovanni Navarrele

Quien me enseñó que no hay límites, que lo que me proponga lo puedo lograr y que sólo depende de mí.

A mis compañeros de facultad

Con quien construí conocimiento, compartimos mañanas, tardes y noches de estudio, momentos de nerviosismo en parciales y finales. Que estuvieron conmigo y compartimos tantas aventuras, experiencias y desveladas.

A mis maestros

Gracias a cada uno de los maestros que participaron en mi desarrollo profesional durante mi carrera, sin su ayuda y conocimientos no estaría en donde me encuentro ahora.

Al personal de la Biblioteca (FFMA)

Por su colaboración, ayuda y una sonrisa amable durante el tiempo de la investigación de esta tesis.

Y antes de dar fin a estas líneas es preciso mirar atrás y recordar con afecto y agradecimiento a todos los que conmigo empezaron, que continúan o que han seguido otros rumbos, por eso quiero dar las gracias al colega desconocido en nombre de todos los que he olvidado por descuido.

Dedicatoria

*Esta tesis es una parte de mi vida y el comienzo de otras etapas,
por esto y más, la dedico a Dios y a mi familia, por brindarme
su amor incondicional y lo mejor de sus vidas.*

GRACIAS.

Prefacio

La graduación y sobre-estimación de tasas de mortalidad juegan un papel importante en la actuaria. Los métodos actuariales tradicionales tratan cada uno de estos aspectos separadamente y por lo general ignoran su naturaleza estadística.

En este trabajo se propone un método estadístico Bayesiano que toma en cuenta, de manera simultánea, los dos aspectos antes mencionados. El enfoque empleado permite además seleccionar una tabla de mortalidad modificada con base en criterios naturales y bien definidos. Concretamente, se ilustra el uso de técnicas estadísticas en la construcción de tablas de mortalidad para seguros de vida. Con este fin, en el Capítulo 2 discutimos primero la estimación de las tasas de mortalidad individuales, que son la materia prima de cualquier método de graduación. También se hace referencia a la Estadística Bayesiana y a algunos métodos computacionales como el muestreo de Gibbs, los cuales se basan en simulaciones y nos permiten hacer inferencias de una manera relativamente sencilla. Por otra parte, se da una breve descripción del análisis de regresión lineal y de los modelos lineales generalizados, con énfasis en el modelo logístico.

En el Capítulo 3 damos una descripción de los datos de mortalidad correspondientes a la Experiencia Mexicana de 1982 a 1989 y describimos con detalle la metodología propuesta. La idea fundamental es que tanto la graduación de tablas de mortalidad como la construcción de las correspondientes tablas modificadas son problemas de predicción. Por lo tanto, el primer objetivo del análisis debe ser la obtención de la distribución predictiva de la tasa de mortalidad futura para cada edad.

Para efectos de este trabajo, suponemos que el interés principal radica en una adecuada predicción del número agregado de reclamaciones (muertes) con el fin de garantizar que la probabilidad de que dicho número exceda un

límite determinado¹ en el futuro sea pequeña. Una vez establecido un valor en riesgo adecuado, el método aquí propuesto permite determinar una tabla modificada eligiendo, para cada edad, el cuantil de la distribución predictiva de las tasas de mortalidad que, en conjunto con la estructura de los expuestos, produce el valor en riesgo elegido para el número agregado de reclamaciones. Por otro lado, es posible construir una tabla graduada con base en la mediana de la distribución predictiva de la tasa de mortalidad futuras para cada edad.

El objetivo principal de este trabajo es ilustrar el uso de esta metodología predictiva para resolver los problemas de graduación y de construcción de tablas modificadas. Por simplicidad, utilizamos dos clases de modelos sencillos, a pesar de que no se satisfacen necesariamente todos sus supuestos. Específicamente, usamos modelos de regresión lineal y modelos logísticos. Cabe señalar que nuestro fin no es explicar la relación entre la variables, sino predecir el comportamiento de la mortalidad, de manera que sólo requerimos de un mecanicismo de predicción razonable.

En el Capítulo 4 se presenta un análisis basado en un modelo de regresión con datos transformados y otro análisis basado en un modelo logístico. Asimismo, se hace una comparación entre los distintos modelos propuestos con el propósito de determinar el más adecuado.

Finalmente, en el Capítulo 5 damos a conocer los resultados y conclusiones de este trabajo.

¹Que en la literatura financiera es conocido como el "valor en riesgo"; véase, por ejemplo, Grayling (1997).

Índice general

1. Introducción	1
1.1. La Estadística y la Actuaría	1
1.1.1. El papel de los modelos lineales	3
1.1.2. ¿Qué es la graduación?	4
1.2. Experiencia Mexicana 82-89	4
1.2.1. Antecedentes	4
1.2.2. Resultados	5
1.3. El enfoque Bayesiano	11
2. Preliminares Técnicos	13
2.1. Estimación de las tasas individuales	13
2.2. Inferencia Bayesiana	16
2.2.1. Fundamentos	16
2.2.2. Inferencia	19
2.2.3. Métodos computacionales	21
2.3. Regresión	27
2.3.1. Análisis de regresión lineal	27
2.3.2. Análisis Bayesiano del modelo de regresión	34
2.4. Modelos lineales generalizados	39
2.4.1. Los componentes de un modelo lineal generalizado	39
2.4.2. Funciones de verosimilitud para los MLG	42
2.4.3. Cuasi-verosimilitud	45
2.4.4. Sobre-dispersión	46
2.4.5. Función liga	47
2.4.6. Estadísticas suficientes y ligas canónicas	48
2.4.7. Medida de bondad de ajuste	48
2.4.8. Análisis Bayesiano de los MLG	50

3. Modelos Estadísticos de Mortalidad	53
3.1. Introducción	53
3.1.1. Descripción del problema	53
3.1.2. Análisis descriptivo de los datos: 1982-1989	57
3.2. Supuestos para el análisis Bayesiano	57
3.2.1. Supuestos básicos	59
3.2.2. Análisis Bayesiano de modelos de regresión logística	60
3.2.3. Pronósticos para la siniestralidad agregada	63
3.3. Pronósticos basados en MLG	65
4. Análisis Bayesiano de Mortalidad	67
4.1. Regresión lineal normal	67
4.1.1. Análisis preliminar del modelo lineal	67
4.1.2. Análisis del modelo lineal: WinBUGS	72
4.1.3. Análisis preliminar del modelo cuadrático	75
4.1.4. Análisis del modelo cuadrático: WinBUGS	79
4.2. Modelos lineales generalizados	83
4.2.1. Análisis preliminar del modelo lineal	84
4.2.2. Análisis del modelo lineal: WinBUGS	87
4.2.3. Análisis preliminar del modelo cuadrático	89
4.2.4. Análisis del modelo cuadrático: WinBUGS	93
4.3. Comparación de modelos	95
4.3.1. Introducción	95
4.3.2. Medidas de Ajuste y Complejidad	95
4.4. Análisis de residuos	98
5. Conclusiones	105
A. Resultados Preliminares	107
A.1. Distribuciones más comunes	107
A.2. Especificación de la distribución inicial	112
A.2.1. Especificación subjetiva	112
A.2.2. Especificación a través de formas funcionales	113
A.2.3. Familia exponencial	114
A.2.4. Distribución inicial no informativa	120
A.3. Análisis de residuos	123
A.3.1. Métodos tradicionales	123
A.3.2. Un método Bayesiano	127

ÍNDICE GENERAL

v

A.4. Componentes de la verosimilitud para los MLG	130
B. Programas	135
B.1. Regresión lineal normal	135
B.1.1. Análisis preliminar del modelo lineal	135
B.1.2. Análisis del modelo lineal: WinBUGS	142
B.1.3. Análisis preliminar del modelo cuadrático	146
B.1.4. Análisis del modelo cuadrático: WinBUGS	152
B.2. Modelos lineales generalizados	156
B.2.1. Análisis preliminar del modelo lineal	156
B.2.2. Análisis del modelo lineal: WinBUGS	160
B.2.3. Análisis preliminar del modelo cuadrático	162
B.2.4. Análisis del modelo cuadrático: WinBUGS	166

Capítulo 1

Introducción

1.1. La Estadística y la Actuaría

Históricamente, la teoría de la probabilidad y los métodos estadísticos han jugado un papel central en la ciencia actuarial, tanto en la teoría como en la práctica. El temprano desarrollo de estos temas está ligado con muchos de los pioneros de la teoría actuarial, tales como Abraham de Moivre, Thomas Simpson, Daniel y Nicolas Bernoulli. En realidad, algunos modelos estadísticos modernos tienen antecedentes actuariales poco conocidos.

Comenzaremos esta revisión con una breve consideración de la naturaleza actuarial. La ciencia actuarial está relacionada con la administración financiera de sistemas de seguridad, los cuales pueden ser descritos como "mecanismos para reducir el impacto financiero desfavorable de eventos aleatorios que previenen el cumplimiento de expectativas razonables".

Tales sistemas tienen ciertas limitaciones fundamentales. Por ejemplo, están restringidos a reducir las consecuencias de eventos aleatorios que crean pérdidas que pueden ser medidas en términos monetarios. Segundo, tales sistemas no reducen directamente la probabilidad de que una pérdida ocurra.

Entre los ejemplos más comunes de situaciones donde ciertos eventos aleatorios pueden causar pérdidas financieras pueden incluirse los siguientes:

1. La destrucción de propiedad por fuego o catástrofe natural (tormenta, granizo, inundación, derrumbe, terremoto o erupción volcánica) es usualmente considerada un evento aleatorio en el cual las pérdidas pueden ser medidas en términos monetarios.

2. Una sentencia impuesta por daño por un juzgado como resultado de un evento negligente es muchas veces considerado como un evento aleatorio con resultados en pérdidas monetarias.
3. Una enfermedad prolongada puede ocurrir de improviso y resultar en pérdidas financieras en términos de reducción de ingresos y un gasto extra en salud.
4. La supervivencia a una edad avanzada puede agotar un recurso individual para mantener el costo de manutención, incluyendo cuidados a largo plazo.

Una de las tareas claves para un trabajo actuarial en sistemas de seguridad financiera es la administración de la incertidumbre. Este proceso puede ser separado en distintas etapas; por ejemplo, una clasificación sería: la identificación de fuentes de información, recolección de datos, análisis, construcción del modelo, pruebas de sensibilidad, predicción, etc.

Los modelos empleados por los actuarios como parte de este proceso tienen ciertas características importantes, comunes a través de muchas ramas de sistemas de seguridad financiera (por ejemplo, seguros de vida y pensiones). Notamos que hay un consenso en la literatura relativo a los siguientes elementos:

1. Una o más variables aleatorias, las cuales cuantifican las principales características del riesgo, tales como duración, tamaño, número o demora.
2. Un conjunto bien definido de estados de la naturaleza, separados por un evento o épocas de transición observables, junto con una ley determinística o estocástica de movimiento entre los estados.
3. Una función económica asociada con las variables fundamentales y/o los estados y eventos de transición, los cuales pueden ser también determinísticos o aleatorios, pero la mayoría de las veces ligados a factores económicos externos no controlables, tales como el crecimiento del mercado o la inflación, pero también el funcionamiento económico bajo el control de la compañía tal como el margen de ganancias de la inversión o ejecución del portafolio.

Es clara la importancia de la probabilidad y la estadística en las etapas descritas. En particular, los modelos lineales han sido utilizados con éxito para resolver diversos problemas actuariales.

1.1.1. El papel de los modelos lineales

Los modelos lineales generalizados (MLG) son una extensión natural de los muy conocidos modelos lineales clásicos. La clase de MLG incluye, como caso especial, el modelo de regresión lineal, modelos de análisis de varianza, modelos log-lineales para el análisis de tablas de contingencia, modelos logísticos para datos binarios y muchos otros.

El uso de modelos lineales clásicos en el trabajo actuarial no es nuevo. Tales modelos se han establecido como herramientas básicas para la descripción de la tasa de reclamaciones y del costo promedio por reclamación en el seguro de autos, por ejemplo, como es evidente dado el gran número de artículos sobre el tema, incluyendo Johnson y Hey (1971), Grimes (1971), Bennett (1978), Baxter *et al.* (1980) y Coutts (1984). Por otra parte, el uso de los MLG en actuaría se inició durante los 80's. Los MLG tienen una gran aplicabilidad dentro de las ciencias actuariales, específicamente en situaciones como las siguientes.

- a). Modelos de supervivencia. Por ejemplo: graduación con respecto a la edad; con respecto a la edad y al tiempo; con respecto a la edad y duración de la póliza.
- b). Modelos de estados múltiples en seguros de salud.
- c). Ajuste de ciertas distribuciones de pérdidas para la severidad de las reclamaciones en seguros de no-vida.
- d). Clasificación de riesgo. Por ejemplo, modelando el exceso de mortalidad entre fumadores.
- e). Clasificación de primas en seguros de no-vida. Frecuencia de las reclamaciones, severidad de las reclamaciones.

Cada una de estas aplicaciones involucran diferentes tipos de datos y diferentes tipos de modelos.

Uno de los propósitos de este trabajo es ilustrar el uso de los modelos lineales en la construcción de tablas de mortalidad para seguros de vida.

Una descripción más completa del uso de modelos lineales generalizados en este contexto puede encontrarse en Haberman y Renshaw (1996).

1.1.2. ¿Qué es la graduación?

La graduación puede ser vista como una colección de principios y métodos por los cuales un conjunto de observaciones (o probabilidades) es ajustado para proveer una base apropiada para hacer inferencias y además para hacer cálculos prácticos. Una de las principales aplicaciones de la graduación es la construcción de un modelo de supervivencia, usualmente presentado en la forma de una tabla de mortalidad.

Se considerará por el momento un conjunto de tasas de mortalidad, \hat{q}_x (para distintas edades de x) las cuales han sido calculadas a partir de un conjunto de observaciones. Estos valores pueden verse como una muestra de una población mayor, conteniendo así algunas fluctuaciones aleatorias. Un supuesto común acerca de las tasas "verdaderas" es que cada una está fuertemente relacionada con sus vecinos. Esta relación es expresada por la creencia de que las tasas verdaderas avanzan con suavidad de una edad a otra. Los métodos de graduación sugeridos en la literatura y usados en la práctica tienden a caer en una de dos categorías: paramétrica o no paramétrica. En este trabajo nos concentraremos en un método particular de graduación paramétrica.

El proceso de graduación es un paso esencial en la construcción de un modelo de supervivencia, asegurando que el modelo despliegue el grado requerido de suavidad. De esta manera, los cálculos prácticos basados en el modelo comparten esta importante propiedad de suavidad.

1.2. Experiencia Mexicana 82-89

1.2.1. Antecedentes

La tabla de mortalidad es la base para el cálculo de primas y reservas en el seguro de vida, y representa un registro de la mortalidad observada, durante un periodo de tiempo, de una población en particular.

Las diferentes condiciones de vida y su influencia en la mortalidad humana hace necesario que periódicamente se realicen estudios que permitan validar la mortalidad de los asegurados mexicanos. Para tal efecto, la Co-

misión Nacional de Seguros y Fianzas a través de su Dirección General de Desarrollo e Investigación, realizó un estudio con base a la experiencia de las compañías aseguradoras mexicanas durante el periodo de 1982 a 1989, cuyos resultados se presentan a continuación (ver Alatorre 1991).

Se consideró un conjunto de datos razonablemente representativos de la situación del sector asegurador. Los datos se refieren al seguro de vida individual y combinan la información de hombres y mujeres ya que no fue posible obtener resultados representativos por sexo. Por sus características, la información no distingue entre seguros con o sin examen médico.

La tabla de mortalidad que se utiliza se denomina "Tabla de mortalidad última" e incluye la mortalidad ocurrida los tres primeros años de vigencia del seguro.

La información que las instituciones de seguros proporcionaron se refiere a los siguientes conceptos:

1. Pólizas de edad alcanzada en vigor al 31 de diciembre de cada año.
2. Muertes ocurridas a edad alcanzada, durante el periodo de observación.

El número de expuestos se calcula con base en la siguiente expresión

$$E_x = P_{x+1} + d_x$$

donde

- E_x = Expuestos de edad x al principio del año.
 P_{x+1} = Pólizas en vigor de edad $x + 1$ reportadas al 31 de diciembre de cada año.
 d_x = Número de asegurados de edad x que mueren en el transcurso del año.

1.2.2. Resultados

Se obtuvieron los expuestos por edad y compañía, reuniendo así la experiencia de todas las instituciones para cada año del periodo objeto de análisis.

Para el periodo de 1982 a 1989, se obtuvo como resultado para el sector de vida individual un total de 6'688,006 expuestos y 23,918 siniestros. Los datos por edad y el análisis descriptivo correspondiente, se presentan en la Sección 3.1.

Las tasas brutas obtenidas a partir de estos datos (Tabla 1.1) fueron graduadas utilizando la fórmula de Makeham, obteniéndose la tabla básica y la tabla modificada presentadas en la Tabla 1.2. La Figura 1.1 muestra las tasas brutas así como las tablas mencionadas.

A manera de referencia, cabe mencionar que para la tabla de mortalidad "Experiencia Mexicana 1962 - 1967" se tomaron como base 649,463 expuestos y 6,320 siniestros.

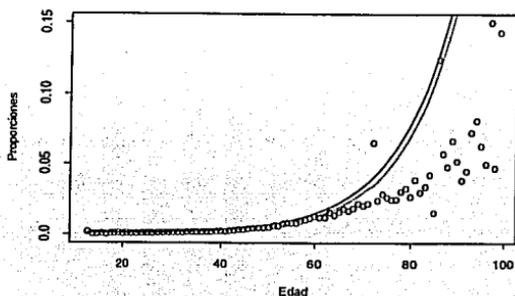


Figura 1.1: *Tasas brutas (o), Tabla básica (· · ·), Tabla modificada (-).*

Tabla 1.1: Tasas Brutas de Mortalidad

Indice	Edad	Tasa Bruta	Indice	Edad	Tasa Bruta
1	12	0.001738564	31	42	0.002407667
2	13	0.000285673	32	43	0.002843794
3	14	0.000178912	33	44	0.003345494
4	15	0.000382958	34	45	0.003392204
5	16	0	35	46	0.003800501
6	17	0.000434729	36	47	0.004113324
7	18	0.000488016	37	48	0.004438394
8	19	0.000735006	38	49	0.004518061
9	20	0.000556151	39	50	0.004782775
10	21	0.000689469	40	51	0.005892548
11	22	0.000535991	41	52	0.005450618
12	23	0.000496238	42	53	0.00736751
13	24	0.000804182	43	54	0.007758566
14	25	0.000866424	44	55	0.007774458
15	26	0.000729957	45	56	0.007863997
16	27	0.001104732	46	57	0.009173536
17	28	0.000986851	47	58	0.010105825
18	29	0.001156557	48	59	0.011132184
19	30	0.001215123	49	60	0.012562197
20	31	0.001193049	50	61	0.011702996
21	32	0.001143825	51	62	0.011695421
22	33	0.001250311	52	63	0.015313745
23	34	0.001251584	53	64	0.013471356
24	35	0.001328397	54	65	0.015569797
25	36	0.001397968	55	66	0.017928383
26	37	0.001503212	56	67	0.016536861
27	38	0.00176965	57	68	0.018177621
28	39	0.00184099	58	69	0.021535642
29	40	0.002229207	59	70	0.020044866
30	41	0.002032384	60	71	0.021586768

Índice	Edad	Tasa Bruta
61	72	0.064872895
62	73	0.0238795
63	74	0.028201996
64	75	0.025936929
65	76	0.024220562
65	77	0.024523161
66	78	0.030350438
67	79	0.032536765
68	80	0.026691887
69	81	0.038730751
70	82	0.02974467
71	83	0.033519553
72	84	0.042117465
73	85	0.015065913
74	86	0.123907863
75	87	0.056955093
76	88	0.047741935
78	89	0.066141732
79	90	0.051779935
80	91	0.037946429
81	92	0.044510386
82	93	0.072164948
83	94	0.080645161
84	95	0.0625
85	96	0.04950495
86	97	0.15
87	98	0.046875
88	99	0.142857143

Tabla 1.2: Tabla de Mortalidad, Seguro Individual

Edad	Básica	Modificada	Edad	Básica	Modificada
12	0.00060000	0.00112000	42	0.00265000	0.00384000
13	0.00061000	0.00114000	43	0.00287000	0.00411000
14	0.00061000	0.00116000	44	0.00311000	0.00440000
15	0.00062000	0.00119000	45	0.00338000	0.00472000
16	0.00063000	0.00121000	46	0.00367000	0.00507000
17	0.00064000	0.00124000	47	0.00399000	0.00545000
18	0.00065000	0.00127000	48	0.00435000	0.00586000
19	0.00067000	0.00130000	49	0.00473000	0.00631000
20	0.00068000	0.00134000	50	0.00550000	0.00680000
21	0.00070000	0.00138000	51	0.00562000	0.00733000
22	0.00072000	0.00142000	52	0.00612000	0.00791000
23	0.00074000	0.00147000	53	0.00668000	0.00855000
24	0.00077000	0.00152000	54	0.00729000	0.00924000
25	0.00079000	0.00157000	55	0.00796000	0.01000000
26	0.00086000	0.00164000	56	0.00869000	0.01082000
27	0.00086000	0.00170000	57	0.00949000	0.01172000
28	0.00090000	0.00177000	58	0.01036000	0.01269000
29	0.00094000	0.00185000	59	0.01132000	0.01376000
30	0.00099000	0.00194000	60	0.01237000	0.01492000
31	0.00104000	0.00203000	61	0.01351000	0.01619000
32	0.00110000	0.00214000	62	0.01476000	0.01757000
33	0.00138000	0.00225000	63	0.01613000	0.01907000
34	0.00147000	0.00237000	64	0.01762000	0.02070000
35	0.00158000	0.00250000	65	0.01924000	0.02249000
36	0.00169000	0.00265000	66	0.02102000	0.02443000
37	0.00181000	0.00281000	67	0.02296000	0.02654000
38	0.00195000	0.00298000	68	0.02507000	0.02884000
39	0.00210000	0.00317000	69	0.02739000	0.03134000
40	0.00227000	0.00338000	70	0.02992000	0.03406000
41	0.00245000	0.00360000	71	0.03270000	0.03702000

Edad	Básica	Modificada
72	0.03433000	0.04024000
73	0.03768000	0.04375000
74	0.04129000	0.04755000
75	0.04520000	0.05169000
76	0.04943000	0.05618000
77	0.05402000	0.06105000
78	0.05899000	0.06634000
79	0.06439000	0.07208000
80	0.07022000	0.07829000
81	0.07654000	0.08503000
82	0.08337000	0.09232000
83	0.09076000	0.10021000
84	0.09877000	0.10874000
85	0.10750000	0.11796000
86	0.11701000	0.12790000
87	0.12725000	0.13862000
88	0.13817000	0.15017000
89	0.14981000	0.16259000
90	0.16213000	0.17593000
91	0.17510000	0.19025000
92	0.18866000	0.20558000
93	0.20259000	0.22198000
94	0.23948000	0.23948000
95	0.25813000	0.25813000
96	0.27795000	0.27795000
97	0.29898000	0.29898000
98	0.32121000	0.32121000
99	1.00000000	1.00000000

1.3. El enfoque Bayesiano

La Estadística Bayesiana ha tenido un desarrollo importante en los últimos años y un gran impacto en las más diversas áreas de aplicación de la estadística. La Estadística Bayesiana es una herramienta poderosa ya que nos permite cuantificar la información que un sujeto posee sobre un determinado suceso, basándose en la interpretación subjetiva de la probabilidad.

El Teorema de Bayes provee de un procedimiento inferencial mediante el cual, a partir de una distribución de probabilidad "inicial", se obtiene la distribución "final" que describe nuestro estado de conocimiento a la luz de toda la información disponible. El Teorema de Bayes nos permite incorporar la información contenida en un conjunto de datos, produciendo una descripción conjunta de la incertidumbre sobre los valores de los parámetros del modelo a través de la distribución final. La implementación de técnicas Bayesianas requiere de un esfuerzo computacional muy alto. En el siguiente capítulo se dará una breve descripción de algunos métodos computacionales.

Capítulo 2

Preliminares Técnicos

2.1. Estimación de las tasas de mortalidad individuales

En esta sección consideraremos el problema de estimar la tasa de mortalidad q_x , individualmente para cada edad x . Puesto que q_x es un parámetro de la población, su verdadero valor nunca podrá ser conocido con exactitud, pero puede ser estimado a partir de una muestra. La aproximación más simple es considerar un experimento binomial, en el que se comienza con una muestra aleatoria de n individuos (todos de edad x , seleccionados de la población de interés) y se observan por un periodo de un año. Si d denota el número (aleatorio) de muertes observadas entre estos n individuos, entonces d/n proporciona un estimador apropiado para q_x .

En estudios basados en vidas aseguradas, esta solución ideal no es aplicable. Los datos usados para estimar la mortalidad son tomados de los registros que las compañías aseguradoras mantienen de sus asegurados. Algunos de los asegurados que están bajo observación a la edad x pueden dejar sus pólizas antes de cumplir la edad $x + 1$, y consecuentemente no se tiene un seguimiento de lo que pasa con ellos. Estos asegurados salen prematuramente de observación y cualquier muerte entre ellos que ocurra después de su salida pero antes de la edad $x + 1$ no es observable y por lo tanto no puede contribuir al cálculo de d . En este caso, d será menor que el número real de muertes que resultan de los n individuos originalmente considerados y por tanto d/n subestimarà a q_x . Este problema surge porque existen dos formas de decremento de la población de interés (muertes y retiros). La observación de un

individuo puede terminar sólo por una de estas dos causas, las cuales están operando simultáneamente. La proporción d/n está estimando la probabilidad de una muerte *observada*, pero q_x , es la probabilidad de una muerte *real*, haya sido observada o no. Una posible solución es usar el llamado estimador actuarial

$$\hat{q}_x = \frac{d}{(n - \frac{1}{2}w)} \quad (2.1)$$

donde w es el número (aleatorio) de individuos que se retiran.

Notemos que, en la proporción d/n , cada individuo contribuye con una unidad en el denominador (número de expuestos). Cuando un individuo se retira, pierde el potencial para contribuir a d y, consecuentemente, no debería contribuir con una unidad completa al número de expuestos. Suponiendo que, en promedio, los retiros ocurren en la edad $x + \frac{1}{2}$ entonces podemos sustraer media unidad al número expuestos por cada retiro. Esta es justamente la idea detrás del estimador (2.1). Una explicación alternativa se obtiene al suponer que la mitad de los retiros ocurre en la edad x y la otra mitad en la edad $x + 1$. Lo anterior es equivalente a considerar un experimento binomial con una muestra de tamaño $n - \frac{1}{2}w$. Dado que se observan d siniestros, el estimador de q_x es (2.1).

La razón principal para considerar estimadores máximo verosímil son sus bien conocidas propiedades asintóticas. A continuación consideraremos la estimación de q_x bajo distintos supuestos.

Modelo básico de retiros aleatorios

Dado un intervalo de edad $[x, x + 1]$, cada individuo en la muestra es observado desde la edad x , hasta que ocurre una de las siguientes situaciones

- (a) el individuo muere,
- (b) el individuo se retira,
- (c) el individuo alcanza la edad $x + 1$.

Para cada individuo, se conoce la forma en la que su observación termina, así como el tiempo T , medido desde la edad x , en el que el individuo deja de observarse. Para modelar a T , asociamos a cada individuo un par de variables aleatorias continuas e independientes, Y y Z , que representan al tiempo de *muerte* y *retiro*, respectivamente. Entonces $T = \min(Y, Z, 1)$. Por ejemplo, si $Z < Y$ y $Z < 1$, entonces $T = Z$ y el retiro será observado en la edad $x + Z$.

El parámetro de interés es $q_x = \Pr(Y \leq 1)$. Sea $Q_x = \Pr(Y \leq Z, Y \leq 1)$, i.e. la probabilidad de una muerte observada. Claramente $q_x \geq Q_x$.

Llamaremos a q_x la tasa de mortalidad y a Q_x la probabilidad de muerte. De manera análoga $r_x = \Pr(Z \leq 1)$ y $R_x = \Pr(Z \leq Y, Z \leq 1)$ serán la tasa y la probabilidad de retiro, respectivamente. Las funciones de distribución para Y y Z son denotadas como $F(\cdot)$ y $H(\cdot)$, respectivamente. Por ejemplo ${}_tq_x \equiv \Pr(Y \leq t) = F(t)$ y ${}_tr_x \equiv \Pr(Z \leq t) = H(t)$. Las funciones de densidad, cuando existan, serán denotadas de manera respectiva como $f(\cdot)$ y $h(\cdot)$.

Sea D_i una variable indicadora (aleatoria) que toma el valor 1 si el i -ésimo individuo muere y el valor 0 en caso contrario, de manera que $D = \sum_i D_i$ es el número total de muertes observadas. De forma similar, W_i denotará la variable indicadora de los retiros y $W = \sum_i W_i$ será el número total de retiros.

Supongamos por el momento que cada individuo es observado desde la edad x hasta que muere o hasta que alcanza la edad $x+1$, lo que ocurra primero, de manera que es imposible que se retire y por lo tanto $T = \min(Y, 1)$. Este es un caso particular del modelo básico de retiros aleatorios con $Z = \infty$. Bajo estas condiciones, los datos consisten de las variables D_i, T_i ($i = 1, \dots, n$), cuyos valores observados serán denotados por d_i y t_i , respectivamente. La contribución del i -ésimo individuo a la verosimilitud es entonces

$$\Pr[t_i < T_i \leq t_i + dt_i, D_i = d_i] \approx \begin{cases} (1 - q_x) & \text{si } d_i = 0, t_i = 0 \\ (1 - {}_tq_x) \mu_{x+t_i} dt_i & \text{si } d_i = 1, t_i < 1, \end{cases}$$

donde μ_{x+t_i} denota la fuerza de mortalidad para la edad x (Broffitt, 1984). La función de verosimilitud para q_x está dada entonces por

$$L(q_x) = (1 - q_x)^{n-d} \prod_{i \in \mathcal{D}} \{(1 - {}_tq_x) \mu_{x-t_i}\}, \quad (2.2)$$

donde \mathcal{D} es el subconjunto de índices correspondiente a las muertes observadas.

Para encontrar el valor de q_x que maximiza (2.2), primero debemos expresar ${}_tq_x$ en términos de q_x . Con este fin, es necesario imponer algunas restricciones adicionales en el modelo. Tres supuestos comunes en la literatura actuarial son:

(A) distribución uniforme de las muertes

$${}_{t-a}q_{x+a} = \frac{(t-a)q_x}{1-aq_x};$$

(B) la hipótesis de Balducci

$${}_{t-a}q_{x+a} = \frac{(t-a)q_x}{1 - (1-t)q_x};$$

(C) fuerza de mortalidad constante

$${}_{t-a}q_{x+a} = 1 - \exp\{-\mu(t-a)\};$$

con $0 \leq a \leq 1$, $0 \leq t \leq 1$ y $0 \leq t+a \leq 1$.

Los supuestos (A) y (C) son generalmente considerados como razonables. Sin embargo, el supuesto (B) puede producir resultados poco razonables, aunque bajo ciertas condiciones da lugar a estimaciones relativamente simples (Batten, 1978). En este trabajo haremos uso del supuesto (A), bajo el cual la verosimilitud (2.2) toma la forma

$$L(q_x) = q_x^d (1 - q_x)^{n-d}$$

y, por lo tanto, el estimador de máxima verosimilitud en este caso es $\hat{q}_x = d/n$.

2.2. Inferencia Bayesiana

2.2.1. Fundamentos

La metodología Bayesiana se deriva de la teoría de decisiones. Al describir un problema de decisión en ambiente de incertidumbre es necesario:

- i. Especificar el conjunto D de posibles alternativas d o espacio de decisión.
- ii. Especificar el conjunto de sucesos inciertos θ , de los cuales pueden depender las consecuencias de las decisiones. En problemas de decisión estadísticos, este conjunto corresponde generalmente al espacio paramétrico y se denota por Θ .
- iii. Asignar una distribución de probabilidad, $p(\theta)$, que describa la información del interesado sobre los sucesos inciertos en el momento de tomar la decisión. En un problema de decisión estadístico, la variable θ frecuentemente se identifica con el parámetro del modelo.

- iv. Construir una función $U : \Theta \times D \rightarrow \mathfrak{R}$, llamada función de utilidad, que describa las preferencias del interesado entre las posibles consecuencias.

Una vez que el decisor ha contemplado los puntos i a iv, el paso final es seleccionar aquella decisión d^* que maximiza la utilidad esperada, es decir

$$\int_{\Theta} U(\theta, d) p(\theta) d\theta.$$

La teoría Bayesiana de la decisión está fundamentada en una serie de postulados conocidos como "axiomas de coherencia". Estos postulados se refieren al comportamiento racional que debe observar un tomador de decisiones cuando se encuentra frente a un problema de decisión. Si el tomador de decisiones (al cual llamaremos "interesado") aceptara todos los postulados de coherencia, entonces debe seguir una metodología bien definida para tomar la decisión. Específicamente, debe de construir una distribución de probabilidad para los sucesos inciertos que reflejen su conocimiento acerca de dichos sucesos, debe construir una función de utilidad que describa sus preferencias entre las posibles consecuencias y debe utilizar el criterio de maximizar la utilidad esperada para tomar la decisión.

Dado que el concepto de probabilidad es fundamental para el desarrollo de los métodos estadísticos, se discutirán algunas de las ideas básicas sobre este concepto.

La teoría de probabilidad se ha venido desarrollando desde el siglo XVII y es ampliamente aplicada en diversos campos de la ciencia. Además de muchas aplicaciones formales, el concepto de probabilidad aparece en nuestra vida cotidiana. A menudo oímos y decimos expresiones semejantes a: "es probable que llueva en la tarde", "probablemente llegará por la noche", o "tiene posibilidades de pasar el examen". Cada una de estas expresiones refleja la noción de la verosimilitud de la ocurrencia de algún suceso específico.

Existen diversas interpretaciones del concepto de probabilidad, pero las más importantes son tres: la clásica, la frecuentista y la subjetiva.

La interpretación clásica está basada en el concepto de resultados *igualmente verosímiles*. Por ejemplo cuando se lanza una moneda existen dos resultados posibles: cara o cruz. Si se puede suponer que la ocurrencia de estos resultados es igualmente verosímil, entonces deben tener la misma probabilidad.

Dos dificultades básicas aparecen cuando se intenta desarrollar una definición formal de probabilidad desde el punto de vista clásico. En primer

lugar, el considerar resultados igualmente verosímiles es lo mismo que afirmar que los resultados tienen la misma probabilidad, lo que da lugar a una definición circular. En segundo lugar, no proporciona un método sistemático para asignar probabilidades a resultados que no sean igualmente verosímiles. Cuando se lanza una moneda o se arroja un dado equilibrado o se escoge un carta de una baraja bien mezclada, los diferentes resultados posibles pueden en general considerarse igualmente verosímiles debido a la naturaleza del proceso. Sin embargo, cuando el problema es predecir si una persona se casará o si un proyecto de investigación tendrá éxito, los resultados posibles no suelen considerarse igualmente verosímiles, y es necesario un método diferente para asignar probabilidades a estos resultados.

Por otra parte, la probabilidad frecuentista puede interpretarse en el sentido de la frecuencia relativa con la que se obtendría un resultado si el proceso se repitiera un número grande de veces en condiciones similares. La interpretación frecuentista de la probabilidad no puede aplicarse, por ejemplo, a la probabilidad de que una persona determinada contraiga matrimonio en los próximos dos años. El concepto de frecuencia relativa de la probabilidad tiene sentido intuitivamente, pero no proporciona una definición general de la probabilidad.

Finalmente, la interpretación subjetiva considera a la probabilidad como el grado de creencia que un sujeto tiene acerca de la ocurrencia de algún suceso, de acuerdo al conocimiento que dicho sujeto tiene del suceso en cuestión. Otra persona, que puede tener diferentes opiniones o información distinta, puede tener un grado de creencia mayor o menor acerca del mismo suceso. Por esta razón conviene hablar de *probabilidad subjetiva* en lugar de la *verdadera probabilidad* de ese suceso.

La interpretación subjetiva de la probabilidad puede formalizarse. En general, si los juicios de una persona acerca de las verosimilitudes relativas a diversas combinaciones de resultados satisfacen ciertas condiciones de consistencia, entonces puede demostrarse que sus probabilidades subjetivas para los diferentes sucesos posibles pueden ser determinadas de forma única. Esta interpretación sin embargo, tiene dos dificultades. En primer lugar, el requisito de que los juicios de una persona sobre las verosimilitudes relativas a un número infinito de sucesos sean completamente consistentes y libres de contradicciones no parece humanamente posible. En segundo lugar, la interpretación subjetiva no proporciona bases "objetivas" para que dos o más científicos obtengan una evaluación conjunta de su estado de conocimientos en una área científica de interés común. Por otro lado, aceptar la interpreta-

ción subjetiva tiene el efecto positivo de subrayar algunos aspectos subjetivos de la ciencia. La evaluación de un determinado científico acerca de la probabilidad de algún resultado incierto debe ser, en última instancia, su propia evaluación, basada en todas las evidencias de que dispone. Esta evaluación puede estar parcialmente basada en la interpretación frecuentista de la probabilidad, ya que el científico puede tener en cuenta la frecuencia relativa de la ocurrencia de este resultado o de resultados similares en el pasado; incluso puede basarse en la interpretación clásica de probabilidad, puesto que el científico puede tener en cuenta el número total de resultados posibles que considera igualmente verosímiles. Sin embargo, la asignación final de probabilidades numéricas es responsabilidad del propio científico y refleja su estado de conocimiento.

La metodología Bayesiana se basa en la interpretación subjetiva de la probabilidad. Es decir, mediante el concepto de probabilidad se pretende cuantificar la información que el sujeto posee sobre un determinado suceso.

2.2.2. Inferencia

El problema fundamental alrededor del cual el estudio de la estadística está dirigido, es la inferencia. La inferencia estadística estudia los métodos mediante los cuales se aprovecha la información disponible para mejorar nuestro conocimiento del mundo real. Con base en una serie de datos observados, se quiere información acerca de una o más características desconocidas del sistema físico que produjo dichas observaciones.

El problema de inferencia ha sido un tema de interés considerable desde que comenzó el estudio sistemático de la teoría de la probabilidad en el siglo *XVIII*.

En la estadística Bayesiana, la aplicación del Teorema de Bayes provee de un procedimiento inferencial mediante el cual, a partir de una distribución de probabilidad "inicial" se obtiene la distribución "final" que describe nuestro estado de conocimiento a la luz de toda la información disponible. A continuación se describe dicho procedimiento.

Supóngase que se tiene un problema de decisión y que θ es la cantidad aleatoria relevante. La información disponible sobre θ se expresa a través de una distribución de probabilidad descrita en términos de su función de densidad $p(\theta)$, llamada distribución inicial.

En el caso de que la información sobre θ contenida en $p(\theta)$ no sea suficiente para los propósitos del interesado (es decir, cuando el interesado

juzga que no tiene la información suficiente para decidir adecuadamente), éste generalmente obtiene información adicional mediante la realización de un experimento.

Sea x el resultado del experimento. Si x proporciona información sobre θ es debido a que la función de densidad de x depende del valor de θ . En la situación general, la distribución de la variable x depende tanto de θ como de otras cantidades. De acuerdo con esto, la distribución de los datos puede denotarse como $p(x | \theta)$ donde $\theta = (\theta, \omega)$ lo cual significa que la variable x no sólo depende del valor de θ sino también del *parámetro de ruido* ω .

De acuerdo con el Teorema de Bayes se tiene que

$$p(\theta, \omega | x) = \frac{p(x | \theta, \omega) p(\theta, \omega)}{\int_{\Theta} p(x | \theta, \omega) p(\theta, \omega) d\theta d\omega}$$

donde Θ es el conjunto de valores que puede tomar la variable aleatoria θ .

La función $p(\theta, \omega | x)$ es llamada la distribución final de $\theta = (\theta, \omega)$.

Desde el punto de vista Bayesiano un problema de inferencia queda resuelto cuando se obtiene la distribución final de la cantidad de interés θ , dada por $p(\theta | x) = \int p(\theta, \omega | x) d\omega$. Cabe mencionar que los términos distribución inicial y distribución final son relativos a la información adquirida a través de la realización del experimento dado.

Existen ocasiones en las que el propósito de un análisis estadístico es predecir el valor de una observación futura Y^* , basándose en la información de que se dispone. De esta manera, el problema que se venía desarrollando acerca de la inferencia sobre θ puede considerarse como un paso previo para la solución del problema de predicción, que en este trabajo es de principal interés. Además, debido a resultados de consistencia, un parámetro puede verse como el límite de una sucesión de estadísticas (funciones de las observaciones), cuando el tamaño de la muestra tiende a infinito. De modo que, al hacer inferencias sobre el parámetro θ , de alguna manera puede considerarse como una forma límite de hacer inferencias predictivas acerca de las observaciones.

Dado el valor de θ , la distribución que describe el comportamiento de la observación futura Y^* esta dada por $p(y^* | \theta)$; pero no hay que olvidar que el valor de θ es desconocido. En los métodos estadístico tradicionales, este problema se ataca generalmente estimando a θ con base en la muestra observada, y simplemente sustituyendo el valor de θ por su valor estimado, no tomando en cuenta la incertidumbre sobre el valor de θ . Desde la perspectiva

Bayesiana, el modelo $p(y^* | \theta)$, junto con la distribución inicial $p(\theta)$, inducen una distribución conjunta para (Y^*, θ) , dada por,

$$p(y^*, \theta) = p(y^* | \theta) p(\theta).$$

La distribución marginal

$$p(y^*) = \int p(y^* | \theta) p(\theta) d\theta$$

describe nuestro conocimiento acerca del valor de Y^* dada la información inicial disponible. Esta distribución es conocida comúnmente como la distribución *predictiva inicial*.

Análogamente, una vez obtenida una muestra $\mathbf{Y} = (Y_1, \dots, Y_n)$, el modelo $p(y^* | \theta)$ y la distribución final inducen una distribución conjunta para (Y^*, θ) condicional en los valores observados $\mathbf{y} = (y_1, \dots, y_n)$, dada por

$$\begin{aligned} p(y^*, \theta | \mathbf{y}) &= p(y^* | \theta, \mathbf{y}) p(\theta | \mathbf{y}) \\ &= p(y^* | \theta) p(\theta | \mathbf{y}) \end{aligned}$$

donde la última igualdad se debe a la independencia condicional de Y^* y \mathbf{Y} dado θ . De manera que, la distribución

$$p(y^* | \mathbf{y}) = \int p(y^* | \theta) p(\theta | \mathbf{y}) d\theta$$

describe el comportamiento de Y^* dada la información disponible y se conoce como la distribución *predictiva final*.

2.2.3. Métodos computacionales

En el enfoque Bayesiano de la estadística, la incertidumbre presente en un modelo dado, $p(x | \theta)$, es representada a través de una distribución de probabilidad $p(\theta)$ sobre los posibles valores de parámetro desconocido θ (generalmente multidimensional) que define al modelo. El Teorema de Bayes permite entonces incorporar la información contenida en un conjunto de datos $\mathbf{x} = (x_1, x_2, \dots, x_k)$, produciendo una descripción conjunta de la incertidumbre sobre los valores de los parámetros del modelo a través de la distribución final $p(\theta | \mathbf{x})$. Desgraciadamente, la implementación de técnicas Bayesianas usualmente requiere de un esfuerzo computacional muy alto.

La mayor parte de este esfuerzo se concentra en el cálculo de ciertas características de la distribución final del parámetro de interés (que llamaremos *resúmenes inferenciales*). Así, por ejemplo, para pasar de una distribución conjunta a una colección de distribuciones y momentos marginales que sean útiles para hacer inferencias sobre subconjuntos de parámetros, se requiere integrar. En la mayoría de los casos los resúmenes inferenciales básicos se reducen a integrales de la forma

$$S_I \{g(\theta)\} = \int g(\theta) p(\theta) p(x | \theta) d\theta_{I^c}, \quad (2.3)$$

donde $g: \mathbb{R}^d \rightarrow \mathbb{R}$, $\theta = (\theta_1, \dots, \theta_d)$, $I \subseteq \{1, \dots, d\}$, $I^c \subseteq \{1, \dots, d\} \setminus I$ y $\theta_{I^c} = \{\theta_i : i \in I^c\}$.

Bajo ciertas condiciones de regularidad, conforme el tamaño de la muestra tiende a infinito la distribución final de θ converge en probabilidad a una distribución normal cuya media es el estimador de máxima verosimilitud, $\hat{\theta}$, y cuya matriz de varianzas-covarianzas es la inversa de la matriz de información de Fisher evaluada en $\hat{\theta}$. Resultados asintóticos similares dan lugar a la aproximación de Laplace, la cual permite aproximar integrales de la forma (2.3); ver, por ejemplo, Bernardo y Smith (1994, Cap. 5).

En la práctica, sin embargo, es común que la dimensión de θ sea muy grande. Por otro lado, excepto en aplicaciones muy sencillas, tanto $p(x | \theta)$ como $p(\theta)$ pueden llegar a tener formas muy complicadas. En la gran mayoría de los problemas las integrales requeridas no pueden resolverse analíticamente, por lo que es necesario contar con métodos numéricos eficientes que permitan calcular o aproximar integrales en varias dimensiones.

Algunos de los métodos computacionales de más rápido desarrollo en los últimos años se basan en técnicas de simulación. A partir de muestras simuladas de la distribución final es posible, en principio, aproximar cualquier característica de dicha distribución. A continuación revisaremos brevemente algunos de los métodos más importantes. El lector interesado puede consultar Gamerman (1997) y Gutiérrez-Peña (1997).

Monte Carlo: muestreo-remuestreo

Supongamos que se tiene una densidad $s(\theta)$ de la cual es posible simular observaciones fácilmente, pero que se requiere una muestra de la densidad

$$f(\theta) = \frac{f_0(\theta)}{\int f_0(\theta) d\theta}$$

donde sólo la forma funcional de $f_0(\theta)$ está especificada. El problema es entonces obtener una muestra de $f(\theta)$ a partir de $f_0(\theta)$ y de una muestra de $s(\theta)$.

Caso 1. Si existe una constante $M > 0$ tal que $f_0(\theta)/s(\theta) \leq M$ para todo θ , entonces es posible utilizar el método de *aceptación y rechazo* para generar variables aleatorias (ver Ripley 1987). El algoritmo es el siguiente:

1. generar una observación $\bar{\theta}$ de $s(\theta)$;
2. generar una variable $u \sim U(0, 1)$;
3. si $u \leq f_0(\bar{\theta}) / \{Ms(\bar{\theta})\}$, aceptar $\bar{\theta}$ como una observación de $f(\theta)$, en caso contrario, repetir los pasos 1 a 3.

Caso 2. Si no existe la constante M , o no puede encontrarse fácilmente, entonces podemos obtener muestras aproximadas de $f(\theta)$ de la siguiente manera. Supongamos que $\theta_1, \dots, \theta_n$ es una muestra de $s(\theta)$. Sea

$$w_i = \frac{\nu_i}{\sum_{j=1}^n \nu_j}$$

donde $\nu_i = f_0(\theta_i) / s(\theta_i)$, $i = 1, \dots, n$.

Lo anterior induce una distribución discreta sobre $\{\theta_1, \theta_2, \dots, \theta_n\}$ tal que $\Pr(\theta = \theta_i) = w_i$. Si ahora generamos una observación $\bar{\theta}$ de dicha distribución entonces $\bar{\theta}$ se distribuye aproximadamente como $f(\theta)$. La aproximación es mejor a medida que $n \rightarrow \infty$.

A diferencia del caso anterior, aquí es posible muestrear con reemplazo, por lo que la muestra generada puede ser tan grande como se quiera. Es claro sin embargo, que la aproximación será más adecuada a medida que $s(\theta)$ se parezca a $f(\theta)$. El lector que se interese en el tema puede consultar el artículo de Smith y Gelfand (1992).

Los métodos descritos anteriormente permiten simular distribuciones de manera directa. Sin embargo, en dimensiones grandes estos métodos no son muy eficientes.

Monte Carlo vía cadenas de Markov

En términos simples, una cadena de Markov es un proceso estocástico $\{\theta^t : t \in T\}$ tal que, dado el estado presente, el pasado y el futuro del proceso son independientes. En otras palabras,

$$\begin{aligned} \Pr(\theta^{(n+1)} \in A \mid \theta^{(n)} = z, \theta^{(n-1)} \in A_{n-1}, \dots, \theta^{(0)} \in A_0) \\ = \Pr(\theta^{(n+1)} \in A \mid \theta^{(n)} = z) \end{aligned}$$

para todos los conjuntos $A_0, \dots, A_{n-1} \subset S$, donde S denota el espacio de estados de la cadena.

Las técnicas de Monte Carlo vía cadenas de Markov permiten generar, de manera iterativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos directos. La idea básica es muy simple: construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución final que nos interesa.

Proposición 2.1 Sea $\theta^{(1)}, \theta^{(2)}, \dots$ una cadena de Markov homogénea, irreducible y aperiódica, con espacio de estados Θ y distribución de equilibrio $p(\theta \mid x)$. Entonces, conforme $t \rightarrow \infty$,

- i) $\theta^{(t)} \xrightarrow{D} \theta$, donde $\theta \sim p(\theta \mid x)$;
- ii) $\frac{1}{t} \sum_{i=1}^t g(\theta^{(i)}) \rightarrow E(g(\theta) \mid x)$.

Muestreo de Gibbs

El algoritmo de Gibbs permite simular una cadena de Markov $\theta^{(1)}, \theta^{(2)}, \dots$ con distribución de equilibrio $p(\theta \mid x)$. En este caso, sin embargo, cada valor nuevo de la cadena se obtiene a través de un proceso iterativo que sólo requiere generar muestras de distribuciones cuya dimensión es menor que d y que en la mayoría de los casos tiene una forma más sencilla que la de $p(\theta \mid x)$.

Sea $\theta = (\theta_1, \dots, \theta_n)$ una partición del vector θ , donde $\theta_i \in \mathbb{R}^{d_i}$ y $\sum_{i=1}^n d_i = d$. Las densidades

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

se conocen como *densidades condicionales completas* y en general pueden identificarse fácilmente al inspeccionar la forma de la distribución final $p(\theta | \mathbf{x})$. De hecho para cada $i = 1, 2, \dots, n$,

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

donde $p(\theta | \mathbf{x})$ es vista sólo como función de θ_i .

Dado un valor inicial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$, el algoritmo de Gibbs simula una cadena de Markov en la que $\theta^{(t+1)}$ se obtiene a partir de $\theta^{(t)}$ de la siguiente manera:

- generar una observación $\theta_1^{(t+1)}$ de $p(\theta_1 | \theta_2^{(t)}, \dots, \theta_n^{(t)}, \mathbf{x})$;
- generar una observación $\theta_2^{(t+1)}$ de $p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_n^{(t)}, \mathbf{x})$;
- generar una observación $\theta_3^{(t+1)}$ de $p(\theta_3 | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \dots, \theta_n^{(t)}, \mathbf{x})$;

⋮

generar una observación $\theta_n^{(t+1)}$ de $p(\theta_n | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, \dots, \theta_{n-1}^{(t+1)}, \mathbf{x})$.

La sucesión obtenida $\theta^{(1)}, \theta^{(2)}, \dots$, es una realización de una cadena de Markov cuya distribución de transición está dada por

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

Convergencia

Supongamos que se desea generar una muestra de tamaño n de la distribución $p(\theta | \mathbf{x})$. Si para cada uno de n valores iniciales $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)}$, corremos el algoritmo de Gibbs, entonces, de acuerdo con la Proposición 2.1, después de cierto número de iteraciones suficientemente grande los valores $\theta_1^{(T)}, \theta_2^{(T)}, \dots, \theta_n^{(T)}$ pueden considerarse como una muestra de tamaño n de la distribución final de θ . Alternativamente podemos generar una sola cadena y tomar los valores $\theta_1^{(T+K)}, \theta_2^{(T+2K)}, \dots, \theta_n^{(T+nK)}$ como una muestra de $p(\theta | \mathbf{x})$, donde K se elige de manera que la correlación entre las observaciones sea pequeña.

En general no es fácil determinar en qué momento la(s) cadena(s) ha(n) convergido. Un método empírico comúnmente utilizado, basado en la Proposición 2.1 (ii), consiste en graficar los promedios ergódicos de algunas funciones de θ contra el número de iteraciones y elegir el valor T a partir del cual las gráficas se estabilizan. En este caso es frecuente omitir los primeros valores de la(s) cadena(s) al calcular los promedios ergódicos. La idea de este *periodo de calentamiento* es permitir que la(s) cadena(s) salga(n) de una primera fase de inestabilidad. En este caso en particular la velocidad de convergencia depende fuertemente de la correlación entre los componentes del vector θ bajo la distribución final $p(\theta | \mathbf{x})$: entre más alta sea la correlación más lenta será la convergencia.

¿Cómo se utilizan las muestras para hacer inferencias?

Como se mencionó anteriormente, una vez que se tiene una muestra de la distribución final de θ es posible aproximar esencialmente cualquier característica de dicha distribución. Por ejemplo, ciertas medidas de tendencia central (generalmente utilizadas como estimadores puntuales de θ) pueden aproximarse utilizando las cantidades muestrales correspondientes. De manera similar los cuantiles de la distribución final de θ pueden utilizarse para construir intervalos de credibilidad para θ .

WinBUGS

WinBUGS es un programa que realiza inferencias Bayesianas de modelos estadísticos utilizando el algoritmo conocido como muestreo de Gibbs, descrito en la sección anterior. El programa requiere de la especificación de un modelo de probabilidad, en el que todas las cantidades desconocidas son

tratadas como variables aleatorias. El modelo consiste de una distribución conjunta sobre todas las cantidades relevantes no observadas (tanto parámetros como observaciones faltantes). El proceso de marginalización de esta distribución, con el objeto de hacer inferencias sobre parámetros específicos u observaciones faltantes, se lleva a cabo a través de técnicas de Monte Carlo.

La base del programa es el lenguaje *BUGS*, diseñado con el propósito de facilitar la especificación del modelo. Existe un pequeño conjunto de instrucciones que controlan la sesión, en la que se analiza un modelo estadístico expresado mediante el lenguaje *BUGS*. Posteriormente, un compilador procesa el modelo y los datos disponibles para transformarlos en un estructura de datos interna, apropiada para realizar los cálculos de manera eficiente. Finalmente, el programa opera sobre esta estructura para generar valores de las cantidades desconocidas de interés. Para una descripción detallada de este lenguaje ver, por ejemplo, Spiegelhalter *et al.* (1996).

Las primeras versiones de *BUGS* fueron implementadas en diversos sistemas, incluyendo DOS y las variantes más importantes de UNIX. Sin embargo, desde hace algunos años la única versión desarrollada con regularidad está disponible sólo para el sistema Windows y es conocida como *WinBUGS*.

2.3. Regresión

2.3.1. Análisis de regresión lineal

El propósito de esta sección es dar una breve introducción al análisis de regresión lineal. El lector interesado en el tema puede consultar Mendenhall *et al.* (1986) y DeGroot (1988).

El análisis de regresión puede definirse ampliamente como el análisis de la relación entre variables. Esta es una de las herramientas estadísticas más usadas porque provee uno de los métodos más simples para establecer una posible relación funcional entre variables. La relación se expresa en la forma de una ecuación conectando el valor esperado de la variable de respuesta o dependiente Y , y una o más variables explicativas o independientes x_1, x_2, \dots, x_k .

Dicha ecuación generalmente toma la forma

$$E(Y) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

donde $\beta_0, \beta_1, \dots, \beta_k$ son llamados los coeficientes de regresión.

Una ecuación de regresión que contiene sólo una variable independiente es llamada una ecuación de regresión simple, mientras que una ecuación de regresión que contiene más de una variable independiente es llamada una ecuación de regresión múltiple.

Un ejemplo de regresión simple puede ser un análisis en el cual el tiempo de reparación de una máquina es estudiado con relación al número de componentes a ser reparadas. Aquí nosotros tenemos una variable dependiente (tiempo de reparación de la máquina) y una variable independiente (número de componentes a ser reparados). Por otra parte, un ejemplo de una situación de regresión múltiple puede ser una prueba para explicar el porcentaje de mortalidad en diferentes regiones geográficas (variable dependiente) por un gran número de factores socioeconómicos y ambientales (variables independientes).

La determinación explícita de una ecuación de regresión es, en un sentido, el producto final del análisis. Esto es un resumen de la relación entre Y (variable dependiente) y el conjunto de variables independientes (las x 's). La ecuación puede usarse con varios propósitos. Por ejemplo, para evaluar la importancia de las x 's individuales, para analizar los efectos en la respuesta al realizar ciertos cambios en los valores de las x 's, o para pronosticar valores de Y para un conjunto dado de x 's.

Nosotros veremos el análisis de regresión como un conjunto de técnicas para analizar datos con el propósito de entender la relación entre variables de cierto ambiente. Esto es, suponemos que los datos del ambiente están disponibles. Algunas veces los datos podrán ser reunidos en un escenario controlado, así que los factores que no son de interés primario pueden mantenerse constantes. En otras ocasiones los datos son recolectados bajo condiciones no experimentales, donde muy pocos pueden ser controlados por el investigador.

La tarea del análisis de regresión es aprender tanto como sea posible acerca del proceso que generó los datos, así como predecir observaciones futuras de dicho proceso.

El método de mínimos cuadrados

Un procedimiento para estimar los parámetros de cualquier modelo lineal es el método de mínimos cuadrados, que se puede ilustrar de manera sencilla aplicándolo para ajustar una línea recta a través de un conjunto de puntos que representan los datos. Supóngase que se desea ajustar el modelo

$$E(Y) = \beta_0 + \beta_1 x.$$

Se postula que $Y = \beta_0 + \beta_1 x + \varepsilon$, donde ε es un error aleatorio con una distribución de probabilidad tal que $E(\varepsilon) = 0$. Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores de los parámetros β_0 y β_1 respectivamente, $\hat{\beta}_0 + \hat{\beta}_1 x$ es un estimador de $E(Y)$.

El procedimiento de mínimos cuadrados para ajustar una línea recta a través de un conjunto de n puntos es similar al método que podríamos utilizar para ajustar una recta a simple vista, es decir, se pretende que las desviaciones sean "pequeñas" en cierto sentido. Una manera conveniente de lograr ésto, y que proporciona propiedades adecuadas, es minimizar la suma de los cuadrados de las desviaciones verticales de la recta ajustada. Por lo tanto, si

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

es el valor que se predice del i -ésimo valor de y (cuando $x = x_i$), entonces la desviación del valor observado de y a partir de la recta \hat{y} (llamada a veces *el error*) es

$$e = y - \hat{y}$$

y la suma de cuadrados de las desviaciones que debe minimizarse es

$$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

Es decir, debe resolverse el siguiente sistema de ecuaciones:

$$\frac{\partial SCE}{\partial \hat{\beta}_0} = \frac{\partial \left\{ \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \right\}}{\partial \hat{\beta}_0} = 0$$

$$\frac{\partial SCE}{\partial \hat{\beta}_1} = \frac{\partial \left\{ \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \right\}}{\partial \hat{\beta}_1} = 0$$

con respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$.

Equivalentemente,

$$\begin{aligned}
 -2 \left(\sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i \right) &= 0 \\
 -2 \left(\sum_{i=1}^n x_i y_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \right) &= 0.
 \end{aligned}
 \tag{2.4}$$

Estas ecuaciones son llamadas *las ecuaciones normales*. Nótese que las ecuaciones son lineales en $\hat{\beta}_0$ y $\hat{\beta}_1$ y por tanto se pueden resolver simultáneamente de manera sencilla. Después de realizar un poco de álgebra se obtienen las siguientes soluciones para $\hat{\beta}_0$ y $\hat{\beta}_1$,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

El método de máxima verosimilitud

La *función de verosimilitud* de n variables aleatorias Y_1, Y_2, \dots, Y_n se define a partir de su densidad conjunta, es decir,

$$f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n; \theta),$$

al considerar ésta como una función de θ . En particular, si Y_1, Y_2, \dots, Y_n es una muestra aleatoria con densidad $f(y; \theta)$, entonces la función de verosimilitud es proporcional a $f(y_1; \theta) f(y_2; \theta) \cdots f(y_n; \theta)$. La función de verosimilitud es una función de θ , y la denotaremos como $L(\theta; y_1, y_2, \dots, y_n)$.

Estimación Puntual Sean Y_1, Y_2, \dots, Y_n variables aleatorias normales e independientes, con media $\beta_0 + \beta_1 x_i$, para cada y_i , y varianza σ^2 .

La función de verosimilitud es entonces

$$L(\beta_0, \beta_1, \sigma^2) = L(\beta_0, \beta_1, \sigma^2; y_1, y_2, \dots, y_n) \\ = \prod_{i=1}^n \left\{ \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \exp \left[-\frac{1}{2} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma^2} \right)^2 \right] \right\}$$

y

$$\ln(L(\beta_0, \beta_1, \sigma^2)) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Obteniendo las derivadas parciales de $\ln(L(\beta_0, \beta_1, \sigma^2))$ con respecto a β_0, β_1 y σ^2 e igualándolas a 0 se tiene las siguientes ecuaciones

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \\ \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = n\hat{\sigma}^2.$$

Notemos que, en este caso, las dos primeras ecuaciones son las mismas que las ecuaciones normales (2.4) cuya solución produce los estimadores de mínimos cuadrados para β_0 y β_1 .

Denotaremos por $\hat{\beta}_0, \hat{\beta}_1$ y $\hat{\sigma}^2$ a los estimadores de los parámetros β_0, β_1 y σ^2 respectivamente. Después de un poco de álgebra se llega a la siguiente solución

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Estos son los *estimadores máximo verosímiles* de β_0, β_1 y σ^2 respectivamente.

Supuestos básicos del modelo de regresión lineal Los supuestos bajo los cuales trabaja el modelo de regresión lineal son los siguientes:

1. $E(\varepsilon_i) = 0 \quad \forall i$.
2. $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$.
3. $Var(\varepsilon_i) = \sigma^2$.
4. Las X_i son fijas.

En muchas situaciones se hace el siguiente supuesto adicional

5. Los ε_i tienen una distribución normal. Este supuesto permite hacer inferencias estadísticas sobre los parámetros del modelo. Por ejemplo, es posible estimar β_0 , β_1 y σ^2 a través del método de máxima verosimilitud, o construir intervalos de confianza para estos parámetros.

Notación matricial para el modelo de regresión lineal En el caso de regresión múltiple, es conveniente introducir la notación matricial. Supóngase que se tiene el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$$

y hacemos n observaciones independientes y_1, y_2, \dots, y_n de Y . Podemos escribir y_i como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$$

donde x_{ij} es el valor de la j -ésima variable independiente para la i -ésima observación, $i = 1, \dots, n$. Si se definen ahora las siguientes matrices

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

donde β es un vector de parámetros a ser estimado, las n ecuaciones se pueden escribir de la siguiente manera:

$$Y = X\beta + \varepsilon. \quad (2.5)$$

Si $\varepsilon \sim N_n(0, \sigma^2 \mathbf{I}_n)$, donde \mathbf{I}_n denota a la matriz identidad de orden n . La función de verosimilitud para el modelo (2.5) es de la forma

$$L(\beta, \sigma^2; \mathbf{y}) \propto (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) \right\}.$$

Los estimadores de máxima verosimilitud para β y σ^2 están dados por

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \\ \hat{\sigma}^2 &= \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta}) \end{aligned}$$

y notemos que

$$(\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) = (\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta}) + (\beta - \hat{\beta})' \mathbf{X}'\mathbf{X}(\beta - \hat{\beta}).$$

Así, la función de verosimilitud puede escribirse como

$$L(\beta, \sigma^2; \mathbf{y}) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} [(\beta - \hat{\beta})' \mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) + n\hat{\sigma}^2] \right\}.$$

Para facilitar la notación y el desarrollo subsecuente, es conveniente trabajar en términos de la precisión $\tau = \frac{1}{\sigma^2}$ en lugar de la varianza σ^2 . La verosimilitud toma entonces la forma

$$L(\beta, \tau; \mathbf{y}) \propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} [(\beta - \hat{\beta})' \mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) + n\hat{\sigma}^2] \right\}. \quad (2.6)$$

Cabe señalar que el estimador máximo-verosímil de β coincide con el estimador obtenido por el método de mínimos cuadrados. En este caso, además,

$$E(\hat{\beta}) = \beta$$

de tal modo que $\hat{\beta}$ resulta ser un estimador insesgado para β . Por otro lado, la varianza de $\hat{\beta}$ queda expresada por

$$\text{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

2.3.2. Análisis Bayesiano del modelo de regresión

Como se ha venido mencionando, en la estadística Bayesiana la aplicación del Teorema de Bayes provee de un procedimiento inferencial mediante el cual, a partir de una distribución de probabilidad "inicial" se obtiene la distribución "final" que describe nuestro estado de conocimiento a la luz de toda la información disponible. El Teorema de Bayes nos permite incorporar la información contenida en un conjunto de datos, produciendo una descripción conjunta de la incertidumbre sobre los valores de los parámetros del modelo a través de la distribución final. En esta sección damos una breve descripción del análisis Bayesiano del modelo de regresión. El lector interesado puede consultar, por ejemplo, Gutiérrez-Peña (1998) y las referencias que ahí se encuentran.

Distribución inicial conjugada

Al trabajar con el Teorema de Bayes, las dificultades computacionales surgen desde el momento en que es necesario calcular la constante de proporcionalidad que aparece en el denominador de la expresión para el cálculo de la distribución final (dada una distribución inicial determinada). Por fortuna, existen casos en los que una selección cuidadosa de la distribución inicial simplifica el cálculo de la distribución final. En cada uno de estos casos es posible identificar una *familia conjugada*, \mathcal{C} , de distribuciones iniciales tales que las correspondientes distribuciones finales también pertenezcan a \mathcal{C} . Dichas familias de distribuciones deben verse siempre como lo que son: un artificio matemático que resulta conveniente en ciertas situaciones. En el apéndice A.2.2 se describe con más detalle cuál es la motivación para el uso de las familias conjugadas y se describe de manera general cómo pueden construirse.

Dada la forma de la verosimilitud (2.6), la familia conjugada tiene densidades de la forma

$$p(\beta, \tau) \propto \tau^{\frac{n_0}{2}} \exp \left\{ -\frac{\tau}{2} [(\beta - b_0)^t \mathbf{B}_0 (\beta - b_0) + s_0] \right\},$$

donde $n_0, s_0 \in \mathbb{R}$, $b_0 \in \mathbb{R}^p$ y \mathbf{B}_0 es una matriz de $p \times p$ simétrica y definida semi-positiva.

Notemos que, dado el valor de τ , el *kernel* de la densidad condicional $p(\beta | \tau)$ es proporcional al de la densidad $N(\beta | b_0, \tau^{-1} \mathbf{B}_0^{-1})$, i.e.

$$p(\beta | \tau) \propto \tau^{\frac{p}{2}} \exp \left\{ -\frac{\tau}{2} (\beta - b_0)^t \mathbf{B}_0 (\beta - b_0) \right\},$$

de manera que los factores restantes corresponden a la densidad marginal $p(\tau)$ i.e.

$$p(\tau) = \tau^{\frac{n_0-p}{2}} \exp\left\{\frac{-s_0\tau}{2}\right\}.$$

Haciendo $a = n_0 - p + 2$ y $d = s_0$ se tiene entonces que

$$\begin{aligned} p(\beta, \tau) &= p(\beta | \tau) p(\tau) \\ &= N_p(\beta | b_0, \tau^{-1} B_0^{-1}) Ga(\tau | a/2, d/2). \end{aligned}$$

Esta distribución es conocida como normal-gamma y es propia si $a > 0$, $d > 0$ y B_0 es definida positiva.

Distribución inicial no informativa

En situaciones en las que se desea representar un estado de información inicial vaga acerca de (β, τ) , es común utilizar algún tipo de distribución inicial "no informativa". Uno de los métodos más populares para obtener dichas distribuciones es la *Regla de Jeffreys*. Puede demostrarse fácilmente que en el caso del modelo de regresión normal la distribución inicial de *Jeffreys* es

$$\pi(\beta, \tau) \propto \tau^{\frac{p-2}{2}}.$$

Esta distribución es impropia y puede obtenerse a partir de la familia conjugada haciendo $a = 0$, $d = 0$ y $B_0 = 0$.

Otra distribución inicial no informativa comúnmente usada en modelos de localización y escala es

$$\pi(\beta, \tau) \propto \tau^{-1},$$

la cual corresponde a la distribución inicial de referencia obtenida a partir del método de Bernardo (1979). Al igual que la distribución de *Jeffreys*, esta distribución es impropia y es un caso límite de la familia conjugada cuando $a = -p$, $d = 0$ y $B_0 = 0$.

A lo largo de este trabajo denotaremos por $\pi(\cdot)$ tanto a las distribuciones iniciales no informativas (en particular a la distribución de referencia), como a las correspondientes distribuciones finales.

Nota: En ambos casos $B_0 = 0$ implica que la varianza de la distribución inicial de β es infinita, lo que generalmente se interpreta como una forma de representar un estado de información inicial vaga acerca del valor de β .

Distribución final conjugada

Proposición 2.2 *La distribución final de (β, τ) para el modelo normal si se utiliza una distribución inicial conjugada es*

$$p(\beta, \tau | y) = N(\beta | b_1, \tau^{-1}(\mathbf{B}_1)^{-1}) Ga(\tau | a_1/2, d_1/2)$$

donde

$$\begin{aligned} b_1 &= (\mathbf{X}'\mathbf{X} + \mathbf{B}_0)^{-1}(\mathbf{X}y + \mathbf{B}_0b_0), \\ \mathbf{B}_1 &= \mathbf{X}'\mathbf{X} + \mathbf{B}_0, \\ a_1 &= n + a, \\ d_1 &= (y - \mathbf{X}b_1)'(y - \mathbf{X}b_1) + (b_1 - b_0)'\mathbf{B}_0(b_1 - b_0) + d. \end{aligned}$$

De esta manera, si se desea hacer inferencias sobre β entonces es necesario calcular su distribución marginal final, dada por

$$\begin{aligned} p(\beta | y) &= \int p(\beta, \tau | y) d\tau \\ &= \frac{\Gamma\left(\frac{a_1+p}{2}\right)}{\Gamma\left(\frac{a_1}{2}\right) \pi^{\frac{p}{2}}} \det\left\{\frac{1}{d_1}\mathbf{B}_1\right\}^{1/2} \left\{1 + \frac{1}{d_1}(\beta - b_1)'\mathbf{B}_1(\beta - b_1)\right\}^{-(a_1+p)/2} \\ &= St(\beta | b_1, \mathbf{T}_1^{-1}, a_1) \end{aligned}$$

donde $\mathbf{T}_1 = \left(\frac{a_1}{d_1}\right) \mathbf{B}_1$, de esta forma la distribución final de β es una *t* de Student con a_1 grados de libertad, parámetro de localización b_1 y parámetro de escala \mathbf{T}_1^{-1} .

Distribución final de referencia

Recordemos que la distribución inicial de referencia $\pi(\beta, \tau) \propto \tau^{-1}$, corresponde a un caso límite de la familia conjugada descrita anteriormente, con $a = -p$, $d = 0$ y $\mathbf{B}_0 = 0$. Aunque es impropia, da lugar a una distribución final propia siempre y cuando $n > p$. De hecho en este caso se tiene

$$\begin{aligned} b_1 &= \hat{\beta}, \\ B_1 &= X^t X, \\ a_1 &= n - p, \\ d_1 &= (n - p) \hat{\sigma}^2, \end{aligned}$$

donde $\hat{\sigma}^2$ es el estimador insesgado usual para σ^2 . Por lo tanto,

$$\pi(\beta, \tau | y) = N(\beta | \hat{\beta}, \tau^{-1}(X^t X)^{-1}) Ga(\tau | \frac{n-p}{2}, \frac{(n-p)\hat{\sigma}^2}{2}),$$

donde

$$\pi(\beta | y) = St(\beta | \hat{\beta}, \hat{\sigma}^2(X^t X)^{-1}, n - p).$$

Inferencia y predicción

De acuerdo con la sección anterior, la distribución final de β está dada por

$$St(\beta | b_1, T_1^{-1}, a_1)$$

la cual cumple con

$$\begin{aligned} E(\beta | y) &= b_1 && \text{si } a_1 > 1, \\ Var(\beta | y) &= \frac{a_1}{a_1 - 2} T_1^{-1} \\ &= \frac{d_1}{a_1 - 2} B_1^{-1} && \text{si } a_1 > 2. \end{aligned}$$

En particular, si se utiliza la distribución de referencia se tiene

$$\begin{aligned} E(\beta | y) &= \hat{\beta} && \text{si } n > p + 1, \\ Var(\beta | y) &= \frac{(n-p)\hat{\sigma}^2}{(n-p-2)} (X^t X)^{-1} && \text{si } n > p + 2. \end{aligned}$$

Estas cantidades pueden servir como base para hacer inferencias sobre β .

Predicción

Supongamos que se desea predecir Y^* , un nuevo valor de la variable de respuesta, dado el vector de covariables $\mathbf{x}^{*t} = (1, x_1^*, \dots, x_k^*)$. De acuerdo con el modelo

$$\begin{aligned} Y^* &= \beta_0 + \beta_1 x_1^* + \dots + \beta_k x_k^* + \varepsilon^* \\ &= \mathbf{x}^{*t} \boldsymbol{\beta} + \varepsilon^*, \end{aligned}$$

donde $\varepsilon^* \sim N(0, \sigma^2)$ es independiente de ε . Entonces

$$\mu^* = E(Y^* | \boldsymbol{\beta}, \sigma^2) = \mathbf{x}^{*t} \boldsymbol{\beta}$$

y

$$\text{Var}(Y^* | \boldsymbol{\beta}, \sigma^2) = \sigma^2.$$

Al querer hacer inferencias sobre Y^* , se necesita calcular su distribución final, i.e.

$$p(y^* | \mathbf{y}) = \int \int p(y^* | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) d\boldsymbol{\beta} d\sigma^2.$$

Recordemos que $Y^* = \mathbf{x}^{*t} \boldsymbol{\beta} + \varepsilon^*$. Trabajando condicionalmente en σ^2 , tenemos que $\boldsymbol{\beta}$ y ε^* son independientes y

$$\begin{aligned} p(\boldsymbol{\beta} | \sigma^2, \mathbf{y}) &= N(\boldsymbol{\beta} | \mathbf{b}_1, \sigma^2 \mathbf{B}_1^{-1}). \\ p(\varepsilon^* | \sigma^2, \mathbf{y}) &= N(\varepsilon^* | 0, \sigma^2). \end{aligned}$$

Esto implica que

$$p(y^* | \sigma^2, \mathbf{y}) = N(y^* | \mathbf{x}^{*t} \mathbf{b}_1, \sigma^2 \{1 + \mathbf{x}^{*t} \mathbf{B}_1^{-1} \mathbf{x}^*\}, n - p).$$

Finalmente, integrando con respecto a la distribución final de σ^2 , se tiene que

$$p(y^* | \mathbf{y}) = St\left(y^* | \mathbf{x}^{*t} \mathbf{b}_1, \frac{d_1}{a_1} \{1 + \mathbf{x}^{*t} \mathbf{B}_1^{-1} \mathbf{x}^*\}, a_1\right).$$

Si se utiliza la distribución de referencia entonces la distribución predictiva final toma la forma

$$p(y^* | y) = St \left(y^* | x^{*t} \hat{\beta}, \hat{\sigma}^2 \{1 + x^{*t} (X^t X)^{-1} x^*\}, n - p \right).$$

En este caso, el intervalo de máxima densidad del $(1 - \alpha) \times 100\%$ está dado por

$$x^{*t} \hat{\beta} \pm t_{n-p}^{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + x^{*t} (X^t X)^{-1} x^*}.$$

2.4. Modelos lineales generalizados

En esta Sección se da una breve introducción a los modelos lineales generalizados (MLG). Un tratado completo de la teoría y aplicaciones puede encontrarse, por ejemplo, en McCullagh y Nelder (1989).

2.4.1. Los componentes de un modelo lineal generalizado

La clase de los MLG extiende el modelo usual de regresión en tres aspectos. Sea y un vector de observaciones de dimensión n que consiste en la realización de un vector de variables aleatorias Y independientemente distribuidas con media μ . El vector de medias μ constituye la parte sistemática del modelo y suponemos la existencia de variables independientes x_1, x_2, \dots, x_k como variables conocidas tales que $\mu = \sum_{j=1}^k \beta_j x_j$ donde las β_j son parámetros usualmente desconocidos y tienen que ser estimados de los datos. Si fijamos el índice i de las observaciones entonces la parte sistemática del modelo puede escribirse como

$$E(Y_i) = \mu_i = \sum_{j=1}^k \beta_j x_{ij} \quad i = 1, \dots, n$$

donde x_{ij} es el valor de la j -ésima variable independiente para la i -ésima observación. En notación matricial lo anterior puede escribirse como

$$\mu = X\beta$$

donde μ es un vector de $n \times 1$, X es una matriz de $n \times k$ y β es un vector de $k \times 1$. De modo que X es la matriz del modelo y β es el vector de parámetros.

Un modelo lineal generalizado tiene los siguientes componentes:

1. *El componente aleatorio:* Y_i tiene una distribución en una familia exponencial, y $E(Y_i) = \mu_i$. (Ver Sección 2.4.2)
2. *El componente sistemático:* covariables x_1, x_2, \dots, x_k producen un predictor lineal dado por

$$\eta_i = \sum_{j=1}^k \beta_j x_{ij} \quad i = 1, \dots, n.$$

3. *La liga* entre el componente aleatorio y el componente sistemático:

$$g(\mu_i) = \eta_i.$$

Esta forma produce un nuevo símbolo η para el predictor lineal.

Los modelos lineales clásicos tienen una distribución normal en el primer punto y la función identidad en el tercero.

La determinación de una escala de medición apropiada es un aspecto importante en la selección del modelo y está directamente relacionada con la elección de una función liga ¿deberíamos, por ejemplo, analizar y o $\ln y$? La cuestión es qué caracteriza a una "buena" escala. En el análisis de regresión clásico, una buena escala produce errores que tienen una varianza relativamente constante y cuya distribución es aproximadamente normal.

Si y tiene una distribución Poisson entonces la escala de raíz cuadrada nos da aproximadamente una varianza constante, mientras que $\sqrt[3]{y^3}$ es mejor para aproximar normalidad, finalmente $\ln y$ puede ser mejor para aditividad en efectos sistemáticos.

La extensión de modelos de regresión clásicos a MLG remueve en muchos casos el problema de escala; normalidad y varianza constante no son un requerimiento para el componente del error; mientras que la aditividad de efectos sistemáticos puede ser especificada para tener una escala transformada si se considera necesario.

Estimación y bondad de ajuste

Dado un modelo particular, tenemos que estimar los parámetros desconocidos y obtener alguna medida de la precisión con la que los hemos estimado.

Es común definir una medida de bondad de ajuste entre los datos y un conjunto correspondiente de valores ajustados generados por el modelo, y escoger entonces los valores de los estimadores como aquellos que minimizan el criterio de bondad de ajuste seleccionado. Nosotros consideraremos estimadores obtenidos por el método de máxima verosimilitud.

Si $f(y; \theta)$ es la función de densidad bajo un modelo dado, para la observación y con parámetro θ , entonces la log-verosimilitud expresada como una función de la media $\mu = E(y)$ está dada por $l(\mu; y) = \ln(f(y; \theta))$.

Existen ventajas al usar, como criterio de bondad de ajuste, una función lineal del logaritmo de la verosimilitud tal como

$$D(y; \mu) = -2[l(\mu; y) - l(y; y)],$$

donde $l(y; y)$ es el máximo de la verosimilitud del modelo saturado, en el cual los valores ajustados son iguales a los datos. Esta última expresión se conoce como la *devianza* del modelo. Dado que $l(y; y)$ no depende de los parámetros, maximizar $l(\mu; y)$ o minimizar $D(y; \mu)$ es equivalente. Así, para regresión clásica con varianza conocida σ^2 , tenemos para una observación y con media μ

$$f(y; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\},$$

así que

$$l(\mu; y) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(y - \mu)^2}{2\sigma^2}.$$

$$\text{Si } \mu = y \text{ entonces } l(y; y) = -\frac{1}{2} \ln(2\pi\sigma^2),$$

de modo que

$$D(y; \mu) = -2[l(\mu; y) - l(y; y)] = \frac{(y - \mu)^2}{\sigma^2}$$

(ver Sección 2.4.7).

2.4.2. Funciones de verosimilitud para los MLG

Primero describiremos algunas de las propiedades en términos de una sola observación y .

La variable aleatoria y tiene una función de densidad de probabilidad de la forma

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\} \quad (2.7)$$

para algunas funciones $b(\cdot)$ y $c(\cdot)$. Si ϕ es conocido, esta es una familia exponencial con parámetro canónico θ . Puede o no ser una familia exponencial si ϕ es desconocido.

Una familia exponencial tiene una log-verosimilitud de la forma

$$l = (\{y\theta - b(\theta)\} / \phi) + c(y, \phi),$$

donde θ es el parámetro canónico y ϕ es el parámetro de dispersión, el cual se asume conocido.

$l(\theta, \phi; y) = \ln f(y; \theta, \phi)$ es la función de log-verosimilitud, considerada como una función de θ , para y y ϕ dadas. Entonces la media y la varianza de y pueden derivarse fácilmente de las siguientes relaciones

$$E \left(\frac{\partial l}{\partial \theta} \right) = 0$$

y

$$E \left(\frac{\partial^2 l}{\partial \theta^2} \right) + E \left(\frac{\partial l}{\partial \theta} \right)^2 = 0.$$

Ahora,

$$l = \frac{y\theta - b(\theta)}{\phi} + c(y, \phi),$$

de donde

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{\phi}$$

y

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{\phi}.$$

Así,

$$0 = E \left(\frac{\partial l}{\partial \theta} \right) = \frac{\mu - b'(\theta)}{\phi}.$$

Ahora, dado que $\phi \neq 0$, resulta que $\mu - b'(\theta) = 0$, de donde

$$\mu = E(y) = b'(\theta).$$

Análogamente,

$$0 = E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = -\frac{b''(\theta)}{\phi} + \frac{\text{var}(y)}{\phi^2},$$

así que

$$\text{var}(y) = \phi b''(\theta),$$

donde las primas denotan diferenciación respecto a θ . Así, la varianza de y es el producto de dos funciones: $b''(\theta)$ que depende solamente del *parámetro canónico* y será llamada *función de varianza*, $V(\mu)$ cuando se escriba en términos de μ ; mientras que la otra es independiente de θ y depende sólo del parámetro de dispersión ϕ , el cual es comúnmente de la forma $\phi = \sigma^2/\omega$, donde ω es un peso conocido.

La log-verosimilitud de algunas de las distribuciones más comunes, y sus propiedades, se dan a continuación.

Normal

$$l = \frac{\mu y - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2),$$

$$\theta = \mu,$$

$$V(\mu) = 1,$$

$$\phi = \sigma^2.$$

Poisson

$$l = y \ln \mu - \mu - \ln y!,$$

$$\theta = \ln \mu,$$

$$V(\mu) = \mu,$$

$$\phi = 1.$$

Binomial

Si suponemos que $d \sim \text{Binomial}(n, \mu)$, y definimos $Y = d/n$, tenemos entonces

$$l = ny \ln \{ \mu / (1 - \mu) \} + n \ln (1 - \mu) + \ln \binom{n}{ny},$$

$$\theta = \ln \{ \mu / (1 - \mu) \},$$

$$V(\mu) = \mu(1 - \mu),$$

$$\phi = 1/n.$$

Gamma

$$l = \frac{-y/\mu + \ln(1/\mu)}{1/v} + v \ln y + v \ln v - \ln \Gamma(v),$$

$$\theta = -1/\mu,$$

$$V(\mu) = \mu^2,$$

$$\phi = v^{-1}.$$

Como se mencionó anteriormente, un MLG considera variables de respuesta independientes $\{Y_i : i = 1, 2, \dots, n\}$, con distribución en una familia exponencial, para las cuales

$$E(Y_i) = \mu_i$$

$$\text{var}(Y_i) = \sigma^2(\mu_i) / \omega_i,$$

con una función de varianza $V(\cdot)$, un parámetro de escala $\sigma^2 > 0$ y pesos ω_i .

Las covariables entran vía un predictor lineal,

$$\eta_i = \sum_{j=1}^k x_{ij} \beta_j = \mathbf{x}_i^t \boldsymbol{\beta}$$

donde $\boldsymbol{\beta}$ es un vector de parámetros desconocidos.

La respuesta media y el predictor lineal se relacionan a través de una función liga denotada por g y tal que

$$g(\mu_i) = \eta_i.$$

La función liga canónica es aquella donde $g(\mu) = \theta$, de manera que $\theta(\mu) = \eta$. Ejemplos de esta función son: la identidad, el logaritmo, el logit y la función

liga recíproca para las distribuciones normal, Poisson, binomial y gamma, respectivamente.

Si se utiliza la liga canónica, entonces la función de verosimilitud para β está dada por

$$l(\beta | y) = \exp \left\{ \sum_{i=1}^n \frac{\omega_i}{\sigma^2} [y_i (\mathbf{x}_i^T \beta) - b(\mathbf{x}_i^T \beta)] \right\}. \quad (2.8)$$

2.4.3. Cuasi-verosimilitud

La motivación principal detrás de la idea de cuasi-verosimilitud consiste en relajar el supuesto de que la forma de la distribución del componente aleatorio se conoce con exactitud. En lugar de eso, se supone que sólo se conoce la forma en que la varianza de la respuesta cambia cuando cambia la media.

Supóngase que el vector de respuestas Y , de dimensión n tiene media μ y matriz de covarianza $\sigma^2 V(\mu)$ donde $V(\mu)$ es una matriz definida semipositiva cuyos elementos son funciones conocidas de μ . Es necesario asumir que la parte sistemática del modelo está especificada en términos del parámetro del valor medio. Usualmente la parte sistemática involucra una matriz X , de orden $n \times p$ la cual puede ser incluida dentro del modelo general

$$\mu = \mu(\beta),$$

donde β es el vector de parámetros desconocidos que se desea estimar. Para asegurar la identificabilidad de los estimadores, se asume que la matriz $D = \partial \mu / \partial \beta$ tiene rango p para toda β . En el caso de los modelos lineales generalizados es equivalente a suponer que X tiene rango completo. La función de cuasi-verosimilitud $l(\cdot; y)$ es considerada inicialmente como una función de μ y es definida a través del sistema de ecuaciones diferenciales parciales

$$\frac{\partial l(\mu; y)}{\partial \mu} = V^{-1}(\mu) (y - \mu),$$

donde V^{-1} denota una inversa generalizada de la matriz V . Usualmente no se requiere resolver este sistema de ecuaciones, pero cuando la solución es requerida es suficiente encontrar funciones $\theta(\mu)$ y $b(\theta)$ que satisfagan $\mu = b'(\theta)$ y

$$l(\mu; \mathbf{y}) = \mathbf{y}'\boldsymbol{\theta} - b(\boldsymbol{\theta}) + f_0(\mathbf{y}),$$

con $f_0(\cdot)$ una función que no depende de μ . Cuando no existe ambigüedad, es común escribir $l(\beta; \mathbf{y})$ en lugar de $l(\mu(\beta); \mathbf{y})$. Las ecuaciones de cuasi-verosimilitud usadas para determinar el valor de $\hat{\beta}$ son

$$\frac{\partial l(\beta; \mathbf{y})}{\partial \beta} = 0$$

y pueden escribirse como

$$D'V^{-1}[\mathbf{y} - \mu(\hat{\beta})] = 0.$$

Los estimadores de los coeficientes de regresión se obtienen resolviendo este sistema de ecuaciones.

2.4.4. Sobre-dispersión

El problema de sobre-dispersión surge cuando la varianza estimada de las respuestas es más grande de lo que se esperaría bajo el modelo. Con el fin de ilustrar con más detalle este problema, en el resto de esta subsección nos concentraremos en el modelo binomial.

Si, para cada $i = 1, \dots, n$, las observaciones sobre n_i individuos son independientes y si la probabilidad μ_i de una respuesta positiva es constante para todos los individuos, entonces la distribución de $n_i Y_i$ es binomial con parámetros n_i y μ_i . La contribución de la i -ésima observación en la función de log-verosimilitud bajo estas condiciones es

$$l(\mu_i; y_i) = n_i y_i \ln \{ \mu_i / (1 - \mu_i) \} + n_i \ln(1 - \mu_i).$$

Sin embargo, en muchas aplicaciones las observaciones pueden estar agrupadas de alguna manera. En casos como éste las observaciones en grupos distintos pueden ser independientes o por lo menos no correlacionadas pero pueden existir razones para esperar una correlación positiva dentro de cada uno de los grupos. Bajo estas circunstancias la varianza de y_i será más grande que bajo el supuesto de independencia, a saber $n_i \mu_i (1 - \mu_i)$. Este fenómeno se conoce como sobre-dispersión. Es conveniente entonces suponer que

$$\text{var}(Y_i) = \sigma^2 n_i \mu_i (1 - \mu_i) \quad i = 1, \dots, n,$$

con $\sigma^2 > 1$, donde σ^2 será estimado a partir de los datos.

La expresión anterior, junto con

$$E(y_i) = n_i \mu_i \quad i = 1, \dots, n,$$

determina los dos primeros momentos de Y_i .

Bajo estos supuestos, la cuasi-verosimilitud de μ_i es igual a la log-verosimilitud del modelo binomial, de manera que se obtienen los mismos estimadores para β independientemente del valor de σ^2 . Esta es una propiedad del método de estimación de mínimos cuadrados. Así, la matriz de covarianza asintótica de $\hat{\beta}$ será multiplicada ahora por un factor igual a σ^2 .

2.4.5. Función liga

La función liga se refiere al predictor lineal η para el valor esperado μ de un dato y . En el modelo lineal clásico η y μ son idénticos, y la liga identidad es razonable en el sentido que ambos η y μ toman cualquier valor en la recta real. Sin embargo, cuando se está tratando con conteos y la distribución Poisson, debemos tener $\mu > 0$, así que la liga identidad es poco atractiva en parte porque η puede ser negativa.

Los modelos basados en la independencia de las probabilidades asociadas con diferentes clasificaciones de datos categóricos llevan naturalmente a considerar efectos multiplicativos, y esto es expresado a través de la liga logarítmica, $\eta = \ln \mu$, con inversa $\mu = e^\eta$. Así, los efectos aditivos que contribuyen a η se convierten en efectos multiplicativos que contribuyen a μ .

Para la distribución binomial tenemos $0 < \mu < 1$ y la liga debería satisfacer la condición que mapce el intervalo $(0, 1)$ en toda la recta real.

Consideraremos 3 funciones:

1. *Logit*

$$\eta = \ln \frac{\mu}{1 - \mu}$$

2. *Probit*

$$\eta = \Phi^{-1}(\mu)$$

donde Φ es la función de distribución normal.

3. *Log-log complementaria*

$$\eta = \ln [-\ln(1 - \mu)].$$

2.4.6. Estadísticas suficientes y ligas canónicas

Cada una de las distribuciones normal, Poisson, binomial y gamma tiene una función liga especial para la cual existen estadísticas suficientes para los parámetros en el predictor lineal $\eta_i = \sum_{j=1}^n \beta_j x_{ij}$. Estas *ligas canónicas* ocurren cuando $\theta = \eta$, donde θ es el parámetro canónico. La Tabla 2.1 muestra algunas de las principales distribuciones y su respectiva liga canónica.

Tabla 2.1: Ligas Canónicas

Distribución	Liga Canónica
Normal	$\eta = \mu$
Poisson	$\eta = \ln \mu$
Binomial	$\eta = \ln [\mu / (1 - \mu)]$
Gamma	$\eta = -\mu^{-1}$

2.4.7. Medida de bondad de ajuste

La discrepancia de un modelo

El ajuste de un modelo a un conjunto de datos puede ser visto como una forma de reemplazar el conjunto de valores de los datos y por un conjunto de valores ajustados \hat{y} derivado del modelo, involucrando (usualmente) un número relativamente pequeño de parámetros. En general los \hat{y} s no igualarán a las y s exactamente y la pregunta entonces surge de cómo discrepan entre ellos, porque mientras una pequeña discrepancia puede ser tolerable una gran discrepancia no. Las medidas de discrepancia o bondad de ajuste puede ser construidas de varias maneras, pero en este trabajo sólo se discutirá la *devianza*.

Dadas n observaciones podemos ajustar modelos que contienen más de n parámetros. El modelo más simple el *modelo nulo*, tiene un parámetro representado comúnmente por μ para todas las y s; el modelo nulo lleva toda la variación entre las y s al componente aleatorio. El otro extremo, el *modelo completo o saturado*, tiene n parámetros (uno por observación) y los \hat{y} s derivados de esto igualan los datos exactamente; el modelo completo lleva toda la variación en las y s al componente sistemático dejando nada al componente aleatorio. En la práctica, el modelo nulo es demasiado simple y el

modelo completo no es informativo porque no resume los datos sino que los reproduce por completo. Sin embargo, el modelo completo nos da bases para medir la discrepancia para un modelo intermedio con p parámetros.

Es conveniente expresar la log-verosimilitud en términos del parámetro μ (que es el valor medio) en vez del parámetro canónico θ .

Sea $l(\hat{\mu}, \phi; y)$ la log-verosimilitud maximizada sobre las β s y no sobre el parámetro ϕ . La verosimilitud máxima alcanzable en un modelo completo con n parámetros es $l(y, \phi; y)$ la cual es generalmente finita. La discrepancia de un ajuste es proporcional a 2 veces la diferencia entre el máximo de la log-verosimilitud alcanzable y el alcanzado por el modelo bajo investigación. Si denotamos $\hat{\theta} = \theta(\hat{\mu})$ y $\tilde{\theta} = \theta(y)$ a los estimadores de los parámetros canónicos bajo los dos modelos, la discrepancia puede ser escrita como $D(y; \hat{\mu})/\phi$, donde

$$D(y; \hat{\mu}) = \sum 2w_i \left[y_i (\hat{\theta}_i - \tilde{\theta}_i) - b(\hat{\theta}_i) + b(\tilde{\theta}_i) \right]$$

es conocida como la *devianza* para el modelo común y es una función de los datos solamente. Las devianzas para las principales distribuciones se presentan en la Tabla 2.2.

Tabla 2.2: Devianzas de las principales distribuciones

Distribución	Devianza
Normal	$\sum (y - \hat{\mu})^2$
Poisson	$2 \left\{ \sum [y \ln (y/\hat{\mu}) - (y - \hat{\mu})] \right\}$
Binomial	$2 \sum \{ [y \ln (y/\hat{\mu})] + (n - y) \ln [(n - y) / (n - \hat{\mu})] \}$
Gamma	$2 \sum [- \ln (y/\hat{\mu}) + (y - \hat{\mu}) / \hat{\mu}]$

Para la distribución normal la devianza resulta ser justo la suma de los cuadrados de los residuos, mientras que para la Poisson esto es el estadístico llamado G^2 por Bishop *et al.* (1975) y otros. Se puede notar que los segundos términos en las expresiones de las devianzas para las distribuciones Poisson y gamma son por lo común idénticamente cero. (Para una revisión más detallada ver Nelder y Wedderburn, 1972).

2.4.8. Análisis Bayesiano de los modelos lineales generalizados

Recordemos que el Teorema de Bayes provee un procedimiento inferencial mediante el cual se obtiene una distribución final con la que podemos incorporar la información contenida en un conjunto de datos, produciendo una descripción conjunta de la incertidumbre sobre los valores de los parámetros del modelo.

Verosimilitud

De acuerdo con (2.7) la densidad para cada variable de respuesta Y_i es de la forma

$$f_{Y_i}(y_i; \theta_i, \phi_i) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\} \quad i = 1, \dots, n$$

con $\phi_i = \sigma^2 / \omega_i$ donde ω_i es un peso previo supuestamente conocido.

Por otra parte, el predictor lineal se supone de la forma

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad i = 1, \dots, n.$$

Es importante notar que la log-verosimilitud (2.8) es siempre unimodal y log-cóncava, lo que generalmente facilita su análisis, mientras que el modelo general no necesariamente produce verosimilitudes log-cóncavas si se utiliza una liga distinta a la canónica.

Si $p(\beta)$ es una distribución inicial arbitraria para β , entonces la distribución final puede escribirse como

$$p(\beta | y) \propto \exp \left\{ \sum_{i=1}^n \frac{\omega_i}{\sigma^2} [y_i (x_i^t \beta) - b(x_i^t \beta)] + \ln p(\beta) \right\}.$$

Distribución inicial y distribución final

Distribución inicial no informativa.

Una distribución inicial no informativa comúnmente utilizada es

$$\pi(\beta) \propto 1.$$

En este caso la distribución final toma la siguiente forma

$$p(\beta | y) \propto f_Y(y; \theta, \sigma^2),$$

donde

$$f_Y(\mathbf{y}; \theta, \sigma^2) = \prod_{i=1}^n f_{Y_i}(y_i; \theta_i, \phi_i).$$

Distribución inicial (no conjugada).

Con frecuencia es más natural y sencillo expresar la información inicial directamente en términos de β . En este caso es común usar un modelo normal, de manera que

$$p(\beta) = N(\beta | b_0, B_0^{-1})$$

donde b_0 es un vector de $p \times 1$ y B_0 es una matriz de $n \times p$ simétrica y definida positiva, ambos constantes.

Este modelo puede considerarse como una aproximación a la distribución inicial que verdaderamente describe nuestro estado de conocimiento acerca de β . En este caso

$$p(\beta | \mathbf{y}) \propto \exp \left\{ \sum_{i=1}^n \frac{\omega_i}{\sigma^2} [y_i(x_i^t \beta) - b(x_i^t \beta)] - \frac{1}{2} (\beta - b_0)^t B_0 (\beta - b_0) \right\}.$$

Es claro de las expresiones para las distribuciones finales de β , que en este caso no es posible analizar estos modelos de manera analítica. Una posibilidad es tratar de obtener una solución aproximada utilizando la aproximación asintótica normal descrita brevemente en la Sección 2.2.3. Por otra parte, es relativamente sencillo obtener un análisis esencialmente exacto utilizando las técnicas de simulación discutidas en esa misma sección.

En el caso particular del modelo logístico, que será utilizado en este trabajo, resulta conveniente describir el modelo de la siguiente forma:

$$\begin{aligned} Y_i &= D_i/n_i, \text{ con } D_i \sim \text{Bin}(\mu_i, n_i) \\ \eta_i &= \ln \left(\frac{\mu_i}{1 - \mu_i} \right) = x_i^t \beta \\ \beta &\sim N(b_0, B_0^{-1}). \end{aligned} \quad (2.9)$$

Tratamiento Bayesiano de la sobre-dispersión

En los casos en los que los datos presentan sobre-dispersión, es posible extender el modelo anterior para modelarla de manera explícita. Específicamente, el modelo toma la siguiente forma:

$$Y_i = D_i/n_i, \text{ con } D_i \sim \text{Bin}(\theta_i, n_i)$$

$$\begin{aligned}\theta_i &\sim \text{Beta}(\lambda\mu_i, \lambda(1 - \mu_i)) \\ \eta_i &= \ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{x}_i^t\boldsymbol{\beta} \\ \boldsymbol{\beta} &\sim N(\mathbf{b}_0, \mathbf{B}_0^{-1}).\end{aligned}\tag{2.10}$$

Como en el caso anterior, bajo este modelo se tiene que $E(Y_i) = \mu_i$. Sin embargo, la varianza es ahora de la forma $\text{var}(Y_i) = \phi_i\mu_i(1 - \mu_i)$, donde $\phi_i = (\lambda + n_i)/(n_i\lambda + n_i)$, a diferencia del modelo (2.9), para el cual se tiene $\text{var}(Y_i) = \mu_i(1 - \mu_i)$. Nótese que, conforme λ tiende a infinito, el modelo (2.10) corresponde al modelo (2.9).

Este modelo representa una contraparte Bayesiana del análisis clásico, vía cuasi-verosimilitud, discutido en la Sección 2.4.3.

Capítulo 3

Modelos Estadísticos de Mortalidad

3.1. Introducción

En esta sección se discutirá el ajuste de modelos estadísticos a los datos de mortalidad correspondientes a la operación del Seguro de Vida del Sector Asegurador en México para el periodo de 1982 a 1989.

3.1.1. Descripción del problema

Como ya se mencionó, el propósito del presente trabajo es la construcción de tablas de mortalidad. La determinación tanto de las primas netas de riesgo como de las reservas matemáticas se produce a partir de estas tablas, de manera que la solvencia y la estabilidad financiera de las empresas que operan el ramo de vida dependen, entre otros factores, de la adecuada medición de la mortalidad.

Las revisiones más recientes de las tablas de mortalidad en México han ocurrido con intervalos de por lo menos diez años. En el caso de vida individual, la tabla vigente en México hasta el 31 de marzo de 2000 utilizó datos del periodo 1982-1989, los cuales analizaremos en el presente trabajo.

La información utilizada en este trabajo para el ajuste de modelos de mortalidad corresponde a los datos que las compañías de seguros reportaron a la Comisión Nacional de Seguros y Fianzas para el periodo de 1982-1989. Esta información se muestra en la Tabla 3.1

Tabla 3.1: Expuestos y muertos por edad

Índice	Edad	Expuestos	Muertos	Índice	Edad	Expuestos	Muertos
1	12	9203.0	16	31	42	211408.5	509
2	13	7001.0	2	32	43	201843.0	574
3	14	16768.0	3	33	44	188313.5	630
4	15	10445.0	4	34	45	174223.0	591
5	16	9654.0	0	35	46	162347.5	617
6	17	16102.5	7	36	47	151459.5	623
7	18	18442.0	9	37	48	141042.5	626
8	19	20408.0	15	38	49	128595.5	581
9	20	25173.0	14	39	50	117505.5	562
10	21	33359.5	23	40	51	106745.5	629
11	22	50374.5	27	41	52	98154.0	535
12	23	74561.0	37	42	53	87818.0	647
13	24	99480.5	80	43	54	79525.0	617
14	25	122342.5	106	44	55	72288.0	562
15	26	147954.5	108	45	56	65234.5	513
16	27	165651.0	183	46	57	59083.5	542
17	28	184425.5	182	47	58	52445.0	530
18	29	201460.0	233	48	59	48508.0	540
19	30	218085.5	265	49	60	44817.5	563
20	31	231340.5	276	50	61	40417.0	473
21	32	242170.0	277	51	62	36168.0	423
22	33	249538.5	312	52	63	32128.5	492
23	34	254078.5	318	53	64	28208.5	380
24	35	261217.0	347	54	65	24342.0	379
25	36	260378.0	364	55	66	20805.0	373
26	37	258114.5	388	56	67	18746.5	310
27	38	250332.5	443	57	68	17329.0	315
28	39	242261.5	446	58	69	15277.5	329
29	40	231921.0	517	59	70	13819.0	277
30	41	224367.0	456	60	71	14268.5	308

Índice	Edad	Expuestos	Muertos
61	72	12116.5	786
62	73	10888.0	260
63	74	9822.5	277
64	75	8752.5	227
65	76	7762.5	188
66	77	6973.5	171
67	78	6392.0	194
68	79	5440.0	177
69	80	4758.5	127
70	81	4286.5	166
71	82	3799.0	113
72	83	3222.0	108
73	84	2588.0	109
74	85	5841.5	88
75	86	1259.5	156
76	87	913.5	52
77	88	775.0	37
78	89	635.5	42
79	90	618.0	32
80	91	448.0	17
81	92	337.5	15
82	93	291.0	21
83	94	186.5	15
84	95	160.0	10
85	96	101.0	5
86	97	60.5	9
87	98	64.5	3
88	99	35.5	5

Con ayuda de la experiencia acumulada se puede estimar la probabilidad de muerte de cada persona que compra una póliza de seguro de vida. Existen diversos factores que afectan esta probabilidad, siendo la edad el más comúnmente utilizado en la construcción de las tablas. Otras características que también influyen son el sexo, registros clínicos, ocupación y hábitos de vida tales como fumar y la práctica de actividades deportivas, entre otras.

Revisaremos distintos modelos y herramientas estadísticas para tratar la información de mortalidad disponible.

Las variables relevantes incluyen el número de Pólizas en Vigor (Expuestos) y el número de Siniestros por edad (Muertos).

De aquí en adelante utilizaremos la siguiente notación:

$$\begin{aligned}x &= \text{Edad,} \\E_x &= \text{Expuestos o Pólizas,} \\d_x &= \text{Muertos o Siniestros.}\end{aligned}$$

Las tasas brutas de mortalidad (q_x) para las edades de 12 a 99 años se obtienen, a partir de la relación:

$$q_x = \frac{d_x}{E_x}.$$

Una tabla de Mortalidad típica es un arreglo de probabilidades q_x conforme a la edad x del individuo. Esta tabla muestra diferentes patrones dependiendo de la naturaleza de la población bajo estudio. En realidad, las probabilidades "verdaderas" siguen, en principio, una tendencia monótona positiva (personas de edad avanzada tienen mayor probabilidad de morir que personas jóvenes). En la tabla, las probabilidades estimadas muestran los patrones reales afectados por errores muestrales. Así, una técnica de "suavizamiento" es muchas veces usada para producir una tabla graduada la cual elimina esas variaciones aleatorias. En la práctica, como una estrategia conservadora, estas probabilidades de muerte graduadas son modificadas (sobre-estimadas) para minimizar el riesgo de tener una reserva insuficiente.

Hoy en día, la graduación y la sobre-estimación son muchas veces implementadas ignorando la naturaleza estadística del problema.

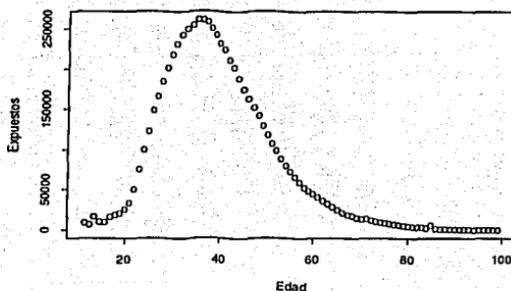


Figura 3.1: *Distribución de frecuencias por edad de los expuestos*

3.1.2. Análisis descriptivo de los datos: 1982-1989

Nuestro conjunto de datos incluye un rango de edad que va desde los 12 hasta los 99 años, como se puede apreciar en la Tabla 3.1.

La Figura 3.1 muestra la distribución de frecuencias por edad de los 6,688,006 expuestos. Como es usual en este tipo de datos, la mayoría de los casos yacen entre 20 y 65 años. En resumen, la distribución es claramente asimétrica. Incidentalmente, un modelo gamma provee un ajuste adecuado.

Respecto a las 23,918 muertes registradas, la Figura 3.2 presenta la correspondiente distribución de frecuencias por edad. El patrón asimétrico es menos obvio aquí, sugiriendo que la probabilidad de muerte no es constante a través de la edad.

Finalmente en la Figura 3.3 es claro que las tasas brutas de mortalidad observadas se comportan de una forma ascendente, tienen una tendencia no lineal y desviaciones importantes aparecen al final de la tabla. Los datos se pueden apreciar en la Tabla 1.1.

3.2. Supuestos para el análisis Bayesiano

Aquí se presentarán los supuestos básicos para el análisis Bayesiano de los modelos de regresión logística, así como el procedimiento para obtener pronósticos para la siniestralidad agregada con base en dichos datos.

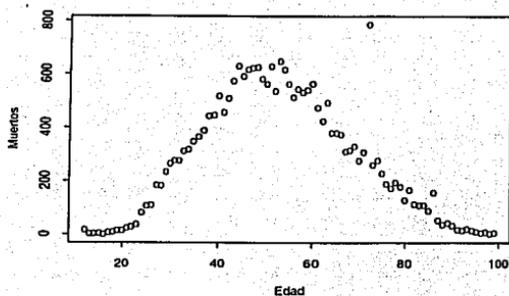


Figura 3.2: *Distribución de frecuencias por edad de los muertos*

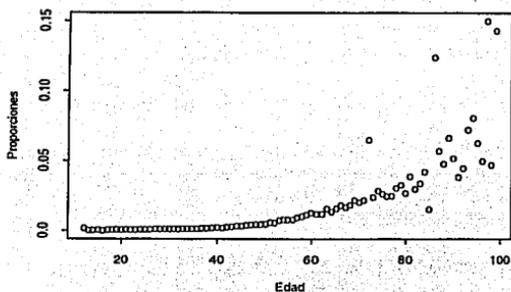


Figura 3.3: *Tasas brutas de muerte*

3.2.1. Supuestos básicos

Para cada edad x y dado un número E_x de asegurados expuestos al principio del año, se tiene que, bajo los supuestos de independencia y homogeneidad, el número de muertos correspondientes d_x , puede ser modelado con una distribución de probabilidad binomial con probabilidad q_x , es decir

$$p(d_x | E_x, q_x) \approx \text{Binomial}(E_x, q_x).$$

Nos interesa describir la mortalidad que se puede observar para cada edad, en determinado periodo. Es decir, dado un patrón de exposición, descrito por E_x , nos interesa pronosticar la cantidad futura d_x^* (i.e. el número de muertos en el siguiente periodo) lo que equivale a pronosticar $\hat{q}_x^* = d_x^*/E_x$ (tasa de mortalidad observada futura), siendo importante su cálculo para efectos de protección frente a desviaciones extremas en la siniestralidad.

Ahora bien, desde la perspectiva Bayesiana, se debe asignar una distribución inicial al parámetro desconocido q_x con el fin de obtener la distribución predictiva final del número futuro de muertos. Debido a que la tasa de mortalidad observada futura es proporcional a d_x^* , la distribución predictiva para \hat{q}_x^* se puede obtener directamente a partir de la de d_x^* . En cuanto a la distribución inicial y ya que se pretende producir inferencias basadas esencialmente en los datos, se optó por utilizar una distribución inicial de referencia para q_x (Bernardo y Smith, 1994; Sección 5.4). Resumiendo los resultados importantes en la Tabla 3.2.

Tabla 3.2: Análisis del modelo binomial

Función	Expresión	Modelo
Verosimilitud	$p(d_x E_x, q_x)$	Binomial (E_x, q_x)
Inicial para q_x	$p(q_x)$	Beta(0.5,0.5)
Final para q_x	$p(q_x d_x, E_x)$	Beta ($d_x + 0.5, E_x - d_x + 0.5$)
Predictiva final para d_x^*	$p(d_x^* d_x, E_x)$	Beta-Binomial ($d_x + 0.5, E_x - d_x + 0.5$)
Predictiva final para \hat{q}_x^*	Se obtiene directamente de la relación $\hat{q}_x^* = d_x^*/E_x$	

Los intervalos predictivos (de probabilidad 95%) de las tasas de mortalidad observadas futuras para cada edad se muestran en la Figura 3.4.

Puede observarse mayor incertidumbre en la estimación de las tasas para las edades más avanzadas. La figura muestra claramente el rango de posibles modelos que podrían considerarse para ajustar estos datos. Más aún, hay unas cuantas observaciones extremas que no pueden describirse a través de un modelo monótono simple. Tales observaciones pueden considerarse "aberrantes" (*ouliers*). Estas observaciones sólo incrementan la variabilidad pero no parecen tener un efecto significativo en la tendencia.

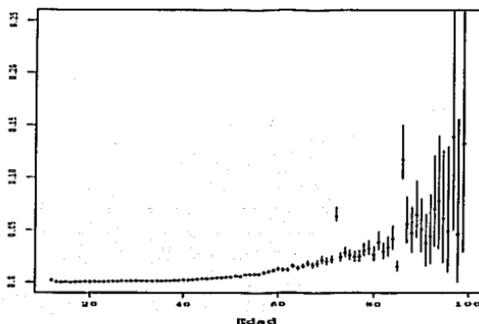


Figura 3.4: Intervalos predictivos para cada edad

3.2.2. Análisis Bayesiano de modelos de regresión logística

Para este análisis es de interés determinar una curva suave que describa la tendencia creciente de las tasas conforme la edad aumenta. De este modo, el objetivo final es obtener una distribución predictiva final para \hat{q}_x^* ($x = 12, \dots, 99$) que considere la propiedad básica de mortalidad: *a mayor edad, mayor probabilidad de muerte*.

Hasta la fecha, en el ámbito actuarial, este problema se ha resuelto utilizando técnicas de suavizamiento y graduación determinísticas para obtener las tablas de mortalidad. Sin embargo estas técnicas, en general, no permiten la medición de la incertidumbre en las estimaciones tanto a lo largo de las

edades como a través de los años. Los modelos de regresión constituyen una herramienta estadística de uso frecuente para la producción de pronósticos.

Como se discutió en la Sección 2.3, un modelo de regresión describe la relación entre la variable aleatoria de respuesta Y y una o más variables explicativas X_j ; $j = 1, \dots, k$. Así un modelo de regresión normal queda definido como:

$$Y_i \sim \text{Normal} \left(\sum \beta_j X_{ij}, \sigma^2 \right) \text{ independientes para } i = 1, \dots, n$$

o bien, en notación matricial

$$Y \sim \text{Normal}(X\beta, \tau^{-1}I),$$

donde X representa a la matriz de dimensión $n \times k$ de variables explicativas, β es el vector de coeficientes de dimensión $k \times 1$ y $\tau = 1/\sigma^2$ es la precisión común.

Observemos que aunque las tasas de mortalidad tienen una relación creciente respecto a la edad, no exhiben una tendencia lineal y se encuentran situadas en el intervalo $(0,1)$. Por otra parte, la distribución normal tiene como soporte toda la recta real; así que será necesario aplicar una transformación a las tasas para obtener un ajuste razonable.

La transformación *logística* resuelve la falta de linealidad y mapea el intervalo $(0,1)$ en toda la recta real, por ello la utilizaremos en lo que resta del presente trabajo. Aplicando esta transformación a las tasas observadas se tiene la siguiente variable de respuesta

$$Y = \ln \left(\frac{\hat{q}}{1 - \hat{q}} \right).$$

Con estos nuevos datos (datos transformados) ajustamos un modelo de regresión lineal Bayesiano lo que significa que se determina la distribución predictiva para el vector Y dada una edad x . A partir de esta distribución podemos determinar la distribución predictiva para las tasas de mortalidad regresando así a la escala original, mediante la transformación inversa

$$\hat{q} = \frac{\exp(Y)}{1 + \exp(Y)}.$$

El análisis Bayesiano requiere de la asignación de distribuciones iniciales para los parámetros desconocidos. Con la intención de fundamentar las inferencias en la información provista por los datos exclusivamente, se utilizan distribuciones de referencia.

Como se discutió en la Sección 2.3.2, si z representa la información aportada por los datos, β y τ representan los estimadores de máxima verosimilitud de β y τ respectivamente, y Y^* el vector asociado a la matriz X^* sobre la cual se desea realizar pronósticos, se obtienen las distribuciones siguientes:

Verosimilitud

$$p(y|\beta, \tau) = N(y | X\beta, \tau^{-1}I),$$

Inicial para (β, τ)

$$p(\beta | \tau) \propto \tau^{-1},$$

Final para (β, τ)

$$p(\beta, \tau | z) = N(\underline{\beta} | \hat{\beta}, \hat{\tau}^{-1}(X^t X)^{-1}) Ga(a, b\hat{\tau}^{-1}),$$

*Predictiva final para Y^**

$$p(y^*|X^*, z) = St(y^* | X^* \hat{\beta}, \hat{\tau}^{-1}(X^* (X^t X)^{-1} X^{*t} + I), n - p),$$

donde $St(\cdot | \mu, \Sigma, k)$ es una distribución t de Student multivariada con parámetro de localización μ , matriz de escala Σ y k grados de libertad (e.g. Bernardo y Smith, 1994). Además, $\hat{\beta}$ y $\hat{\tau}$ son los estimadores de máxima verosimilitud de β y τ respectivamente, mientras que $a = (n-2)/2$, $b = n/2$ y $\hat{\tau} = 1/\hat{\sigma}^2$, donde $\hat{\sigma}^2$ es el estimador insesgado usual de σ^2 .

Un resultado bien conocido establece que una variable aleatoria t de Student converge, en distribución, a una normal si el número de grados de libertad tiende a infinito. En este caso 88 grados de libertad de la distribución Student puede considerarse suficientemente grande para aproximarla adecuadamente con una normal.

De este modo, el resultado del ajuste Bayesiano puede resumirse con la distribución predictiva aproximada

$$p(y^*|X^*, z) = N(X^* \hat{\beta}, \hat{\tau}^{-1}\{X^* (X^t X)^{-1} X^{*t} + I\}).$$

Aunque el valor esperado no es invariante ante transformaciones no lineales, sí lo es la mediana y en general cualquier cuantil. Esto es, el cuantil de nivel α de la distribución predictiva para las tasas observadas se obtiene de manera directa al transformar el correspondiente cuantil de la distribución predictiva.

3.2.3. Pronósticos para la siniestralidad agregada

En aplicaciones de la vida real, al usar proporciones de mortalidad para los cálculos financieros, hay dos componentes que deben tomarse en cuenta: la tasa de mortalidad y el perfil demográfico de la población específica.

Las tablas de mortalidad constituyen la base técnica para el cálculo de primas y reservas. Sus aplicaciones en términos de siniestralidad esperada proporcionan una herramienta útil para la determinación de las reservas que se deben mantener para afrontar las obligaciones.

El pronóstico del número agregado de muertes o suma de siniestros corresponde a la previsión del número de reclamaciones que deberían ser pagadas durante un año bajo un esquema de un seguro temporal a un año para cierto patrón de asegurados expuestos.

Para un patrón de tasas de mortalidad $\{q_x : x = 12, \dots, 99\}$ y un patrón de expuestos $\{E_x : x = 12, \dots, 99\}$, el número esperado de la suma de siniestros (SS) que se observa durante un año se puede obtener a partir de la siguiente relación

$$SS = \sum_x E_x q_x.$$

La distribución predictiva del número agregado de muertes, suponiendo que se observa el mismo patrón de exposición E_x , se puede obtener a partir de la correspondiente distribución predictiva de \hat{q}_x usando la relación

$$SS^* = \sum_x E_x \hat{q}_x.$$

Si el patrón de asegurados expuestos está fijo, la variabilidad asociada al pronóstico de la suma de siniestros proviene de la incertidumbre que se tiene sobre las tasas de mortalidad. Así, la distribución predictiva final para la suma de siniestros observados depende y puede ser derivada de la distribución predictiva final conjunta de \hat{q}^* que, a pesar de no contar con una forma algebraica cerrada, puede analizarse fácilmente vía simulación.

Como ya se mencionó anteriormente, la distribución predictiva conjunta de Y^* se puede aproximar satisfactoriamente con una normal multivariada tal que

$$\begin{aligned} E(Y^* | X^*) &= X^* \hat{\beta}. \\ \text{Var}(Y^* | X^*) &= \hat{\tau}^{-1} \{X^* (X^{*t} X^*)^{-1} X^{*t} + I\}. \end{aligned}$$

Con el fin de contar con una muestra de la distribución predictiva de las tasas observadas de muerte \hat{q}^* , se simulan observaciones del vector \mathbf{Y}^* a partir del modelo ajustado y se utiliza la transformación inversa

$$\hat{q}_x^* = \frac{\exp\{y_x^*\}}{1 + \exp\{y_x^*\}}$$

aplicada a cada entrada del vector. De esta manera, cada simulación representa un posible patrón de tasa de mortalidad a observarse en el periodo de estudio.

Finalmente, se multiplica cada una de las simulaciones de \hat{q}^* por el patrón de expuestos por edad, para obtener una muestra de la distribución predictiva del número agregado de siniestros, SS^* .

El procedimiento descrito sugiere inmediatamente una primera forma de establecer un grado razonable de sobre estimación, con el propósito de obtener una tabla graduada que proporcione un nivel adecuado de protección ante el riesgo de que la reserva sea insuficiente. Específicamente, la tabla modificada basada en el cuantil de orden p de la distribución predictiva final de \hat{q}_x^* garantiza que las tasas de mortalidad que pudieran observarse en el futuro solo excederán ese valor, para cada edad x , con probabilidad $1 - p$.

Sin embargo, debe recordarse que la cantidad de interés es el número agregado de siniestros, pues éste tiene un impacto directo sobre el monto de las reclamaciones. Por lo tanto, para determinar un nivel de protección adecuado ante el riesgo de una reserva insuficiente, es necesario examinar la distribución predictiva final del número agregado de siniestros SS^* .

Específicamente, dicha distribución puede usarse para determinar el nivel de protección global proporcionado por una tabla de mortalidad específica al aplicarse a una población asegurada en particular.

La idea es seleccionar un modelo modificado en términos de una medida global de riesgo. La medida propuesta es análoga a lo que en la literatura financiera se conoce como "valor en riesgo" (ver, por ejemplo, Grayling, 1997).

Sean $\{q_x^*\}$ las tasas de mortalidad correspondientes al modelo modificado que se desea evaluar y sea $SS^+ = \sum_x q_x^+ E_x$. Entonces

$$\alpha = \Pr [SS^* > SS^+ \mid \text{Datos}]$$

es una medida global de riesgo en el que incurre la compañía al utilizar la tabla modificada $\{q_x^+\}$ en una población de asegurados con un patrón de expuestos dado por $\{E_x\}$.

Notemos que SS^+ puede interpretarse como el valor en riesgo al nivel α para este problema.

3.3. Pronósticos basados en modelos lineales generalizados

Dada la naturaleza binomial de los datos presentados en la Tabla 3.1, es posible considerar un análisis alternativo basado en un modelo logístico. Una vez ajustado dicho modelo de acuerdo al procedimiento descrito en la Sección 2.4.8, los pronósticos para el número agregado de siniestros se puede realizar de una manera totalmente análoga a la descrita en la Sección 3.2.3. Esta idea se ilustra en la Sección 4.2.

Capítulo 4

Análisis Bayesiano de Mortalidad

4.1. Regresión lineal normal

Con la finalidad de seleccionar un modelo que describa adecuadamente los datos y que permita predecir el comportamiento del número agregado de muertos, se ajustaron en principio dos modelos de regresión: uno lineal y otro cuadrático (en la edad).

4.1.1. Análisis preliminar del modelo lineal

Considerando el conjunto de datos de la Tabla 3.1 y omitiendo el dato correspondiente a la edad 16, dado que no se reportaron muertes y la transformación que se usará no está definida para este caso, se ajustó un primer modelo de regresión logística dado por

$$Y_x = \ln \left(\frac{q_x}{1 - q_x} \right) = \beta_0 + \beta_1 x + \varepsilon$$

donde x representa a la edad y es la única variable independiente en el modelo, y ε es un error aleatorio.

Para cualquier edad x y dado un patrón de expuestos E_x , se supone que el número de muertos d_x sigue una distribución binomial con probabilidad q_x .

Aplicando la transformación *logística* a las tasas de mortalidad y ajustando el modelo lineal con la ayuda del paquete estadístico *S-Plus*, se obtuvo el análisis preliminar que se presenta a continuación.

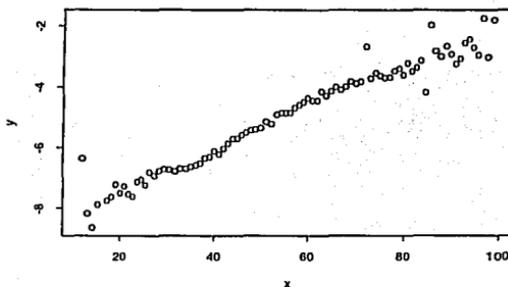


Figura 4.1: *Tasas transformadas*

Se observa una mayor dispersión en las tasas que corresponden a las edades extremas, tanto pequeñas como grandes, debido a que en estas edades es menor el número de expuestos. La gráfica muestra una tendencia lineal, como se muestra en la Figura 4.1, así que ahora podremos utilizar una regresión clásica para ajustar los datos, el programa con el que fueron obtenidos los resultados se muestra en el Apéndice B.

El modelo ajustado con los datos transformados es

$$\hat{E}(Y_x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

y se muestra en la Figura 4.2. Aquí,

$$\begin{aligned}\hat{\beta}_0 &= -8.716076 \\ \hat{\beta}_1 &= 0.066913 \\ \hat{\sigma}^2 &= 0.137619\end{aligned}$$

y

$$\widehat{\text{var}}(\hat{\beta}) = \begin{bmatrix} 0.0093868334 & -0.0001394896 \\ -0.0001394896 & 0.0000024929 \end{bmatrix}$$

Utilizando la aproximación descrita al final de la Sección 3.2.2, se simuló una muestra de tamaño 5000 de la distribución predictiva de Y mediante el paquete estadístico *S-Plus*. Se construyó una banda predictiva para las tasas transformadas (Y^*) mediante los cuantiles de 0.025 y 0.975 como se puede apreciar en la Figura 4.3.

Aplicando la transformación inversa para regresar a la escala original obtenemos el siguiente ajuste a través de la mediana de \hat{q}_x^* , como se aprecia en la Figura 4.4. Así como la banda predictiva en la escala original mostrada en la Figura 4.5.

Se realizó también una comparación entre la tabla actual (Tabla 1.2), utilizada por el sector asegurador para el cálculo de las primas y las reservas, y la tabla producida por el ajuste Bayesiano con base en el cuantil de orden 0.70 de la distribución predictiva de \hat{q}_x^* para cada x , como se muestra en la Figura 4.6. Dicho cuantil se calculó de manera que el "valor en riesgo" de la suma de siniestros (SS) fuera aproximadamente del 2.5%, tal como se describe al final de la Sección 3.2.3. Puede observarse en la Figura 4.7 que dicho valor es de 27164 muertos o siniestros al final del periodo.

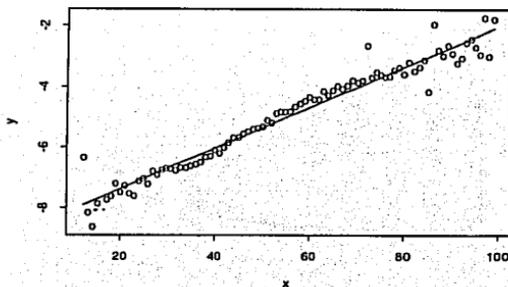


Figura 4.2: Ajuste lineal sobre las tasas transformadas

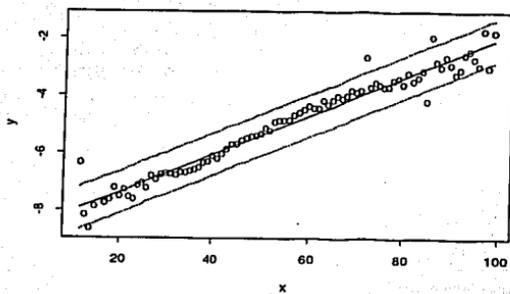


Figura 4.3: Banda predictiva para las tasas transformadas

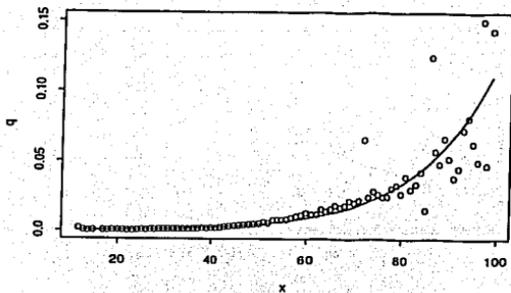


Figura 4.4: Ajuste en la escala original

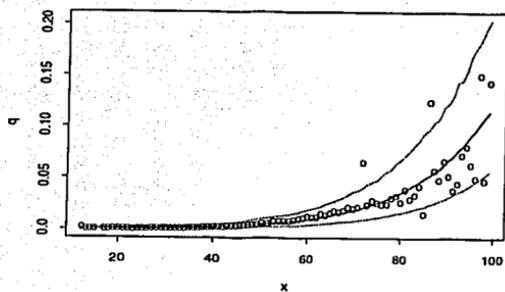


Figura 4.5: Banda predictiva en la escala original

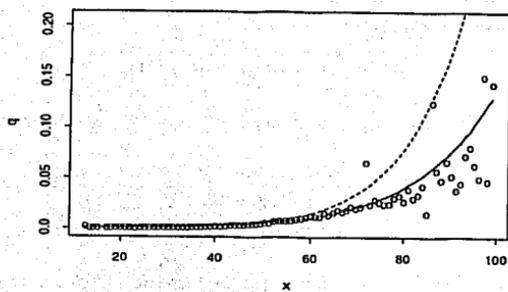


Figura 4.6: Ajuste lineal (—) y tabla actual (- - -)

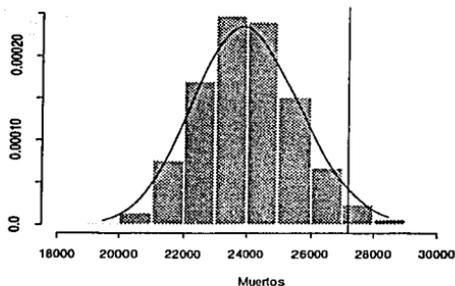


Figura 4.7: Número agregado de siniestros al final del periodo

4.1.2. Análisis del modelo lineal: WinBUGS

Por otra parte, y con propósitos comparativos, se realizó un segundo análisis del mismo modelo pero en esta ocasión utilizando el programa *WinBUGS*. Este programa utiliza la técnica de simulación discutida en la Sección 2.2.3 conocida como muestreo de Gibbs.

Para el análisis del modelo

$$p(Y|\beta, \tau) = N(y | x\beta, \tau^{-1}I)$$

se consideraron las siguientes distribuciones iniciales "no informativas"

$$\beta_0 \sim N(0, 1.0e-3)$$

$$\beta_1 \sim N(0, 1.0e-3)$$

$$\tau \sim Ga(1.0e-3, 1.0e-3).$$

Para este caso se simularon muestras de tamaño 10000 mediante el programa *WinBUGS* para los parámetros β_0 , β_1 y σ^2 .

Los resultados obtenidos con este nuevo análisis fueron muy similares a los obtenidos con la aproximación descrita al principio de esta sección. Los

estimadores para β_0 , β_1 y σ^2 fueron

$$\begin{aligned}\hat{\beta}_0 &= -8.717 \\ \hat{\beta}_1 &= 0.066 \\ \hat{\sigma}^2 &= 0.141.\end{aligned}$$

A partir de observaciones simuladas de Y^* , se utilizó la transformación logística inversa para obtener una muestra de la distribución predictiva de las tasas de muerte por observar, \hat{q}^* . Posteriormente, esta muestra fue utilizada, junto con los valores de los expuestos para cada edad, para obtener la muestra correspondiente de la distribución final del número agregado de muertos.

Se construyeron bandas predictivas para \hat{q}_x^* mediante los cuantiles de 0.025 y 0.975 de la correspondiente distribución predictiva final. La Figura 4.8 muestra un ajuste basado en la mediana, y la Figura 4.9 basado en la media.

También se realizó la comparación entre la tabla actual (ver Tabla 1.2) y la tabla producida por el análisis Bayesiano reproducido en esta sección con base en el cuantil de orden 0.70, como se muestra en la Figura 4.10. Este cuantil corresponde a un "valor en riesgo" para la suma de siniestros (SS) del 2.5% y resultó ser de 27419 muertos o siniestros al final del periodo, como se aprecia en la Figura 4.11.

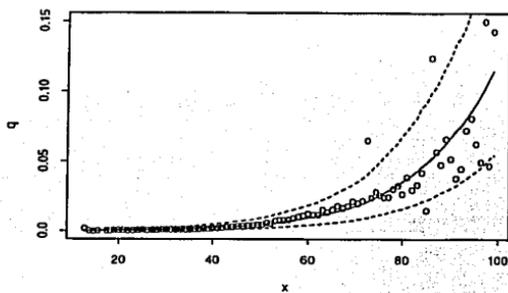


Figura 4.8: Ajuste (mediana) y banda predictiva para q

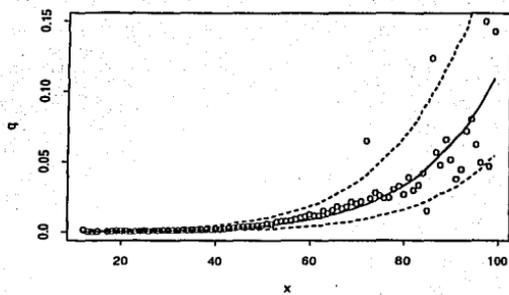
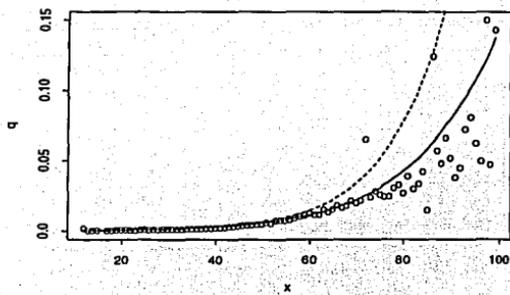
Figura 4.9: Ajuste (media) y banda predictiva para q 

Figura 4.10: Ajuste lineal (—) y tabla actual (- - -)

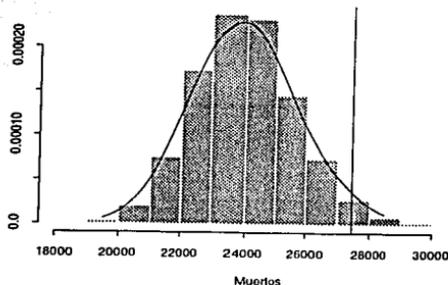


Figura 4.11: Número agregado de siniestros al final del periodo

4.1.3. Análisis preliminar del modelo cuadrático

Con los mismos supuestos que se hacen en el modelo lineal, se ajustó un modelo cuadrático, el cual incluye un término constante, un componente lineal y otro cuadrático en la edad, de la siguiente manera:

$$Y_x = \ln \left(\frac{q_x}{1 - q_x} \right) = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon. \quad (4.1)$$

Con la ayuda del paquete estadístico *S-Plus* se realizó el ajuste cuadrático. En este caso los parámetros ajustados fueron

$$\begin{aligned} \hat{\beta}_0 &= -9.120 \\ \hat{\beta}_1 &= 0.085 \\ \hat{\beta}_2 &= -0.0001 \\ \hat{\sigma}^2 &= 0.1302 \end{aligned}$$

$$\widehat{\text{var}}(\hat{\beta}) = \begin{bmatrix} 0.03712991999 & -0.00140241113 & 0.00001137683 \\ -0.00140241113 & 0.00005949887 & -0.00000051171 \\ 0.00001137683 & -0.00000051171 & 0.00000000458 \end{bmatrix}.$$

El ajuste de este modelo se muestra en la Figura 4.12. Mediante la transformación logística inversa se obtiene el ajuste en la escala original, mostrado en la Figura 4.13.

Al igual que en el caso anterior, se simuló una muestra de tamaño 5000 de la distribución predictiva de Y mediante el paquete estadístico *S-Plus*. Se construyeron bandas predictivas tanto en la escala original como en la escala *logit* mediante los cuantiles de 0.025 y 0.975 como se puede apreciar en las Figuras 4.14 y 4.15. Asimismo, se realizó la comparación entre la tabla actual y un ajuste Bayesiano basado en el cuantil de orden 0.70 de la distribución predictiva de \hat{q}_x^* para cada x (ver Figura 4.16). Finalmente, se construyó el histograma de la suma de siniestros, el cual se muestra en la Figura 4.17. El "valor en riesgo" en este caso fue de 29132 y corresponde a un nivel del 2.5 %.

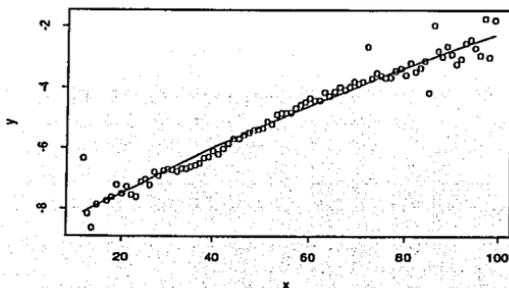
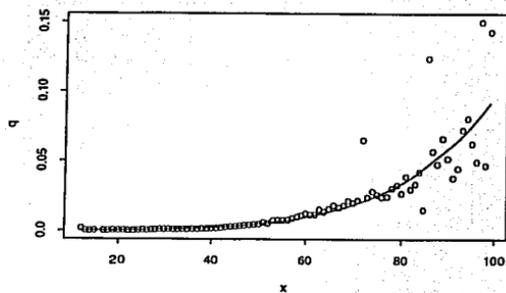
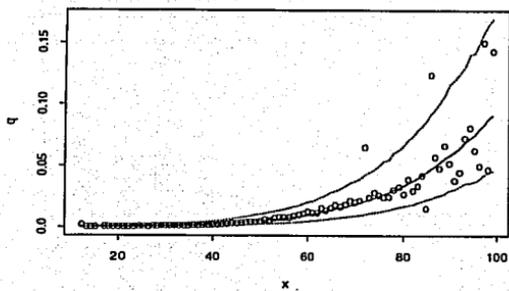


Figura 4.12: Ajuste cuadrático sobre las tasas transformadas

Figura 4.13: *Ajuste cuadrático en la escala original*Figura 4.14: *Banda predictiva en la escala original*

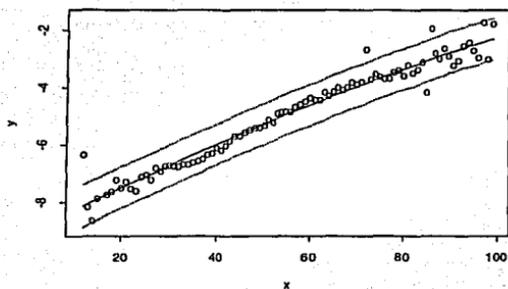


Figura 4.15: Banda predictiva en la escala transformada

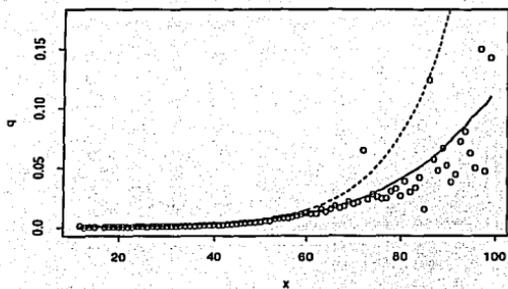


Figura 4.16: Ajuste cuadrático (—) y tabla actual (- - -)

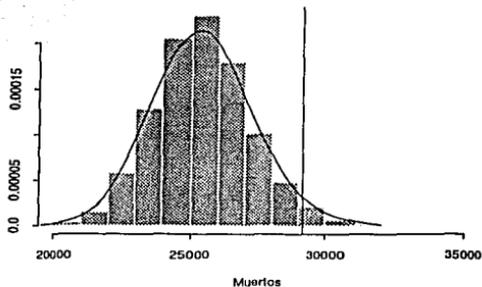


Figura 4.17: Histograma de la suma de siniestros para el ajuste cuadrático

4.1.4. Análisis del modelo cuadrático: WinBUGS

Al igual que para el modelo lineal, para el caso cuadrático también se realizó un nuevo análisis utilizando el paquete estadístico *WinBUGS*, consiguiendo los siguientes resultados.

Para el análisis del modelo

$$p(Y|\beta, \tau) = N(y | x\beta, \tau^{-1}I)$$

se consideraron las siguientes distribuciones iniciales "no informativas"

$$\beta_0 \sim N(0, 1.0e - 3)$$

$$\beta_1 \sim N(0, 1.0e - 3)$$

$$\beta_2 \sim N(0, 1.0e - 3)$$

$$\tau \sim Ga(1.0e - 3, 1.0e - 3).$$

Se simuló una muestra de tamaño 10000, con la que se obtuvieron los siguientes estimadores para los parámetros β_0 , β_1 , β_2 y σ^2 respectivamente

$$\hat{\beta}_0 = -9.090$$

$$\hat{\beta}_1 = 0.083$$

$$\hat{\beta}_2 = -0.00014$$

$$\hat{\sigma}^2 = 0.1332.$$

Se contruyó una banda predictiva mediante los cuantiles de orden 0.025 y 0.975 de la correspondiente distribución predictiva final y se incluyeron los ajustes basados en la media y la mediana como se aprecia en las Figuras 4.18 y 4.19 respectivamente.

También se realizó la comparación entre la tabla actual utilizada por el sector asegurador para el cálculo de las primas y las reservas (ver Tabla 1.2) y la tabla producida por el análisis Bayesiano reproducido en esta sección con base en el cuantil de orden 0.70, como se muestra en la Figura 4.20.

Aún para las edades de 12 a 65 años la tabla producida por este análisis está por debajo de la tabla actual, como se puede apreciar en la Figura 4.21.

En este caso el "valor en riesgo" de la suma de siniestros (SS) resultó ser de 29100 muertos o siniestros al final del periodo y corresponde a un nivel del 2.5%, como se muestra en la Figura 4.22.

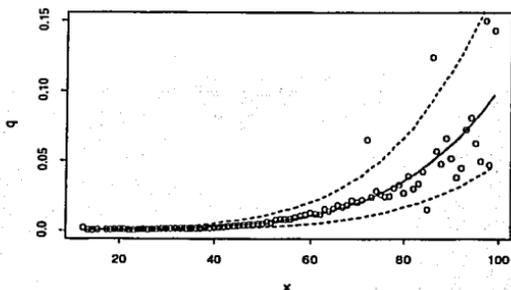


Figura 4.18: Ajuste (media) y banda predictiva

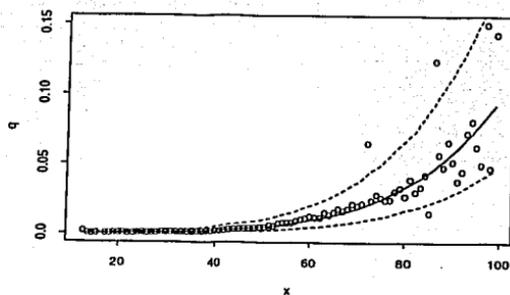


Figura 4.19: Ajuste (mediana) y banda predictiva

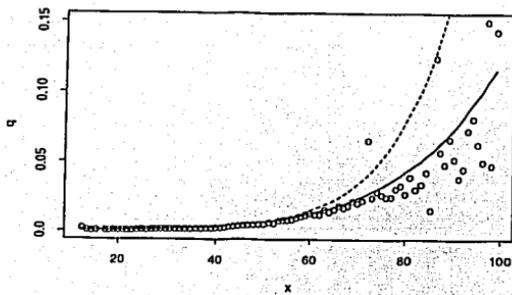


Figura 4.20: Ajuste cuadrático (—) y tabla actual (- -)

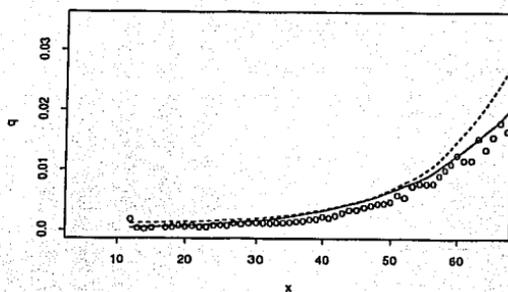


Figura 4.21: Detalle del análisis, para las edades entre 12 y 65 años

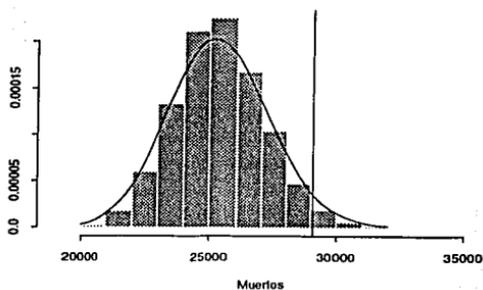


Figura 4.22: Número agregado de siniestros al final del periodo

4.2. Modelos lineales generalizados

En esta sección analizaremos los datos de la Tabla 3.1 utilizando un modelo lineal generalizado. Específicamente, ajustaremos un modelo logístico dada la naturaleza binomial de los datos. De manera similar al análisis de la Sección 4.1, con el fin de predecir el comportamiento del número agregado de siniestros se ajustarán dos modelos: uno lineal y otro cuadrático (en la edad). Primero discutiremos un análisis preliminar basado en una aproximación normal (asintótica) a la distribución final de los parámetros involucrados. Posteriormente, se presentará el análisis Bayesiano discutido brevemente en la Sección 2.4.8, haciendo uso de las técnicas descritas en la Sección 2.2.3.

En las Secciones 4.2.1 y 4.2.3 se presentarán análisis preliminares de los modelos lineal y cuadráticos, respectivamente, utilizando el paquete estadístico *S-Plus*. Estos análisis utilizan la noción de cuasi-verosimilitud para abordar el problema de sobre-dispersión tal como se describió en las Secciones 2.4.3 y 2.4.4.

Por otra parte, en las Secciones 4.2.2 y 4.2.4 se presentarán los análisis Bayesianos correspondientes, utilizando el paquete *WinBUGS* para ajustar cada uno de los modelos de la manera discutida en la Sección 2.4.8. Cabe mencionar que estos modelos parecen ser bastante sensibles a la presencia de "outliers" y por lo tanto presentaron dificultades para su análisis. Concretamente, la correlación entre los parámetros bajo la distribución final es relativamente alta, lo que ocasiona que la convergencia del algoritmo sea lenta. Con el fin de mitigar este problema, se ajustaron de nuevo estos modelos omitiendo las observaciones correspondientes a las edades 12, 72, 85 y 86, las cuales fueron detectadas como "outliers" en un análisis exploratorio previo (ver Figura 3.4). Es interesante observar que estos "outliers" coinciden con los detectados a través de un análisis de residuos del modelo normal cuadrático (ver Sección 4.4).

A pesar de lo anterior, la correlación entre los parámetros del modelo siguió siendo alta y, en consecuencia, los ajustes (que se basan en aproximaciones de Monte Carlo) no resultaron suficientemente suaves. Con el propósito de comparar estos resultados con los de los otros modelos considerados en esta sección, fue necesario suavizar las curvas correspondientes. Esto se hizo utilizando las funciones "smooth" y "lowess" de *S-Plus*.

4.2.1. Análisis preliminar del modelo lineal

Para cualquier edad x y dado un patrón de expuestos E_x , se supone que el número de muertos d_x sigue una distribución binomial con probabilidad q_x .

El modelo que se ajusta es de la forma:

$$\ln \left(\frac{q_x}{1 - q_x} \right) = \beta_0 + \beta_1 x$$

donde la *función liga* que utiliza es la denominada *logit* (i.e. la liga canónica para este caso).

Con la ayuda del paquete estadístico *S-Plus*, se ajustó este modelo, obteniendo los siguientes estimadores para los parámetros β_0 , β_1 y σ^2 respectivamente

$$\begin{aligned} \hat{\beta}_0 &= -8.972 \\ \hat{\beta}_1 &= 0.073 \\ \hat{\sigma}^2 &= 17.932 \end{aligned}$$

$$\widehat{\text{var}}(\hat{\beta}) = \begin{bmatrix} 0.0094837281 & -0.0001640408 \\ -0.0001640408 & 0.0000030839 \end{bmatrix}$$

Los valores de estos estimadores fueron utilizados como base para construir una aproximación normal (asintótica) a la distribución final de β_0 y β_1 , a partir de la cual se simuló, para cada edad x , una muestra de tamaño 5000 de la distribución predictiva de \hat{q}_x^* , así como de la media, la mediana y de los cuantiles de orden 0.025 y 0.975.

Con base en estos cuantiles se construyó una banda predictiva y se realizaron ajustes basados en la mediana y en la media, para las tasas de mortalidad presentadas en las Figuras 4.23 y 4.24 respectivamente. Nótese que el modelo trata de capturar la sobre-dispersión produciendo un valor grande para el estimador de σ^2 ; sin embargo, la banda de predicción no cubre adecuadamente a los datos observados.

También se realizó una comparación entre la tabla actual utilizada por el sector asegurador y la tabla producida por este análisis con base en el cuantil de orden 0.80. La Figura 4.25 muestra este análisis. Dicho cuantil se calculó de manera que el "valor en riesgo" de la suma de siniestros fuera aproximadamente del 2.5%, tal como se describe al final de la Sección 3.2.3. Se puede observar en la Figura 4.26 que este valor es de 25301 siniestros al final del periodo.

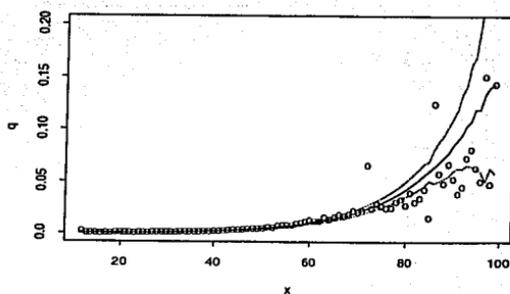


Figura 4.23: Ajuste (mediana) y banda predictiva

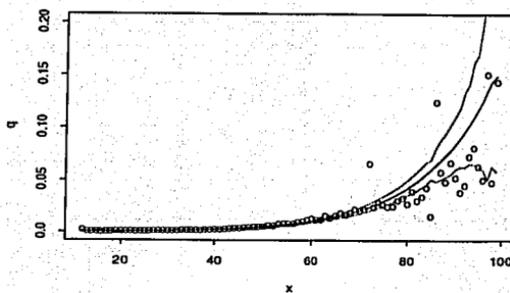


Figura 4.24: Ajuste (media) y banda predictiva

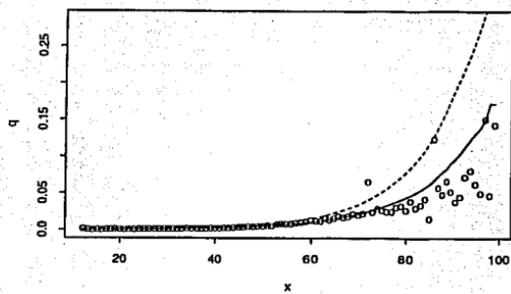


Figura 4.25: Ajuste lineal (—) y tabla actual (- - -)

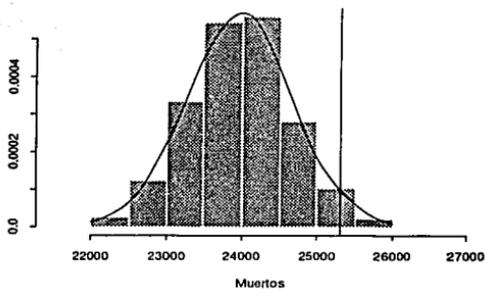


Figura 4.26: Histograma del número de siniestros al final del periodo

4.2.2. Análisis del modelo lineal: WinBUGS

En esta sección se presenta el análisis del mismo modelo pero ahora utilizando la técnica de Muestreo de Gibbs descrita en la Sección 2.2.3.

Con el fin de reducir la correlación entre los parámetros del modelo bajo la distribución final, se utilizó una formulación equivalente dada por

$$\ln \left(\frac{q_x}{1 - q_x} \right) = \alpha + \beta_1 (x - \bar{x})$$

donde $\alpha = \beta_0 + \beta_1 \bar{x}$, y se consideraron las siguientes distribuciones iniciales "no informativas" para α y β_1

$$\begin{aligned}\alpha &\sim N(0, 1.0E-3) \\ \beta_1 &\sim N(0, 1.0E-3),\end{aligned}$$

se realizó entonces el siguiente análisis.

Mediante el programa *WinBUGS* se simuló una muestra de tamaño 100000 de la distribución final de β_0 y β_1 , obteniendo como resultado los siguientes estimadores para los parámetros β_0 y β_1

$$\begin{aligned}\hat{\beta}_0 &= -4.463 \\ \hat{\beta}_1 &= 0.0239.\end{aligned}$$

Asimismo, se construyó la banda predictiva para \hat{q}_2^* mediante los cuantiles de 0.025 y 0.975 de la correspondiente distribución predictiva final (ver Figura 4.27). Nótese que, a diferencia del análisis preliminar, en este caso la banda de predicción cubre a la mayoría de las observaciones.

Se realizó una comparación entre la tabla actual y la tabla producida por el análisis Bayesiano de este modelo con base en el cuantil de orden 0.80. Se puede apreciar en la Figura 4.28 que esta última tabla está por debajo de la utilizada por el sector asegurador.

El "valor en riesgo" en este caso fue de 23310 y corresponde a un nivel del 2.5%, como se presenta en la Figura 4.29.

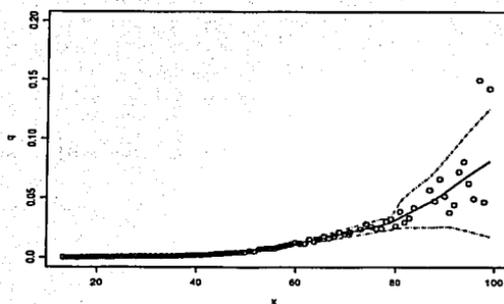


Figura 4.27: Ajuste (media) y banda predictiva

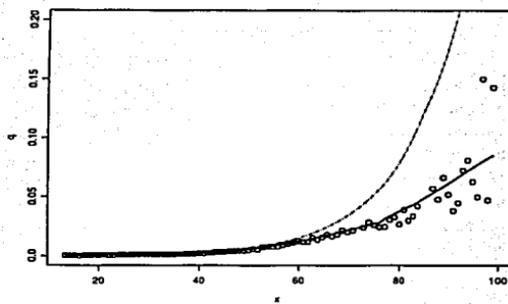


Figura 4.28: Ajuste lineal (-) y tabla actual (- -)

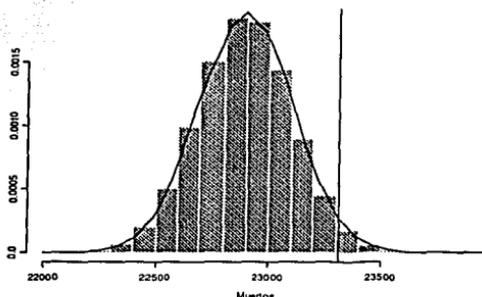


Figura 4.29: *Histograma del número de siniestros al final del periodo*

4.2.3. Análisis preliminar del modelo cuadrático

Con los mismos supuestos que se hacen para el modelo lineal, se ajustará ahora un modelo cuadrático de la forma:

$$\ln \left(\frac{q_x}{1 - q_x} \right) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

Con la ayuda del paquete estadístico *S-Plus* se obtienen los siguientes resultados.

Al ajustar el modelo cuadrático obtenemos los estimadores para los parámetros β_0 , β_1 , β_2 y σ^2 que se muestran a continuación

$$\begin{aligned} \hat{\beta}_0 &= -10.047 \\ \hat{\beta}_1 &= 0.114 \\ \hat{\beta}_2 &= -0.0003 \\ \hat{\sigma}^2 &= 15.370 \end{aligned}$$

$$\widehat{\text{var}}(\hat{\beta}) = \begin{bmatrix} 0.0888534415 & -0.0032416998 & 0.0000276244 \\ -0.0032416998 & 0.0001221229 & -0.0000010672 \\ 0.0000276244 & -0.0000010672 & 0.0000000095 \end{bmatrix}.$$

Los valores de estos estimadores fueron utilizados como base para construir una aproximación normal (asintótica) a la distribución final de β_0, β_1 y β_2 , a partir de la cual se simuló, para cada edad x , una muestra de tamaño 5000 para la distribución predictiva de \hat{q}_x^* , así como para la media, mediana y para los cuantiles de orden 0.025 y 0.975. De esta manera se construyó la banda predictiva basándose en la mediana, mostrada en la Figura 4.30. Análogamente se construyó la banda predictiva basándose en la media, la cual se muestra en la Figura 4.31.

También se realizó una comparación entre la tabla actual utilizada por el sector asegurador y la tabla producida por el análisis con modelos lineales generalizados para el caso del modelo cuadrático, con base en el cuantil de orden 0.80. (Ver Figura 4.32). Podemos ver que aún para las edades de 12 a 65 años la tabla actual está por arriba de la tabla producida por nuestro análisis, como se puede apreciar en la Figura 4.33.

Por último, se calculó el "valor en riesgo" de la suma de siniestros. Este valor resultó ser de 25301 al final del periodo y corresponde a un nivel de 2.5 %, como se observa en la Figura 4.34.

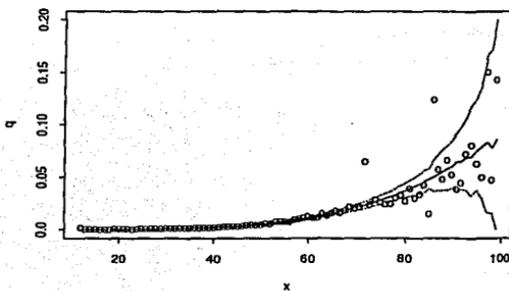


Figura 4.30: Ajuste (mediana) y banda predictiva

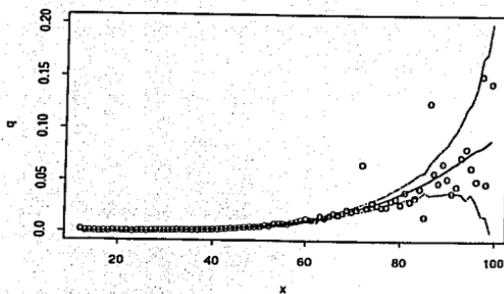


Figura 4.31: Ajuste (media) y banda predictiva

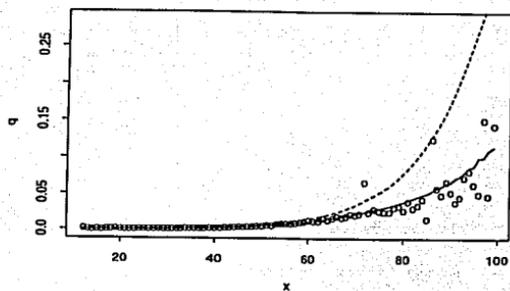


Figura 4.32: Ajuste cuadrático (—) y tabla actual (- - -)

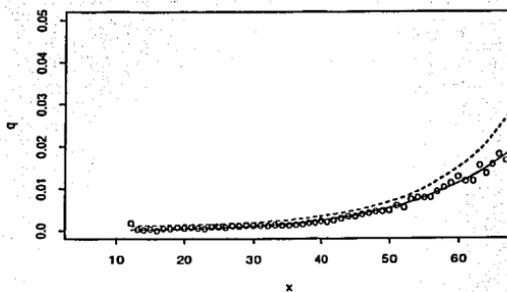


Figura 4.33: Detalle de la figura anterior para las edades entre 12 y 65 años

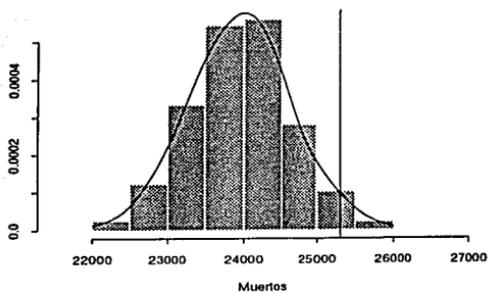


Figura 4.34: Histograma del número de siniestros al final del periodo

4.2.4. Análisis del modelo cuadrático: WinBUGS

En esta sección se presenta el análisis Bayesiano del modelo cuadrático, utilizando la técnica de muestreo de Gibbs descrita en la Sección 2.2.3.

Se utilizaron las siguientes distribuciones iniciales "no informativas" para β_0 , β_1 y β_2 respectivamente

$$\beta_0 \sim N(0, 1.0e - 3)$$

$$\beta_1 \sim N(0, 1.0e - 3)$$

$$\beta_2 \sim N(0, 1.0e - 3)$$

y se consiguió el análisis que se presenta a continuación.

Mediante el programa *WinBUGS* se simuló una muestra de tamaño 100000, para los parámetros del modelo, obteniendo como resultado los siguientes estimadores para los parámetros β_0 , β_1 y β_2

$$\hat{\beta}_0 = -4.336$$

$$\hat{\beta}_1 = 0.0178$$

$$\hat{\beta}_2 = 5.512e - 5.$$

Para este caso también, se construyó la banda predictiva para \hat{q}_x^* mediante los cuantiles de orden 0.025 y 0.975 de la correspondiente distribución predictiva final (ver Figura 4.35).

De igual forma, se comparó la tabla actual utilizada por el sector asegurador contra la tabla producida por este análisis Bayesiano con base en el cuantil de orden 0.80. En la Figura 4.36 se puede ver que también en este caso la tabla actual está por arriba de la producida por este análisis.

El "valor en riesgo" en este caso fue de 23310 y corresponde a un nivel del 2.5 %, como se presenta en la Figura 4.37.

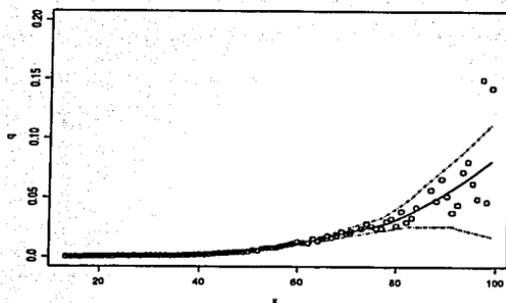


Figura 4.35: Ajuste (media) y banda predictiva

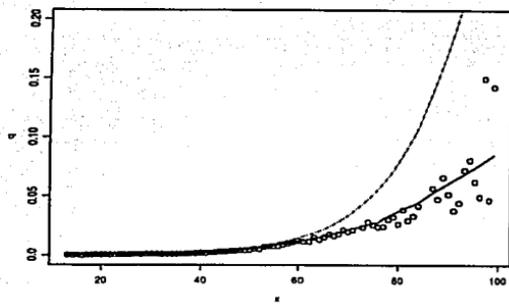


Figura 4.36: Ajuste lineal (—) y tabla actual (- - -)

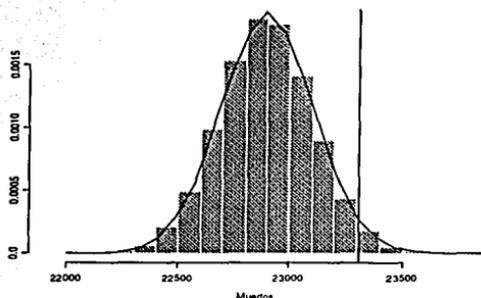


Figura 4.37: *Histograma del número de siniestros al final del periodo*

4.3. Comparación de modelos

4.3.1. Introducción

El desarrollo de los métodos de Monte Carlo vía cadenas de Markov ha hecho posible ajustar modelos cada vez más generales con el objeto de explorar la complejidad de los datos (Gilks *et al.*, 1996). La posibilidad de ajustar tales modelos nos lleva a la necesidad de comparar formulaciones alternativas con el objetivo de identificar una clase de modelos que sean verosímiles como una descripción adecuada de los datos.

Dentro de la estructura de modelos clásicos, en la comparación de modelos se tiende a definir medidas de ajuste, como son la devianza y la complejidad en términos del número de parámetros libres en el modelo. Una mayor complejidad se ve reflejada en un mejor ajuste, pero modelos demasiado complejos pueden dar lugar a predicciones muy pobres.

Los modelos son comparados combinando estas dos cantidades.

4.3.2. Medidas de Ajuste y Complejidad

Discutiremos la comparación de modelos Bayesianos, basándonos en la comparación de la distribución final de la devianza bajo cada modelo.

Identificaremos el "ajuste" como la media posterior de la devianza y la "complejidad" (i.e. el número efectivo de parámetros, p_D) como la diferencia entre la media posterior de la devianza y la devianza basada en la media posterior de los parámetros. Estas cantidades pueden ser fácilmente obtenidas a través de un análisis basado en las técnicas de Monte Carlo vía cadenas de Markov.

El ajuste y la complejidad forman parte del Criterio de Información de Devianza (DIC) el cual se describe a continuación. El lector interesado puede consultar el trabajo de Spiegelhalter *et al.* (2002).

Análisis

Supongamos que estamos ajustando un modelo con datos observados y y cantidades desconocidas θ , que pueden incluir parámetros en diferentes niveles del modelo, variables latentes, datos perdidos, etc.

El enfoque Bayesiano requiere de la especificación de una distribución conjunta $p(y, \phi)$ que generalmente consiste de un producto de muchos términos a causa de suposiciones de independencia condicional.

Esta distribución condicional puede ser escrita como:

$$p(y, \phi) = p(y | \theta)p(\theta | \psi)p(\psi)$$

donde $\phi = (\theta, \psi)$ y y es condicionalmente independiente de ψ dado θ . Así θ es el parámetro que influye directamente en y (i.e. la verdadera media) aun cuando ψ es el hiper-parámetro que gobierna la forma de la distribución previa de θ .

Nuestro interés se enfocará en el parámetro θ ya que éste influye directamente en el ajuste y la habilidad predictiva del modelo.

Dempster (1974) sugirió considerar la distribución final de la log-verosimilitud de los datos, lo cual equivale a examinar la distribución posterior de

$$D(\theta) = -2 \ln p(y, \theta) + 2 \ln f(y)$$

donde $f(y)$ es algún término de estandarización especificado completamente y es una función que depende sólo de los datos, no afectando la comparación de los modelos.

Resumiremos el "ajuste" de un modelo a través de la esperanza de la devianza

$$\bar{D} = E_{\theta|y}[D]$$

y la medida de "complejidad" de un modelo como el número efectivo de parámetros p_D , definida como la esperanza de la devianza menos la devianza evaluada en la esperanza de los parámetros

$$\begin{aligned} p_D &= E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) \\ &= \bar{D} - D(\bar{\theta}). \end{aligned}$$

Finalmente, los modelos pueden ser comparados usando un *Criterio de Información de Devianza*, definido como

$$\begin{aligned} DIC &= \bar{D} + p_D \\ &= D(\bar{\theta}) + 2p_D. \end{aligned}$$

El *DIC* puede calcularse durante una corrida de *MCMC* monitoreando θ y $D(\theta)$, y al finalizar la corrida simplemente se toma la media muestral de los valores simulados de D menos el estimado de la devianza usando las medias de los valores simulados de θ . Los valores más pequeños del *DIC* indican un mejor ajuste. Como en muchas otras herramientas de comparación de modelos, el *DIC* consiste de dos términos, uno representando la bondad de ajuste y el otro penalizando el incremento en la complejidad del modelo.

Resultados

Se realizó una comparación entre los modelos generados para el análisis Bayesiano y otra para los modelos del análisis con Modelos Lineales Generalizados, basándose en ambos casos en un Criterio de Información de Devianza (*DIC*) con la ayuda del paquete estadístico *WinBUGS*.

Los resultados obtenidos se resumen en la Tabla 4.1, los programas y las corridas de éstos pueden consultarse en el Apéndice B.

Se puede observar que, tanto para el modelo de regresión como para el modelo logístico, la diferencia entre el caso lineal y el cuadrático (por lo que respecta al valor del *DIC*) es muy pequeña. Cabe señalar que el *DIC* no es un criterio apropiado para discriminar entre modelos de distintas clases, por lo que la comparación entre el modelo de regresión y el logístico debe hacerse con base en otro criterio. En nuestro caso, debido a las dificultades computacionales para analizar el modelo logístico y a la mayor flexibilidad del modelo de regresión, decidimos utilizar el modelo de regresión cuadrático, el cual se analizará en la siguiente sección.

Tabla 4.1: Comparación de modelos

Regresión normal	Lineal	Cuadrático
Modelo	Sección 4.1.2	Sección 4.1.4
DIC	78.35	74.60

Modelo logístico	Lineal	Cuadrático
Modelo	Sección 4.2.2	Sección 4.2.4
DIC	724.53	724.65

4.4. Análisis de residuos

El análisis de residuos es un método simple y efectivo para detectar posibles deficiencias de un modelo de regresión (ver Apéndice A.3). Nosotros realizamos este análisis para el modelo cuadrático

$$Y_x = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

el cual consideramos es el que mejor describe a los datos, ya que bajo la comparación que se llevó a cabo con el Criterio de Información de la Devianza, el *DIC* resultó ser de 74.6, el cual es menor que el resultante para modelo lineal que fue de 78.35.

Al hacer el análisis del modelo a través del examen de los residuos se obtienen los resultados mostrados en la Tabla 4.2, la cual muestra los residuos e_i para cada edad i . Se puede observar la presencia de "outliers" para las edades 12, 72, 85 y 86, tal como se muestra en la Figura 4.38. Al examinar la gráfica de los residuos e_i contra los valores ajustados se llega a resultados análogos, como se muestra en la Figura 4.39.

Finalmente, al analizar este mismo modelo bajo el enfoque Bayesiano y utilizando el método descrito en el Apéndice A.3.2, se obtiene la Figura 4.40, en la que se puede apreciar la presencia de "outliers", como en el análisis anterior, en las edades 12, 72, 85 y 86.

Se volvió a realizar el análisis de este modelo, pero en esta ocasión omitiendo las observaciones que se detectaron como "outliers".

El modelo analizado es

$$p(\mathbf{Y}|\beta, \tau) = N(\mathbf{y} | \mathbf{x}\beta, \tau^{-1}\mathbf{I})$$

Tabla 4.2: Residuos por edad

Indice	Edad	Residuo	Indice	Edad	Residuo
1	12	4.9782088761	31	43	-0.2759055474
2	13	-0.3509916584	32	44	-0.0135232324
3	14	-1.8976436324	33	45	-0.1746195847
4	15	0.0248311606	34	46	-0.0502831708
5	17	-0.0688155304	35	47	-0.0237192755
6	18	0.03478733	36	48	-0.0028728403
7	19	0.9682515856	37	49	-0.1499859671
8	20	-0.039265913	38	50	-0.183274792
9	21	0.3473540443	39	51	0.2168401772
10	22	-0.5837876695	40	52	0.2168401772
11	23	-1.0214640375	41	53	0.4676662242
12	24	0.1244916263	42	54	0.4228295973
13	25	0.1177525864	43	55	0.2395717375
14	26	-0.5835957705	44	56	0.0830442736
15	27	0.3723771039	45	57	0.3323661714
16	28	-0.1608762329	46	58	0.4219350095
17	29	0.0729498241	47	59	0.5115800428
18	30	0.0008550792	48	60	0.670424404
19	31	-0.2625939683	49	61	0.2846369499
20	32	-0.5904646243	50	62	0.0993556849
21	33	-0.5506201924	51	63	0.6869192486
22	34	-0.7565757479	52	64	0.1392819562
23	35	-0.7970946482	53	65	0.3732716917
24	36	-0.8588522479	54	66	0.5974858685
25	37	-0.8597327254	55	67	0.187623561
26	38	-0.6044930025	56	68	0.2815863209
27	39	-0.6959469657	57	69	0.5917599423
28	40	-0.3580552332	58	70	0.2073242466
29	41	-0.8226567181	59	71	0.245836674
30	42	-0.5437767369	60	72	3.3037747731

Indice	Edad	Residuo
61	73	0.1903016803
62	74	0.5000435291
63	75	0.0850366363
64	76	-0.2853114183
65	77	-0.4187404563
66	78	0.0306925449
67	79	0.0668626763
68	80	-0.6757354903
69	81	0.2434238269
70	82	-0.6939658912
71	83	-0.5094691543
72	84	-0.0035642409
73	85	-3.1472190888
74	86	2.9677004366
75	87	0.405157961
76	88	-0.2794556781
77	89	0.53514975
78	90	-0.3544035089
79	91	-1.4304797319
80	92	-1.1205008315
81	93	0.1752489004
82	94	0.3512512864
83	95	-0.5673718425
84	96	-1.4164118452
85	97	1.8474310126
86	98	-1.9060787203
87	99	1.3650534758

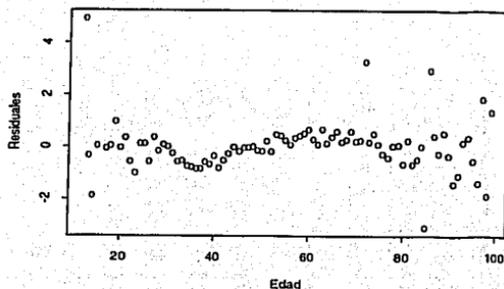


Figura 4.38: Residuos estandarizados vs variable independiente

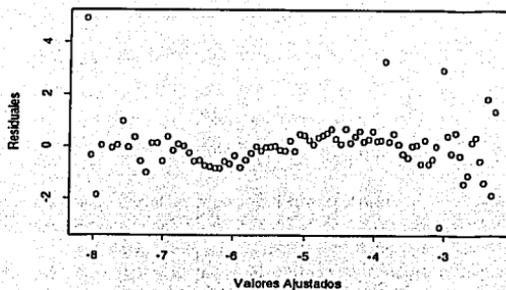


Figura 4.39: Residuos estandarizados vs valores ajustados

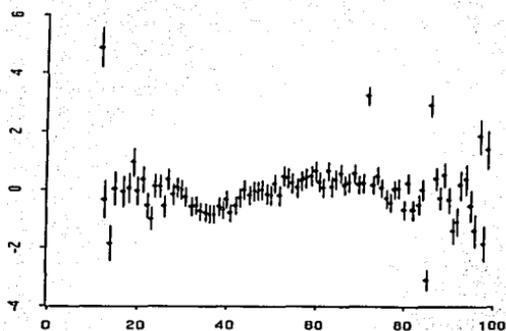


Figura 4.40: *Análisis de residuos Bayesiano*

y se consideraron las siguientes distribuciones iniciales "no informativas"

$$\beta_0 \sim N(0, 1.0e-3)$$

$$\beta_1 \sim N(0, 1.0e-3)$$

$$\beta_2 \sim N(0, 1.0e-3)$$

$$\tau \sim Ga(1.0e-3, 1.0e-3).$$

En esta ocasión los estimadores para los parámetros β_0 , β_1 , β_2 y σ^2 fueron

$$\hat{\beta}_0 = -9.439$$

$$\hat{\beta}_1 = 0.09598$$

$$\hat{\beta}_2 = -0.0002$$

$$\hat{\sigma}^2 = 0.04804.$$

Se hizo un ajuste basado en la media, y se contruyó una banda predictiva mediante los cuantiles de orden 0.025 y 0.975 de la correspondiente distribución predictiva final los cuales se muestran en la Figura 4.41.

En esta ocasión, también se realizó la comparación entre la tabla actual utilizada por el sector asegurador para el cálculo de las primas y las reservas (ver Tabla 1.2) y la tabla producida por el análisis Bayesiano sin outliers con base en el cuantil de orden 0.70, como se muestra en la Figura 4.42.

Como se puede apreciar en la Figura 4.43 aún para las edades de 13 a 65 años, donde esta el mayor número de la población, la tabla producida por este nuevo análisis se encuentra por debajo de la tabla actual. Al calcular el "valor en riesgo" de la suma de siniestros (SS) a un nivel del 97.5% éste resultó ser de 25710 muertos o siniestros al final del periodo, como se muestra en el histograma de la Figura 4.44.

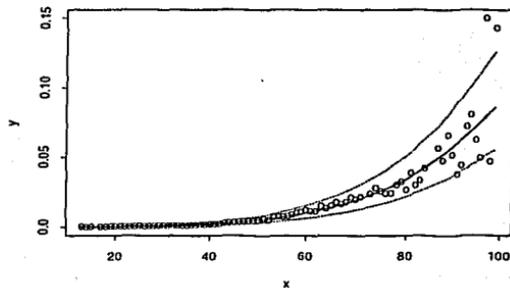


Figura 4.41: Ajuste (media) y banda predictiva

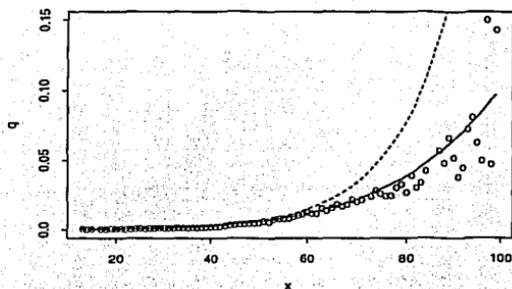


Figura 4.42: Ajuste cuadrático (—) y tabla actual (- - -)

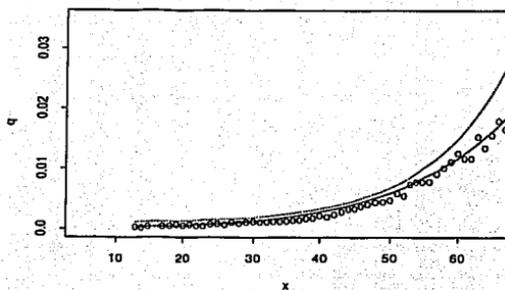


Figura 4.43: Detalle del análisis, para las edades entre 13 y 65 años

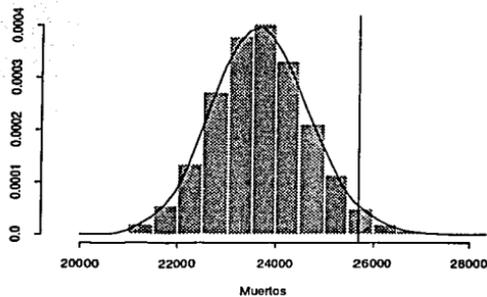


Figura 4.44: Número agregado de siniestros al final del periodo

Capítulo 5

Conclusiones

La graduación y sobre-estimación de las tasas de mortalidad son dos de los problemas más importantes en la actuaría. Los métodos actuariales tradicionales han tratado cada uno de estos aspectos separadamente y por lo general han ignorado su naturaleza estadística.

El método propuesto en este trabajo permite la construcción de tablas de mortalidad por medio de un algoritmo que aborda los temas de graduación y de sobre-estimación de manera integral, reconociendo la naturaleza estadística del problema y haciendo uso de un enfoque predictivo, siempre desde un punto de vista Bayesiano. Los pasos básicos son:

1. El ajuste de un modelo de regresión para las tasas de mortalidad observadas transformadas.
2. El cálculo de la distribución predictiva final conjunta para las tasas de mortalidad por observarse en el siguiente periodo.
3. La selección de un modelo modificado con base en un "valor en riesgo" determinado, calculado respecto a la distribución predictiva final del número agregado de reclamaciones por observarse en el siguiente periodo, dado un patrón específico de expuestos.

La consecuencia más importante de este resultado es que no es necesario utilizar un método extremadamente conservador para obtener un nivel razonable de protección global.

Por simplicidad, en este trabajo se utilizaron dos clases de modelos sencillos a pesar de que, para el conjunto de datos analizado, algunos de los

supuestos eran claramente violados. Cabe señalar, sin embargo, que nuestro fin no es explicar la relación entre las variables, sino predecir el comportamiento de la mortalidad, de manera que sólo requerimos de un mecanismo de predicción razonable. De los cuatro modelos considerados, el modelo de regresión normal con un término cuadrático es el que describe mejor los datos observados.

Se debe hacer hincapié en que el método es muy general en el sentido que, primero, podría utilizarse otra transformación en lugar de la logística; segundo, es posible utilizar otra estructura monótona para describir la relación entre la edad y las probabilidades de muerte desconocidas. Todos los conceptos estadísticos involucrados han sido discutidos ampliamente y están bien establecidos en la literatura.

La distribución predictiva final del número agregado de reclamaciones fue producida con base en el número observado de expuestos (el mismo que fue usado para ajustar el modelo). Sin embargo, un modelo más realista debería reconocer la naturaleza incierta de la estructura de los expuestos a través de una distribución de probabilidad.

Nuestra propuesta es usar el número agregado de reclamaciones como una cantidad clave a fin de calcular un nivel apropiado de protección provisto por una tabla de mortalidad. Sin embargo, las pérdidas sufridas por un sistema de seguridad no se miden a través del número de reclamaciones sino a través del monto total de éstas. El procedimiento aquí propuesto es equivalente a considerar el monto de las reclamaciones sólo si se supone que la suma asegurada es la misma para todos los individuos sin importar su edad. Este problema puede ser abordado de una manera similar, aunque se requeriría de mayor información que la disponible para el presente análisis.

Apéndice A

Resultados Preliminares

El objetivo de este apéndice es hacer una recopilación de algunos de los resultados básicos que pueden ser importantes para la comprensión de la teoría matemática que utilizan los modelos lineales generalizados.

A.1. Distribuciones más comunes

A continuación se listan las familias paramétricas de las densidades discretas y continuas que fueron utilizadas a lo largo del presente trabajo.

Distribución binomial

Definición A.1 Una variable aleatoria discreta X tiene una distribución binomial con parámetros n y p , si su función de probabilidad $Bi(x|p, n)$ es:

$$Bi(x|p, n) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n \quad (\text{A.1})$$

donde los parámetros n y p satisfacen $0 \leq p \leq 1$, $n \in N$.

La media y la varianza de X están dadas por

$$\begin{aligned} E(x) &= np, \\ Var(x) &= np(1-p). \end{aligned}$$

Si $n=1$, X tiene una distribución Bernoulli, con función de probabilidad denotada por $Br(x|\theta)$. La suma de k variables aleatorias independientes binomial con parámetros (p, n_i) , $i = 1, \dots, k$, es una cantidad aleatoria binomial con parámetro p y $n_1 + n_2 + \dots + n_k$.

Distribución de Poisson

Definición A.2 Una variable aleatoria discreta X tiene una distribución Poisson con parámetro λ , ($\lambda > 0$) si su función de probabilidad es:

$$P_n(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (\text{A.2})$$

La media y la varianza de X están dadas por

$$\begin{aligned} E(x) &= \lambda, \\ \text{Var}(x) &= \lambda. \end{aligned}$$

La suma de k variables aleatorias independientes Poisson con parámetros λ_i , $i = 1, \dots, k$, es una variable aleatoria Poisson con parámetro $\lambda_1 + \dots + \lambda_k$.

Distribución beta

Definición A.3 Una variable aleatoria continua X tiene una distribución beta, con parámetros α y β ($\alpha > 0, \beta > 0$), si su función de densidad es:

$$Be(x|\alpha, \beta) = cx^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1 \quad (\text{A.3})$$

donde

$$c = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

La cantidad $\Gamma(\alpha)$ se conoce como la *función gamma* y se define como:

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

Además, para cualquier entero n se tiene

$$\Gamma(n) = n\Gamma(n-1) = (n-1)!.$$

La media y la varianza de X están dadas por:

$$\begin{aligned} E(x) &= \frac{\alpha}{\alpha + \beta}, \\ \text{Var}(x) &= \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}. \end{aligned}$$

Distribución normal

Definición A.4 Una variable aleatoria continua X tiene una distribución normal con parámetros μ y σ^2 , si su densidad $N(x|\mu, \sigma^2)$ es:

$$N(x|\mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \text{para } -\infty < x < \infty, \quad (\text{A.4})$$

donde los parámetros μ y σ satisfacen $-\infty < \mu < \infty$, $\sigma > 0$.

La media y la varianza de X están dadas por

$$\begin{aligned} E(x) &= \mu, \\ \text{Var}(x) &= \sigma^2. \end{aligned}$$

Si $\mu = 0$ y $\sigma^2 = 1$, X tiene una función de distribución normal estándar Φ dada por:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}t^2\right\} dt.$$

Distribución gamma

Definición A.5 Una variable aleatoria continua X tiene una distribución gamma con parámetros α y β , si su función de densidad $Ga(x|\alpha, \beta)$ es:

$$Ga(x|\alpha, \beta) = cx^{\alpha-1}e^{-\beta x}, \quad x > 0 \quad (\text{A.5})$$

donde

$$c = \frac{\beta^\alpha}{\Gamma(\alpha)}.$$

La media y la varianza de X están dadas por

$$\begin{aligned} E(x) &= \frac{\alpha}{\beta}, \\ \text{Var}(x) &= \frac{\alpha}{\beta^2}. \end{aligned}$$

Si $\alpha = 1$, X tiene una distribución exponencial $Ex(x|\beta)$ con parámetro β y densidad

$$Ex(x|\beta) = \beta e^{-\beta x}, \quad x \geq 0.$$

Todas las distribuciones descritas hasta este punto son familias exponenciales (ver la sección siguiente).

Distribución beta-binomial

Definición A.6 Una variable aleatoria discreta X tiene una distribución beta - binomial con parámetros α , β y n ($\alpha > 0, \beta > 0, n = 1, 2, \dots$) si su función de probabilidad $Bb(x|\alpha, \beta, n)$ es

$$Bb(x|\alpha, \beta, n) = c \binom{n}{x} \Gamma(\alpha + x) \Gamma(\beta + n - x), \quad x = 0, \dots, n \quad (\text{A.6})$$

donde

$$c = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\beta + n + x)}.$$

La distribución es generada de la siguiente relación

$$Bb(x|\alpha, \beta, n) = \int_0^1 Bi(x|\alpha, \beta, n) Be(\theta|\alpha, \beta) d\theta.$$

La media y la varianza de X están dadas por

$$E(x) = n \frac{\alpha}{\alpha + \beta}, \quad Var(x) = \frac{n\alpha\beta}{(\alpha + \beta)^2} \frac{(\alpha + \beta + n)}{(\alpha + \beta + 1)}.$$

Distribución normal-gamma

Definición A.7 Un vector aleatorio continuo (X, Y) tiene una distribución normal-gamma con parámetros μ, λ, α y β , ($\mu \in \mathfrak{R}, \lambda > 0, \alpha > 0, \beta > 0$) si su densidad $Ng(x, y|\mu, \lambda, \alpha, \beta)$ es

$$Ng(x, y|\mu, \lambda, \alpha, \beta) = N(x|\mu, \lambda y) Ga(y|\alpha, \beta), \quad x \in \mathfrak{R}, y > 0 \quad (\text{A.7})$$

donde las distribuciones normal y gamma fueron definidas anteriormente.

Es claro de la definición que, la densidad condicional de x dado y es $N(x|\mu, \lambda y)$ y la densidad marginal de y es $Ga(y|\alpha, \beta)$. Sin embargo la densidad marginal de x es $St(x|\mu, \lambda\alpha/\beta, 2\alpha)$.

La media y la varianza de X y Y están dadas por

$$E(x) = \mu, \quad Var(x) = \frac{\beta}{\lambda(\alpha - 1)}, \quad E(y) = \frac{\alpha}{\beta}, \quad Var(y) = \frac{\alpha}{\beta^2}.$$

Distribución t de Student

Definición A.8 Una variable aleatoria continua X tiene una distribución *t* - Student con parámetros μ, λ y α ($\mu \in \mathfrak{R}, \lambda > 0, \alpha > 0$) si su densidad $St(x|\mu, \lambda, \alpha)$ es

$$St(x|\mu, \lambda, \alpha) = c \left[1 + \frac{\lambda}{\alpha} (x - \mu)^2 \right]^{-\frac{\alpha+1}{2}} \quad x \in \mathfrak{R}, \quad (\text{A.8})$$

donde

$$c = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{1}{2})} \left(\frac{\lambda}{\alpha} \right)^{\frac{1}{2}}$$

La media y la varianza de x están dadas por

$$\begin{aligned} E(x) &= \mu, & \text{si } \alpha > 1 \\ \text{Var}(x) &= \frac{1}{\lambda} \frac{\alpha}{\alpha - 2}, & \text{si } \alpha > 2. \end{aligned}$$

Donde α son los grados de libertad de la distribución.

La distribución es generada de la siguiente relación

$$St(x|\mu, \lambda, \alpha) = \int_0^{\infty} N(x|\mu, \lambda y) Ga(y|\frac{\alpha}{2}, \frac{\alpha}{2}) dy$$

e incluye a la distribución normal como un caso límite, ya que

$$N(x|\mu, \lambda) = \lim_{\alpha \rightarrow \infty} St(x|\mu, \lambda, \alpha).$$

Si $y = \lambda^{1/2} (x - \mu)$, donde x tienen una distribución $St(x|\mu, \lambda, \alpha)$, entonces y tiene una distribución Student estándar $St(x|0, 1, \alpha)$. Si $\alpha = 1$, se dice que x tiene una distribución Cauchy con densidad $Ca(x|\mu, \lambda)$.

Si x tiene una distribución normal estándar, y tiene una distribución χ^2 , y x y y son mutuamente independientes entonces

$$z = \frac{x}{(y/\nu)^{1/2}}$$

tiene una distribución estándar Student $St(x|0, 1, \nu)$.

A.2. Especificación de la distribución inicial

En este apéndice se discutirán diferentes formas de especificar una distribución inicial, partiendo de la interpretación subjetiva de la probabilidad. La especificación de la distribución inicial tiene el propósito de describir la información acerca de un parámetro desconocido θ antes de observar los datos; por ejemplo, los resultados de un nuevo experimento.

A.2.1. Especificación subjetiva

Cuando el espacio parametral Θ es finito (i.e. si θ es una variable aleatoria discreta) en principio la probabilidad inicial para cada posible valor de θ puede evaluarse directamente. En el caso continuo es ligeramente más complicado. Algunas sugerencias para este caso son:

Aproximación por medio de un histograma

Un primer paso consiste en construir una partición del espacio parametral Θ en regiones (e.g. intervalos) y asignar probabilidad iniciales a cada uno de ellos como en el caso discreto. La finalidad de un histograma para θ es construir una curva paramétrica suave que pueda ser utilizada como aproximación a la distribución inicial. El número de intervalos es seleccionado arbitrariamente. Este método presenta problemas cuando Θ no está acotado, pues asignar probabilidades a las colas de la distribución no es una tarea sencilla en este caso. Aunque la probabilidad en las colas de la distribución inicial es a menudo muy pequeña, el comportamiento en las colas puede llegar a influir en las inferencias subsecuentes.

Aproximación de la función de distribución

Se define θ_α como el percentil de orden $100\alpha\%$ (cuantil de orden α) de la distribución de θ si $\Pr(\theta \leq \theta_\alpha) = \alpha$, $\alpha \in [0, 1]$. La mediana de θ , denotada por m , es el percentil 50%, tal que $\Pr(\theta \leq m) = 0.5$. La colección de todos los percentiles de θ describe la función de distribución de θ . En este procedimiento, se asignan subjetivamente algunos de los percentiles, como en el caso discreto, y posteriormente se ajusta una curva suavizada para obtener una aproximación de la función de distribución de θ .

A.2.2. Especificación a través de formas funcionales

Una alternativa usada frecuentemente en la práctica para construir distribuciones iniciales consiste en restringir la selección de la densidad inicial a una familia paramétrica y posteriormente determinar los valores de los parámetros a través de la especificación de ciertas características de la distribución tales como momentos y/o cuantiles.

Aunque muchas veces esta familia se construye de manera que sea fácil de analizar, se debe tener cuidado y estar seguros que la densidad seleccionada realmente representa la información disponible. Por ejemplo, podríamos hacer las siguientes suposiciones acerca de θ :

- θ está simétricamente distribuida con respecto a la moda.
- Su densidad decae rápido (e.g. exponencialmente) cuando se encuentra lejos de la moda.
- Intervalos lejos de la moda tienen probabilidades despreciables.

Estas consideraciones caracterizan aproximadamente a una distribución normal.

Estas ideas pueden llevarse a una estructura más general, la cual proporciona una forma sistemática de determinar distribuciones iniciales. El caso más relevante corresponde a las llamadas *familias conjugadas* de distribuciones. Por ejemplo, se sabe que si la distribución muestral es $p(x | \theta) = N(x | \theta, \sigma^2)$ (σ^2 conocida) y la distribución inicial de θ es $p(\theta) = N(\theta | \mu, \tau^2)$ entonces la distribución final también es normal, y está dada por $p(\theta | x) = N(\theta | \mu_1, \tau_1^2)$; ver Ejemplo A.2.

En otras palabras, si se inicia con una distribución normal se finalizará con una normal donde sólo los valores de los *hiperparámetros* cambiarán; no requiere de nuevos cálculos analíticos. La ventaja principal de este enfoque es la simplicidad en el análisis.

Definición A.9 Sea $\mathcal{F} = \{p(x | \theta), \theta \in \Theta\}$ una familia de distribuciones muestrales. Se dice que una clase \mathcal{P} de distribuciones es conjugada con respecto a \mathcal{F} si para toda $p(x | \theta) \in \mathcal{F}$ y $p(\theta) \in \mathcal{P}$ entonces $p(\theta | x) \in \mathcal{P}$.

Así, se puede decir que la clase de distribuciones normal es conjugada con respecto a la clase de distribuciones (muestrales) normales.

Ejemplo A.1 Bernoulli

Sea $(x_i | \theta) \sim \text{Bernoulli}(\theta)$, para $i = 1, \dots, n$. La densidad conjunta es $p(x | \theta) = \theta^t (1 - \theta)^{n-t}$, donde $t = \sum_{i=1}^n x_i$; $x_i = 0, 1$, definiendo una clase de distribuciones parametrizadas por $\theta \in (0, 1)$, por el teorema de Bayes, se conoce que la distribución final de θ dada x esta dada por

$$\begin{aligned} p(\theta | x) &\propto p(x | \theta) p(\theta) \\ &\propto \theta^t (1 - \theta)^{n-t} p(\theta). \end{aligned}$$

La conjugada inicial puede ser obtenida del kernel de la función de verosimilitud. En este ejemplo el kernel de la verosimilitud es de la forma

$$\theta^a (1 - \theta)^b,$$

que es también el kernel de la familia de distribuciones beta. Tomando la distribución inicial como una $\text{Beta}(\alpha, \beta)$ y combinándola con la verosimilitud, la distribución final es

$$\begin{aligned} p(\theta | x) &\propto \theta^t (1 - \theta)^{n-t} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &\propto \theta^{\alpha+t-1} (1 - \theta)^{\beta+n-t-1}. \end{aligned}$$

Por lo tanto $(\theta | x) \sim \text{Beta}(\alpha + t, \beta + n - t)$ la cual cae dentro de la familia de distribuciones de la distribución inicial. Así la familia beta es conjugada con respecto al modelo muestral Bernoulli. No es difícil mostrar que lo mismo sucede con las distribuciones binomial, geométrica y binomial Negativa. La constante de proporcionalidad de la distribución final es

$$\frac{1}{B(\alpha + t, \beta + n - t)} = \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + t)\Gamma(\beta + n - t)}.$$

En la siguiente sección se describe la forma general de la familia conjugada para una clase importante de distribuciones muestrales.

A.2.3. Familia exponencial

La familia exponencial incluye las principales distribuciones de probabilidad utilizadas en Estadística. Una característica esencial de esta familia es

que posee estadísticas suficientes de dimensión fija. La clase conjugada \mathcal{P} para la familia exponencial con un parámetro es fácil de caracterizar. Los miembros de esta familia tienen densidades de la forma

$$p(x | \theta) = h(x) \exp \{u(x)a(\theta) - b(\theta)\}.$$

Si consideramos densidades iniciales de la forma

$$p(\theta) \propto \exp \{\alpha a(\theta) - \beta b(\theta)\}$$

entonces

$$p(\theta | x) \propto \exp \{[\alpha + u(x)]a(\theta) + [\beta + 1]b(\theta)\}.$$

Denotando a la constante involucrada en $p(\theta)$ por $k(\alpha, \beta)$ la constante asociada para $p(\theta | x)$ es $k(\alpha + u(x), \beta + 1)$.

Entonces de acuerdo con el teorema de Bayes se tiene

$$p(x) = \frac{p(x | \theta)p(\theta)}{p(\theta | x)}$$

así, sustituyendo se llega a

$$p(x) = \frac{h(x) \exp \{u(x)a(\theta) - b(\theta)\} k(\alpha, \beta) \exp \{\alpha a(\theta) - \beta b(\theta)\}}{k(\alpha + u(x), \beta + 1) \exp \{[\alpha + u(x)]a(\theta) - [\beta + 1]b(\theta)\}}$$

después de un poco de álgebra se obtiene

$$p(x) = \frac{h(x) k(\alpha, \beta)}{k(\alpha + u(x), \beta + 1)}.$$

Para el modelo binomial en este caso se tiene

$$\begin{aligned} p(x) &= \frac{\binom{n}{x} \theta^x (1-\theta)^{n-x} B^{-1}(\alpha, \beta) \theta^{\alpha-1} (1-\theta)^{\beta-1}}{B^{-1}(\alpha+x, \beta+n-x) \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}} \\ &= \binom{n}{x} \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \end{aligned}$$

la cual es una distribución beta-binomial para $x = 0, 1, \dots, n, n \geq 1$.

En general, para una muestra de tamaño n de la familia exponencial se obtiene la densidad conjunta

$$p(x | \theta) = \left[\prod_{i=1}^n h(x_i) \right] \exp \left\{ \left[\sum_{i=1}^n u(x_i) \right] a(\theta) + nb(\theta) \right\}.$$

El uso de una distribución inicial conjugada $p(\theta) = k(\alpha, \beta) \exp \{ \alpha a(\theta) + \beta b(\theta) \}$ da como resultado una distribución final de la forma

$$p(\theta | x) = k \left(\alpha + \sum_{i=1}^n u(x_i), \beta + n \right) \times \exp \left\{ \left[\alpha + \sum_{i=1}^n u(x_i) \right] a(\theta) + [\beta + n] b(\theta) \right\}$$

y la distribución marginal o predictiva es

$$p(x) = \frac{[\prod h(x_i)] k(\alpha, \beta)}{k(\alpha + \sum u(x_i), \beta + n)}.$$

Ejemplo A.2 Normal (media desconocida y varianza conocida)

La familia de distribución normal es conjugada para el modelo normal. Para el caso de una muestra de tamaño n se tiene que

$$L(\theta; x) \propto \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \theta)^2 \right\}$$

donde los términos que involucran a σ^2 fueron incorporados en la constante de proporcionalidad. De este modo la verosimilitud tiene la misma forma que la que se basa en una sola observación x , sustituyendo x por \bar{x} y σ^2 por $\frac{\sigma^2}{n}$. Otra forma de ver esto es notando que \bar{X} es una estadística suficiente para θ y así la verosimilitud basada en el valor observado de \bar{X} , la cual se distribuye como una $N\left(\theta, \frac{\sigma^2}{n}\right)$, es proporcional a la verosimilitud obtenida con las observaciones individuales X . Por tanto, si la distribución inicial de θ es $N(\mu, \tau^2)$, entonces la distribución final de θ , dada x es $N(\mu_1, \tau_1^2)$, donde

$$\mu_1 = \frac{n\sigma^{-2}\bar{x} + \tau^{-2}\mu}{n\sigma^{-2} + \tau^{-2}} \quad \text{y} \quad \tau_1^{-2} = n\sigma^{-2} + \tau^{-2}.$$

Ejemplo A.3 Poisson

Supóngase que $X = (X_1, \dots, X_n)$ es una muestra aleatoria de una distribución Poisson con parámetro θ . Su función de probabilidad conjunta es

$$p(x | \theta) = \prod_{i=1}^n p(x_i | \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

así que la función de verosimilitud tiene la forma

$$L(\theta; x) \propto e^{-n\theta} \theta^{\sum x_i}$$

Su *kernel* tiene la forma $\theta^\alpha e^{-\theta}$ el cual caracteriza a la familia de distribuciones gamma. La distribución inicial conjugada de θ es gamma con parámetros α y β , $G(\alpha, \beta)$ y la densidad final es

$$p(\theta | x) \propto \theta^{\alpha + \sum x_i - 1} \exp\{-(\beta + n)\theta\}$$

la cual corresponde a una distribución gamma $G(\alpha + \sum x_i, \beta + n)$.

Ejemplo A.4 Exponencial

Supóngase que $X = (X_1, \dots, X_n)$ es una muestra aleatoria de una distribución exponencial con parámetro θ . Su función de probabilidad conjunta es

$$p(x | \theta) = \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\}.$$

La forma de la verosimilitud permite reconocer del *kernel* de la familia gamma como la distribución conjugada para θ . Asumiendo a $G(\alpha, \beta)$ como la distribución inicial, la distribución final tiene la forma

$$\begin{aligned} p(\theta | x) &\propto \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\} \theta^{\alpha-1} \exp\{-\beta\theta\} \\ &\propto \theta^{n+\alpha-1} \exp\left\{-\left(\beta + \sum x_i\right)\theta\right\} \end{aligned}$$

la cual es la densidad de la distribución $G(n + \alpha, \beta + \sum x_i)$.

Ejemplo A.5 Normal (media y varianza desconocidas)

Sea $X = (X_1, \dots, X_n)$ una muestra aleatoria de tamaño n de una distribución normal $N(\theta, \sigma^2)$, con θ conocida, y $\phi = \frac{1}{\sigma^2}$. La distribución inicial conjugada para (θ, ϕ) se presentará en dos etapas. Primero, considerese la distribución condicional de θ como

$$p(\theta | \phi) = N(\mu_0, [c_0\phi]^{-1})$$

y la inicial marginal para ϕ está dada por

$$\phi \sim G\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right).$$

Esta distribución es usualmente llamada normal-gamma y su densidad conjunta esta dada por

$$\begin{aligned} p(\theta, \phi) &= p(\theta | \phi) p(\phi) \\ &\propto \theta^{\frac{1}{2}} \exp\left\{-\frac{c_0\phi}{2}(\theta - \mu_0)^2 \phi^{\frac{n_0}{2}-1}\right\} \exp\left\{-\frac{n_0\sigma_0^2\phi}{2}\right\} \\ &= \theta^{\frac{(n_0+1)}{2}-1} \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + c_0(\theta - \mu_0)^2]\right\}. \end{aligned}$$

La distribución inicial marginal de θ se puede obtener por integración, usando el siguiente resultado

$$\int_0^{\infty} \phi^{a-1} e^{-b\phi} d\phi = \frac{\Gamma(a)}{b^a}.$$

Aplicando los resultados obtenidos, se tiene que

$$\begin{aligned} p(\theta) &\propto \int_0^{\infty} \theta^{\frac{(n_0+1)}{2}-1} \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + c_0(\theta - \mu_0)^2]\right\} d\phi \\ &= \frac{\Gamma\left(\frac{(n_0+1)}{2}\right)}{\left\{\frac{[n_0\sigma_0^2 + c_0(\theta - \mu_0)^2]}{2}\right\}^{\frac{(n_0+1)}{2}}} \\ &\propto [n_0\sigma_0^2 + c_0(\theta - \mu_0)^2]^{-\frac{(n_0+1)}{2}} \end{aligned}$$

ya que $\Gamma(\cdot)$ no depende de θ . Reagrupando términos se tiene

$$p(\theta) \propto \left[1 + \frac{(\theta - \mu_0)^2}{n_0 \left(\frac{\sigma_0^2}{c_0} \right)} \right]$$

el cual es el kernel de una distribución t-Student con n_0 grados de libertad, parámetro de localización μ_0 y parámetro de escala $\frac{\sigma_0^2}{c_0}$, denotada por $t_{n_0} \left(\mu_0, \frac{\sigma_0^2}{c_0} \right)$.

La distribución condicional de ϕ | θ puede obtenerse de la distribución conjunta $p(\theta, \phi)$ y es una gamma con parámetros $\frac{n_0+1}{2}$ y $\frac{n_0\sigma_0^2+c_0(\theta-\mu_0)^2}{2}$.

La distribución conjunta de una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ es

$$\begin{aligned} p(x | \theta, \phi) &= \prod_{i=1}^n \phi^{\frac{1}{2}} \exp \left\{ -\frac{\phi}{2} (x_i - \theta)^2 \right\} \\ &\propto \phi^{\frac{n}{2}} \exp \left\{ -\frac{\phi}{2} [ns^2 + n(\bar{x} - \theta)^2] \right\} \end{aligned}$$

donde $s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$. Esta expresión tiene el mismo kernel que el de una densidad normal-gamma para (θ, ϕ) . El siguiente paso es verificar que la familia normal-gamma es cerrada bajo muestreo. La distribución final está dada por

$$\begin{aligned} p(\theta, \phi | x) &\propto p(x | \theta, \phi) p(\theta, \phi) \\ &\propto \phi^{\left(\frac{n+n_0+1}{2} \right) - 1} \\ &\quad \times \exp \left\{ -\frac{\phi}{2} [n_0\sigma_0^2 + ns^2 + c_0(\theta - \mu_0)^2 + n(\bar{x} - \theta)^2] \right\}. \end{aligned}$$

No es difícil mostrar que

$$c_0(\theta - \mu_0)^2 + n(\bar{x} - \theta)^2 = (c_0 + n)(\theta - \mu_1)^2 + \frac{c_0 n}{c_0 + n}(\mu_0 - \bar{x})^2$$

donde $\mu_1 = \frac{c_0\mu_0 + n\bar{x}}{c_0 + n}$. De esto se sigue que la distribución final para (θ, ϕ) es proporcional a

$$\begin{aligned} &\phi^{\left(\frac{n+n_0+1}{2} \right) - 1} \\ &\times \exp \left\{ -\frac{\phi}{2} \left[n_0\sigma_0^2 + ns^2 + (c_0 + n)(\theta - \mu_1)^2 + \frac{c_0 n}{c_0 + n}(\mu_0 - \bar{x})^2 \right] \right\}. \end{aligned}$$

Por lo tanto la distribución final conjunta es normal-gamma con parámetros $(\mu_1, c_1, n_1, \sigma_1^2)$ dados por

$$\begin{aligned}\mu_1 &= \frac{c_0\mu_0 + n\bar{x}}{c_0 + n} \\ c_1 &= c_0 + n \\ n_1 &= n_0 + n \\ \sigma_1^2 &= n_0\sigma_0^2 + ns^2 + \frac{c_0n}{c_0 + n}(\mu_0 - \bar{x})^2.\end{aligned}$$

Tanto la distribución inicial como la final son miembros de la misma familia. De esta forma, la familia normal-gamma es conjugada con respecto al modelo muestral normal cuando θ y σ^2 son desconocidas.

A.2.4. Distribución inicial no informativa

Las distribuciones iniciales conjugadas son de gran utilidad como descripciones (aproximadas) a la verdadera información inicial. Sin embargo, cuando no hay información inicial disponible, o cuando por alguna razón no se desea incorporar dicha información en el análisis, su justificación es sólo analítica, pues generalmente producen expresiones cerradas para algunas características de la distribución final tales como momentos. En este caso no es posible justificar la selección de la distribución inicial con bases subjetivas y por lo tanto los parámetros de la distribución inicial conjugada no pueden determinarse.

En este contexto, la idea de una distribución inicial *no informativa*, que represente un estado de ignorancia y que permita que los datos "hablen por sí mismos", es muy atractiva. A menudo las inferencias basadas en este tipo de distribuciones iniciales son consideradas "objetivas".

Describiremos a continuación algunas de las técnicas más importantes en la derivación de distribuciones iniciales no informativas; el lector interesado puede consultar Kass y Wasserman (1996). No se debe esperar que las distribuciones iniciales no informativas representen siempre un estado de ignorancia total acerca del problema en cuestión, pero en la práctica, en términos generales, basta considerar las llamadas *distribuciones iniciales de referencia* (en particular, la distribución de Jeffreys), las cuales se describen a continuación.

Distribución de Jeffreys

Jeffreys (1946, 1961) propone una distribución inicial no informativa que es invariante ante reparametrizaciones del modelo.

La distribución inicial no informativa de Jeffreys se basa en la información de Fisher, dada por

$$I(\theta) = E_{\theta} \left[\left(\frac{\partial \ln f(x|\theta)}{\partial \theta} \right)^2 \right].$$

Bajo ciertas condiciones de regularidad, esta cantidad también puede escribirse como

$$I(\theta) = -E_{\theta} \left(\frac{\partial^2 \ln f(x|\theta)}{\partial \theta^2} \right)$$

dando lugar a la distribución inicial no informativa de Jeffreys

$$\pi^*(\theta) \propto I(\theta)^{\frac{1}{2}}$$

definida módulo un coeficiente de normalización cuando π^* es propia. Esta distribución es típicamente impropia y es invariante en el sentido de que, dada una transformación h , uno a uno, se tiene que

$$I(\theta) = I(h(\theta)) (h'(\theta))^2.$$

Cuando θ es un parámetro multidimensional, la matriz de información de Fisher está definida por

$$I_{ij}(\theta) = -E_{\theta} \left(\frac{\partial^2 \ln f(x|\theta)}{\partial \theta_i \partial \theta_j} \right)^2 \quad i, j = 1, \dots, k$$

y la distribución inicial no informativa de Jeffreys se define como

$$\pi^*(\theta) \propto \det \{I(\theta)\}^{\frac{1}{2}}.$$

Distribución de Referencia

Bernardo (1979) propone una extensión al método de Jeffreys, llamado *análisis de referencia*. La principal diferencia es que este método distingue entre parámetros de interés y parámetros de ruido. Por lo tanto, la distribución inicial resultante depende no sólo de la distribución muestral sino

también del problema inferencial en cuestión. A continuación se describirá la construcción de la distribución inicial de referencia.

Consideremos el modelo $f(x|\theta)$ y supongamos que la matriz de información de Fisher, $I(\theta)^{-1}$, existe. Sea $S = I(\theta)^{-1}$ y sean

$$\theta_{(1)} = (\theta_1, \dots, \theta_{n_1}), \dots, \theta_{(m)} = (\theta_{N_{m-1}+1}, \dots, \theta_k)$$

m grupos separados según su nivel de importancia inferencial, donde $N_i = \sum_{j=i}^m n_j$. El método produce una distribución inicial sobre $(\theta_{(1)}, \dots, \theta_{(m)})$, que distingue entre parámetros de ruido y parámetros de interés. Sea

$$\theta_{[j]} = (\theta_{(1)}, \dots, \theta_{(j)}) \quad y \quad \theta_{[\sim j]} = (\theta_{(j+1)}, \dots, \theta_{(m)}).$$

La matriz S se descompone de acuerdo a los m grupos y tiene la siguiente forma

$$S = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{21}^t & \dots & \mathbf{A}_{m1}^t \\ \mathbf{A}_{21} & \mathbf{A}_{22} & & \mathbf{A}_{m2}^t \\ & & \dots & \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & & \mathbf{A}_{mm} \end{pmatrix}$$

donde S_j es la matriz de orden $N_j \times N_j$ en la esquina superior izquierda de S , de manera que $S_1 = \mathbf{A}_{11}$. Sea $\mathbf{H}_j = S_j^{-1}$ y denotemos por h_j a la matriz de orden $n_j \times n_j$ en la esquina inferior derecha de \mathbf{H}_j , de manera que, en particular $h_1 = \mathbf{A}_{11}^{-1}$. Entonces la construcción de la distribución inicial de referencia es la siguiente:

Inicialización:

$$\pi_m(\theta_{(m)}|\theta_{[m-1]}) = \frac{|\mathbf{h}_m(\theta)|^{\frac{1}{2}}}{\int |\mathbf{h}_m(\theta)|^{\frac{1}{2}} d\theta_{(m)}}.$$

Iteración: Para $j = m-1, \dots, 1$

$$\pi_j(\theta_{[\sim j-1]}|\theta_{[j-1]}) = \frac{\pi_{j+1}(\theta_{[\sim j]}|\theta_{[j]}) \exp\{\frac{1}{2}\mathbb{E}_j[\ln(|h_j(\theta)|)|\theta_{[j]}]\}}{\int \exp\{\frac{1}{2}\mathbb{E}_j[\ln(|h_j(\theta)|)|\theta_{[j]}]\} d\theta_{[j]}}$$

donde $\mathbb{E}_j[g(\theta)|\theta_{[j]}] = \int g(\theta) \pi_{j+1}(\theta_{[\sim j]}|\theta_{[j]}) d\theta_{[\sim j]}$.

Conclusión: La distribución inicial de referencia es

$$\pi(\theta) = \pi_1(\theta_{[\sim 0]}|\theta_{[0]}).$$

A.3. Análisis de residuos

A.3.1. Métodos tradicionales

Un método simple y efectivo para detectar posibles deficiencias de un modelo de regresión es por medio de un examen de los residuos; ver, por ejemplo, Chatterjee y Price (1977). El i -ésimo residuo se define como

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n$$

donde y_i es una observación y \hat{y}_i es el correspondiente valor ajustado. Dado que un residuo puede verse como la desviación entre el dato y su valor ajustado, es una medida de variabilidad no explicada por el modelo de regresión. También se define el i -ésimo residuo estandarizado e_{is} , de la siguiente manera

$$e_{is} = e_i / s \quad i = 1, 2, \dots, n$$

donde $s^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2$.

Los residuos estandarizados tienen media cero y desviación estándar uno. Con una muestra moderadamente grande los residuos deberían estar aproximadamente distribuidos independientemente como una normal. Los residuos no están estrictamente distribuidos independientemente, pero con un gran número de observaciones, la falta de independencia puede ser ignorada.

El estudio de gráficas de residuos es una de las muchas herramientas en nuestro análisis. Podemos considerar el uso de estas gráficas para examinar diversas situaciones para el modelo de regresión simple con errores normales. Algunas de ellas son las siguientes:

1. La función de regresión no es lineal.
2. El término de error no tiene varianza constante.
3. Los errores no son independientes.
4. El modelo ajusta a la gran mayoría de las observaciones, sin embargo una o pocas observaciones "se disparan".
5. El término de error no está distribuido normalmente.
6. Una o varias variables independientes importantes han sido omitidas en el modelo.

Algunas de las gráficas usadas comúnmente para proveer información de los casos anteriores son aquellas en las cuales los residuos estandarizados son graficados contra:

1. El valor ajustado \hat{y} .
2. La variable independiente x .
3. El orden de tiempo en el cual las observaciones ocurren.
4. Variables independientes omitidas.

En general, cuando el modelo es correcto, los residuos, tienden a caer entre -2 y 2, estando distribuidos aleatoriamente alrededor de cero.

Cuando el modelo no es válido, las gráficas de residuos tienen características que son diferentes a las descritas. Se hace énfasis de que una parte esencial de cualquier análisis de regresión incluye un cuidadoso examen de residuos para asegurar que las suposiciones de la teoría se satisfacen.

Como una regla general, las gráficas de residuos brindan más información acerca de las deficiencias del modelo que las gráficas de los datos. Las gráficas de los residuos son herramientas útiles para detectar posibles violaciones a los supuestos del modelo, tales como falta de linealidad, varianza no constante (heteroscedasticidad), presencia de "outliers", y errores correlacionados.

Cuando se analiza un conjunto de datos es necesario estar seguros que las conclusiones no dependen solamente de uno o dos observaciones extremas. Los "outliers" pueden aparecer en el momento de recolectar o capturar los datos. Por otra parte, también pueden ser observaciones genuinas, altamente significativas y sugestivas para las cuales se deberá hacer un examen más detallado.

Transformación de variables

La necesidad de transformar los datos surge en situaciones donde la variable original, o el modelo en términos de la variable original, viola una o más de las suposiciones estándar.

Para satisfacer los supuestos usuales del modelo de regresión, algunas veces se trabaja con variables transformadas. Las transformaciones pueden ser necesarias por varias razones; incluyendo:

1. Consideraciones teóricas pueden especificar que la relación entre dos variables no es lineal. Una transformación apropiada de las variables puede hacer que la relación entre las variables transformadas sea lineal. Consideremos por ejemplo el modelo $T_i = \alpha\beta^i$, $\alpha > 0$, $0 < \beta < 1$, la relación entre T_i e i no es lineal, así que no podemos usar directamente las técnicas de regresión lineal; por otra parte, si nosotros tomamos logaritmos de ambos lados obtenemos: $\ln T_i = \ln \alpha + i \ln \beta$ mostrando que $\ln T_i$ e i están relacionados linealmente. La transformación nos capacita para usar métodos de regresión lineal estándar. Aunque la relación entre las variables originales no es lineal, la relación entre las variables transformadas sí lo es.
2. La variable dependiente y , puede tener una distribución de probabilidad cuya varianza esté relacionada con su media. Si la media está relacionada con el valor de la variable independiente x , entonces la varianza de y cambiará con x y la varianza no será constante. La distribución de y puede no ser normal bajo estas condiciones. La falta de normalidad invalida las pruebas estándar de significancia, ya que están basadas en el supuesto de normalidad. La varianza desigual de los errores produce estimadores insesgados pero no son adecuados en el sentido de precisión. En estas situaciones, es frecuente que se transformen los datos con el fin de lograr tanto la normalidad como la homocedasticidad que el ajuste requiere.
3. Existen razones no probabilísticas para suponer que las transformaciones son requeridas. La evidencia llega al examinar los residuos del ajuste de un modelo lineal.

Transformaciones para alcanzar linealidad

Uno de los supuestos más comunes en el análisis de regresión es que el modelo que describe los datos es lineal. Por consideraciones teóricas o para un examen de gráficas de dispersión de y contra x , la relación entre x y y puede no ser lineal. Hay, sin embargo, muchos modelos no lineales que bajo transformaciones apropiadas pueden ser lineales.

Transformaciones para estabilizar la varianza

Algunas transformaciones son también usadas para estabilizar la varianza del error, esto es, para hacer la varianza del error constante para todas

las observaciones. Que la varianza del error sea constante es uno de los supuestos estándar de la teoría de mínimos cuadrados (comúnmente llamado homocedasticidad). Cuando la varianza del error no es constante sobre todas las observaciones, el error mencionado es heteroscedástico. La heteroscedasticidad puede en ocasiones removerse por medio de una transformación adecuada. La variable dependiente y , en un problema de regresión, puede seguir una distribución de probabilidad cuya varianza es una función de la media de tal distribución. La binomial y la Poisson son dos distribuciones de probabilidad comunes que tienen esta característica. Sabemos, por ejemplo, que una variable que está distribuida binomialmente con parámetros n y p , tiene media np y varianza $np(1-p)$. Es también conocido que la media y la varianza de la distribución Poisson son iguales. Cuando la relación entre la media y la varianza de una variable aleatoria es conocida, es posible encontrar una transformación simple de la variable, la cual tendrá una varianza aproximadamente constante (estabiliza la varianza).

Por otro lado, en muchas aplicaciones de la industria, la economía y la biología, cuando se encuentra que la varianza del error no es constante, se sabe que la desviación estándar de los residuos se incrementa a medida que la variable exploratoria crece. Basados en su observación empírica, se tiene la hipótesis en este ejemplo de que la desviación estándar de los residuos es proporcional a X .

$$\text{Var}(\varepsilon_i) = k^2 x_i^2, \quad k > 0.$$

Consideremos el modelo

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}$$

el cual se obtiene al multiplicar ambos lados del modelo original por $1/x_i$. Definiendo un nuevo conjunto de variables y coeficientes

$$y' = \frac{y}{x}, \quad x' = \frac{1}{x}, \quad \beta_0' = \beta_0, \quad \beta_1 = \beta_1, \quad \varepsilon' = \frac{\varepsilon}{x}.$$

En términos de las nuevas variables, el modelo se reduce a

$$y'_i = \beta_0' + \beta_1 x'_i + \varepsilon'_i.$$

Nótese que para el modelo transformado, la varianza de ε'_i es constante e igual a k^2 .

Si el supuesto acerca del término del error está dado por $Var(\varepsilon_i) = k^2 x_i^2$, entonces para ajustar el modelo propiamente se debe trabajar con las variables transformadas: $y/x, 1/x$ como variables dependientes e independientes respectivamente. Si el modelo ajustado para los datos transformados es

$$\frac{y}{x} = a' + \frac{b'}{x}$$

entonces el modelo ajustado en términos de las variables originales es

$$y = b' + a'x.$$

A.3.2. Un método Bayesiano

Chaloner y Brant (1988) proponen una definición precisa de un *outlier* en un modelo lineal, la cual proporciona una forma simple de exploración de los datos para determinar existencia de un posible *outlier*. Su definición es tal que, si los parámetros del modelo fueran conocidos, entonces se sabría cuales observaciones son *outliers*. Alternativamente, si los parámetros son desconocidos la distribución final puede usarse para calcular la probabilidad final de que cualquier observación sea un *outlier*.

En un modelo lineal con errores aleatorios normalmente distribuidos, con media 0 y varianza σ^2 , la i -ésima observación es un *outlier* si $|\varepsilon_i| > k\sigma$, para alguna k . El valor de k puede ser seleccionado de manera tal que la probabilidad inicial de un *outlier* sea pequeña; así, estos *outliers* serán observaciones más extremas de que lo usualmente se espera. Recordemos que una de las definiciones más comunes de un *outlier* es "una observación que no ha sido generada por el mecanismo que generó a la mayoría de las observaciones en el conjunto de datos".

Un método de detección de *outliers*

Supongamos que el modelo bajo consideración es un modelo lineal, con parámetros $\theta^t = (\theta_1, \dots, \theta_p)$ normalmente distribuidos $N(0, \sigma^2)$, errores aleatorios independientes $\varepsilon^t = (\varepsilon_1, \dots, \varepsilon_n)$, X una matriz de $n \times p$ con observaciones x_1, \dots, x_n , y tal que

$$y = X\theta + \varepsilon.$$

Para calcular la probabilidad final de que $|\varepsilon_i|$ sea mayor que $k\sigma$, necesitamos la distribución final de ε , $p(\varepsilon | y)$. Esta distribución final se deriva

suponiendo una distribución inicial no-informativa, y puede derivarse de forma similar asumiendo una distribución inicial normal-gamma.

Sea $\tau = \sigma^{-2}$ y sea \mathbf{R} una matriz definida positiva. La distribución inicial conjugada normal-gamma lleva a que θ tenga una distribución normal condicional de τ con media m_0 y varianza $\tau^{-1}\mathbf{R}^{-1}$. La distribución inicial de τ es una distribución gamma con parámetros a y b , de la siguiente forma

$$p(\tau) = \tau^{a-1} e^{-b\tau}$$

con media inicial ab^{-1} . La distribución final de θ , condicional en τ es una distribución normal con media $m_1 = (\mathbf{R} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{R}m_0 + \mathbf{X}'y)$ y matriz de covarianza $\mathbf{V} = \tau^{-1}(\mathbf{R} + \mathbf{X}'\mathbf{X})^{-1}$. La distribución final de τ es una gamma con parámetros a_1 y b_1 , donde $a_1 = a + \frac{1}{2}n$ y $b_1 = b + \frac{1}{2}\{(y - \mathbf{X}m_1)'y + (m_0 - m_1)'\mathbf{R}m_0\}$.

La distribución final de ϵ se deriva fácilmente de la relación

$$\epsilon = y - \mathbf{X}\theta$$

y notando que ϵ es una función lineal de θ .

Sea $\mathbf{H} = \mathbf{X}(\mathbf{R} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$. Entonces la distribución final de ϵ condicional de τ , es una normal multivariada singular con media $\hat{\epsilon} = y - \mathbf{X}m_1$ y matriz de covarianza $\tau^{-1}\mathbf{H}$.

$$p(\epsilon | \tau) \sim N(\hat{\epsilon}, \mathbf{V}).$$

Sean h_{ij} los elementos de \mathbf{H} . Cada $\epsilon_i \sim St(\hat{\epsilon}_i, \frac{a_1}{(b_1 h_{ij})}, 2a_1)$. La matriz de covarianza de ϵ es proporcional a \mathbf{H} .

Para calcular la distribución final correspondiente a la distribución inicial no-informativa $p(\theta, \tau) = \tau^{-1}$, sea $\mathbf{R} \rightarrow 0$, $a = \frac{1}{2}p$ y $b \rightarrow 0$. Si denotamos

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y \quad y \quad s^2 = \frac{(y - \mathbf{X}\hat{\theta})'(y - \mathbf{X}\hat{\theta})}{n - p}$$

entonces la distribución final de ϵ es otra vez una distribución t multivariada p -dimensional con

$$\hat{\epsilon} = y - \mathbf{X}\hat{\theta} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} \quad a_1 = \frac{1}{2}(n - p) \quad b_1 = \frac{1}{2}(n - p)s^2.$$

Para determinar cuáles observaciones son *outliers*, se define la probabilidad p_i de la siguiente forma

$$\Pr(|\varepsilon_i| > k\sigma | y)$$

como la probabilidad final de que la i -ésima observación sea un *oullier*.

Sea $\Phi(z)$ la función de distribución normal estándar y

$$z_1 = \frac{(k - \hat{\varepsilon}_i\sqrt{\tau})}{\sqrt{h_{ii}}} \quad z_2 = \frac{(-k - \hat{\varepsilon}_i\sqrt{\tau})}{\sqrt{h_{ii}}} \quad (\text{A.9})$$

de donde

$$p_i = \Pr(|\varepsilon_i| > k\sigma | y) = \int \{1 - \Phi(z_1) + \Phi(z_2)\} p(\tau | y) d\tau.$$

Se pueden comparar estas p_i con la probabilidad inicial $2\Phi(-k)$. Los puntos con probabilidad final p_i alta, serán tales que $|\hat{\varepsilon}_i|$, h_{ii} , o ambos tendrán valores grandes. Si $|\hat{\varepsilon}_i|$ es grande, se induce que $|\varepsilon_i|$ también lo es. Por otra parte, si h_{ii} es grande, existe incertidumbre acerca de ε_i . A menudo, se hace referencia a la magnitud de h_{ii} como una medida de la influencia de la i -ésima observación.

Los residuos usuales $\hat{\varepsilon}_i$ son la media final de los ε_i y pueden pensarse como una estimación puntual de ε_i . Los intervalos de estimación de ε_i son fáciles de construir. Sea $t(\frac{1}{2}\alpha, n-p)$ el cuantil superior de orden $\frac{1}{2}\alpha$ de una distribución t con $n-p$ grados de libertad. Entonces el intervalo $(1-\alpha)$ para ε_i es

$$\hat{\varepsilon}_i \pm t(\frac{1}{2}\alpha, n-p)s\sqrt{h_{ii}}.$$

Estos intervalos pueden añadirse a la gráfica de los residuos $\hat{\varepsilon}_i$. Este tipo de gráficas pueden interpretarse como descripción de la distribución final marginal de ε_i , los "errores reales". Esta es una interpretación diferente a la del enfoque usual que ve a las gráficas de residuos como una representación de la distribución muestral de $\hat{\varepsilon}_i$.

La distribución final de los ε_i es bastante diferente a la distribución muestral de los $\hat{\varepsilon}_i$. Notemos que la distribución final de ε_i es p -dimensional, mientras que la distribución muestral de los $\hat{\varepsilon}_i$ es $(n-p)$ -dimensional. La distribución final de los ε_i considera a $\hat{\varepsilon}_i$ y $\hat{\theta}$ como fijos y la incertidumbre recae en θ . En cambio, la distribución muestral de $\hat{\varepsilon}_i$ tiene media 0 y matriz de covarianza proporcional a $(\mathbf{I} - \mathbf{H})$.

Selección de k

El valor de k puede seleccionarse de manera tal que la probabilidad de que un punto no sea un *outlier* sea grande, digamos 0.95. Así, $k = \Phi^{-1} \left\{ 0.5 + \frac{1}{2} (0.95)^{\frac{1}{n}} \right\}$, donde para $n = 20$, $k = 3$, para $n = 50$, $k = 3.3$, para $n = 100$, $k = 3.5$ y para $n = 1000$, $k = 4$. Cualquier observación con probabilidad final mayor que la probabilidad inicial $2\Phi(-k)$ es candidata a ser un *outlier*.

Alternativamente, si el modelo bajo consideración es requerido para describir los datos, en vez de ser considerado como un modelo estocástico, entonces puede utilizarse el valor $k = 2$ para determinar cuáles observaciones no son bien descritas, no importando el tamaño de la muestra. Si más del 5 % tiene probabilidad alta de ser más grandes que 2 desviaciones estándar, entonces esta es una causa para pensar que el modelo no describe adecuadamente el conjunto completo de datos.

A.4. Componentes de las funciones de verosimilitud para los MLG

En la Sección 2.4.2 se describió brevemente la forma de la función de verosimilitud para las principales familias exponenciales utilizadas en la construcción de los MLG. Aquí se presenta una descripción más detallada.

Recordemos que un elemento de la familia exponencial tiene una log-verosimilitud de la forma

$$l = \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}, \quad (\text{A.10})$$

donde θ es el parámetro canónico y ϕ es el parámetro de dispersión, el cual se asume conocido.

$l(\theta, \phi; y) = \ln f(y; \theta, \phi)$ es la función de log-verosimilitud considerada como una función de θ , para y y ϕ dadas.

Binomial

Sea $D \sim \text{Binomial}(n, \mu)$, y definimos $Y = \frac{d}{n}$, tenemos entonces $d = ny$ de tal modo que

$$l = ny \ln \left(\frac{\mu}{1-\mu} \right) + n \ln (1-\mu) + \ln \binom{n}{ny},$$

ya que

$$f(d) = \binom{n}{d} \mu^d (1-\mu)^{n-d},$$

la log-verosimilitud de $f(d)$ es

$$l = \ln \binom{n}{d} + \ln \mu^d + (n-d) \ln (1-\mu)$$

en términos de ny

$$\begin{aligned} l &= \ln \binom{n}{ny} + \ln \mu^{ny} + (n-ny) \ln (1-\mu) \\ &= ny \ln \mu + n(1-y) \ln (1-\mu) + \ln \binom{n}{ny} \\ &= \frac{y \ln \mu + (1-y) \ln (1-\mu)}{n^{-1}} + \ln \binom{n}{ny} \\ &= \frac{y \ln \mu + \ln (1-\mu) - y \ln (1-\mu)}{n^{-1}} + \ln \binom{n}{ny} \\ &= \frac{y (\ln \mu - \ln (1-\mu)) + \ln (1-\mu)}{n^{-1}} + \ln \binom{n}{ny} \\ &= \frac{y \ln \left(\frac{\mu}{1-\mu} \right) + \ln (1-\mu)}{n^{-1}} + \ln \binom{n}{ny} \end{aligned}$$

de modo que

$$\begin{aligned} \theta &= \ln \left(\frac{\mu}{1-\mu} \right), \\ b(\theta) &= \ln (1 + \exp \theta), \\ \phi &= n^{-1}, \\ c(y, \phi) &= \ln \binom{n}{ny}. \end{aligned}$$

Por tanto la distribución binomial pertenece a la familia exponencial.

Para ver a la distribución binomial como un elemento de la familia exponencial, la vemos en términos de la tasa.

$$\begin{aligned} Y &= \frac{d}{n} = \text{tasa de mortalidad.} \\ d &= \text{Sinistros.} \\ n &= \text{Expuestos.} \\ \mu &= \text{Probabilidad del Sinistro.} \end{aligned}$$

Normal

Sabemos que, para una v.a. que se distribuye normal $f(y | \mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$, donde μ es el parámetro de interés y σ^2 es visto como un parámetro de ruido. Esto puede escribirse de la siguiente forma

$$f(y | \mu, \sigma^2) = \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\},$$

ya que, de la ecuación (A.4),

$$\begin{aligned} f(y | \mu, \sigma^2) &= \exp\left\{\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right)\right\} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\ln\left((2\pi\sigma^2)^{-\frac{1}{2}}\right)\right\} \exp\left\{-\frac{1}{2\sigma^2}(y^2 - 2\mu y + \mu^2)\right\} \\ &= \exp\left\{-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} + \ln\left((2\pi\sigma^2)^{-\frac{1}{2}}\right)\right\} \end{aligned}$$

Agrupando y por propiedades de logaritmos, tenemos:

$$f(y | \mu, \sigma^2) = \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\},$$

por tanto la log-verosimilitud de la distribución normal, dada en la forma de la ecuación (A.10) es

$$l = \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]$$

así

$$\begin{aligned} \theta &= \mu \\ b(\theta) &= \frac{\mu^2}{2} \\ \phi &= \sigma^2 \\ c(y, \phi) &= -\frac{1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]. \end{aligned}$$

De tal forma que la distribución normal, es un elemento de la familia exponencial.

Poisson

La función de probabilidad de una variable aleatoria Poisson es

$$f(y | \lambda) = \frac{\lambda^y \exp(-\lambda)}{y!} \quad (\text{A.11})$$

lo anterior puede ser escrito de la siguiente manera

$$f(y | \lambda) = \exp \{y \ln \lambda - \lambda - \ln y!\}$$

de tal forma que pertenece a la familia exponencial, dado que

$$\begin{aligned} f(y | \lambda) &= \frac{\exp \{ \ln(\lambda^y) \} \exp \{-\lambda\}}{\exp \{ \ln(y!) \}} \\ &= \frac{\exp \{ \ln(\lambda^y) - \lambda \}}{\exp \{ \ln(y!) \}} \\ &= \exp \{ \ln(\lambda^y) - \lambda - \ln(y!) \} \\ &= \exp \{ y \ln(\lambda) - \lambda - \ln(y!) \}. \end{aligned}$$

Así, la log-verosimilitud es

$$l = y \ln(\lambda) - \lambda - \ln(y!)$$

donde

$$\begin{aligned} \theta &= \ln(\lambda) \\ b(\theta) &= \exp(\theta) \\ \phi &= 1 \\ c(y, \phi) &= -\ln(y!). \end{aligned}$$

Gamma

La función de probabilidad de una variable aleatoria Y con distribución gamma $Ga(\alpha, \beta)$ es,

$$f(y | \lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\} \quad y, \alpha, \beta \in \mathfrak{R}^+$$

donde lo anterior puede ser escrito de la siguiente manera:

$$f(y | \lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\{(\alpha - 1) \ln y - \beta y\}$$

así, la log-verosimilitud

$$l = -y\beta + \alpha \ln \beta + (\alpha - 1) \ln y - \ln \Gamma(\alpha)$$

$$l = (\{[y\theta - b(\theta)]/\phi\} + c(y, \phi))$$

$$\begin{aligned} \theta &= -\beta \\ b(\theta) &= -\alpha \ln \beta \\ \phi &= 1 \\ c(y, \phi) &= (\alpha - 1) \ln y - \ln \Gamma(\alpha). \end{aligned}$$

Apéndice B

Programas

B.1. Regresión lineal normal

B.1.1. Análisis preliminar del modelo lineal

Este programa fue realizado con el paquete estadístico S-Plus.

Datos

Lee los datos de mortalidad correspondientes a la operación del Seguro de Vida del Sector Asegurador en México para el periodo de 1982 a 1989, desde el archivo "Mort8289.txt", y se asignan a la variable "Datos".

```
Datos_matrix(scan('c:\\Mort8289.txt'),byrow=T,ncol=3)
```

Se asigna a las variables "edad", "expuestos" y "muertos" sus correspondientes valores, obtenidos anteriormente.

```
edad_Datos[,1]
```

```
expuestos_Datos[,2]
```

```
muertos_Datos[,3]
```

Debido a que en la edad 16 no se presentan siniestros, para este análisis se omite ese valor.

```
edad.ok_edad[-5]
```

```
expuestos.ok_expuestos[-5]
```

```
muertos.ok_muertos[-5]
```

Se obtienen las tasas brutas i.e. Muertos / Expuestos.

```
q.x_muertos.ok/expuestos.ok
```

Posteriormente, mediante la transformación logística ($y = \log \frac{q}{1-q}$) se obtiene una nueva variable de respuesta.

```
y.m_log(q.x)-log(1-q.x)
```

Ajuste

Con la ayuda de la función `glm`, se ajusta un modelo lineal generalizado vía máxima verosimilitud. En este caso se hace una regresión normal, pues se tiene una distribución Gaussiana y como función liga, la función identidad (que por defecto toma la función `glm`).

De esta manera se realiza el ajuste para los nuevos datos transformados (`y.m`).

```
Lin.norm.1_glm(edad.ok,y.m)
```

Se obtienen los valores ajustados mediante la matriz $X\hat{\beta}$, donde X es la matriz de variables independientes (`edad`) y $\hat{\beta}$ es el vector de parámetros estimados anteriormente.

```
y.fit.1_cbind(1,edad.ok)%*%Lin.norm.1$coef
```

Gráficas

Se grafican las proporciones transformadas

```
win.graph()
plot(edad.ok,y.m)
```

Posteriormente los valores ajustados

```
lines(edad.ok,y.fit.1)
```

Se grafica las tasas en su escala original

```
win.graph()
plot(edad.ok,q.x)
```

Y su ajuste, aplicando la transformación inversa a los valores ajustados

```
lines(edad.ok,exp(y.fit.1)/(1+exp(y.fit.1)))
```

Simula una muestra de la distribución predictiva de Y

Para realizar esta simulación es necesario una media y una varianza, para la distribución predictiva de Y', la cual es una t-Student.

Se propone una media, utilizando los estimadores de $\hat{\beta}_0$ y $\hat{\beta}_1$.

```
Media.pred.1_cbind(1,edad.ok)%*%Lin.norm.1$coef
```

Se propone una matriz de varianza.

```
Var.pred.1_Lin.norm.1$scale*diag(87) +
cbind(1,edad.ok)%*%Lin.norm.1$var%*%t(cbind(1,edad.ok))
```

Se simula una muestra de tamaño 5000 de una distribución Gaussiana para las tasas transformadas (Y's) mediante la función `rmultnorm`, con media `Media.pred.1`, y varianza `Var.pred.1`.

```
N_5000
```

```
Lin.norm.pred.1_rmultnorm(N,Media.pred.1,Var.pred.1)
```

Gráficas en la escala transformada

Se construye una matriz de 87 x 3

```
y.pred.quan.1_matrix(0,87,3)
```

la cual será llenada de la siguiente manera: para cada uno de los 87 renglones

- la primera columna será llenada con los cuantiles de 0.025
- la segunda columna será llenada con los cuantiles de 0.5
- la tercer columna será llenada con los cuantiles de 0.975

```
for(i in 1:87)
```

```
{y.pred.quan.1[i,]_c(quantile(Lin.norm.pred.1[,1],0.025),
mean(Lin.norm.pred.1[,1]),
quantile(Lin.norm.pred.1[,1],0.975))}
```

Se grafican las tasas transformadas (Y's), limitando el eje y

```
win.graph()
plot(edad.ok,y.m,ylim=c(min(y.pred.quan.1[,1]),
                        max(y.pred.quan.1[,3])))
```

Se grafica el ajuste con base en la mediana

```
lines(edad.ok,y.pred.quan.1[,2],lty=1)
```

Posteriormente se grafica una banda con los cuantiles de 0.975 y 0.025

```
lines(edad.ok,y.pred.quan.1[,3],lty=2)
lines(edad.ok,y.pred.quan.1[,1],lty=2)
```

Gráficas en la escala original

Mediante la transformación inversa, se obtiene los cuantiles 0.025, 0.5 y 0.975 para las tasas en su escala original

```
q.1_exp(y.pred.quan.1[,1])/(1+exp(y.pred.quan.1[,1]))
q.2_exp(y.pred.quan.1[,2])/(1+exp(y.pred.quan.1[,2]))
q.3_exp(y.pred.quan.1[,3])/(1+exp(y.pred.quan.1[,3]))
```

Para comodidad se construye una matriz con estos cuantiles

```
q.pred.quan.1_cbind(q.1,q.2,q.3)
```

Se construye la banda con base en la mediana, pero en esta ocasión en la escala original

```
win.graph()
plot(edad.ok,q.x,ylim=c(0,max(q.pred.quan.1[,3])))
lines(edad.ok,q.pred.quan.1[,2],lty=1)
lines(edad.ok,q.pred.quan.1[,3],lty=2)
lines(edad.ok,q.pred.quan.1[,1],lty=2)
```

Gráficas en la escala original

Mediante la transformación inversa se obtiene una muestra de tamaño 5000 predictiva para las tasas (q 's)

```
q.norm.pred.1_exp(Lin.norm.pred.1)/(1+exp(Lin.norm.pred.1))
```

Se construye una matriz de 87 x 3

```
q.quan.1_matrix(0,87,3)
```

la cual será llenada de la siguiente manera: para cada uno de los 87 renglones

- la primera columna será llenada con los cuantiles de 0.025
- la segunda columna será llenada con la media
- la tercer columna será llenada con los cuantiles de 0.975

```
for(i in 1:87)
{q.quan.1[i,]_c(quantile(q.norm.pred.1[,i],0.025),
               mean(q.norm.pred.1[,i]),
               quantile(q.norm.pred.1[,i],0.975))}
```

Se construye una banda con los cuantiles 0.975 y 0.025 con base en la media

```
win.graph()
plot(edad.ok,q.x,ylim=c(0,max(q.quan.1[,3])))
lines(edad.ok,q.quan.1[,3],lty=2)
lines(edad.ok,q.quan.1[,2],lty=1)
lines(edad.ok,q.quan.1[,1],lty=2)
```

Calcula la curva de mortalidad correspondiente al cuantil p

Se propone el cuantil de orden 0.70 para hacer una comparación con la tabla actual

```
p_0.70
```

```

tablas_function(p)
{
  q.quan_rep(-99,87)
  for(i in 1:87){q.quan[i]_quantile(q.norm.pred.1[,i],p)}
}

```

Se realiza una comparación con la tabla actual

Lee los datos de la Experiencia Mexicana

```
exmex_read.table('c:\DatExMex.txt',header=T)
```

Asigna a la variable "estimacion", los valores estimados utilizados por el sector asegurador

```
estimacion_exmex[,3]
```

Se omite la observación de edad 16, pues no se presentaron siniestros

```
estimacion.ok\_estimacion[-5]
```

Se grafica la comparación entre la tabla actual y la tabla producida por este análisis

```

win.graph()
plot(edad.ok,q.x,ylim=c(0,max(q.quan)))
lines(edad.ok,q.quan,lty=1)
lines(edad.ok,estimacion.ok,lty=2)

```

Se calcula el número de muertos que se esperaría con el cuantil de 0.70

```

muertos.p_sum(q.quan*expuestos.ok)
drop(muertos.p)
}
tablas(p)

```

Se calcula el número agregado de muertos al final del periodo

Se calcula una muestra de la distribución predictiva del número agregado de siniestros.

```
muertos.sam.1_q.norm.pred.1%*%expuestos.ok
```

Realiza el histograma del número agregado de siniestros.

```
win.graph()
```

```
hist(muertos.sam.1,probability=T,xlab='Muertos')
```

Dibuja la densidad del número agregado de siniestros.

```
lines(density(muertos.sam.1,from=19500,to=28500,width=3000))
```

Se calcula el "valor en riesgo" correspondiente al cuantil $\alpha = 0.975$

```
alfa = 0.975
```

```
v.r.1_quantile(muertos.sam.1,alfa)
```

```
abline(v=v.r.1)
```

Devuelve el valor del número de muertos

```
v.r.1
```

Se compara el valor anterior con el producido mediante el cuantil de 0.975

```
tablas(0.70)
```

B.1.2. Análisis del modelo lineal: WinBUGS

Análisis realizado con el programa WinBUGS

Se define el modelo lineal desde el enfoque Bayesiano.

Se define cada una de las N observaciones, con distribución normal.

Se propone el modelo $\mu_i \sim \alpha + \beta x_i$ para ajustar los datos

Asignamos distribuciones iniciales a los parámetros alpha, beta, y tau.

Ajuste

```

model
{
  for (i in 1:N){
    yp[i] ~ dnorm(mu[i],tau);
    mu[i] <- alpha + beta*x[i];
    expu[i] <- expuestos[i];
  }
  alpha ~ dnorm(0,1.0E-3);
  beta ~ dnorm(0,1.0E-3);
  tau ~ dgamma(1.0E-3,1.0E-3);
  sigma2 <- 1/tau;
}

```

Mediante observaciones simuladas de Y^* a partir del modelo ajustado, se utiliza la función inversa para contar con una muestra de la distribución predictiva de las tasas observadas de muerte q^* . De esta manera, se calcula la distribución del número agregado de muertos

```

for (i in 1:N){
  yp[i] ~ dnorm(mu[i],tau);
  logit(qp[i]) <- yp[i];
  muertos[i] <- qp[i]*expuestos[i];
}
mag <- sum(muertos[]);
}

```

Datos

Se declara el número de datos

```
list(N=87)
```

Se inicializan los valores de los parámetros

```
list(alpha=-8.717,beta=0.06692, tau=7.259)
```

DIC

Con el fin de comparar los modelos se calculó DIC (Criterio de información de la Devianza). El cual se define como 2 veces la esperanza de la devianza menos la devianza evaluada en la esperanza de los parámetros. Para el cálculo de la Devianza se realizó el siguiente procedimiento.

Ajuste

```
model
```

```
{
```

```
  for (i in 1:N){
```

```
    y[i] ~ dnorm(mu[i],tau);
```

```
    mu[i] <- alpha + beta*x[i];
```

```
    basura[i] <- expuestos[i];
```

```
  }
```

```
  alpha ~ dnorm(-8.717,10000);
```

```
  beta ~ dnorm(0.06692,10000);
```

```
  alf <- (7.259*7.259)/1.0E-10;
```

```
  bet <- 7.259/1.0E-10;
```

```
  tau ~ dgamma(alf,bet);
```

Cálculo de la devianza evaluada en la media de los parámetros

```

for (i in 1:N){
  res[i] <- y[i] - mu.gorro[i];
  sres[i] <- res[i]*sqrt(tau.gorro);
  lik[i] <- sqrt(tau.gorro / (2*3.14159)) *
    exp(-0.5*pow(sres[i],2));
  log.lik[i] <- log(lik[i]);
}
devianze.e <- -2.0*sum(log.lik[]);

```

Cálculo directo de la devianza Baycsiana

```

for (i in 1:N){
  resid[i] <- y[i] - mu[i];
  sresid[i] <- resid[i]*sqrt(tau);
  like[i] <- sqrt(tau / (2*3.14159)) *
    exp(-0.5*pow(sresid[i],2));
  log.like[i] <- log(like[i]);
}
devianze <- -2.0*sum(log.like[]);
dico <- 2*devianze - devianze.e
}

```

Datos

Se declara el número de datos, así como los valores iniciales para μ , y $\hat{\tau}$

```

list(N=87, mu.gorro=c(-7.914, -7.847, -7.78, -7.713, -7.58, -7.513,
-7.446, -7.379, -7.312, -7.245, -7.178, -7.111, -7.044, -6.977, -6.911,
-6.844, -6.777, -6.71, -6.643, -6.576, -6.509, -6.442, -6.375, -6.308,
-6.241, -6.175, -6.108, -6.041, -5.974, -5.907, -5.84, -5.773, -5.706,
-5.639, -5.572, -5.505, -5.439, -5.372, -5.305, -5.238, -5.171, -5.104,
-5.037, -4.97, -4.903, -4.836, -4.769, -4.703, -4.636, -4.569, -4.502,
-4.435, -4.368, -4.301, -4.234, -4.167, -4.1, -4.034, -3.967, -3.9,

```

B.1. REGRESIÓN LINEAL NORMAL

145

-3.833, -3.766, -3.699, -3.632, -3.565, -3.498, -3.431, -3.364, -3.298,
-3.231, -3.164, -3.097, -3.03, -2.963, -2.896, -2.829, -2.762, -2.695,
-2.628, -2.562, -2.495, -2.428, -2.361, -2.294, -2.227, -2.16, -2.093),
tau.gorro=7.259)

Se inicializan los valores de los parámetros

list(alpha=-8.717,beta=0.06692, tau=7.259)

B.1.3. Análisis preliminar del modelo cuadrático

Este programa fue realizado con el paquete estadístico S-Plus.

Datos

Lee los datos de mortalidad correspondientes a la operación del Seguro de Vida del Sector Asegurador en México para el periodo de 1982 a 1989, desde el archivo "Mort8289.txt", y se asignan a la variable "Datos".

```
Datos_matrix(scan("c:\\Mort8289.txt"), byrow=T, ncol=3)
```

Se asigna a las variables "edad", "expuestos" y "muertos" sus correspondientes valores, obtenidos anteriormente.

```
edad_Datos[,1]
expuestos_Datos[,2]
muertos_Datos[,3]
```

Debido a que en la edad 16 no se presentan siniestros, para este análisis se omite ese valor.

```
edad.ok_edad[-5]
expuestos.ok_expuestos[-5]
muertos.ok_muertos[-5]
```

Se obtienen las tasas brutas i.e. Muertos / Expuestos.

```
q.x_muertos.ok/expuestos.ok
```

Posteriormente, mediante la transformación logística ($y = \log \frac{q}{1-q}$) se obtiene una nueva variable de respuesta.

```
y.m_log(q.x)-log(1-q.x)
```

Ajuste

Con la ayuda de la función `glim`, se ajusta un modelo cuadrático generalizado vía máxima verosimilitud. En este caso se hace una regresión normal, pues se tiene una distribución Gaussiana y como función liga, la función identidad (que por defecto toma la función `glim`).

De esta manera se realiza el ajuste para los nuevos datos transformados (`y.m`).

```
Lin.norm.2_glim(cbind(edad.ok, edad.ok^2),y.m)
```

Se obtienen los valores ajustados mediante la matriz $X\hat{\beta}$, donde X es la matriz de variables independientes (`edad`) y $\hat{\beta}$ es el vector de parámetros estimados anteriormente.

```
y.fit.2_cbind(1,edad.ok,edad.ok^2)%*%Lin.norm.2$coef
```

Gráficas

Se grafican las proporciones transformadas

```
win.graph() plot(edad.ok,y.m)
```

Posteriormente los valores ajustados

```
lines(edad.ok,y.fit.2)
```

Se grafica las tasas en su escala original

```
win.graph()
plot(edad.ok,q.x)
```

Y su ajuste, aplicando la transformación inversa a los valores ajustados

```
lines(edad.ok,exp(y.fit.2)/(1+exp(y.fit.2)))
```

Simula una muestra de la distribución predictiva de Y

Para realizar esta simulación es necesario una media y una varianza, para la distribución predictiva de Y^* , la cual es una t-Student.

Se propone una media, utilizando los estimadores de $\hat{\beta}_0$ y $\hat{\beta}_1$.

```
Media.pred.2_cbind(1,edad.ok,edad.ok^2) %*% Lin.norm.2$coef
```

Se propone una matriz de varianza.

```
Var.pred.2_Lin.norm.2$scale*diag(87) +
  cbind(1,edad.ok,edad.ok^2) %*%
  Lin.norm.2$var %*%
  t(cbind(1,edad.ok,edad.ok^2))
```

Se simula una muestra de tamaño 5000 de una distribución Gaussiana para las tasas transformadas (Y's) mediante la función `rmultnorm`, con media `Media.pred.2`, y varianza `Var.pred.2`.

```
N_5000
```

```
Lin.norm.pred.2_rmultnorm(N,Media.pred.2,Var.pred.2)
```

Gráficas en la escala transformada

Se construye una matriz de 87 x 3

```
y.pred.quant.2_matrix(0,87,3)
```

la cual será llenada de la siguiente manera: para cada uno de los 87 renglones

- la primera columna será llenada con los cuantiles de 0.025
- la segunda columna será llenada con los cuantiles de 0.5
- la tercer columna será llenada con los cuantiles de 0.975

```
for(i in 1:87)
  {y.pred.quant.2[i,]_c(quantile(Lin.norm.pred.2[,1],0.025),
                      mean(Lin.norm.pred.2[,1]),
                      quantile(Lin.norm.pred.2[,1],0.975))}
```

Se grafican las tasas transformadas (Y's), limitando el eje y

```
win.graph()
plot(edad.ok,y.m,ylim=c(min(y.pred.quant.2[,1]),
                        max(y.pred.quant.2[,3])))
```

Se grafica el ajuste con base en la mediana

```
lines(edad.ok,y.pred.quant.2[,2],lty=1)
```

Posteriormente se grafica una banda con los cuantiles de 0.975 y 0.025

```
lines(edad.ok,y.pred.quant.2[,1],lty=2)
```

```
lines(edad.ok,y.pred.quant.2[,3],lty=2)
```

Gráficas en la escala original

Mediante la transformación inversa, se obtiene los cuantiles 0.025, 0.5 y 0.975 para las tasas en su escala original

```
q.1_exp(y.pred.quant.2[,1])/(1+exp(y.pred.quant.2[,1]))
```

```
q.2_exp(y.pred.quant.2[,2])/(1+exp(y.pred.quant.2[,2]))
```

```
q.3_exp(y.pred.quant.2[,3])/(1+exp(y.pred.quant.2[,3]))
```

Para comodidad se construye una matriz con estos cuantiles

```
q.pred.quant.2_cbind(q.1,q.2,q.3)
```

Se construye la banda con base en la mediana, pero en esta ocasión en la escala original

```
win.graph()
```

```
plot(edad.ok,q.x,ylim=c(0,max(q.pred.quant.2[,3])))
```

```
lines(edad.ok,q.pred.quant.2[,2],lty=1)
```

```
lines(edad.ok,q.pred.quant.2[,1],lty=2)
```

```
lines(edad.ok,q.pred.quant.2[,3],lty=2)
```

Gráficas en la escala original

Mediante la transformación inversa se obtiene una muestra de tamaño 5000 predictiva para las tasas (q's)

```
q.norm.pred.2_exp(Lin.norm.pred.2)/(1+exp(Lin.norm.pred.2))
```

Se construye una matriz de 87 x 3

```
q.quant.2_matrix(0,87,3)
```

la cual será llenada de la siguiente manera: para cada uno de los 87 renglones

- la primera columna será llenada con los cuantiles de 0.025
- la segunda columna será llenada con la media
- la tercer columna será llenada con los cuantiles de 0.975

```
for(i in 1:87)
{q.quant.2[i,]_c(quantile(q.norm.pred.2[,i],0.025),
                mean(q.norm.pred.2[,i]),
                quantile(q.norm.pred.2[,i],0.975))}
```

Se construye una banda con los cuantiles 0.975 y 0.025 con base en la media

```
win.graph()
plot(edad.ok,q.x,ylim=c(0,max(q.quant.2[,3])))

lines(edad.ok,q.quant.2[,2],lty=1)
lines(edad.ok,q.quant.2[,1],lty=2)
lines(edad.ok,q.quant.2[,3],lty=2)
```

Calcula la curva de mortalidad correspondiente al cuantil p

Se propone el cuantil de orden 0.70 para hacer una comparación con la tabla actual

```
p_0.70

tablas_function(p)
{
q.quan2_rep(-99,87)
  for(i in 1:87){q.quan2[i]_quantile(q.norm.pred.2[,i],p)}
```

Se realiza una comparación con la tabla actual

Lee los datos de la Experiencia Mexicana

```
exmex_read.table("c:\\DatExMex.txt",header=T)
```

Asigna a la variable "estimacion", los valores estimados utilizados por el sector asegurador

```
estimacion_exmex[,3]
```

Se omite la observación de edad 16, pues no se presentaron siniestros

```
estimacion.ok_estimacion[-5]
```

Se grafica la comparación entre la tabla actual y la tabla producida por este análisis

```
win.graph()
plot(edad.ok,q.x,ylim=c(0,max(q.quan2)))
lines(edad.ok,q.quan2,lty=1)
lines(edad.ok,estimacion.ok,lty=3)
muertos2.p_sum(q.quan2*expuestos.ok)
drop(muertos2.p)
}
tablas(p)
```

Se calcula el número agregado de muertos al final del periodo

Se calcula una muestra de la distribución predictiva del número agregado de siniestros.

```
muertos.sample.2_q.norm.pred.2%*expuestos.ok
```

Realiza el histograma del número agregado de siniestros.

```
win.graph()
hist(muertos.sample.2,probability=T,xlab="Muertos")
```

Dibuja la densidad del número agregado de siniestros.

```
lines(density(muertos.sample.2,from=19500,to=32000,width=3000))
```

Se calcula el "valor en riesgo" correspondiente al cuantil $\alpha = 0.975$.

```
alfa_0.975
v.r.i_quantile(muertos.sample.2,alfa)
aline(v=v.r.i)
v.r.i
```

B.1.4. Análisis del modelo cuadrático: WinBUGS

Análisis realizado con el programa WinBUGS

Se define el modelo cuadrático desde el enfoque Bayesiano.

Se define cada una de las N observaciones, con distribución normal.

Se propone el modelo $\mu_i \sim \alpha + \beta x_i + \omega x_i^2$ para ajustar los datos

Asignamos distribuciones iniciales a los parámetros alpha, beta, y tau.

Ajuste

```

model
{
    for(i in 1:N){
        y[i]~dnorm(mu[i],tau);
        mu[i]<-alpha + beta*x[i] + omega*x[i]*x[i]
        basura[i] <- expuestos[i]
    }
    alpha~dnorm(0,1.0E-3);
    beta~dnorm(0,1.0E-3);
    omega~dnorm(0,1.0E-3);
    tau~dgamma(1.0E-3,1.0E-3);
    sigma2<-1/tau;

```

Mediante observaciones simuladas de Y^* a partir del modelo ajustado, se utiliza la función inversa para contar con una muestra de la distribución predictiva de las tasas observadas de muerte q^* . Y con ello se calcula la distribución del número agregado de muertos

```

    for (i in 1:N) {
        yp2[i]~dnorm(mu[i],tau);
        logit(qp2[i])<-yp2[i];
        muertos2[i]<-qp2[i]*expuestos[i]
    }
    mag2<-sum(muertos2[])
}

```

Datos

Se declara el número de datos

`list(N=87)`

Se inicializan los valores de los parámetros

```
list(alpha=-9.120451,beta= 0.08510127,omega=-0.0001628853,  
tau= 7.675863630607, yp2=c(-8.122691, -8.041662, -7.960958,  
-7.880581, -7.720803, -7.641403, -7.562328, -7.483579,  
-7.405156, -7.327059, -7.249288, -7.171842, -7.094722,  
-7.017928, -6.941460, -6.865317, -6.789500, -6.714009,  
-6.638844, -6.564004, -6.489491, -6.415303, -6.341441,  
-6.267904, -6.194694, -6.121809, -6.049250, -5.977016,  
-5.905109, -5.833527, -5.762271, -5.691341, -5.620736,  
-5.550457, -5.480505, -5.410877, -5.341576, -5.272600,  
-5.203950, -5.135626, -5.067628, -4.999956, -4.932609,  
-4.865588, -4.798893, -4.732523, -4.666479, -4.600761,  
-4.535369, -4.470303, -4.405562, -4.341147, -4.277058,  
-4.213295, -4.149858, -4.086746, -4.023960, -3.961500,  
-3.899365, -3.837557, -3.776074, -3.714916, -3.654085,  
-3.593580, -3.533400, -3.473546, -3.414017, -3.354815,  
-3.295938, -3.237387, -3.179162, -3.121263, -3.063689,  
-3.006441, -2.949519, -2.892923, -2.836652, -2.780707,  
-2.725088, -2.669795, -2.614827, -2.560186, -2.505870,  
-2.451880, -2.398215, -2.344877, -2.291864))
```

DIC

Con el fin de comparar los modelos se calculó DIC (Criterio de información de la Devianza). El cual se define como 2 veces la esperanza de la devianza menos la devianza evaluada en la esperanza de los parámetros. Para el cálculo de la Devianza se realizó el siguiente procedimiento.

Ajuste

```

model
{
  for (i in 1:N){
    y[i] ~ dnorm(mu[i],tau);
    mu[i]<-alpha + beta*x[i] + omega*x[i]*x[i]
    basura[i] <- expuestos[i];
  }
  alpha ~ dnorm(-8.717,10000);
  beta ~ dnorm(0.06692,10000);
  alf <- (7.259*7.259)/1.0E-10;
  bet <- 7.259/1.0E-10;
  tau ~ dgamma(alf,bet);

```

Cálculo de la devianza evaluada en la media de los parámetros

```

for (i in 1:N){
  res[i] <- y[i] - mu.gorro[i];
  sres[i] <- res[i]*sqrt(tau.gorro);
  lik[i] <- sqrt(tau.gorro / (2*3.14159)) *
    exp(-0.5*pow(sres[i],2));
  log.lik[i] <- log(lik[i]);
}
devianze.e <- -2.0*sum(log.lik[]);

```

Cálculo directo de la devianza Bayesiana

```

for (i in 1:N) {
  resid[i] <- y[i] - mu[i];
  sresid[i] <- resid[i]*sqrt(tau);
  like[i] <- sqrt(tau / (2*3.14159)) *
    exp(-0.5*pow(sresid[i],2));
  log.like[i] <- log(like[i]);
}
devianze <- -2.0*sum(log.like[]);
dico<- 2*devianze - devianze.e
}

```

Datos

Se declara el número de datos, así como los valores iniciales para $\hat{\mu}$, y $\hat{\tau}$

```

list(N=87,mu.gorro=c(-8.117,-8.037,-7.956,-7.876,-7.717,-7.638,
-7.559,-7.481,-7.402,-7.325,-7.247,-7.17,-7.093,-7.016,
-6.94,-6.864,-6.789,-6.713,-6.638,-6.564,-6.489,-6.415,
-6.342,-6.268,-6.195,-6.123,-6.05,-5.978,-5.906,-5.835,
-5.764,-5.693,-5.622,-5.552,-5.482,-5.413,-5.344,-5.275,
-5.206,-5.138,-5.07,-5.002,-4.935,-4.868,-4.802,-4.735,
-4.669,-4.604,-4.538,-4.473,-4.408,-4.344,-4.28,-4.216,
-4.153,-4.09,-4.027,-3.964,-3.902,-3.84,-3.779,-3.718,
-3.657,-3.596,-3.536,-3.476,-3.416,-3.357,-3.298,-3.239,
-3.181,-3.123,-3.065,-3.008,-2.951,-2.894,-2.838,-2.782,
-2.726,-2.67,-2.615,-2.561,-2.506,-2.452,-2.398,-2.344,
-2.291),tau.gorro=7.668)

```

B.2. Modelos lineales generalizados

B.2.1. Análisis preliminar del modelo lineal

Realizado con el paquete estadístico S-Plus.

Análisis preliminar basado en una aproximación normal asintótica a la distribución final de los parámetros.

Suponemos que para cada edad x_i , $\mu_i \sim \text{Binomial}(q, n)$.

Datos

Lee los datos desde el archivo "Mort8289.txt", y los asigna a la variable "Datos".

```
datos_matrix(scan('c:\Mort8289.txt'),byrow=T,ncol=3)
```

Se asigna a las variables "edad", "expuestos" y "muertos" sus correspondientes valores, leídos anteriormente.

```
edad_datos[,1]
```

```
expuestos_datos[,2]
```

```
muertos_datos[,3]
```

Ajuste

La función `glim`, ajusta un modelo lineal generalizado vía máxima verosimilitud. En este caso con un modelo binomial dada la naturaleza de los datos, la correspondiente liga `logit` y como escala se utiliza la devianza.

```
Ajuste1_glim(edad,muertos,expuestos,error='binomial',
             link='logit',scale='deviance')
```

Se asigna a las variables "Coef.ajust", "Varianza", los coeficientes ajustados y la matriz de covarianza respectivamente.

```
Coef.ajust_Ajuste1$coef
```

```
Varianza_Ajuste1$var
```

Se simula una muestra de tamaño 5000 de una distribución Gaussiana mediante la función `rmulnorm`, con media `Coef.ajust` y varianza `Varianza`.

```
N_5000
```

```
muestra.coef_rmultnorm(N,Coef.ajust,Varianza)
```

Se obtiene una muestra de las tasas transformadas Y 's

```
muestra.logit.q_muestra.coef%*t(cbind(1,edad))
```

Mediante la función inversa se obtiene una muestra de las tasas de mortalidad

```
muestra.q_exp(muestra.logit.q)/(1+exp(muestra.logit.q))
```

Se construyen dos matrices de 5000 x 88 y otra de 88 x 4.

```
q.pred_matrix(0,N,88)
```

```
m.pred_matrix(0,N,88)
```

```
q.quantil_matrix(0,88,4)
```

Se obtiene una muestra predictiva de los muertos y de las tasas

```
for(i in 1:88)
```

```
{m.pred[,i]_rbinom(rep(1,N),rep(expuestos[i],N),muestra.q[,i])
  q.pred[,i]_m.pred[,i]/expuestos[i]}
```

```
for(i in 1:88)
```

```
{q.quantil[i,]_c(quantile(q.pred[,i],0.025),
  median(q.pred[,i]),
  quantile(q.pred[,i],0.975),
  mean(q.pred[,i]))}
```

Gráficas

Se construye una banda predictiva del 95% de confianza, basándose en la mediana

```
win.graph()
q.x_(muertos/expuestos)
plot(edad,q.x,ylim=c(0,0.2))
lines(edad,q.quantil[,2],lty=1)
lines(edad,q.quantil[,1],lty=2)
lines(edad,q.quantil[,3],lty=2)
```

Se contruye una banda predictiva del 95% de confianza, basándose en la media

```
win.graph()
q.x_(muertos/expuestos)
plot(edad,q.x,ylim=c(0,0.2))
lines(edad,q.quantil[,4],lty=1)
lines(edad,q.quantil[,1],lty=2)
lines(edad,q.quantil[,3],lty=2)
```

Calcula la curva de mortalidad correspondiente al cuantil p

```
p_0.80
tablas1_function(p)
{
  q.quant_rep(-99,88)

  for(i in 1:88){q.quant[i]_quantile(q.pred[,i],p)}
```

Se realiza una comparación de la tabla actual
Lee los datos de la Experiencia Mexicana

```
exmex_read.table('c:\\DatExMex.txt',header=T)
```

Asigna a la variable estimacion, los valores estimados utilizados por el sector asegurador

```
estimacion_exmex[,3]
```

Se grafica la comparación entre la tabla actual y la tabla producida por este análisis

```
win.graph()
plot(edad,q.x,ylim=c(0,max(q.quant)))
lines(edad,q.quant,lty=1)
lines(edad,estimacion,lty=3)
muertos.p_sum(q.quant*expuestos)
drop(muertos.p)
}
tablas1(p)
```

Número agregado de muertos al final del periodo

Se calcula una muestra de la distribución predictiva del número agregado de siniestros.

```
muertos.agre.1_sumc(t(m.pred))
```

Realiza el histograma del número agregado de siniestros

```
win.graph()
hist(muertos.agre.1,probability=T,xlab='Muertos')
```

Dibuja la densidad del número agregado de siniestros

```
lines(density(muertos.agre.1,from=22000,to=26000,width=1000))
```

Calcula el "valor en riesgo" correspondiente al cuantil $\alpha = 0.975$

```
alfa_0.975
v.r.1_quantile(muertos.agre.1,alfa)
abline(v=v.r.1)
v.r.1
```

B.2.2. Análisis del modelo lineal: WinBUGS

Análisis realizado con el programa WinBUGS

Modelo logístico lineal con errores beta-binomiales, con parámetro de dispersión común.

Para este análisis se omiten las observaciones que parecen ser outliers

Se define el modelo lineal desde el enfoque Bayesiano. Con el fin de reducir la correlación entre los parámetros del modelo bajo la distribución final se utilizó un modelo equivalente

Se propone el modelo $\ln\left(\frac{q_x}{1-q_x}\right) = \alpha + \beta_1(x - \bar{x})$ para ajustar los datos, donde $\alpha = \beta_0 + \beta_1 \bar{x}$.

Asignamos distribuciones iniciales a los parámetros alpha, beta, y tau.

Ajuste

model

{

```

    for (i in 1:N){
        r[i] ~ dbin(theta[i],n[i]);
        theta[i] ~ dbeta(a[i],b[i]);
        a[i] <- lam*p[i];
        b[i] <- lam*(1.0-p[i]);
        logit(p[i]) <- alph0 + beta * (x[i] - xbar);
    }

```

```

    alpha <- alph0 + beta*xbar;
    alph0 ~ dnorm(0,1.0E-3);

```

```

    beta ~ dnorm(0,1.0E-3);
    lam ~ dunif(0.1,10);

```

```

    for (k in 1:N) {
        ee[k] ~ dbin(theta[k],n[k])
        rp[k] <- ee[k]/n[k]
    }

```

```

    magre<-sum(ee [ ]);

```

}

Datos

Se declara el número de datos, así como \bar{x}

```
list(N = 84, xbar=55.5)
```

Se inicializan los valores de los parámetros

```
list(alpha= -4.92, beta=0.073, lam=5.0, theta= c(0.0002856735,  
0.0001789122, 0.0003829584, 0.00040295, 0.0004347286, 0.0004880165,  
0.0007350059, 0.0005561514, 0.0006894691, 0.0005359908, 0.0004962380,  
0.0008041817, 0.0008664236, 0.0007299566, 0.0011047322, 0.0009868510,  
0.0011565571, 0.0012151225, 0.0011930492, 0.0011438246, 0.0012503106,  
0.0012515842, 0.0013283975, 0.0013979676, 0.0015032118, 0.0017696499,  
0.0018409897, 0.0022292074, 0.0020323844, 0.0024076667, 0.0028437944,  
0.0033454939, 0.0033922042, 0.0038005014, 0.0041133244, 0.0044383942,  
0.0045180606, 0.0047827752, 0.0058925477, 0.0054506184, 0.0073675101,  
0.0077585665, 0.0077744577, 0.0078639973, 0.0091735355, 0.0101058252,  
0.0111321844, 0.0125621974, 0.0117029963, 0.0116954214, 0.0153137450,  
0.0134713556, 0.0155697971, 0.0179283826, 0.0165368612, 0.0181776213,  
0.0215356418, 0.0200448658, 0.0215867676, 0.0238795004, 0.0282019955,  
0.0259369287, 0.0242205617, 0.0245231608, 0.0303504380, 0.0325367647,  
0.0266918873, 0.0387307513, 0.0297446696, 0.0335195531, 0.0421174652,  
0.0569550931, 0.0477419355, 0.0661417323, 0.0517799353, 0.0379464286,  
0.0445103858, 0.0721649485, 0.0806451613, 0.0625000000, 0.0495049505,  
0.1500000000, 0.0468750000, 0.1428571429))
```

B.2.3. Análisis preliminar del modelo cuadrático

Realizado con el paquete estadístico S-Plus.

Análisis preliminar basado en una aproximación normal asintótica a la distribución final de los parámetros.

Suponemos que para cada edad x_i , $\mu_i \sim \text{Binomial}(q, n)$.

Datos

Lee los datos desde el archivo "Mort8289.txt", y los asigna a la variable "Datos".

```
datos_matrix(scan("c:\\Mort8289.txt"), byrow=T, ncol=3)
```

Se asigna a las variables "edad", "expuestos" y "muertos" sus correspondientes valores, leídos anteriormente.

```
edad_datos[,1]
expuestos_datos[,2]
muertos_datos[,3]
```

Ajuste

La función glm, ajusta un modelo lineal generalizado vía máxima verosimilitud. En este caso con un modelo binomial dada la naturaleza de los datos, la correspondiente liga logit y como escala se utiliza la devianza.

```
Ajuste2_glim(cbind(edad, edad^2), muertos, expuestos,
  error="binomial", link="logit", scale="deviance")
```

Se asigna a las variables "Coef.ajust2", "Varianza2", los coeficientes ajustados y la matriz de covarianza respectivamente.

```
Coef.ajust2_Ajuste2$coef
Varianza2_Ajuste2$var
```

Se simula una muestra de tamaño 5000 de una distribución Gaussiana mediante la función rmulnorm, con media Coef.ajust2 y varianza Varianza2 .

```
N_5000
muestra.coef2_rmulnorm(N, Coef.ajust2, Varianza2)
```

Se obtiene una muestra de las tasas transformadas Y^* 's

```
muestra.logit.q2_muestra.coef2*%t(cbind(1,edad,edad^2))
```

Mediante la función inversa se obtiene una muestra de las tasas de mortalidad

```
muestra.q2_exp(muestra.logit.q2)/(1+exp(muestra.logit.q2))
```

Se construyen dos matrices de 5000 x 88 y otra de 88 x 4.

```
q.pred2_matrix(0,N,88)
```

```
m.pred2_matrix(0,N,88)
```

```
q.quantil2_matrix(0,88,4)
```

Se obtiene una muestra predictiva de los muertos y de las tasas

```
for(i in 1:88){
  m.pred2[,i]_rbinom(rep(1,N),rep(expuestos[i],N),muestra.q2[,i])
  q.pred2[,i]_m.pred2[,i]/expuestos[i]}

```

```
for(i in 1:88)
  {q.quantil2[i,]_c(quantile(q.pred2[,i],0.025),
                  median(q.pred2[,i]),
                  quantile(q.pred2[,i],0.975),
                  mean(q.pred2[,i]))}

```

```
q.x_(muertos/expuestos)
```

Gráficas

Se construye una banda predictiva del 95% de confianza, basándose en la mediana

```
win.graph()
plot(edad,q.x,ylim=c(0,0.2))
lines(edad,q.quantil2[,2],lty=1)
lines(edad,q.quantil2[,1],lty=2)
lines(edad,q.quantil2[,3],lty=2)
```

```

win.graph()
plot(edad,q.x,ylim=c(0,0.2))
lines(edad,q.quant112[,4],lty=1)
lines(edad,q.quant112[,1],lty=2)
lines(edad,q.quant112[,3],lty=2)

```

Calcula la curva de mortalidad correspondiente al cuantil p

```

p_0.80
tablas2_function(p)
{
  q.quant_rep(-99,88)
  for(i in 1:88){q.quant[i]-quantile(q.pred2[,i],p)}
}

```

Se realiza una comparación de la tabla actual

Lee los datos de la Experiencia Mexicana

```
exmex_read.table("c:\\DatExMex.txt",header=T)
```

Asigna a la variable "estimacion", los valores estimados utilizados por el sector asegurador

```
estimacion_exmex[,3]
```

Se grafica la comparación entre la tabla actual y la tabla producida por este análisis

```

win.graph()
plot(edad,q.x,ylim=c(0,max(q.quant)))
lines(edad,q.quant,lty=1)
lines(edad,estimacion,lty=3)

```

Detalle para las edades de 10 a 60 años

```

exmex_read.table("c:\\DatExMex.txt",header=T)
estimacion_exmex[,3]
win.graph()
plot(edad,q.x,xlim=c(5,65),ylim=c(0,0.05))

```

```
lines(edad,q.quant,lty=1)
lines(edad,estimacion,lty=3)
muertos.p_sum(q.quant*expuestos)
drop(muertos.p)
}
tablas2(p)
```

Número agregado de muertos al final del periodo

Se calcula una muestra de la distribución predictiva del número agregado de siniestros.

```
muertos.agre.2_sumc(t(m.pred))
Realiza el histograma del número agregado de siniestros
win.graph()
hist(muertos.agre.2,probability=T,xlab="Muertos")
Dibuja la densidad del número agregado de siniestros
lines(density(muertos.agre.2,from=23500,to=24500,width=300))
Calcula el "valor en riesgo" correspondiente al cuantil  $\alpha = 0.975$ 
alfa_0.975
v.r.2_quantile(muertos.agre.2,alfa)
abline(v=v.r.2)
v.r.2
```

B.2.4. Análisis del modelo cuadrático: WinBUGS

Análisis realizado con el programa WinBUGS

Modelo logístico cuadrático con errores beta-binomiales, con parámetro de dispersión común.

Para este análisis también se omiten las observaciones que parecen ser outliers

Se define el modelo cuadrático desde el enfoque Bayesiano.

Asignamos distribuciones iniciales a los parámetros alpha, beta, omega y tau.

Ajuste

model

```
{
  for (i in 1:N){
    r[i] ~ dbin(theta[i],n[i]);
    theta[i] ~ dbeta(a[i],b[i]);
    a[i] <- lam*p[i];
    b[i] <- lam*(1.0-p[i]);
    logit(p[i]) <- alpha + beta * x[i]* x[i];
  }
  alpha ~ dnorm(0,1.0E-3);
  beta ~ dnorm(0,1.0E-3);
  omega ~ dnorm(0,1.0E-3);
  lam ~ dunif(0.1,10);
  for (k in 1:N) {
    ee[k] ~ dbin(theta[k],n[k])
    rp[k] <- ee[k]/n[k]
  }
  magre<-sum(ee[ ]);
}
```

Datos

Se declara el número de datos

```
list(N = 84)
```

Se inicializan los valores de los parámetros

```
list(alpha=-10.05, beta=0.1147, omega=-3.727E-4, lam=5.0,  
theta=c(0.0002856735, 0.0001789122, 0.0003829584, 0.00040295, 0.0004347286,  
0.0004880165, 0.0007350059, 0.0005561514, 0.0006894691, 0.0005359908,  
0.0004962380, 0.0008041817, 0.0008664236, 0.0007299566, 0.0011047322,  
0.0009868510, 0.0011565571, 0.0012151225, 0.0011930492, 0.0011438246,  
0.0012503106, 0.0012515842, 0.0013283975, 0.0013979676, 0.0015032118,  
0.0017696499, 0.0018409897, 0.0022292074, 0.0020323844, 0.0024076667,  
0.0028437944, 0.0033454939, 0.0033922042, 0.0038005014, 0.0041133244,  
0.0044383942, 0.0045180606, 0.0047827752, 0.0058925477, 0.0054506184,  
0.0073675101, 0.0077585665, 0.0077744577, 0.0078639973, 0.0091735355,  
0.0101058252, 0.0111321844, 0.0125621974, 0.0117029963, 0.0116954214,  
0.0153137450, 0.0134713556, 0.0155697971, 0.0179283826, 0.0165368612,  
0.0181776213, 0.0215356418, 0.0200448658, 0.0215867676, 0.0238795004,  
0.0282019955, 0.0259369287, 0.0242205617, 0.0245231608, 0.0303504380,  
0.0325367647, 0.0266918873, 0.0387307513, 0.0297446696, 0.0335195531,  
0.0421174652, 0.0569550931, 0.0477419355, 0.0661417323, 0.0517799353,  
0.0379464286, 0.0445103858, 0.0721649485, 0.0806451613, 0.0625000000,  
0.0495049505, 0.1500000000, 0.0468750000, 0.1428571429))
```


Bibliografía

- [1] Alatorre, R.M. (1991). Tabla de Mortalidad "Experiencia Mexicana 1982-1989". *Documento de Trabajo* No. 2. Dirección General de Desarrollo e Investigación, Comisión Nacional de Seguros y Fianzas.
- [2] Batten, R.W. (1978). *Mortality Table Construction*. Englewood Cliffs: Prentice Hall.
- [3] Baxter, L.A., Coultts, S.M. and Ross, G.A.F. (1980). Applications of linear models in motor insurance. In *Proc. 21st Int. Congr. Actuaries*, Vol. 2, pp. 11-29. Bern: Stämpfi.
- [4] Bennett, M. (1978). Models in motor insurance. *J. Inst. Act. Student Soc.*, 22, 87-148.
- [5] Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society B*, 41, 113-147.
- [6] Bernardo, J.M. y Smith. A.F.M. (1994). *Bayesian Theory*. New York: Wiley.
- [7] Bishop, Y.M.M., Fienberg, S.E. y Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.
- [8] Broffitt, J.D. (1984). Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates (with discussion). *Transactions of the Society of Actuaries* 36, 77-142.
- [9] Chaloner, K. y Brant, R. (1988). A Bayesian Approach to Outlier Detection and Residual Analysis. *Biometrika*, 75, 651-659.
- [10] Chatterjee, S. y Price B. (1977). *Regression Analysis By Examples*. New York: Wiley.

- [11] Coutts, S. M. (1984). Motor insurance rating, an actuarial approach. *J. Inst. Act.*, 111, 87-148.
- [12] DeGroot, M. H. (1988). *Probabilidad y Estadística*. México: Addison-Wesley Iberoamericana.
- [13] Dempster, A. P. (1974). The direct use of likelihood for significance testing. In *Proc. Conf. Foundational Questions in Statistical Inference*, (eds. O. Barndorff-Nielsen, P. Blaesild and G. Schou), pp. 335-352. Aarhus: University of Aarhus.
- [14] Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall.
- [15] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.
- [16] Grayling, S. (1997). *VaR: Understanding and Applying Value-at-Risk*. London: Risk Publications.
- [17] Grimes, T. (1971). Claim frequency analysis in motor insurance. *J. Inst. Act. Student Soc.*, 19, 147-154.
- [18] Gutiérrez-Peña, E. (1997). Métodos Computacionales en la Inferencia Bayesiana. Series Monografías, Vol.6, No. 15. México: IIMAS, UNAM.
- [19] Gutiérrez-Peña, E. (1998). Análisis Bayesiano de Modelos Jerárquicos Lineales. Series Monografías, Vol.7, No. 16. México: IIMAS, UNAM.
- [20] Haberman, S. y Renshaw, A.E. (1996). Generalized Linear Models and Actuarial Science. *The Statistician*, 45, 407-436.
- [21] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London (Series A)* 186, 453-461.
- [22] Jeffreys, H. (1961). **Theory of Probability** (3a. edición). Oxford University Press, Oxford.
- [23] Johnson, P.D. and Hey, G.B. (1971). Statistical studies in motor insurance (with discussion). *J. Inst. Act.*, 97, 199-232.

- [24] Kass, R. E. y Wasserman, L. (1996). Formal rules of selecting prior distribution: a review and annotated bibliography. *J. Amer. Statist. Assoc.* **91**, 343-1370.
- [25] O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference*, Vol. 2B. London: Arnold.
- [26] McCullagh, P. y Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall (2a. Edición).
- [27] Mendenhall, W. Schaffer R.L. y Wackerly, D.D. (1986). *Estadística Matemática con Aplicaciones*. México: Iberoamérica.
- [28] Mendoza, M., Madrigal, A. M. y Martínez, E. (1999). Modelos Estadísticos de Mortalidad: Análisis de Datos 1991-1998. *Documento de Trabajo No. 77*. México: Comisión Nacional de Seguros y Fianzas.
- [29] Migon, H.S. y Gamerman, D. (1999). *Statistical Inference: an Integrated Approach*. London: Arnold.
- [30] Mood, A.M., Graybill, F.A. y Boes, D.C. (1974). *Introduction to the Theory of Statistics*. Singapore: McGraw-Hill.
- [31] Nelder, J.A. y Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370-84.
- [32] Ripley, D.B. (1987). *Stochastic Simulation*. New York: Wiley.
- [33] Robert, C.P. (2001). *The Bayesian Choice*. New York: Springer.
- [34] Smith, A. F. y Gelfand, A. E. (1992). Bayesian Statistic without Tears: a Sampling-resampling perspective. *American Statistician*, **46**, 84-88.
- [35] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. y van der Linde, A. (2002). Bayesian Measure of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society B*, **64**, 1-34.
- [36] Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996). BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5, (version ii). MRC Biostatistics Unit, Cambridge.

Software

- **BUGS (Bayesian Inference Using Gibbs Sampling)**

Disponible en

<http://www.mrc-bsu.cam.ac.uk/bugs/>

- **S-PLUS**

Disponible en

<http://insightful.com/>