Tarificación y provisiones

(Segunda edición)

Eva Boj del Val, M. Mercè Claramunt Bielsa y Teresa Costa Cor

Departamento de Matemática Económica, Financiera y Actuarial.
Universidad de Barcelona.

14 de marzo de 2018

Prefacio

Esta segunda edición de la publicación constituye un material de apoyo para la impartición y estudio de la asignatura Matemática Actuarial del Máster de Ciencias Actuariales y Financieras de la Universidad de Barcelona (UB). Tiene su origen en la experiencia de las autoras dentro de la asignatura Matemática Actuarial No Vida de la ya extinta Licenciatura de segundo ciclo en Ciencias Actuariales y Financieras de la UB y recoge parte del material incluido en la publicación que se utilizaba en dicha asignatura, Claramunt, M.M. y Costa, T. (2003). *Matemática Actuarial No Vida. Un enfoque práctico.* Colección de Publicaciones del Departamento de Matemática Económica, Financiera y Actuarial de la Universidad de Barcelona, n. 63.

La publicación se estructura en dos partes. La primera, con dos capítulos, se dedica a los sistemas de tarificación, profundizando en los sistemas conocidos como Bonus-Malus que cumplen las propiedades necesarias para ser analizados mediante cadenas de Markov. La segunda, incluye el cálculo de provisiones en los seguros no-vida, haciendo hincapié en los modelos estocásticos que proporcionan no sólo una estimación puntual de dichas provisiones, sino también del error que lleva asociado y en determinados casos, incluso permiten obtener toda la distribución para dichas provisiones. En todos los temas, la publicación incluye, junto con los desarrollos teóricos, numerosos ejemplos prácticos solucionados analíticamente y con el lenguaje de programación `R` ([R Development Core Team (2018)])), utilizando paquetes ya existentes y programas de elaboración propia. También se incluyen enunciados de ejercicios sin resolver.

Índice general

Índice de tablas

Capítulo 1

Sistemas de tarificación

La tarificación es el proceso de determinación de primas de tal forma que se correspondan con los siniestros a pagar. Por ello, la tarificación constituye una actividad clave del negocio asegurador.

Se entiende por tarifas a las tablas o cuadros que contienen las primas comerciales de los distintos riesgos, así como las normas de aplicación. Los principios técnicos en que se basa la elaboración de una tarifa constituyen el sistema de tarificación.

El objetivo de un sistema de tarificación es que las primas que se calculen sean equitativas a cada riesgo, teniendo en cuenta la solvencia y solidaridad entre los asegurados. La equidad implica que cada asegurado pague según el riesgo que comporta. Esto conlleva tener en cuenta los factores de riesgo más significativos, es decir, los que expliquen mejor la siniestralidad. La solvencia implica que las primas deben ser suficientes, es decir, deben permitir la rentabilidad de la empresa aseguradora con una estabilidad a largo plazo. Por último, la solidaridad implica un reparto del riesgo y de la prima total, de manera que los asegurados que incorporan menos riesgo paguen más de lo que estrictamente les correspondería y compensen una prima menor a pagar por los que incorporan más riesgo. En determinadas ocasiones estos objetivos de equidad y solidaridad pueden entrar en conflicto, como ocurre en los seguros de automóviles respecto de las motocicletas.

Los sistemas de tarificación pueden clasificarse en:

- *A priori* o *Class-rating*. Se establecen primas diferentes por clases. Se denominan *a priori* ya que nos permitirá asignar una prima a una póliza o riesgo que se incorpora en nuestra cartera sin tener experiencia sobre la siniestralidad de ese riesgo en concreto, excepto la proporcionada en el caso del seguro de automóviles por el Fichero SINCO (Fichero de siniestralidad de conductores, con experiencia de los últimos cinco años). Únicamente conociendo determinadas características de la póliza determinaremos su prima asignándole una siniestralidad esperada.
- *A posteriori* o *Experience rating*. Podemos entenderla en un sentido estricto, por oposición a la tarificación *a priori*, como aquella que presupone la existencia de una prima inicial que se va modificando en base a la experiencia de siniestralidad de aquel individuo (póliza), para dar lugar a las primas de los períodos sucesivos. Da lugar a los sistemas *Bonus-Malus*. En un sentido amplio podría entenderse como la actualización de las tarifas mediante la incorporación de nueva información.

En todo lo que sigue nos referiremos a la prima pura, sin recargos por gastos de gestión interna ni externa. La misma se calculará como la esperanza de la siniestralidad total, coste total en un período, que se reduce al producto del número esperado de siniestros por la cuantía esperada de un siniestro. De esta forma, la variable objeto de estudio detallado en el sistema de tarificación puede ser el coste total en un período, el número de siniestros o la cuantía de un siniestro. En adelante, en las explicaciones de los distintos métodos, cuando hablemos de *prima* nos estaremos refiriendo a la esperanza de la variable concreta objeto de estudio.

1.1. Tarificación *a priori*

En el proceso de tarificación *a priori* podemos distinguir tres fases, más una inicial:

- Fase 0: Datos iniciales. Homogeneización y depuración de los datos. Debemos tener presente que la aplicación de un método estadístico-matemático sofisticado de análisis de datos a datos de poca calidad es una pérdida de tiempo. Ello implica tener en cuenta, entre muchas otras las siguientes cuestiones:
    - La inclusión de los siniestros que ocurrieron en períodos pasados pero que se han declarado ahora y de los siniestros ocurridos en el período objeto de análisis pero que no se han declarado o no se conoce totalmente su importe.
    - Si se utilizan observaciones de diversos años, debemos eliminar los cambios interanuales.
    - Ver si incluimos o no las pólizas que no han estado vigentes todo el año (o todo el período para el cual tenemos los datos).
- Fase 1: Selección de las variables tarificadoras y de las clases de tarifa. Definimos previamente lo que se entiende por factor de riesgo. Los factores de riesgo son características medibles que habremos observado y que tendrán una posible relación de causa con la siniestralidad objeto de estudio. Podrán hacer referencia tanto a las características del objeto asegurado como a características del asegurado, del tomador, etc.

  Las variables tarificadoras serán aquellos factores de riesgo que el asegurador utiliza para distinguir entre los riesgos, entre los asegurados, por ser las que explican una parte importante de la siniestralidad. Dependiendo de la variable objeto de estudio podemos tener variables tarificadoras que afecten al número de siniestros, al coste de un siniestro o al coste total.

  Las clases de tarifa se definen como los niveles que distinguiremos dentro de cada variable tarificadora. La determinación de las clases de tarifa no es sencilla pues, por un lado, es bueno que sean amplias para que cada clase sea lo más significativa posible, es decir incluya un número elevado de observaciones, y por otro, las clases de tarifa tienen que ser lo más estrechas posible para proporcionar un mejor ajuste.
- Fase 2: Obtención de los grupos de tarifa. Formación de grupos homogéneos de riesgo, exclusivos y exhaustivos, a partir de las clases de tarifa.
- Fase 3: Estimación de las primas.

Estas fases no son independientes, de forma que hay modelos o métodos que sirven para todas o varias al mismo tiempo. Esquematizamos a continuación las principales técnicas que se utilizan dentro del proceso de tarificación *a priori*:

- Selección de las variables y de las clases de tarifa: Análisis *cluster*, análisis discriminante, técnicas de segmentación.
- Selección de las variables y de las clases de tarifa más estimación de las primas: regresión lineal, modelos lineales generalizados (siendo el modelo de Bailey y Simon un caso particular).
- Estimación de las primas: método de Bailey y Simon, método de los mínimos cuadrados, métodos de la teoría de la credibilidad (Bühlmann-Straub, *Two-way* y generalizaciones, ...)

Algunos de estos modelos se aplican sobre datos con clasificación cruzada. En ellos la estructura de los datos es tal que todas las clases de una variable tarificadora se subdividen, respecto de otra variable tarificadora, exactamente de la misma forma.

Pueden establecerse, también, a la hora de estimar las primas, estructuras aditivas o multiplicativas, de forma que la prima correspondiente a un grupo se calcule como suma de los efectos de pertenecer a cada una de las clases o como producto.

Un buen análisis sobre los los métodos de tarificación *a priori* puede encontrarse en [Boj et al. (2004)] en castellano, así como en [Van Eeghen et al. (1983)]. Dedicamos el capítulo 2 al estudio del Modelo Lineal Generalizado (MLG) (ver [McCullagh and Nelder (1989)]) y su aplicación a la tarificación *a priori*. El MLG es la metodología más extendida de tarificación *a priori*. En el capítulo 2 realizamos un repaso teórico del MLG debido a que también lo aplicaremos para el cálculo de provisiones técnicas en el capítulo 4.

1.2. Tarificación *a posteriori*

Vamos a considerar el *experience-rating* en su sentido más estricto como aquel conjunto de métodos que incorporan la información individual de su siniestralidad para proporcionar correcciones *a posteriori* sobre la tarificación inicial.

En los procesos de tarificación *a priori* se determinan una serie de factores de riesgo y niveles de los mismos con el objetivo de clasificar a los individuos en grupos más homogéneos respecto al riesgo. Así, se cobra a todos los individuos del grupo la misma prima común. Sin embargo, dentro de cada grupo no se ha eliminado totalmente la heterogeneidad, los riesgos no son totalmente homogéneos, debido a varios factores (factores de riesgo insuficientes, factores inobservables, etc) y por tanto si aplicamos a todos los individuos del grupo la tasa media de siniestralidad común habrán algunos beneficiados y otros perjudicados, provocando a largo plazo una antiselección. Una manera de conseguir que cada asegurado pague una prima de acuerdo con su riesgo es introducir modificaciones *a posteriori* según la experiencia que vaya teniendo.

Dentro del esquema general de *experience-rating* podemos distinguir, entre otros, los siguientes grupos de métodos:

- Reembolso de beneficios por baja siniestralidad: *profit return* o *premium refund.*
- Teoría de la credibilidad de las fluctuaciones limitadas.
- Teoría de la credibilidad de la máxima precisión:
    - Credibilidad exacta.

  - Credibilidad de distribución libre.

- Sistemas *Bonus-Malus* markovianos.

En el capítulo 3 desarrollamos los sistemas *Bonus-Malus* markovianos.

Capítulo 2

Modelo lineal generalizado: Tarificación *a priori*

Dedicamos este capítulo a una descripción general del MLG y mostramos su aplicación en tarificación *a priori* mediante un ejemplo con `R`.

2.1. El Modelo Lineal Generalizado

2.1.1. Descripción del modelo

En este apartado vamos a describir las características básicas del MLG. Como referencia básica nos referimos a [McCullagh and Nelder (1989)] donde se puede encontrar con detalle su descripción y características.

Supongamos la variable aleatoria $\mathbf{Y}_{(n\times1)}$, (y_i) para $i = 1, 2, \ldots, n$ observaciones de la respuesta independientes, que recogen la siniestralidad a explicar en el modelo. Supongamos P predictores o factores potenciales de la estructura de riesgo $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_P$, vectores $(n \times 1)$: (f_{ij}) para $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, P$. Recordemos el modelo clásico de regresión lineal por mínimos cuadrados ordinarios, en el que suponemos una distribución del error ε_i Normal centrada y con varianza constante, $\varepsilon_i \sim N\left(0, \sigma^2\right)$. La relación lineal de la respuesta con la estructura sistemática dada por los predictores es:

$$y_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij} + \varepsilon_i.$$

Disponemos de observaciones independientes $y_i \sim N\left(\mu_i, \sigma^2\right)$, con esperanza

$$E\left[y_i\right] = \mu_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij},$$

y varianza constante

$$Var\left[y_i\right] = \sigma^2.$$

En el MLG seguimos teniendo (y_i) para $i = 1, 2, \ldots, n$ observaciones independientes de la respuesta, y un predictor lineal determinista al que simbolizamos por η_i:

$$\eta_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij}.$$

Las dos extensiones respecto al modelo clásico son:

1. La distribución de $\mathbf{Y}$ no tiene porqué ser la Normal. Consideramos en primer lugar que puede ser cualquier distribución derivada de la familia exponencial de McCullagh y Nelder (ver [McCullagh and Nelder (1989)] pp.28). Estas distribuciones se caracterizan por tener una función de densidad o de probabilidad en un punto de la forma:

$$f(y_i; \theta_i, \phi_i) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right\},$$

para funciones especificadas $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$, donde θ_i se denomina parámetro canónico y ϕ_i parámetro de dispersión. Puede deducirse fácilmente a partir de la formulación más general que:

$$E[y_i] = \mu_i = b'(\theta_i),$$
$$Var[y_i] = b''(\theta_i)\, a(\phi_i).$$

Por lo tanto la varianza es el producto de dos componentes:

- la primera, $b''(\theta_i)$, que depende únicamente de θ_i (y por tanto de la esperanza μ_i). A esta componente se le denomina función de varianza y se explicita su dependencia respecto de la esperanza: $b''(\theta_i) = V(\mu_i)$,
- la segunda, $a(\phi_i)$, depende sólo del parámetro de dispersión ϕ_i y usualmente adopta la forma $a(\phi_i) = \frac{\phi}{w_i}$, con parámetro de dispersión constante para todas las observaciones, ϕ, y unos pesos especificados *a priori*, w_i, que varían de observación a observación.

  Teniendo en cuenta lo anterior podemos rescribir:

$$Var[y_i] = \frac{\phi V(\mu_i)}{w_i},$$

  donde ϕ es el parámetro de dispersión, $V(\mu_i)$ es la función de varianza y w_i es el posible peso especificado *a priori* de la observación i. Notamos que suponiendo $\phi = 1$, los recíprocos de los pesos pueden reinterpretarse como parámetros de escala no constantes: $1/w_i = \phi_i$.

2. La respuesta está ligada con el predictor lineal a través de una función F:

$$E[y_i] = \mu_i = F(\eta_i) = F\left(\beta_0 + \sum_{j=1}^{P} \beta_j f_{ij}\right).$$

Para despejar el predictor lineal debemos hacer la función inversa de F, $g = F^{-1}$, a la que denominamos *función de enlace* (*link function*), pues es la que nos enlaza la respuesta con el predictor lineal. A la función de enlace, g, le exigimos que sea monótona y diferenciable:

$$g(E[y_i]) = g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij}.$$

Existen, para algunas distribuciones de la familia exponencial, funciones de enlace "naturales", también denominadas canónicas. Para estos enlaces canónicos se produce que el parámetro canónico coincide con el predictor lineal: $\theta(\mu_i) = \eta_i$. Pero, en general, podemos optar por modelizar con cualquier otro enlace que no sea el canónico. Es usual utilizar uno derivado de la familia de enlaces paramétricos:

$$\eta_i = g(\mu_i) = \begin{cases} \mu_i^\lambda, \lambda \neq 0 \\ \log(\mu_i), \lambda = 0. \end{cases}$$

En la Tabla 2.1 se recogen las distribuciones más conocidas que forman parte de la familia exponencial junto con el enlace canónico asociado.

Función de enlace Identidad y Logarítmica

En el MLG obtenemos un modelo aditivo si combinamos cualquier distribución del error con el enlace identidad, $\lambda = 1$:

$$\mu_i = \eta_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij},$$

y obtenemos un modelo multiplicativo si utilizamos el enlace logarítmico, $\lambda = 0$:

$$\log(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{P} \beta_j f_{ij},$$

$$\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{P} \beta_j f_{ij}\right) = e^{\beta_0} \times e^{\beta_1 f_{i1}} \times e^{\beta_2 f_{i2}} \times \cdots \times e^{\beta_p f_{iP}}.$$

Observamos que si partimos de la base β_0 en el modelo aditivo debemos sumar (o restar) las P cantidades $\sum\limits_{j=1}^{P} \beta_j f_{ij}$. Y si partimos de la base e^{β_0} en el multiplicativo, debemos multiplicar por las P cantidades de incremento (o decremento) $\prod\limits_{j=1}^{P} e^{\beta_j f_{ij}}$. En el caso particular de predictores binarios, partiendo del efecto global, sumaremos β_j o multiplicaremos por e^{β_j}, sólo cuando para el individuo i se dé la característica j, para $j = 1, \ldots, P$, ya que en tal caso se cumplirá $f_{ij} = 1$.

Al igual que con la función de enlace, dónde podemos trabajar con la familia paramétrica anterior, también podemos utilizar para la distribución del error la familia paramétrica

$$V(\mu_i) = \mu_i^\zeta. \tag{2.1}$$

Observamos (ver Tabla 2.1) que con esta familia de distribuciones del error obtenemos algunos casos particulares de la familia exponencial, por ejemplo:

- Si $\zeta = 1$, la estructura de error es Poisson.
- Si $\zeta = 2$, la estructura de error es Gamma.
- Si $\zeta = 3$, la estructura de error es Gaussiana inversa.

Tabla 2.1: Principales características para casos particulares del MLG.

	Gaussiana $N\left(\mu,\sigma^2\right)$	Binomial $B(m,\pi)/m$	Poisson $P(\mu)$	Gamma $G(\mu,\nu)$	Gaussiana inversa $GI\left(\mu,\sigma^2\right)$
Rango(y)	$(-\infty,+\infty)$	$0,1,...,m/m$	$0,1,2,...,\infty$	$(-\infty,+\infty)$	$(-\infty,+\infty)$
w	1	1	1	1	1
ϕ	σ^2	$1/m$	1	υ^{-1}	σ^2
$b(\theta)$	$\theta^2/2$	$\log\left(1+e^\theta\right)$	$\exp(\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y;\theta)$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\log(2\pi\phi)\right)$	$\log\begin{pmatrix} m \\ my \end{pmatrix}$	$-\log y!$	$\nu\log(\nu y)-\log(y)$ $-\log\Gamma(\nu)$	$-\frac{1}{2}\left(\log\left(2\pi\phi y^3\right)+\frac{1}{\phi y}\right)$
$\mu(\theta)=E(Y;\theta)$	θ	$\frac{e^\theta}{(1+e^\theta)}$	$\exp(\theta)$	$\frac{-1}{\theta}$	$(-2\theta)^{-1/2}$
$\theta(\mu)$	μ	$\log\left(\mu/(1-\mu)\right)$	$\log(\mu)$	$1/\mu$	$1/\mu^2$
$V(\mu)$	1	$\mu(1-\mu)$	μ	μ^2	μ^3

2.2. Tarificación *a priori* con modelo lineal generalizado

2.2.1. Cálculo de la prima pura

A continuación recordamos los aspectos básicos de cálculo de la prima pura para poder enlazar con la aplicación del MLG y los métodos de tarificación *a priori*. Para el detalle de todo el proceso hasta la estimación de primas nos referimos a [Boj et al. (2004)] en castellano y a [Van Eeghen et al. (1983)], [Brockman and Wright(1992)] y [Haberman and Renshaw A.E. (1996)] entre otros.

Establecemos la nomenclatura:

n = total de pólizas de la cartera.

m = total de celdas en las que hemos agregado.

e_u = expuestos o número de pólizas de la celda u para $u=1,...,m$, por lo que $\sum_{u=1}^{m} e_u = n$.

N_u = variable aleatoria número de siniestros de la celda u para $u=1,...,m$.

n_u = dato sobre el número de siniestros empírico de la celda u para $u=1,...,m$. Realización de N_u.

S_u = variable aleatoria coste total de los siniestros de la celda u para $u=1,...,m$.

s_u = dato sobre el coste total empírico de los siniestros de la celda u para $u=1,...,m$. Realización de S_u.

N_{iu} = variable aleatoria número de siniestros de la póliza i de la celda u para $i=1,\dots,e_u$, $u=1,...,m$, por lo que $N_u=\sum_{i=1}^{e_u} N_{iu}$.

X_{iu} = variable aleatoria coste del siniestro i de la celda u para $i=1,\dots,N_u$, $u=1,...,m$, por lo que $S_u = \sum_{i=1}^{N_u} X_{iu}$.

La variable aleatoria **frecuencia de siniestralidad** o número medio de siniestros por póliza en la celda u, se define como

$$Y_u = \frac{N_u}{e_u},$$

junto con la ponderación $w_u = e_u$. Empíricamente la obtenemos como: $\frac{n_u}{e_u}$.

Los MLG mayormente estudiados y propuestos en la bibliografía actuarial para la frecuencia de siniestralidad son los de distribución Poisson (sobre-dispersa o no) y Binomial Negativa, combinada generalmente con la función de enlace logarítmica, sin tener por qué ser estas combinaciones las que mejor se ajusten a unos datos determinados. Sus características son:

Poisson:

$$\mu_u = E\left[Y_u\right],\ V\left(\mu_u\right) = \mu_u,\ \phi = 1,\ w_u = e_u.$$

Binomial Negativa:

$$\mu_u = E\left[Y_u\right],\ V\left(\mu_u\right) = \left(1 + \frac{\mu_u}{h_u}\right)\mu_u,\ \phi = 1,\ w_{ij} = e_u.$$

Poisson sobre-disperso:

$$\mu_u = E\left[Y_u\right],\ V\left(\mu_u\right) = \mu_u,\ \phi > 1,\ w_u = e_u.$$

La variable aleatoria **media del coste de un siniestro** para las pólizas de la celda u, se define como

$$Y_u = \frac{S_u}{n_u},$$

junto con la ponderación, $w_u = n_u$. Empíricamente la obtenemos como: $\frac{s_u}{n_u}$.

Los MLG mayormente estudiados y propuestos en la bibliografía actuarial para el coste medio de un siniestro son los de distribución Gamma y Gaussiana inversa, combinada generalmente con la función de enlace logarítmica, sin tener por qué ser estas combinaciones las que mejor se ajusten a unos datos determinados, al igual que ocurre con la frecuencia de siniestralidad. Sus características, teniendo en cuenta la familia 2.1, son:

$$\mu_u = E\left[Y_u\right],\ V\left(\mu_u\right) = \mu_u^{\zeta},\ \hat{\phi},\ w_u = n_u,$$

siendo para la distribución Gamma $\zeta = 2$ y para la Gaussiana inversa $\zeta = 3$.

Finalmente, si aplicamos tarificación *a priori* simultáneamente respecto del número de siniestros y de la cuantía de un siniestro, la prima pura, P_u, se calcularía como el producto de la esperanza de ambos MLG:

$$P_u = E\left[\frac{N_u}{e_u}\right] E\left[\frac{S_u}{n_u}\right].$$

2.2.2. Ejemplo con `R`

Ilustramos la estimación del MLG en aplicación a tarificación *a priori* haciendo uso de la función `glm` del paquete `stats` de `R`. Los datos, *carinsuk*, del ejemplo que vamos a realizar suelen utilizarse para ilustrar el uso del GLM en aplicación a tarificación y hacen referencia a la variable aleatoria coste medio por siniestro. Los datos son originales de la referencia [Baxter et al. (1980)] y se encuentran disponibles en el paquete `forward` de `R`, [Konis et al. (2012)], para su libre uso. A continuación preparamos los datos y estimamos dos modelos. Uno con distribución Gamma y su enlace canónico, el inverso, y otro con distribución Gamma y enlace logarítmico, que daría como resultado una tarifa multiplicativa.

```
R> library(forward)
R> data(carinsuk)
R> carinsuk
    OwnerAge Model CarAge NClaims AvCost
1      17-20     A    0-3       8    289
2      17-20     A    4-7       8    282
3      17-20     A    8-9       4    133
4      17-20     A    10+       1    160
5      17-20     B    0-3      10    372
6      17-20     B    4-7      28    249
7      17-20     B    8-9       1    288
8      17-20     B    10+       1     11
9      17-20     C    0-3       9    189
10     17-20     C    4-7      13    288
11     17-20     C    8-9       1    179
12     17-20     C    10+       0     NA
13     17-20     D    0-3       3    763
...     ...     ...   ...      ...   ...

R> # 'data.frame': 128 observaciones de 5 variables

R> # $OwnerAge: Factor, 8 levels:
R> # 17-20 21-24 25-29 30-34 35-39 40-49 50-59 60+

R> # $Model: Factor, 4 levels:
R> # A B C D

R> # $CarAge: Factor, 4 levels:
R> # 0-3 4-7 8-9 10+

R> # $NClaims: Integer
R> # $AvCost: Integer

R> # 128 observaciones - 5 missings = 123 observaciones

R> carinsuk <- carinsuk[-c(12,15,16,32,80),]
R> nrow (carinsuk)
```

```
[1] 123

# Respuesta y pesos:

R> y <- carinsuk$AvCost
R> w <- carinsuk$NClaims

R> # Gamma(link = "inverse")

R> glm1 <- glm(y~OwnerAge+CarAge+Model,family=Gamma(link = "inverse"),
 data=carinsuk, weights=w); glm1

Call:  glm(formula = y ~ OwnerAge + CarAge + Model,
family = Gamma(link = "inverse"), data = carinsuk, weights = w)

Coefficients:
  (Intercept)  OwnerAge21-24  OwnerAge25-29  OwnerAge30-34  OwnerAge35-39
    0.0034104      0.0001014      0.0003500      0.0004623      0.0013705

  OwnerAge40-49  OwnerAge50-59    OwnerAge60+      CarAge10+      CarAge4-7
      0.0009695      0.0009164      0.0009200      0.0041537      0.0003663

    CarAge8-9         ModelB         ModelC         ModelD
    0.0016525      0.0000377     -0.0006136     -0.0014205

Degrees of Freedom: 122 Total (i.e. Null);  109 Residual
Null Deviance:      649.9
Residual Deviance: 124.7        AIC: 84700


R> # Gamma(link = "log")

R> glm2 <- glm(y~OwnerAge+CarAge+Model,family=Gamma(link = "log"),
data=carinsuk, weights=w); glm2

Call:  glm(formula = y ~ OwnerAge + CarAge + Model,
family = Gamma(link = "log"), data = carinsuk, weights = w)

Coefficients:
  (Intercept)  OwnerAge21-24  OwnerAge25-29  OwnerAge30-34  OwnerAge35-39
    5.6586543      0.0023502     -0.0621003     -0.1120241     -0.3140591

  OwnerAge40-49  OwnerAge50-59    OwnerAge60+      CarAge10+      CarAge4-7
     -0.2392105     -0.2194737     -0.2263459     -0.6989978     -0.0860017

    CarAge8-9        ModelB         ModelC         ModelD
   -0.3431970     0.0004845      0.1555723      0.4005228

Degrees of Freedom: 122 Total (i.e. Null);  109 Residual
```

```
Null Deviance:      649.9
Residual Deviance: 127.1        AIC: 84870
```

Capítulo 3

Sistemas Bonus-Malus markovianos

3.1. Introducción

Tal y como hemos comentado en el apartado anterior, con la tarificación *a priori*, las reducciones o incrementos de prima se otorgan en base a las observaciones realizadas por la compañía en un pasado reciente sobre asegurados de un perfil idéntico. Un asegurado se incluye en un determinado colectivo de asegurados de los que no puede diferenciarse en base a la información de la que dispone la compañía (quedan incluidos en la misma clase de tarifa) y ve como se le aplica una tarificación basada en la experiencia relativa a ese colectivo. Sin embargo, los asegurados dentro de una misma clase de tarifa no son idénticos en cuanto a la propensión a la siniestralidad, no son todos homogéneos. Si nos concentramos en el número de siniestros, la propia experiencia del asegurado (el número de siniestros que ha declarado en el pasado) normalmente es la variable más pertinente para para predecir el número futuro de siniestros.

Los sistemas *Bonus-Malus* de clases tratados con cadenas de Markov, constituyen uno de los métodos más usuales de introducir modificaciones en las tarifas teniendo en cuenta la experiencia de siniestralidad del propio asegurado. En este apartado estudiaremos los principales aspectos de estos sistemas *Bonus-Malus* (SBM). Entre los manuales de referencia mencionamos a [Lemaire (1995)] y a [Denuit y Charpentier (2005)].

Desde un punto de vista práctico, los SBM de clases consisten en un sistema de rebajas e incrementos que dependen del número de siniestros que tiene el asegurado en los distintos períodos. Se denominan de clases ya que se definen un conjunto de clases por las que puede pasar el asegurado. Cada clase lleva asociada un porcentaje de descuento o de penalización. El conjunto de clases que implican un descuento constituye la zona de *Bonus* y el conjunto de clases que implican una penalización la zona de *Malus*.

Prácticamente siempre, los SBM se basan únicamente en el número de siniestros culpables (aunque en algunos casos pueden incluir todos los siniestros), de manera que se hace tarificación *a posteriori* solo respecto del número de siniestros. Dependiendo de la naturaleza del seguro en cuestión puede ser razonable considerar la tarificación *a posteriori* respecto de la cuantía de un siniestro o respecto de la cuantía total de todos los siniestros. De hecho existen numerosos trabajos que tratan de esta cuestión desde el punto de vista teórico. Sin embargo, en la práctica, es difícil incorporar la posibilidad de modificación de la prima en función del importe de los siniestros experimentados.

Uno de los principales inconvenientes es el tiempo que transcurre entre la ocurrencia del siniestro y la determinación definitiva de su importe. Este tiempo puede ser bastante largo, superior a un año, de manera que el sistema debería contemplar importes estimados y ello seria recurrible por el asegurado responsable del siniestro.

La utilización de los SBM crea un fenómeno conocido como hambre de bonus (*bonus hunger*, en inglés). El asegurado sabe que al declarar un siniestro, con un SBM, es muy posible que en el próximo período cambie de clase, resultando más penalizado y pagando una prima superior. Si el importe del siniestro es pequeño, el tomador entonces puede preferir asumir él mismo dicho importe y no declarar el siniestro, de manera que no se le incremente la prima. Si el tomador conoce las reglas del SBM, se le plantea un problema de toma de decisiones. Otras características de la póliza tales como la existencia de franquicias pueden intervenir también en la decisión de declarar o no declarar un siniestro. En la literatura actuarial podemos encontrar diversos trabajos dedicados al hambre de bonus, así por ejemplo, en [Lemaire (1995)] encontramos un algoritmo para determinar la estrategia óptima del tomador. Este fenómeno es beneficioso para la compañía aseguradora ya que los siniestros pequeños, que incorporan unos gastos de gestión normalmente desproporcionados respecto del perjuicio a reparar, no le llegan.

Los SBM tratados con cadenas de Markov necesitan que el sistema que definen sea Markoviano: la clase de un asegurado para el período próximo depende únicamente de la clase del período actual y del número de siniestros que ha provocado en este período. De esta forma el historial previo al período actual no influye en la clase para el período próximo.

Vamos a centrar nuestro análisis en los SBM que incluyen un número finito de clases y que por lo tanto trabajan con Cadenas de Markov finitas. Así, en primer lugar, en este capítulo resumimos las características de dichas cadenas. En segundo lugar, definimos formalmente los SBM y proporcionamos diversos ejemplos prácticos. En tercer lugar, analizamos la obtención de la distribución estacionaria del sistema y en cuarto lugar, estudiamos diversas medidas para evaluar y comparar distintos SBM. El capítulo finaliza con una serie de ejercicios propuestos.

3.2. Cadenas de Markov finitas

Definición 1 (Cadena de Markov finita) *Sea $(X_t)_{t\in N}$ un proceso en tiempo discreto siendo el conjunto de estados (los valores posibles de las variables aleatorias (v.a.) X_t) una parte finita E de N. $(X_t)_{t\in N}$ es una cadena de Markov finita si, para $n \in N$, $(i_0, i_1, \ldots, i_n, i) \in E^{n+2}$, tenemos*

$$P(X_{n+1} = i|X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = i|X_n = i_n).$$

Definición 2 (Cadena de Markov finita homogénea) *Sea $(X_t)_{t\in N}$ una cadena de Markov finita. $(X_t)_{t\in N}$ es homogénea si $P(X_{m+n} = j|X_m = i)$ es independiente de m.*

Trabajaremos con cadenas de Markov homogéneas.

Nomenclatura:

- P_{ij}, es la probabilidad de pasar en un período del estado i al j.

- $M = (P_{ij})_{i,j \in E}$ es la matriz de transición.

La matriz de transición M es una matriz estocástica ya que

$$P_{ij} \geq 0, \forall i, j \; y \sum_{\forall j \in E} P_{ij} = 1, \forall i.$$

Además puede comprobarse fácilmente que 1 es valor propio de M y el vector unitario es su vector propio asociado.

La ley del proceso de Markov $(X_t)_{t \in N}$ está determinada por:

- las probabilidades iniciales $(P_j^{(0)})_{j \in E}$, con $E = \{1, 2, \ldots, s\}$,
- la matriz de transición M,

ya que

$$P(X_0 = i_0, \ldots, X_n = i_n) = P_{i_0}^{(0)} \cdot P_{i_0, i_1} \cdot \ldots \cdot P_{i_{n-1}, i_n}.$$

A partir de las probabilidades iniciales y la matriz de transición, pueden calcularse el resto de probabilidades que puedan interesarnos:

- $P_{ij}^{(m+n)}$ es la probabilidad de transición del estado i al estado j en $m + n$ etapas-épocas. (ecuaciones de Chapman-Kolmogorov)

$$P_{ij}^{(m+n)} = \sum_{k \in E} P_{ik}^{m} \cdot P_{kj}^{n}.$$

- $M^{(n)}$ es la matriz de probabilidades de transición en n etapas

$$M^{(n)} = M^n.$$

- $P_j^{(n)} = P(X_n = j)$ es la probabilidad de que la cadena esté en el estado j en la época n.

$$P_j^{(n)} = \sum_{i \in E} P_i^{(0)} \cdot P_{ij}^{(n)}.$$

- $P^{(n)}$ es el vector de probabilidades de que la cadena esté en los diferentes estados en la época n.

$$P^{(n)} = \left(M^{(n)}\right)^T \cdot P^{(0)} = (M^n)^T \cdot P^{(0)}.$$

3.3. Definición y ejemplos de sistemas Bonus-Malus

Para la aplicación del SBM, nos situamos dentro de un mismo grupo de tarifa *a priori*, con una determinada prima de referencia (o prima base). Dentro de este grupo de tarifa, los asegurados se clasifican en un número finito de clases correspondiendo cada una a un nivel de prima anual, que se expresa como un porcentaje de la prima inicial de referencia del grupo considerado.

Un SBM queda definido por:

1. El número de clases: s.
2. El vector de niveles de prima: $b = (b_1, \ldots, b_s)$, donde b_i es la prima en la clase i como porcentaje de la prima de referencia.
3. La clase de entrada: i_0
4. Las reglas de transición de una clase a otra bajo la forma de clase del próximo período después de un determinado número de siniestros en este período.

Estos elementos de un SBM pueden resumirse en la Tabla 3.1, en la que las clases se ordenan por niveles de prima decrecientes, de s a 1: $b_s > \cdots b_1$.

Tabla 3.1: Tabla esquemática de un SBM

Clase	Nivel de primas	Clase después de … siniestros			
i	b_i	0	1	2	2+
s	b_s				
$\vdots$	$\vdots$				
i_0	b_{i_0}				
$\vdots$	$\vdots$				
1	b_1				

En la actualidad, en casi todos los países, las compañías aseguradoras tienen libertad de elección del SBM que deseen utilizar. Sin embargo, respecto del seguro de obligatorio del automóvil en Europa, dos países (Francia y Luxemburgo) tienen un único SBM obligatorio y Bélgica lo tuvo hasta 2002 (con un período transitorio hasta la libertad total que duró dos años). Para más detalles respecto de las características del seguro obligatorio de automóvil en Europa puede consultarse [Meyer (2000)].

Indicamos a continuación diversos SBM que se utilizan (o se han utilizado en la práctica) en el seguro de automóviles y que nos servirán para el desarrollo de los ejemplos de este capítulo.

SBM clásico del Reino Unido La Tabla 3.2 recoge las características de un SBM clásico en el Reino Unido, siendo la clase de entrada, la sexta, $i_0 = 6$.

Tabla 3.2: SBM clásico del Reino Unido. Fuente: [Lemaire (1995)]

Clase	Nivel de primas	Clase después de … siniestros			
i	b_i	0	1	2	3+
7	100	6	7	7	7
6	75	5	7	7	7
5	65	4	6	7	7
4	55	3	5	7	7
3	45	2	5	7	7
2	40	1	4	6	7
1	35	1	4	6	7

SBM de Irlanda La Tabla 3.3 recoge las características de un SBM irlandés, siendo la clase de entrada, la sexta, $i_0 = 6$.

Tabla 3.3: SBM irlandés. Fuente: [Asmussen (2014)]

Clase	Nivel de primas	Clase después de … siniestros		
i	b_i	0	1	2+
6	100	5	6	6
5	90	4	6	6
4	80	3	6	6
3	70	2	5	6
2	60	1	4	6
1	50	1	3	6

SBM brasileño. La Tabla 3.4 recoge las características de un SBM brasileño, siendo la clase de entrada, la séptima, $i_0 = 7$.

Tabla 3.4: SBM brasileño. Fuente: [Lemaire (1995)]

Clase	Nivel de primas	Clase después de … siniestros						
i	b_i	0	1	2	3	4	5	6+
7	100	6	7	7	7	7	7	7
6	90	5	7	7	7	7	7	7
5	85	4	6	7	7	7	7	7
4	80	3	5	6	7	7	7	7
3	75	2	4	5	6	7	7	7
2	70	1	3	4	5	6	7	7
1	65	1	2	3	4	5	6	7

SBM de Irán. La Tabla 3.5 recoge las características del SBM de Irán, siendo la clase de entrada, la sexta, $i_0 = 7$. En Irán tienen un único SBM obligatorio.

Tabla 3.5: SBM de Irán. Fuente: [Mahmoudvand (2013)]

Clase	Nivel de primas	Clase después de … siniestros					
i	b_i	0	1	2	3	4	5+
11	200	6	7	8	9	10	11
10	160	6	7	8	9	10	11
9	140	6	7	8	9	10	11
8	120	6	7	8	9	10	11
7	100	6	7	8	9	10	11
6	95	5	7	8	9	10	11
5	90	4	7	8	9	10	11
4	85	3	7	8	9	10	11
3	75	2	7	8	9	10	11
2	65	1	7	8	9	10	11
1	65	1	7	8	9	10	11

Tabla 3.6: Ejemplo de la escala $-1/+2$

Clase	Nivel de primas	Clase después de … siniestros		
i	b_i	0	1	2+
4	100	3	4	4
3	80	2	4	4
2	70	1	4	4
1	60	1	3	4

Veamos ahora un ejemplo de cálculo de la matriz de transición y de las probabilidades de que un asegurado esté en los diferentes estados al cabo de n períodos desde su entrada en la cartera. Si no especificamos la distribución para el número de siniestros en un período en el grupo de tarifa considerado, la matriz de transición y el resto de probabilidades quedará en función de las probabilidades de dicho número de siniestros.

Ejemplo 1 (Análisis del SBM irlandés (1)) *Utilizando el SBM irlandés (Tabla 3.3),*

1. *Calcular el vector de probabilidades iniciales.*

2. *Matriz de transición en función de las probabilidades del número de siniestros.*

3. *Probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases el próximo período.*

4. *Si asumimos que el número de siniestros en el grupo de tarifa con el que estamos trabajando sigue una distribución de Poisson (caso HOMOGÉNEO), para los valores de λ = 0.04, 0.1, 0.2 y 0.4, calcular las probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases al cabo de 1, 5, 10 y 30 períodos.*

5. *Repetir el apartado anterior asumiendo en este caso que el número de siniestros en el grupo de tarifa con el que estamos trabajando sigue una distribución Binomial Negativa (caso NO HOMOGÉNEO), con esperanzas 0.04, 0.1, 0.2 y 0.4 y desviaciones típicas 0.08, 0.2, 0.4 y 0.8.*

1. ***Calcular el vector de probabilidades iniciales.***

   Las probabilidades iniciales vienen determinadas por la clase de acceso al sistema, que en este caso es es la sexta, $i_0 = 6$. Por lo tanto tenemos,

$$P^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

2. ***Matriz de transición en función de las probabilidades del número de siniestros.***

   Simbolizando por $PN_j = P\left[N = j\right], j = 0, 1$, la matriz de transición es

$$M = \begin{pmatrix} PN_0 & 0 & PN_1 & 0 & 0 & 1 - PN_0 - PN_1 \\ PN_0 & 0 & 0 & PN_1 & 0 & 1 - PN_0 - PN_1 \\ 0 & PN_0 & 0 & 0 & PN_1 & 1 - PN_0 - PN_1 \\ 0 & 0 & PN_0 & 0 & 0 & 1 - PN_0 \\ 0 & 0 & 0 & PN_0 & 0 & 1 - PN_0 \\ 0 & 0 & 0 & 0 & PN_0 & 1 - PN_0 \end{pmatrix}.$$

3. ***Probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases el próximo período.***

$$P^{(1)} = M^T \cdot P^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ PN_0 \\ 1 - PN_0 \end{pmatrix}.$$

4. ***Si asumimos que el número de siniestros en el grupo de tarifa con el que estamos trabajando sigue una distribución de Poisson (caso HOMOGÉNEO), para los valores de $\lambda =$ 0.04, 0.1, 0.2 y 0.4, calcular las probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases al cabo de 1, 5, 10 y 30 períodos.***

   Si el número de siniestros sigue una distribución de Poisson, $P(\lambda)$, sabemos que $PN_0 = e^{-\lambda}$ y $PN_1 = \lambda e^{-\lambda}$. La matriz de transición para $\lambda = 0.04$, por ejemplo, es la siguiente:

$$M = \begin{pmatrix} 0.9607894 & 0.0000000 & 0.0384316 & 0.0000000 & 0.0000000 & 0.0007790 \\ 0.9607894 & 0.0000000 & 0.0000000 & 0.0384316 & 0.0000000 & 0.0007790 \\ 0.0000000 & 0.9607894 & 0.0000000 & 0.0000000 & 0.0384316 & 0.0007790 \\ 0.0000000 & 0.0000000 & 0.9607894 & 0.0000000 & 0.0000000 & 0.0392106 \\ 0.0000000 & 0.0000000 & 0.0000000 & 0.9607894 & 0.0000000 & 0.0392106 \\ 0.0000000 & 0.0000000 & 0.0000000 & 0.0000000 & 0.9607894 & 0.0392106 \end{pmatrix}.$$

   En la Tabla 3.7 mostramos los valores de las probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases al cabo de 1, 5, 10 y 30 períodos. Las probabilidades al cabo de un 1 período se calculan como $P^{(1)} = M^T \cdot P^{(0)}$, al cabo de 5 como $P^{(5)} = \left(M^5\right)^T \cdot P^{(0)}$, al cabo de 10 como $P^{(10)} = \left(M^{10}\right)^T \cdot P^{(0)}$, y al cabo de 30 como $P^{(30)} = \left(M^{30}\right)^T \cdot P^{(0)}$. A continuación indicamos el código de `R` *utilizado.*

```
R> lambda <- 0.04
R> p0 <- dpois(0,lambda)
R> p1 <- dpois(1,lambda)
R> M <- matrix(c(p0,p0,0,0,0,0,0,0,p0,0,0,0,p1,0,0,p0,0,0,0,p1,0,0,p0,0,0,0,
p1,0,0,p0,1-p0-p1,1-p0-p1,1-p0-p1,1-p0,1-p0,1-p0), nrow = 6, ncol = 6)

R> # Al cabo de un período
R> P0 <- c(0,0,0,0,0,1)
R> P1 <- t(M) %*% P0

R> # Al cabo de 5 períodos
R> P5 <- P0
R> for (k in 1:5) P5 <- t(M) %*% P5

R> # Al cabo de 10 períodos
R> P10 <- P0
R> for (k in 1:10) P10 <- t(M) %*% P10

R> # Al cabo de 30 períodos
R> P30 <- P0
R> for (k in 1:30) P30 <- t(M) %*% P30
```

5. ***Repetir el apartado anterior asumiendo en este caso que el número de siniestros en el grupo de tarifa con el que estamos trabajando sigue una distribución Binomial Negativa (caso NO HOMOGÉNEO), con esperanzas 0.04, 0.1, 0.2 y 0.4 y varianzas 0.08, 0.2, 0.4 y 0.8.***

   Si el número de siniestros sigue una distribución Binomial Negativa, $BN(a,b)$, sabemos que $PN_0 = \left(\frac{1}{1+b}\right)^a$ y $PN_1 = \frac{a \cdot b}{(1+b)} \left(\frac{1}{1+b}\right)^a$. Si suponemos una Binomial Negativa con esperanza

Tabla 3.7: Caso Poisson. $P^{(n)}$, con $n = 1$, 5, 10 y 30.

1	5	10	30
$\lambda = 0.04$:			
0.0000000	0.8187308	0.9058686	0.9162474
0.0000000	0.0334130	0.0466218	0.0373928
0.0000000	0.0347766	0.0388719	0.0389188
0.0000000	0.1016944	0.0042235	0.0038572
0.9607894	0.0062604	0.0033032	0.0025189
0.0392106	0.0051248	0.0011109	0.0010650
$\lambda = 0.1$:			
0.0000000	0.6065307	0.7413293	0.7798474
0.0000000	0.0637894	0.1110754	0.0820172
0.0000000	0.0704982	0.0896482	0.0906441
0.0000000	0.1992187	0.0249436	0.0221915
0.9048374	0.0318325	0.0230812	0.0163237
0.0951626	0.0281306	0.0099221	0.0089761
$\lambda = 0.2$:			
0.0000000	0.3678794	0.4890174	0.5484651
0.0000000	0.0814495	0.1569905	0.1214317
0.0000000	0.0994827	0.1430279	0.1483420
0.0000000	0.2686602	0.0768912	0.0714673
0.8187308	0.0911247	0.0820053	0.0630040
0.1812692	0.0914035	0.0520677	0.0472899
$\lambda = 0.4$:			
0.0000000	0.1353353	0.1842499	0.2119136
0.0000000	0.0665612	0.1169932	0.1042244
0.0000000	0.0992977	0.1481588	0.1555162
0.0000000	0.2564030	0.1473270	0.1472057
0.6703200	0.1934815	0.1940884	0.1779154
0.3296800	0.2489214	0.2091827	0.2032248

0.04 *y varianza* 0.08*, por ejemplo, y utilizamos el método de los momentos para la estimación de parámetros, obtenemos que la estimación de los parámetros es* $a = 0.04$ *y* $b = 1$*, y la matriz de transición es la siguiente:*

$$M = \begin{pmatrix} 0.9726549 & 0.0000000 & 0.0194531 & 0.0000000 & 0.0000000 & 0.007891954 \\ 0.9726549 & 0.0000000 & 0.0000000 & 0.0194531 & 0.0000000 & 0.007891954 \\ 0.0000000 & 0.9726549 & 0.0000000 & 0.0000000 & 0.0194531 & 0.007891954 \\ 0.0000000 & 0.0000000 & 0.9726549 & 0.0000000 & 0.0000000 & 0.027345053 \\ 0.0000000 & 0.0000000 & 0.0000000 & 0.9726549 & 0.0000000 & 0.027345053 \\ 0.0000000 & 0.0000000 & 0.0000000 & 0.0000000 & 0.9726549 & 0.027345053 \end{pmatrix}.$$

En la Tabla 3.8 mostramos los valores de las probabilidades de que un asegurado que entra hoy en la cartera esté en las diferentes clases al cabo de 1, 5, 10 y 30 períodos. A continuación indicamos el código de `R` *utilizado.*

```
R> medn <- 0.04
R> varn <- 0.08
R> agamma <- medn^2 / (varn-medn)
R> bgamma <- (varn-medn) / medn
R> p0 <- dnbinom(0, agamma, 1/(1+bgamma))
R> p1 <- dnbinom(1, agamma, 1/(1+bgamma))
R> M <- matrix(c(p0,p0,0,0,0,0,0,0,p0,0,0,0,p1,0,0,p0,0,0,0,p1,0,0,p0,0,0,
0,p1,0,0,p0,1-p0-p1,1-p0-p1,1-p0-p1,1-p0,1-p0,1-p0), nrow = 6, ncol = 6)

R> # Al cabo de un período
R> P0 <- c(0,0,0,0,0,1)
R> P1 <- t(M) %*% P0

R> # Al cabo de 5 períodos
R> P5 <- P0
R> for (k in 1:5) P5 <- t(M) %*% P5

R> # Al cabo de 10 períodos
R> P10 <- P0
R> for (k in 1:10) P10 <- t(M) %*% P10

R> # Al cabo de 30 períodos
R> P30 <- P0
R> for (k in 1:30) P30 <- t(M) %*% P30
```

3.4. Distribución estacionaria de un SBM

Un elemento crucial para un SBM es el comportamiento asintótico (cuando $n \to \infty$) de las probabilidades de que un asegurado que entra hoy en la cartera esté en las distintas clases al cabo de n períodos, $P_j^{(n)}, j \in E$. Volvamos primero a las cadenas de Markov en general.

Tabla 3.8: Caso Binomial Negativa. $P^{(n)}$, con n = 1, 5, 10 y 30.

1	5	10	30
E[N] = 0.04	y	Var[N] = 0.08:	
0.0000000	0.8705506	0.9186148	0.9214235
0.0000000	0.0244745	0.0285541	0.0259047
0.0000000	0.0251626	0.0266285	0.0266330
0.0000000	0.0606920	0.0090049	0.0089533
0.9726549	0.0096758	0.0087961	0.0086869
0.0273451	0.0094446	0.0084017	0.0083985
E[N] = 0.1	y	Var[N] = 0.2:	
0.0000000	0.7071068	0.7932592	0.8052121
0.0000000	0.0507515	0.0681850	0.0577929
0.0000000	0.0543941	0.0618288	0.0619409
0.0000000	0.1290088	0.0266035	0.0261260
0.9330330	0.0296647	0.0262287	0.0251115
0.0669670	0.0290741	0.0238948	0.0238168
E[N] = 0.2	y	Var[N] = 0.4:	
0.0000000	0.5000000	0.5989631	0.6233837
0.0000000	0.0743492	0.1115648	0.0926961
0.0000000	0.0854048	0.1056543	0.1064802
0.0000000	0.1981043	0.0614686	0.0599750
0.8705506	0.0701272	0.0639525	0.0596235
0.1294494	0.0720145	0.0583966	0.0578414
E[N] = 0.4	y	Var[N] = 0.8:	
0.0000000	0.2500000	0.3152913	0.3383687
0.0000000	0.0798770	0.1232381	0.1081115
0.0000000	0.1053983	0.1401136	0.1426564
0.0000000	0.2390739	0.1218227	0.1205601
0.7578583	0.1494845	0.1450985	0.1374577
0.2421417	0.1761663	0.1544358	0.1528457

Definición 3 (Cadena de Markov finita ergódica) *Sea $(X_t)_{t\in N}$ una cadena de Markov finita. $(X_t)_{t\in N}$ es ERGÓDICA si las probabilidades $P_j^{(n)}$ tienen, $\forall j \in E$, límites Π_j independientes de $P^{(0)}$ tales que $\sum_{j\in E} \Pi_j = 1$.*

Por lo tanto es importante saber determinar fácilmente si una cadena de Markov finita es ergódica y, por lo tanto, si podemos calcular su distribución estacionaria. Para ello definamos los siguientes estados:

1. Un estado j es accesible a partir de i si existe $n > 0$ tal que $P_{ij}^{(n)} > 0$.
2. Los estados j e i se comunican si j es accesible a partir de i e i a partir de j. Una cadena de Markov es irreductible si todos los estados se comunican.
3. El período d de un estado i es el máximo común divisor de $n \geq 1 : P_{ii}^{(n)} > 0$ con $d_i = +\infty$ si $P_{ii}^{(n)} = 0, \forall n$. Un estado i es aperiódico si $d_i = 1$. Una cadena de Markov es **aperiódica** si todos sus estados son aperiódicos.

Teorema 1 (Cadena de Markov ergódica) *Una cadena de Markov es ERGÓDICA si es aperiódica e irreductible.*

Desde el punto de vista matemático:

ERGÓDICA $\Leftrightarrow$ M tiene un único valor propio de valor 1. El resto de valores propios tienen módulo inferior a 1 Si la cadena de Markov es ergódica, el SBM tiene una distribución estacionaria $(\Pi_i)_{i=1,\ldots,s}$ tal que

$$\Pi_i = lim_{n\to\infty} P_i^{(n)}, i = 1, \ldots, s.$$

Por lo que disponemos de un instrumento sencillo para ver si una cadena de Markov tiene distribución estacionaria. Una vez determinado que tiene distribución estacionaria debemos calcularla. A continuación desarrollamos cuatro métodos. En primer lugar, es fácil comprobar que

$$\Pi = M^T \cdot \Pi,$$

por lo que Π es el vector propio de M^T (matriz traspuesta de M) asociado al valor propio 1. Así mismo, sabemos que $\sum_{i=1}^{s} \Pi_i = 1$. Por lo tanto, para encontrar Π, debemos resolver el siguiente sistema de ecuaciones lineal con $s+1$ ecuaciones y s incógnitas:

$$\begin{aligned} \Pi &= M^T \cdot \Pi, \\ 1 &= e^T \cdot \Pi, \end{aligned}$$

siendo e el vector unitario y e^T su traspuesto.

En algunos casos podemos encontrar las expresiones de la distribución estacionaria en función de las probabilidades del número de siniestros en un período que conforman la matriz M, pero en la mayoría de los casos tenemos que recurrir al cálculo numérico. En lugar de resolver el sistema, podemos calcular Π a partir del cálculo del vector propio de valor propio 1 de la matriz traspuesta de M, y quedarnos con la parte real estandarizada a suma.

En [Rolski et al. (1999)] podemos encontrar una expresión explícita para Π

$$\Pi^T = e^T \cdot (I - M + E)^{-1},$$

donde E es una matriz de dimensiones $(s) \times (s)$ en la que todos los elementos son 1 e I es la matriz identidad.

Otra posibilidad es calcular las probabilidades $P^{(n)} = M^T P^{(n-1)}$ para diversos valores de n suficientemente grandes hasta que las probabilidades se mantengan constantes. Es decir, mediante cálculo del límite de $M^{(n)}$ cuando $n \to +\infty$ y seleccionando la fila correspondiente a la clase de entrada. Aunque al converger para todas las clases es indiferente la fila elegida, ya que todas ellas convergen a la distribución estacionaria.

Una cuarta posibilidad es estimar mediante simulación la distribución estacionaria Π. Se trata de suponer una clase de partida, un cierto número de asegurados de la cartera y generar números aleatorios de la distribución del número de siniestros. De cada asegurado se va siguiendo la trayectoria hasta un cierto número de períodos suficientemente grande como para que converja la distribución. La velocidad de convergencia dependerá tanto del número de asegurados como del número de períodos de seguimiento simulados. Este último método es el menos preciso y necesita de un cierto tiempo computacional.

La distribución estacionaria es interesante, pues refleja el comportamiento estable del SBM, pero implica la hipótesis de que la distribución del número de siniestros en un período no cambia a lo largo del tiempo y necesita ser complementada con el estudio de la velocidad de convergencia de $(P_i^{(n)})_{i=1,\ldots,s}$ hacia Π y de las distribuciones de probabilidad transitorias.

Ejemplo 2 (Análisis del SBM irlandés (2)) *Utilizando el SBM irlandés (Tabla 3.3) y continuando con el ejemplo 1,*

1. *Si asumimos que el número de siniestros en el grupo de tarifa con el que estamos trabajando sigue una distribución de Poisson (caso HOMOGÉNEO), con $\lambda = 0.04$, calcular la distribución estacionaria, con los cuatro métodos mencionados anteriormente, y cuántos períodos son necesarios para llegar a la distribución estacionaria.*
2. *Repetir el apartado anterior para otros valores de λ, λ = 0.1, 0.2 y 0.4.*

Si asumimos una distribución de Poisson con $\lambda = 0.04$, la distribución estacionaria es:

$$\Pi = (0.9162474, 0.03739276, 0.03891879, 0.003857203, 0.002518908, 0.001064955)^T .$$

Para calcular cuántos períodos son necesarios para llegar a la distribución estacionaria calcularemos $P^{(n)} = M^T P^{(n-1)}$ para diversos valores de n hasta que no varíe:

$$P^{(33)} = (0.916247376, 0.037392765, 0.038918793, 0.003857203, 0.002518908, 0.001064955)^T ,$$

$$P^{(34)} = (0.916247376, 0.037392765, 0.038918792, 0.003857203, 0.002518908, 0.001064955)^T ,$$

$$P^{(35)} = (0.916247376, 0.037392765, 0.038918792, 0.003857203, 0.002518908, 0.001064955)^T ,$$

y observamos que a partir de $n = 34$ podemos considerar que coverge hacia la distribución estacionaria.

Repetimos los cálculos para $\lambda = 0.1$ y obtenemos:

$$\Pi = (0.77984844, 0.08201738, 0.09064322, 0.02219141, 0.01632356, 0.00897600)^T ,$$

$$P^{(42)} = (0.779848437, 0.082017376, 0.090643221, 0.022191406, 0.016323559, 0.008976001)^T ,$$

$$P^{(43)} = (0.77984844, 0.08201738, 0.09064322, 0.02219141, 0.01632356, 0.00897600)^T ,$$

$$P^{(44)} = (0.77984844, 0.08201738, 0.09064322, 0.02219141, 0.01632356, 0.00897600)^T ,$$

llegando a la distribución estacionaria aproximadamente en $n = 43$ períodos.

Repetimos los cálculos para $\lambda = 0.2$ y obtenemos:

$$\Pi = (0.54849009, 0.12143722, 0.14832375, 0.07146502, 0.06300013, 0.04728379)^T ,$$

$$P^{(54)} = (0.54849008, 0.12143722, 0.14832376, 0.07146502, 0.06300013, 0.04728379)^T ,$$

$$P^{(55)} = (0.54849008, 0.12143722, 0.14832375, 0.07146502, 0.06300013, 0.04728379)^T ,$$

$$P^{(56)} = (0.54849008, 0.12143722, 0.14832375, 0.07146502, 0.06300013, 0.04728379)^T ,$$

llegando a la distribución estacionaria aproximadamente en $n = 55$ períodos.

Repetimos los cálculos para $\lambda = 0.4$ y obtenemos:

$$\Pi = (0.2119359, 0.1042353, 0.1555008, 0.1472056, 0.1779108, 0.2032114)^T ,$$

$$P^{(57)} = (0.2119359, 0.1042353, 0.1555008, 0.1472056, 0.1779108, 0.2032115)^T ,$$

$$P^{(58)} = (0.2119359, 0.1042353, 0.1555008, 0.1472056, 0.1779108, 0.2032114)^T ,$$

$$P^{(59)} = (0.2119359, 0.1042353, 0.1555008, 0.1472056, 0.1779108, 0.2032114)^T ,$$

llegando a la distribución estacionaria aproximadamente en $n = 58$ períodos.

El código de `R` *utilizado para el cálculo de la distribución estacionaria con los cuatro procedimientos es:*

```
R> # Primer procedimiento
R> eigen(t(M))
R> pi <- eigen(t(M))$vectors[,1]
R> pi <- pi/sum(pi); pi <- Re(pi)

R> # Segundo procedimiento
R> lim.distr =
R> function(matrix) {
R> et = matrix(nrow=1, ncol=dim(matrix)[2], data=1)
R> E = matrix(nrow=dim(matrix)[1], ncol=dim(matrix)[2], data=1)
R> mat = diag(dim(matrix)[1]) - matrix + E
R> inverse.mat = solve(mat)
R> p = et %*% inverse.mat
R> return(p)}
R> pi = lim.distr(M)
```

```
R> # Tercer procedimiento
R> Mlim <- M
R> for (k in 1:10) Mlim <- Mlim %*% Mlim ## i.e., Mlim <- Mlim ^ (2^10)
R> Mlim
R> pi <- Mlim[6,]

R> # Cuarto procedimiento
R> Next <- matrix(c(1,3,6,1,4,6,2,5,6,3,6,6,4,6,6,5,6,6), nrow=3, ncol=6)
R> TMax <- 50; NSim <- 100000; FinalBM <- numeric(NSim)
R> for (n in 1:NSim)
R> { cn1 <- rpois(TMax,lambda[1]); cn1 <- pmin(cn1, 2) + 1
R> BM <- 6; for (i in 1:TMax) BM <- Next[cn1[i],BM]
R> FinalBM[n] <- BM }
R> pi <- c(sum(FinalBM==1)/NSim, sum(FinalBM==2)/NSim, sum(FinalBM==3)/NSim,
sum(FinalBM==4)/NSim, sum(FinalBM==5)/NSim, sum(FinalBM==6)/NSim)
```

La distribución estacionaria nos permite ver si el sistema está **equilibrado en prima pura**. Utilizaremos como unidad monetaria el coste medio de un siniestro. Sean:

- P_b la prima pura de referencia (sobre la que se aplica los porcentajes).
- P_m la prima pura media con la distribución estacionaria:

$$P_m = (\sum_{i=1}^{s} \Pi_i \cdot \frac{b_i}{100}) \cdot P_b.$$

El sistema está equilibrado en prima pura si la prima media coincide con la prima *a priori*:

$$P_m = E(N),$$
$$(\sum_{i=1}^{s} \Pi_i \cdot \frac{b_i}{100}) \cdot P_b = E(N).$$

Despejando, obtenemos la prima base necesaria para que el sistema esté equilibrado:

$$P_b = \frac{100 \cdot E(N)}{\sum_{i=1}^{s} \Pi_i \cdot b_i}.$$

Ejemplo 3 (Análisis del SBM irlandés (3)) *Utilizando el SBM irlandés (Tabla 3.3) y continuando con el ejemplo 2,*

1. *Calcular la prima base necesaria para que el sistema esté equilibrado en prima pura en su estado estacionario para distintos valores de λ, λ = 0.04, 0.1, 0.2 y 0.4.*

2. *Utilizando como prima base la calculada en el apartado anterior, calcular la evolución de la prima esperada a cobrar por asegurado durante los veinte primeros períodos de vigencia del SBM.*

Tabla 3.9: Prima esperada durante los veinte primeros períodos y $\lambda = 0.04, 0.1, 0.2 , 0.4$.

	λ			
Período	0.04	0.1	0.2	0.4
1	0.07031392	0.1671889	0.2999974	0.4920358
2	0.06313322	0.1521388	0.2780947	0.4683387
3	0.05623407	0.1385210	0.2601623	0.4524541
4	0.04934029	0.1249668	0.2425442	0.4375472
5	0.04246208	0.1115876	0.2257156	0.4246999
6	0.04172780	0.1085611	0.2198107	0.4189586
7	0.04099409	0.1055488	0.2140092	0.4135708
8	0.04026204	0.1025753	0.2084678	0.4089273
9	0.04018389	0.1019027	0.2065234	0.4068523
10	0.04010580	0.1012332	0.2046131	0.4049049
11	0.04002789	0.1005723	0.2027883	0.4032266
12	0.04001957	0.1004229	0.2021481	0.4024766
13	0.04001126	0.1002741	0.2015190	0.4017728
14	0.04000297	0.1001272	0.2009182	0.4011662
15	0.04000208	0.1000940	0.2007073	0.4008951
16	0.04000120	0.1000609	0.2005002	0.4006408
17	0.04000032	0.1000283	0.2003023	0.4004215
18	0.04000022	0.1000209	0.2002329	0.4003235
19	0.04000013	0.1000135	0.2001647	0.4002316
20	0.04000003	0.1000063	0.2000996	0.4001523

La prima base necesaria es 0.0777877, 0.1838218, 0.3267494 y 0.5273876 para $\lambda = 0.04$, 0.1, 0.2 y 0.4 respectivamente. El código de `R` *utilizado es:*

```
R> b <- c(50,60,70,80,90,100)
R> b <- b/100
R> bmedia <- sum(b*pi)
R> primabase <- lambda / bmedia
```

La evolución de la prima esperada a cobrar por asegurado durante los veinte primeros períodos, para $\lambda = 0.04$, 0.1, 0.2 y 0.4 se recoge en la Tabla 3.9.

El código de `R` *utilizado es:*

```
R> primaclase <- b*primabase
R> Primas <- c(0,20)
R> pis <- matrix(0,20,1)

R> P0 <- c(0,0,0,0,0,1)
R> P1 <- t(M)%*%P0
R> pis[1,1:6] <- P1
R> Primas[1] <- sum(P1*primaclase)

R> for (i in 2:20)
   { tMnext <- t(M)
     for ( k in 1:(i-1) ) tMnext <- t(M) %*% tMnext
```

```
pis[i,1:6] <- tMnext%*%P0
Primas[i] <- sum(primaclase*pis[i,])}
```

3.5. Evaluación de un SBM

Se han propuesto en la literatura actuarial un conjunto de criterios de comparación ligados a la distribución estacionaria. Vamos a comentar tres de ellos. Los dos primeros dan una posición media y una dispersión respecto de ésta; el tercero es más elaborado y se corresponde con una elasticidad.

Las tres medidas para realizar la comparación son:

- Nivel Medio Estacionario Relativo.
- Coeficiente de variación.
- Eficiencia de Loimaranta.

Nivel Medio Estacionario Relativo

En inglés *Relative Stationay Average Level* (RSAL). Esta medida nos da la posición de un asegurado medio, en el estado estacionario, en una escala normalizada de 0 a 1:

$$RSAL = \frac{\bar{b} - b_1}{b_s - b_1},$$

siendo $\bar{b} = \sum_{i=1}^{s} \Pi_i \cdot b_i$ el nivel medio.

Coeficiente de variación:

$$CV = \frac{\sqrt{\sum_{i=1}^{s} \Pi_i \cdot (b_i - \bar{b})^2}}{\bar{b}}.$$

Eficiencia de Loimaranta:

Para su cálculo es necesario suponer que el número de siniestros en la categoría de tarifa que estamos analizando es Poisson de parámetro λ. A partir de las probabilidades $P(N = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$ se deducen las expresiones de las probabilidades estacionarias Π_i, $i = 1, \ldots, s$ en función del parámetro λ. Recordemos que la prima media de un asegurado en el estado estacionario es, expresada en tanto por ciento:

$$P_m(\lambda) = (\sum_{i=1}^{s} \Pi_i \cdot \frac{b_i}{100}) \cdot P_b.$$

Se define la elasticidad del sistema como:

$$\eta(\lambda) = \frac{\frac{dP_m(\lambda)}{P_m(\lambda)}}{\frac{d\lambda}{\lambda}} = \lambda \cdot \frac{P'_m(\lambda)}{P_m(\lambda)},$$

donde $\eta(\lambda)$ es la variación relativa de la prima media generada por una variación relativa de la frecuencia media de siniestro. Sustituyendo la expresión de la prima media y su derivada, tenemos:

$$\eta(\lambda) = \lambda \cdot \frac{\sum_{i=1}^{s} \Pi'_i(\lambda) \cdot b_i}{\sum_{i=1}^{s} \Pi_i(\lambda) \cdot b_i}$$

tomada para ciertos valores de λ o para todo un rango. Un SBM será perfectamente elástico si $\eta(\lambda) = 1$. Para calcular $\eta(\lambda)$ necesitamos las derivadas de la distribución estacionaria respecto de λ, $\Pi_i^{'}(\lambda)$, que pueden obtenerse diferenciando el sistema que define la distribución estacionaria. El sistema es:

$$\begin{aligned} \Pi &= M^T \cdot \Pi, \\ 1 &= e^T \cdot \Pi, \end{aligned}$$

que diferenciando nos da el siguiente sistema lineal:

$$\begin{aligned} \frac{d\Pi(\lambda)}{d\lambda} &= M^T \cdot \frac{d\Pi(\lambda)}{d\lambda} + \frac{dM^T(\lambda)}{d\lambda} \cdot \Pi(\lambda), \\ 0 &= e^T \cdot \frac{d\Pi(\lambda)}{d\lambda}. \end{aligned}$$

Ejemplo 4 (Análisis del SBM irlandés (3)) *Utilizando el SBM irlandés (Tabla 3.3) y continuando con el ejemplo 2, calcular para λ = 0.04, 0.1, 0.2 y 0.4:*

1. *Nivel medio.*
2. *Nivel medio estacionario relativo.*
3. *Coeficiente de variación.*
4. *Eficiencia de Loimaranta.*

1. *El nivel medio:*

   El nivel medio es 0.5142202, 0.5440052, 0.6120899 y 0.7584554 respectivamente para λ = 0.04, 0.1, 0.2 y 0.4.

2. *El nivel medio estacionario relativo:*

   El nivel medio estacionario relativo es 0.0284405, 0.0880104, 0.2241799 y 0.5169109 respectivamente para λ = 0.04, 0.1, 0.2 y 0.4.

3. *El coeficiente de variación:*

   El coeficiente de variación es 0.1018855, 0.1778654, 0.2473171 y 0.2394402 respectivamente para λ = 0.04, 0.1, 0.2 y 0.4.

4. *La eficiencia de Loimaranta:*

   La eficiencia de Loimaranta es 0.0320546, 0.1061762, 0.2468757 y 0.3316555 respectivamente para λ = 0.04, 0.1, 0.2 y 0.4.

   Teniendo en cuenta que para el cálculo de derivadas en la fórmula de la eficiencia de Loimaranta se utiliza una aproximación mediante incrementos, el código de `R` *utilizado es:*

```
R> # Nivel medio
R> bmedia <- sum(b*pi)

R> # Nivel medio estacionario
R> RSAL <- (bmedia - b[1]) / (b[6] - b[1])

R> # Coeficiente de variación
R> CV <- (sum(pi * (b-bmedia)^2)^.5) / bmedia

R> # Loimaranta
R> lambdaori <- lambda
R> bmedia <- c(0,0)
R> lambdaperturbada <- c(lambdaori*(1-0.0001), lambdaori*(1+0.0001));
R> for (i in 1:2)
  { p0 <- dpois(0, lambdaperturbada[i]);
    p1 <- dpois(1, lambdaperturbada[i]);
    M <- matrix(c(p0,p0,0,0,0,0,0,0,p0,0,0,0,p1,0,0,p0,0,0,0,p1,0,0,p0,0,0,0,
    p1,0,0,p0,1-p0-p1,1-p0-p1,1-p0-p1,1-p0,1-p0,1-p0), nrow=6, ncol=6);
    pi <- eigen(t(M))$vectors[,1];
    pi <- pi/sum(pi); pi <- Re(pi);
    bmedia[i] <- sum(pi*b) }
R> eficiencia <- (log(bmedia[2]) - log(bmedia[1])) /
 (log(lambdaperturbada[2]) - log(lambdaperturbada[1]))
```

3.6. Supuestos de Sistemas Bonus-Malus

1. Utilizando el SBM clásico del Reino Unido (Tabla 3.2),

   a) Calcular el vector de probabilidades iniciales.

   b) Matriz de transición en función de las probabilidades del número de siniestros.

   c) Encontrar las expresiones analíticas para las probabilidades estacionarias.

   d) Si asumimos que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una distribución Binomial Negativa (caso NO HOMOGÉNEO), siendo su esperanza 0.1 y su varianza 0.4:

      1) Prima esperada (expresada como porcentaje de la prima base) de un asegurado a los 5 años de su entrada en la cartera.
      2) Calcular la distribución estacionaria.
      3) Calcular la prima de referencia que equilibra el sistema.
      4) Calcular las primas puras que se cobrarán en las diferentes clases en el sistema equilibrado.

2. Sea un SBM con cinco clases, una escala -1/+2 y con valores $b_i = h$, 75, 80, 100 y 110 siendo $0 < h < 75$. La clase de entrada es la cuarta.

   a) Calcular el vector de probabilidades iniciales.

   b) Matriz de transición en función de las probabilidades del número de siniestros.

c) Encontrar las expresiones analíticas para las probabilidades estacionarias.

d) Asumiendo que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una de Poisson de parámetro 0.15, calcular el valor de h para que el sistema esté equilibrado en prima pura en su estado estacionario.

e) Asumiendo la distribución de Poisson del apartado anterior y $h = 60$, se pide:

1) Prima de referencia que equilibra el sistema.
2) Primas puras que se cobrarán en las diferentes clases del sistema equilibrado.
3) Nivel medio estacionario relativo.
4) Coeficiente de variación del sistema
5) Eficiencia de Loimaranta.

3. Un sistema Bonus-Malus tiene tres clases y la clase de entrada es la tercera. Se reduce una clase por cada año sin siniestro y hay una penalización de una sola clase por siniestro, siendo los niveles de prima 100, 70 y 50 (clases 3, 2 y 1 respectivamente). Se pide:

a) Construir la tabla que resume el sistema Bonus-Malus.

b) Vector de probabilidades iniciales y matriz de transición en función de las probabilidades del número de siniestros.

c) Suponiendo que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una distribución de Poisson, $N \sim P(\lambda)$, de parámetro $\lambda = 0.05$, calcular:

1) Distribución estacionaria,
2) Prima de referencia que equilibra el sistema,
3) Primas puras que se cobrarán en las diferentes clases del sistema equilibrado,
4) Nivel medio estacionario relativo y coeficiente de variación del sistema,
5) Eficiencia de Loimaranta.

Repetir el análisis para $\lambda = 0.10$ y $\lambda = 0.20$.

d) Suponiendo que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una distribución Binomial Negativa, $N \sim BN(a, b)$, de media 0.05 y varianza 0.1, calcular:

1) Distribución estacionaria,
2) Prima de referencia que equilibra el sistema,
3) Primas puras que se cobrarán en las diferentes clases del sistema equilibrado,
4) Nivel medio estacionario relativo y coeficiente de variación del sistema.

Repetir el análisis para una distribución Binomial Negativa de media 0.1 y varianza 0.15, y para una Binomial Negativa de media 0.2 y varianza 0.25.

4. Consideremos un sistema Bonus-Malus con cuatro clases y niveles de prima 120, 100, 75, 50 para las clases 4, 3, 2 y 1 respectivamente. La clase de entrada es la clase 3. En el sistema se prevé el descenso de una clase por cada año sin siniestro. Se aplica, por períodos anuales una penalización de una clase por el primer siniestro y de dos clases para cada uno de los siguientes siniestros. Se pide:

a) Construir la tabla que resume el sistema Bonus-Malus.

b) Vector de probabilidades iniciales y matriz de transición en función de las probabilidades del número de siniestros.

c) Suponiendo que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una distribución de Poisson, $N \sim P(\lambda)$, de parámetro $\lambda = 0.1$, estimar:

1) Distribución estacionaria mediante cálculo de valores y vectores propios, mediante la fórmula deducida por Rolski en [Rolski et al. (1999)], mediante el límite del producto de las probabilidades de transición en el tiempo y mediante el método de simulación. Comentar los resultados.
2) Prima de referencia que equilibra el sistema a partir de la distribución estacionaria calculada mediante el cálculo de valores y vectores propios.
3) Primas puras correspondientes que se cobrarán en las diferentes clases del sistema equilibrado.

Realizar el ejercicio suponiendo ahora que el número de siniestros en la clase *a priori* con la que estamos trabajando sigue una distribución Binomial Negativa de media 0.1 y varianza 0.15.

Capítulo 4

Provisiones

4.1. Introducción

Dedicamos la primera parte de este capítulo a una pequeña introducción del concepto y contexto de las provisones técnicas. A continuación realizamos un resumen extraido del apartado 15 del libro *Teoría general del seguro*, [ICEA (2017)].

Las provisiones técnicas reflejan todas las obligaciones derivadas de contratos de seguro y de reaseguro y forman parte de las deudas de las entidades aseguradoras y reaseguradoras, tal como se recoge en la Ley 20/2015, de 14 de julio, de ordenación, supervisión y solvencia de las entidades aseguradoras y reaseguradoras (en adelante, LOSSEAR) en su artículo 69. La LOSSEAR, en el mismo artículo 69, fija el valor de transferencia como el criterio de valoración de dichas obligaciones, de forma que el valor de las provisiones técnicas se corresponderá con el importe actual que las entidades aseguradoras y reaseguradoras tendrían que pagar si transfirieran sus obligaciones de seguro y reaseguro de manera inmediata a otra entidad aseguradora o reaseguradora.

En estos momentos, el cálculo y definición de las provisiones técnicas se encuentra en transición debido a la nueva legislación de solvencia en España, recogida fundamentalmente en la LOSSEAR y en el Real Decreto 1060/2015, de 20 de noviembre, de ordenación, supervisión y solvencia de las entidades aseguradoras y reaseguradoras (en adelante, ROSSEAR). Este enfoque de las provisiones técnicas convive con su definición y cálculo desde el punto de vista contable, recogida en ROSSP. De esta forma, mientras no se modifique el Real Decreto 1317/2008, de 24 de julio, por el que se aprueba el Plan de contabilidad de las entidades aseguradoras, continuan vigentes los artículos 29 a 48bis y las disposiciones adicionales cuarta y décima, así como las disposiciones transitorias primera, segunda y undécima del ROSSP. Así, hasta que se apruebe el nuevo Plan Contable, las entidades de seguros y reaseguros realizaran un doble cómputo de las provisiones técnicas: uno a efectos de solvencia y otro a efectos contables.

Uno de los cambios que, desde el punto de vista práctico, introduce la nueva normativa de solvencia afecta a los tipos de interés a utilizar en el cálculo de las provisiones técnicas, de forma que si hasta ahora se utilizaba un único tipo de interés (con un máximo fijado anualmente por la Dirección General de Seguros y Fondos de Pensiones, DGSFP), ahora debe utilizarse una estructura temporal de tipos de interés. Este cambio cobra más importancia en los seguros de vida, debido a su larga duración en el tiempo.

Las provisiones a calcular a efectos de solvencia son:

- Provisión del seguro de vida.
- Provisión de los seguros de enfermedad con técnicas similares a las de los seguros de vida.
- Provisión para primas (en los seguros distintos al de vida).
- Provisión para siniestros pendientes (en los seguros distintos al de vida):
  - Provisión para siniestos declarados pendientes de liquidación o pago.
  - Provisión para siniestos ocurridos pero no declarados.
  - Provisión por gastos de liquidación de siniestros (*Unallocated Loss Adjustment Expenses* (ULAE)).
- Provisión para primas (en los seguros de enfermedad calculados con técnicas similares a los de los seguros distintos al de vida).
- Provisión para siniestros pendientes (en los seguros de enfermedad calculados con técnicas similares a los de los seguros distintos al de vida):
  - Provisión para siniestos declarados pendientes de liquidación o pago.
  - Provisión para siniestos ocurridos pero no declarados.
  - Provisión por gastos de liquidación de siniestros (*Unallocated Loss Adjustment Expenses* (ULAE)).

Mientras que las provisiones a calcular a efectos contables son:

a) Provisión de primas no consumidas.

b) Provisión de riesgos en curso.

c) Provisión de seguros de vida

d) Provisión de participación en beneficios y para extornos.

e) Provisión de prestaciones:

  a. Provisión de prestaciones pendientes de liquidación y pago.
  b. Provisión de prestaciones pendientes de declaración.
  c. Provisión de gastos internos de liquidación de siniestros.

f) Provisión de estabilización.

g) Provisión del seguro de decesos

h) Provisión del seguro de enfermedad.

i) Provisión de desviaciones en las operaciones de capitalización por sorteo

j) Provisión de gestión de riesgos derivados de la internacionalización asegurados por cuenta del estado.

Desde el punto de vista de la técnica actuarial, y centrándonos en los seguros distintos al de vida (y por extensión directa a los seguros de enfermedad calculados con técnicas similares a los seguros distintos al de vida), en este texto estudiaremos la provisión de prestaciones pendientes de declaración. Esta provisión se calcula ligeramente distinta según los dos enfoques (el de solvencia y el contable); aunque los métodos actuariales que se aplican para la determinación de los elementos básicos que permiten su cálculo son únicos y se aplican en ambos casos. En el presente texto nos centraremos en dichos métodos actuariales, prescindiendo del resto de elementos que la legislación contempla, tales como importes mínimos y cálculos simplificados.

Sin embargo, antes de entrar en el desarrollo de los métodos actuariales, creemos conveniente destacar las diferencias de definición básica entre la normativa de solvencia y la normativa contable en cuanto a las provisiones y en concreto a las provisiones de prestaciones pendientes de liquidación.

Según la normativa de solvencia, las provisiones técnicas estarán formadas por la suma de la mejor estimación y de un margen de riesgo (artículo 77 de la LOSSEAR):

- La mejor estimación se corresponderá con la media de los flujos de caja futuros ponderada por su probabilidad, teniendo en cuenta el valor temporal del dinero mediante la aplicación de la pertinente estructura temporal de tipos de interés sin riesgo, es decir, el valor actual esperado de los flujos de caja futuros.

- El margen de riesgo será tal que se garantice que el valor de las provisiones técnicas sea equivalente al importe que las entidades aseguradoras y reaseguradoras previsiblemente exigirían para poder asumir y cumplir las obligaciones de seguro y reaseguro.

Como regla general, ambas componentes de las provisiones técnicas se calcularan por separado, excepto cuando los flujos de caja futuros asociados a las obligaciones de seguro o reaseguro puedan replicarse con fiabilidad utilizando instrumentos financieros para los cuales exista un valor de mercado fiable; en este caso, el valor de las provisiones técnicas asociadas con esos flujos de caja futuros se determinará a partir del valor de mercado de dichos instrumentos financieros y no será necesario calcular por separado la mejor estimación y el margen de riesgo.

Si el margen de riesgo y la mejor estimación se calculan por separado, el margen de riesgo será igual al coste de financiar el capital de solvencia obligatorio exigible por asumir las obligaciones de seguro y reaseguro durante su período de vigencia (artículo 48 del ROSSEAR). Por lo tanto, para calcular el margen de riesgo es necesario haber calculado previamente el capital de solvencia obligatorio para diversos períodos futuros, con la complejidad que ello supone.

La tasa utilizada para determinar el coste financiero citado en el párrafo anterior, tasa de coste del capital, será igual al tipo adicional, por encima del tipo de interés sin riesgo pertinente, que tendría que satisfacer una entidad aseguradora o reaseguradora por mantener un importe de fondos propios admisibles igual al capital de solvencia obligatorio exigible por asumir las obligaciones de seguro y de reaseguro durante su período de vigencia. Esta tasa será la misma para todas las entidades aseguradoras y reaseguradoras y se revisará periódicamente de acuerdo con la normativa de la Unión Europea de directa aplicación. Esta tasa de coste de capital es se fijó en el 6 % (artículo 39 del Reglamento Delegado).

Según el artículo 37 del Reglamento Delegado, matemáticamente, el márgen de riesgo (MR) se calcula como sigue:

$$MR = CoC \sum_{t \geq 0} \frac{SCR(t)}{(1 + r_{(t+1)})^{(t+1)}},$$

siendo:

- CoC, la tasa de coste de capital (por el momento $CoC = 0,06$).
- $SCR(t)$, el capital de solvencia obligatorio al cabo de t años.
- $r_{(t+1)}$, tipo de interés sin riesgo básico correspondiente al vencimiento de $t + 1$ años.

La mejor estimación se calculará en términos brutos, sin deducir los importes recuperables procedentes de los contratos de reaseguro y, en su caso, de las entidades con cometido especial. Dichos importes se calcularán por separado conforme a las disposiciones aplicables al efecto. En concreto se incluirán todos los flujos de caja siguientes (artículos 28 y 31 del Reglamento Delegado):

a) pagos de prestaciones a tomadores y beneficiarios de seguros;

b) pagos que la empresa de seguros y reaseguros deberá satisfacer al proporcionar prestaciones contractuales que se paguen en especie;

c) pagos de los gastos de administración, de gestión de inversiones y de adquisición. En todos ellos se tendrán en cuenta los gastos generales de administración. Los gastos se proyectarán partiendo del supuesto de que la empresa desarrollará nuevas actividades en el futuro;

d) pagos de primas y cualquier flujo de caja adicional que se derive de tales primas;

e) pagos entre la empresa de seguros o reaseguros e intermediarios en relación con obligaciones de seguro o reaseguro;

f) pagos entre la empresa de seguros y reaseguros y empresas de inversión en relación con contratos con prestaciones vinculadas a índices o a fondos de inversión;

g) pagos por salvamento y subrogación, en la medida en que no puedan considerarse activos o pasivos separados con arreglo a las normas internacionales de contabilidad, aprobadas por la Comisión en virtud del Reglamento (CE) no 1606/2002;

h) pagos de impuestos cobrados, o que se prevea cobrar, a los tomadores de seguros, o que se precisen para liquidar las obligaciones de seguro o reaseguro.

Según el artículo 53 del ROSSEAR, cuando en circunstancias específicas, las entidades aseguradoras y reaseguradoras no dispongan de datos suficientes de calidad adecuada para aplicar un método actuarial fiable a un conjunto o a un subconjunto de sus obligaciones de seguro o de reaseguro, o a los importes recuperables procedentes de los contratos de reaseguro o a las entidades con cometido especial, podrán utilizarse aproximaciones, incluidos métodos caso a caso, para el cálculo de la mejor estimación.

En el caso de la provision para prestaciones pendientes declaración, los flujos de caja futuros estarán relacionados con los siniestros ocurridos con anterioridad al momento de cálculo de la provisión, pero que todavia no han sido declarados a la entidad, de forma que ésta desconoce si han ocurrido y su importe. Evidentemente, como han ocurrido se declarán y deberán ser abonados en un momento futuro. Dicho pago debe ir a cargo de la prima cobrada en el año de ocurrencia del siniestro y no a cargo de la prima que se cobrará cuando el siniestro sea definitivamente pagado. Por esta razón se constituye la provisión técnica. La mejor estimación de esta provisión se refiere al valor esperado actual de dichos pagos futuros y coincide con el concepto que de esta provisión se tiene en la legislación contable, con la diferencia de los tipos de interés a utilizar para la actualización financiera de dichos importes futuros (ya mencionada previamente).

Los métodos actuariales se aplican para la determinación del valor esperado, sin actualizar, de los pagos futuros. En éstos vamos a centrarnos en el presente texto. A efectos contables, no es necesario actualizar financieramente dichos valores esperados de los pagos futuros, de forma que son posibles determinadas simplificaciones en el cálculo. Sin embargo, a efectos de solvencia, es imprescindible la actualización financiera de los valores esperados de los pagos futuros utilizando además una estructura de tipos de interés básicos predeterminada y cambiante en el tiempo, de forma que los importes de la provisión serán distintos.

4.2. Clasificación de los métodos actuariales

Según [De Vylder (1986)] y [Taylor (1986)], podemos clasificar los métodos actuariales a utilizar para el cálculo de las provisiones de prestaciones en:

1. Primera clasificación:
   - Micro-modelos: importes de los siniestros por separado,
   - Macro-modelos: los importes de los siniestros se tratan de forma agregada.
2. Segunda clasificación:
   - Modelos deterministas: no intervienen variables aleatorias,
   - Modelos estocásticos: los siniestros se consideran realizaciones de variables aleatorias. Podemos distinguir:
     - Modelos paramétricos: las distribuciones implicadas se suponen conocidas (Normal, Poisson, etc) pero ciertos parámetros deben ser estimados,
     - Modelos de distribución libre: no se realizan hipótesis particulares sobre las distribuciones implicadas.
3. Otros criterios de clasificación tienen en cuenta:
   - Si el modelo considera (o no) el número de siniestros como variable dependiente,
   - Las variables explicativas incluidas en el modelo: año de desarrollo, volumen de exposición, inflación, velocidad de finalización de los siniestros,...

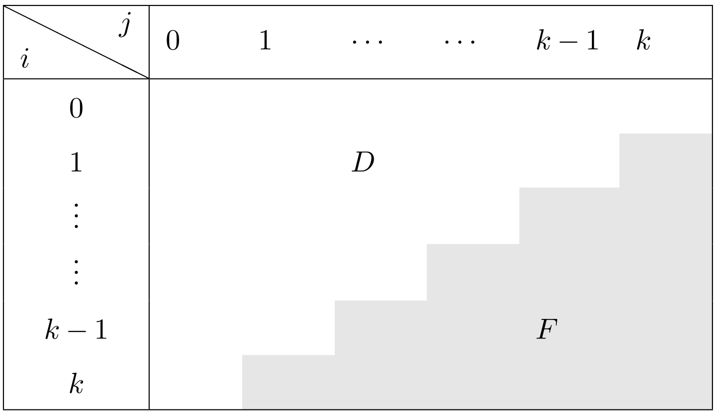
4.3. Datos

Cada método de cálculo de provisiones utiliza distintos tipos de datos de partida, sin embargo, en general, los datos básicos necesarios son:

- c_{ij}: cuantía pagada en el año de desarrollo j, respecto de los siniestros ocurridos en el año de origen i,
- C_{ij} : cuantía acumulada pagada hasta (e incluido) el año de desarrollo j, respecto de los siniestros ocurridos en el año de origen i,
- n_i: número de siniestros ocurridos en el año de origen i (estimación al final del año i)
- s_{ij}: cuantía media pagada en el año de desarrollo j, respecto de los siniestros ocurridos en el año de origen i,
- S_{ij}: cuantía media acumulada pagada hasta (e incluido) el año de desarrollo j, respecto de los siniestros ocurridos en el año de origen i.

Siendo $D = \{i, j | i \in I, j \in J\}$ el dominio (años de origen, años de desarrollo para los que tenemos datos).

Si $i = 0, \dots, k$ y $j = 0, 1, \dots, k$, el dominio D tiene forma triangular, denominándose triángulo de desarrollo. Gráficamente:



En la diagonal del triángulo tenemos los datos del año de cálculo de provisiones.

Por F representamos el conjunto de combinaciones de i, j que permiten completar el cuadrado.

Para un i determinado, podemos simbolizar por J_i: conjunto de años de desarrollo para los cuales tenemos datos. De forma similar se define I_j.

El dominio D puede adoptar otras formas no triangulares como la de trapecio o triángulo sin la esquina superior izquierda.

4.4. Métodos estadísticos deterministas clásicos

En este apartado resumimos las características más importantes de un grupo de métodos estadísticos que consideramos como clásicos, por ser de los primeros planteados, y que por su sencillez son de los más utilizados en la práctica. Nos referimos al libro [Van Eeghen and De Vylder (1981)] para una descripción completa, donde se describen éstos y otros métodos deterministas de cálculo de provisiones. En esta publicación consideramos dentro de este grupo de métodos clásicos a los siguientes:

A) Chain-Ladder

B) Variantes de Chain-Ladder

C) Mínimos cuadrados de De Vylder

D) Métodos separación y de regresión de Taylor

A) CHAIN-LADDER Este es uno de los primeros métodos utilizados en la bibliografía para cálculo de provisiones. Las primeras ideas del método clásico fueron inicialmente apuntadas en [Tarbell (1934)]. Su utilización se remonta a los años 70 y algunas de las referencias en que fue utilizado son [Kramreiter and Straub (1973)], [Skurnick (1973)] y [Clarke and Harland (1974)]. A continuación describimos los datos en que se basa, las hipótesis y la estimación de sus parámetros:

Datos: C_{ij} en D triangular.

Hipótesis-estimadores: Las columnas del triángulo son proporcionales. El estimador del cambio de la columna h a la $h+1$ es:

$$\hat{m}_h = \frac{\sum_{i=0}^{k-h-1} C_{i,h+1}}{\sum_{i=0}^{k-h-1} C_{i,h}},$$

$$C_{i,h+1} = \hat{m}_h \cdot C_{i,h}.$$

El estimador de C_{ij} con $i, j \in F$ es:

$$\hat{C}_{i,j} = C_{i,k-i} \cdot \prod_{h=k-i}^{j-1} \hat{m}_h.$$

Comentarios: A continuación se listan características o comentarios del método:

- Los datos del triángulo deben ser positivos.
- Puede aplicarse aunque supongamos que en el año de origen y los k siguientes no se paga el 100 % de los siniestros de un año de origen.
- Se supone que el esquema de desarrollo de los pagos es estable en el tiempo, a lo largo de los años de desarrollo.
- Sirve en las situaciones de tasas de inflación constantes.
- Puede aplicarse sobre las S_{ij}.
- Puede obtenerse como un caso particular de un planteamiento más general, variantes del método de Chain-Ladder.
- Aunque es un modelo no estocástico, puede considerarse proveniente de modelos estocásticos.

Ejemplo 5 (Estimación de pagos futuros en cálculo de provisiones (1)) *Los datos sobre los pagos por siniestros, en miles de euros, para un determinado ramo en una compañía de seguros generales, son los siguientes:*

Tabla 4.1: Triángulo con los pagos por siniestros, siendo las filas los años de origen y las columnas los años de desarrollo.

	0	1	2	3	4
0	88	43.6	51	54.15	15.6
1	93.2	45	64.2	54.8	
2	109	69.2	57.4		
3	122.4	63.4			
4	136.8				

El número de siniestros en cada año de origen, estimado a final del año correspondiente, es:

Tabla 4.2: Número de siniestros en cada año de origen, estimado a final de año para los datos de la Tabla 4.1.

Año de origen	Número estimado de siniestros
0	630
1	750
2	800
3	805
4	935

Calcular los pagos futuros para el cálculo de provisiones de siniestros pendientes, suponiendo que la inflación futura será del $4.5\,\%$ *anual constante, utilizando los siguientes métodos:*

- *método de Chain-Ladder,*
- *variantes del método de Chain-Ladder,*
- *método de los mínimos cuadrados de De Vylder,*
- *método de separación aritmética de Taylor,*
- *método de separación geométrica de Taylor,*
- *método de regresión de Taylor,*
- *modelo estocástico de Mack,*
- *y modelo lineal generalizado con distribución Poisson (sobredispersa) y función de enlace logarítmica.*

La función de `R` *incorporada en el área de trabajo* `provisio` *(ver [Claramunt et al. (2017)]) creada para el cálculo de las provisiones para siniestros pendientes de declaración o pago según el método de Chain-Ladder que vamos a utilizar es* `ibnrchl`*. La función necesita necesita como datos el vector con las cuantías según el año de origen:*

```
R> c0 < -c(88,43.6,51,54.15,15.6)
R> c1 <- c(93.2,45,64.2,54.8)
R> c2 <- c(109,69.2,57.4)
R> c3 <- c(122.4,63.4)
R> c4 <- 136.8
R> ibnrchl(c(c0,c1,c2,c3,c4))

triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0

triángulo cuantías acumuladas
      [,1]  [,2]  [,3]   [,4]   [,5]
[1,]  88.0 131.6 182.6 236.75 252.35
[2,]  93.2 138.2 202.4 257.20 257.20
[3,] 109.0 178.2 235.6 235.60 235.60
[4,] 122.4 185.8 185.8 185.80 185.80
[5,] 136.8 136.8 136.8 136.80 136.80

factores de desarrollo
[1] 1.536112 1.385268 1.282987 1.065892

rectángulo con cuantías acumuladas
      [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.0 131.6000 182.6000 236.7500 252.3500
[2,]  93.2 138.2000 202.4000 257.2000 274.1475
[3,] 109.0 178.2000 235.6000 302.2717 322.1891
[4,] 122.4 185.8000 257.3828 330.2187 351.9776
[5,] 136.8 210.1402 291.1004 373.4781 398.0874
rectángulo con cuantías desacumuladas
      [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.0 43.60000 51.00000 54.15000 15.60000
[2,]  93.2 45.00000 64.20000 54.80000 16.94750
[3,] 109.0 69.20000 57.40000 66.67174 19.91738
[4,] 122.4 63.40000 71.58277 72.83598 21.75887
[5,] 136.8 73.34018 80.96026 82.37764 24.60933

provisiones (suma de pagos futuros)
[1] 531.0016
```

```
vector pagos futuros
[1] 228.54219 173.71362 104.13651  24.60933
```

B) Variantes de Chain-Ladder A continuación describimos algunas variantes del método de Chain-Ladder (ver [Berquist and Sherman (1977)]).

Datos: Triángulo con los factores de desarrollo d_{ij}:

$$d_{ij} = \frac{C_{i,j+1}}{C_{ij}},$$

con $i = 0, 1, ..., k-1$, y $j = 0, 1, ..., k-1$.

Hipótesis-estimadores: Cada variante del método Chain-Ladder utiliza una hipótesis alternativa. Destacamos dos de ellas:

- Ajuste de tendencias lineales (por mínimos cuadrados) a cada columna de d_{ij}.
- El factor de desarrollo es constante en cada columna. Se estima este valor constante mediante una media aritmética ponderada de los factores de desarrollo empíricos:

$$\hat{d}_h = \frac{\sum_{i=0}^{k-h-1} w_{ih} \cdot d_{ih}}{\sum_{i=0}^{k-h-1} w_{ih}},$$

  con pesos:
  - 1,
  - C_{ij} (Chain-Ladder),
  - $i + j + 1$,
  - $(i + j + 1)^2$.

Ejemplo 6 (Estimación de pagos futuros en cálculo de provisiones (2)) *Cálculo de provisiones según variantes del método de Chain-Ladder.*

Para la variante I del método de Chain-Ladder, la función de `R` *del área de trabajo* `provisio` *a ejecutar es* `ibnrchlvar1`*, y en el caso de la variante II, la función es* `ibnrchlvar2`*. En este último caso, además, existe la opción de poder elegir qué pesos o ponderaciones se aplican en el cálculo de los parámetros del modelo:*

```
R> ibnrchlvar1(c(c0,c1,c2,c3,c4))

triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0
```

```
triángulo cuantías acumuladas
      [,1]  [,2]  [,3]   [,4]   [,5]
[1,]  88.0 131.6 182.6 236.75 252.35
[2,]  93.2 138.2 202.4 257.20 257.20
[3,] 109.0 178.2 235.6 235.60 235.60
[4,] 122.4 185.8 185.8 185.80 185.80
[5,] 136.8 136.8 136.8 136.80 136.80

matriz factores desarrollo
         [,1]     [,2]     [,3]     [,4]
[1,] 1.495455 1.387538 1.296550 1.065892
[2,] 1.482833 1.464544 1.270751 1.000000
[3,] 1.634862 1.322110 1.000000 1.000000
[4,] 1.517974 1.000000 1.000000 1.000000
[5,] 1.000000 1.000000 1.000000 1.000000

matriz factores desarrollo completa
         [,1]     [,2]     [,3]     [,4]
[1,] 1.495455 1.387538 1.296550 1.065892
[2,] 1.482833 1.464544 1.270751 1.065892
[3,] 1.634862 1.322110 1.283650 1.065892
[4,] 1.517974 1.325969 1.283650 1.065892
[5,] 1.587678 1.293255 1.283650 1.065892

rectángulo con cuantías acumuladas
      [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.0 131.6000 182.6000 236.7500 252.3500
[2,]  93.2 138.2000 202.4000 257.2000 274.1475
[3,] 109.0 178.2000 235.6000 302.4280 322.3557
[4,] 122.4 185.8000 246.3651 316.2467 337.0849
[5,] 136.8 217.1943 280.8877 360.5616 384.3199

rectángulo con cuantías desacumuladas
      [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.0 43.60000 51.00000 54.15000 15.60000
[2,]  93.2 45.00000 64.20000 54.80000 16.94750
[3,] 109.0 69.20000 57.40000 66.82804 19.92768
[4,] 122.4 63.40000 60.56511 69.88156 20.83822
[5,] 136.8 80.39432 63.69340 79.67392 23.75823

provisiones (suma de pagos futuros)
[1] 502.508

vector pagos futuros
[1] 224.73496 153.50264 100.51214  23.75823

R> ibnrchlvar2(c(c0,c1,c2,c3,c4))
```

```
triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0

triángulo cuantías acumuladas
      [,1]  [,2]  [,3]   [,4]   [,5]
[1,]  88.0 131.6 182.6 236.75 252.35
[2,]  93.2 138.2 202.4 257.20 257.20
[3,] 109.0 178.2 235.6 235.60 235.60
[4,] 122.4 185.8 185.8 185.80 185.80
[5,] 136.8 136.8 136.8 136.80 136.80

matriz factores desarrollo
         [,1]     [,2]     [,3]     [,4]
[1,] 1.495455 1.387538 1.296550 1.065892
[2,] 1.482833 1.464544 1.270751 1.000000
[3,] 1.634862 1.322110 1.000000 1.000000
[4,] 1.517974 1.000000 1.000000 1.000000
[5,] 1.000000 1.000000 1.000000 1.000000

Opciones de ponderaciones: 1: wij=1  2: wij=i+j+1  3: wij=(i+j+1)^2  4: wij=2^(i+j+1)
entrar opción: 1,2,3 o 4
2

matriz estimaciones de dij
        [,1]     [,2]     [,3]     [,4]
[1,] 1.54376 1.384128 1.281808 1.065892
[2,] 1.54376 1.384128 1.281808 1.065892
[3,] 1.54376 1.384128 1.281808 1.065892
[4,] 1.54376 1.384128 1.281808 1.065892
[5,] 1.54376 1.384128 1.281808 1.065892

rectángulo con cuantías acumuladas
      [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.0 131.6000 182.6000 236.7500 252.3500
[2,]  93.2 138.2000 202.4000 257.2000 274.1475
[3,] 109.0 178.2000 235.6000 301.9939 321.8929
[4,] 122.4 185.8000 257.1709 329.6436 351.3646
[5,] 136.8 211.1864 292.3089 374.6838 399.3726

rectángulo con cuantías desacumuladas
      [,1]    [,2]     [,3]     [,4]     [,5]
[1,]  88.0 43.6000 51.00000 54.15000 15.60000
[2,]  93.2 45.0000 64.20000 54.80000 16.94750
[3,] 109.0 69.2000 57.40000 66.39388 19.89907
```

```
[4,] 122.4 63.4000 71.37091 72.47273 21.72097
[5,] 136.8 74.3864 81.12252 82.37489 24.68877

provisiones (suma de pagos futuros)
[1] 531.3776

vector pagos futuros
[1] 229.09868 173.49432 104.09586  24.68877
```

C) Mínimos cuadrados de De Vylder

Nos refermos a [De Vylder (1978)] como referencia original del método de los mínimos cuadrados de De Vylder que a continuación describimos:

Datos: c_{ij} en D triangular.

Hipótesis: Las hipótesis son:

$$c_{ij} = x_i \cdot p_j,$$

$$\sum_{j=0}^{k} p_j = 1.$$

siendo x_i la cuantía total a pagar de los siniestros ocurridos en el año de origen i, y p_j la proporción de x_i que se paga en el año de desarrollo j.

Estimadores: Los estimadores se obtienen por mínimos cuadrados:

$$Min \sum_{\forall i,j \in D} (c_{ij} - x_i \cdot p_j)^2,$$

$$x_i = \frac{\sum\limits_{\forall j \in J_i} c_{ij} \cdot p_j}{\sum\limits_{\forall j \in J_i} p_j^2} \text{ y } p_j = \frac{\sum\limits_{\forall i \in I_j} c_{ij} \cdot x_i}{\sum\limits_{\forall i \in I_j} x_i^2}.$$

Comentarios: A continuación se listan caracteristcas o comentarios del método:

- Las columnas del triángulo con las cuantías no acumuladas (o con las cuantías acumuladas) también resultan proporcionales.
- El modelo es válido aunque haya una tasa de inflación constante.

Ejemplo 7 (Estimación de pagos futuros en cálculo de provisiones (3)) *Cálculo de provisiones según el método de mínimos cuadrados de De Vylder.*

Para este método la función de `R` del área de trabajo `provisio` a ejecutar es `ibnrvylder`:

```
R> ibnrvylder(c(c0,c1,c2,c3,c4))

triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0

Resultados intermedios

xi
[1] 253.2720 273.8097 319.7737 352.5971 397.6113

pj
[1] 0.34405463 0.18552746 0.20238523 0.20643883 0.06159385

rectángulo con cuantías totales
          [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  87.13942 46.98892 51.25852 52.28518 15.60000
[2,]  94.20548 50.79921 55.41503 56.52494 16.86499
[3,] 110.01961 59.32680 64.71747 66.01370 19.69609
[4,] 121.31266 65.41644 71.36044 72.78973 21.71781
[5,] 136.80000 73.76781 80.47065 82.08240 24.49041

provisiones (suma de pagos futuros)
[1] 529.254

vector pagos futuros
[1] 228.00694 172.95647 103.80022  24.49041
```

D) Separación de Taylor

Datos: s_{ij} en D triangular, siendo:

$$s_{ij} = \frac{c_{ij}}{n_i},$$

con n_i el número de siniestros estimado a final del año de origen i.

Hipótesis: Las hipótesis son:

$$s_{ij} = r_j \cdot \lambda_{i+j},$$

siendo:

r_j: la proporción de la cuantía de los siniestros ocurridos en el año de origen i, que se paga en el año de desarrollo j, y

λ_{i+j}: el índice de las influencias externas exógenas (inflación, ...), correspondiendo el mismo valor a aquellas cuantías que están situadas temporalmente en el mismo momento.

Estimadores: La estimación se puede realizar por:

1. Separación aritmética (ver [Verbeek (1972)]).
2. Separación geométrica (ver [Taylor (1979)]).
3. Mínimos cuadrados ordinarios. Regresión de Taylor (ver [Taylor (1979)]).

Comentarios: A continuación se listan característcas o comentarios del método:

- Para completar el triángulo F es necesario estimar, fuera del modelo, los valores para $h > k$ de $\hat{\lambda}_h$.
- Se estima de forma separada el efecto del año de pago (inflación y otros factores).
- El método de regresión sirve también para triángulos con datos incompletos.
- El método de regresión proporciona los mismos resultados que el de separación geométrica, aunque las estimaciones de los parámetros sean distintas.

Resumimos, a continuación, las fórmulas relacionadas con las diferentes estimaciones de los parámetros del modelo.

1. **Separación aritmética**.

   - Condición adicional:
   $$\sum_{j=0}^{k} r_j = 1.$$
   - Cálculos intermedios:
     - d_j: suma de s_{ij} por la diagonal j,
     $$d_j = \sum_{i=0}^{j} s_{i,j-i},$$
     - v_j: suma de s_{ij} por la columna j,
     $$v_j = \sum_{i=0}^{k-j} s_{i,j}.$$
   - Estimadores de los parámetros para $h = 0, 1, \ldots, k$:
   $$\hat{r}_h = \frac{v_h}{\sum\limits_{j=h}^{k} \hat{\lambda}_j} \text{ y } \hat{\lambda}_h = \frac{d_h}{1 - \sum\limits_{j=h+1}^{k} \hat{r}_j}.$$

2. **Separación geométrica**

   - Condición adicional:
   $$\prod_{j=0}^{k} r_j = 1.$$

- Cálculos intermedios:
  - e_j: productos de s_{ij} por la diagonal j,
  $$e_j = \prod_{i=0}^{j} s_{i,j-i},$$
  - w_j: productos de s_{ij} por la columna j,
  $$w_j = \prod_{i=0}^{k-j} s_{i,j}.$$
- Estimadores de los parámetros para $h = 0, 1, \ldots, k$:
$$\hat{r}_h = \left(\frac{w_h}{\prod_{j=h}^{k} \hat{\lambda}_j} \right)^{\frac{1}{k-h+1}} \text{ y } \hat{\lambda}_h = \left(e_h \cdot \prod_{j=h+1}^{k} \hat{r}_j \right)^{\frac{1}{h+1}}.$$

3. **Mínimos cuadrados ordinarios. Regresión de Taylor.**
   - Se linealiza el modelo:
   $$\ln s_{ij} = \ln r_j + \ln \lambda_{i+j} + U_{ij},$$
   siendo U_{ij} el parámetro que indica las perturbaciones aleatorias del modelo.
   - Se obtienen los estimadores mínimo cuadráticos de los parámetros para $h = 0, 1, \ldots, k$.
   - Condición adicional: $r_0 = 1$.

Ejemplo 8 (Estimación de pagos futuros en cálculo de provisiones (4)) *Cálculo de provisiones según los métodos de Taylor.*

Para el método de separación aritmética, la función de `R` *del área de trabajo* `provisio` *a ejecutar es* `ibnrarit` *y en el caso de separación geométrica, la función que se utiliza es* `ibnrgeom`*. Por último, para el método de regresión de Taylor, la función a aplicar es* `ibnrreg`*. En todos los casos se necesitan como parámetros de la función el vector con las cuantías y el vector con el número de siniestros. A continuación los ejecutamos las funciones para los datos del ejemplo:*

```
R> ni <- c(630,750,800,805,935)
R> ibnrarit(c(c0,c1,c2,c3,c4),ni)

triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0
```

```
triángulo cuantías medias
          [,1]       [,2]       [,3]       [,4]      [,5]
[1,] 0.1396825 0.06920635 0.08095238 0.08595238 0.0247619
[2,] 0.1242667 0.06000000 0.08560000 0.07306667 0.0000000
[3,] 0.1362500 0.08650000 0.07175000 0.00000000 0.0000000
[4,] 0.1520497 0.07875776 0.00000000 0.00000000 0.0000000
[5,] 0.1463102 0.00000000 0.00000000 0.00000000 0.0000000

suma por columnas
[1] 0.6985591 0.2944641 0.2383024 0.1590190 0.0247619

suma por diagonales
[1] 0.1396825 0.1934730 0.2772024 0.4101021 0.3946465

estimaciones de r
[1] 0.35903061 0.18916814 0.19797466 0.19108208 0.06274452

estimaciones de lambda
[1] 0.3890547 0.3529250 0.3714986 0.4375563 0.3946465

Entrar valor de inflación futura constante para
estimación lambdas futuras, en tanto por uno:
0.045

matriz de lambdas completa
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.3890547 0.3529250 0.3714986 0.4375563 0.3946465
[2,] 0.3529250 0.3714986 0.4375563 0.3946465 0.4124056
[3,] 0.3714986 0.4375563 0.3946465 0.4124056 0.4309638
[4,] 0.4375563 0.3946465 0.4124056 0.4309638 0.4503572
[5,] 0.3946465 0.4124056 0.4309638 0.4503572 0.4706233

rectángulo con cuantías totales
          [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.00000 42.06016 46.33481 52.67378 15.60000
[2,]  95.03315 52.70677 64.96880 56.55741 19.40714
[3,] 106.70350 66.21737 62.50400 63.04265 21.63250
[4,] 126.46237 60.09691 65.72491 66.29132 22.74724
[5,] 132.48031 72.94309 79.77412 80.46161 27.60964

provisiones (suma de pagos futuros)
[1] 519.6342

vector pagos futuros
[1] 221.11780 167.69794 103.20885  27.60964


R> ibnrgeom(c(c0,c1,c2,c3,c4),ni)
```

```
triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0

triángulo cuantías medias
          [,1]       [,2]       [,3]       [,4]      [,5]
[1,] 0.1396825 0.06920635 0.08095238 0.08595238 0.0247619
[2,] 0.1242667 0.06000000 0.08560000 0.07306667 0.0000000
[3,] 0.1362500 0.08650000 0.07175000 0.00000000 0.0000000
[4,] 0.1520497 0.07875776 0.00000000 0.00000000 0.0000000
[5,] 0.1463102 0.00000000 0.00000000 0.00000000 0.0000000

producto por columnas
[1] 5.261303e-05 2.828829e-05 4.971933e-04 6.280254e-03 2.476190e-02

producto por diagonales
[1] 1.396825e-01 8.600042e-03 6.617857e-04 9.676835e-05 1.495867e-06

estimaciones de r
[1] 2.0591431 1.0779544 1.1388740 1.0925264 0.3620807

estimaciones de lambda
[1] 0.06783528 0.06224534 0.06397126 0.07693690 0.06838780

Entrar valor de inflación futura constante para
estimación lambdas futuras, en tanto por uno:
0.045

matriz de lambdas completa
           [,1]       [,2]       [,3]       [,4]       [,5]
[1,] 0.06783528 0.06224534 0.06397126 0.07693690 0.06838780
[2,] 0.06224534 0.06397126 0.07693690 0.06838780 0.07146525
[3,] 0.06397126 0.07693690 0.06838780 0.07146525 0.07468119
[4,] 0.07693690 0.06838780 0.07146525 0.07468119 0.07804184
[5,] 0.06838780 0.07146525 0.07468119 0.07804184 0.08155372

rectángulo con cuantías totales
          [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.00000 42.27151 45.89878 52.95503 15.60000
[2,]  96.12904 51.71858 65.71608 56.03661 19.40714
[3,] 105.38079 66.34758 62.30807 62.46214 21.63250
[4,] 127.53139 59.34374 65.51888 65.68089 22.74724
[5,] 131.66695 72.02892 79.52405 79.72069 27.60964

provisiones (suma de pagos futuros)
```

```
[1] 516.3321

vector pagos futuros
[1] 219.41709 166.83744 102.46794  27.60964

R> ibnrreg(c(c0,c1,c2,c3,c4),ni)

triángulo cuantías
      [,1] [,2] [,3]  [,4] [,5]
[1,]  88.0 43.6 51.0 54.15 15.6
[2,]  93.2 45.0 64.2 54.80  0.0
[3,] 109.0 69.2 57.4  0.00  0.0
[4,] 122.4 63.4  0.0  0.00  0.0
[5,] 136.8  0.0  0.0  0.00  0.0

triángulo cuantías medias
          [,1]       [,2]       [,3]       [,4]      [,5]
[1,] 0.1396825 0.06920635 0.08095238 0.08595238 0.0247619
[2,] 0.1242667 0.06000000 0.08560000 0.07306667 0.0000000
[3,] 0.1362500 0.08650000 0.07175000 0.00000000 0.0000000
[4,] 0.1520497 0.07875776 0.00000000 0.00000000 0.0000000
[5,] 0.1463102 0.00000000 0.00000000 0.00000000 0.0000000

estimaciones de r
[1] 1.0000000 0.5234966 0.5530815 0.5305733 0.1758405

estimaciones de lambda
[1] 0.1396825 0.1281721 0.1317260 0.1584241 0.1408203

Entrar valor de inflación futura constante para
estimación lambdas futuras, en tanto por uno:
0.045

matriz de lambdas completa
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.1396825 0.1281721 0.1317260 0.1584241 0.1408203
[2,] 0.1281721 0.1317260 0.1584241 0.1408203 0.1471572
[3,] 0.1317260 0.1584241 0.1408203 0.1471572 0.1537792
[4,] 0.1584241 0.1408203 0.1471572 0.1537792 0.1606993
[5,] 0.1408203 0.1471572 0.1537792 0.1606993 0.1679308

rectángulo con cuantías totales
          [,1]     [,2]     [,3]     [,4]     [,5]
[1,]  88.00000 42.27151 45.89878 52.95503 15.60000
[2,]  96.12904 51.71858 65.71608 56.03661 19.40714
[3,] 105.38079 66.34758 62.30807 62.46214 21.63250
[4,] 127.53139 59.34374 65.51888 65.68089 22.74724
[5,] 131.66695 72.02892 79.52405 79.72069 27.60964
```

```
provisiones (suma de pagos futuros)
[1] 516.3321

vector pagos futuros
[1] 219.41709 166.83744 102.46794  27.60964
```

Tabla 4.3: Resumen métodos estadísticos deterministas de cálculo de provisiones.

Método	Datos	Estimación parámetros	$\hat{C}_{ij},\ i,j \in F$
Chain-Ladder	C_{ij}	$\hat{m}_h = \frac{\sum_{i=0}^{k-h-1} C_{i,h+1}}{\sum_{i=0}^{k-h-1} C_{i,h}}$	$\hat{C}_{i,j} = C_{i,k-i} \cdot \prod_{h=k-i}^{j-1} \hat{m}_h$
Variantes de Chain-Ladder	$d_{ij} = \frac{C_{i,j+1}}{C_{ij}}$ $i,j = 0,1,...,k-1$	Ajuste tendencias lineales $\hat{d}_h = \frac{\sum_{i=0}^{k-h-1} w_{ih} \cdot d_{ih}}{\sum_{i=0}^{k-h-1} w_{ih}}$	$\hat{C}_{i,j} = C_{i,k-i} \cdot \prod_{h=k-i}^{j-1} \hat{d}_h$
Mínimos cuadrados De Vylder	c_{ij} Hipótesis: $c_{ij} = x_i \cdot p_j$ $\sum_{j=0}^{k} p_j = 1$	$Min \sum_{\forall i,j \in D} (c_{ij} - x_i \cdot p_j)^2$ $\hat{x}_i = \frac{\sum_{\forall j \in J_i} c_{ij} \cdot p_j}{\sum_{\forall j \in J_i} p_j^2}$ $\hat{p}_j = \frac{\sum_{\forall i \in I_j} c_{ij} \cdot x_i}{\sum_{\forall i \in I_j} x_i^2}$	$\hat{c}_{ij} = \hat{x}_i \cdot \hat{p}_j$
Separación aritmética de Taylor	$s_{ij} = \frac{c_{ij}}{n_i}$ Hipótesis: $s_{ij} = r_j \cdot \lambda_{i+j}$ $\sum_{j=0}^{k} r_j = 1$	$\hat{r}_h = \frac{v_h}{\sum_{j=h}^{k} \hat{\lambda}_j}$ $\hat{\lambda}_h = \frac{d_h}{1 - \sum_{j=h+1}^{k} \hat{r}_j}$ $d_j = \sum_{i=0}^{j} s_{i,j-i}$ $v_j = \sum_{i=0}^{k-j} s_{i,j}$	$\hat{s}_{ij} = \hat{r}_j \cdot \hat{\lambda}_{i+j}$
Separación geométrica de Taylor	$s_{ij} = \frac{c_{ij}}{n_i}$ Hipótesis: $s_{ij} = r_j \cdot \lambda_{i+j}$ $\prod_{j=0}^{k} r_j = 1$	$\hat{r}_h = \left(\frac{w_h}{\prod_{j=h}^{k} \hat{\lambda}_j} \right)^{\frac{1}{k-h+1}}$ $\hat{\lambda}_h = \left(e_h \cdot \prod_{j=h+1}^{k} \hat{r}_j \right)^{\frac{1}{h+1}}$ $e_j = \prod_{i=0}^{j} s_{i,j-i}$ $w_j = \prod_{i=0}^{k-j} s_{i,j}$	$\hat{s}_{ij} = \hat{r}_j \cdot \hat{\lambda}_{i+j}$
Regresión de Taylor	$\ln s_{ij} = \ln r_j + \ln \lambda_{i+j} + U_{ij}$	$\hat{\beta} = (X \cdot X')^{-1} \cdot X' \cdot Y$ $\hat{\beta} = \begin{bmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_k \\ \hat{\lambda}_0 \\ \vdots \\ \hat{\lambda}_k \end{bmatrix}$ $\hat{r}_0 = 1$	$\hat{s}_{ij} = \hat{r}_j \cdot \hat{\lambda}_{i+j}$

4.5. Versión estocástica del modelo Chain-Ladder: Modelo de Mack

En este apartado resumimos las características más importantes de la versión estocástica del modelo de Chain-Ladder siguiendo a [Mack, T. (1993)]. En este modelo, el factor de cambio de columna $\hat{m}_h$ de Chain-Ladder se corresponde con el valor esperado del factor de desarrollo supuesto conocido el importe total pagado hasta (e incluido) el período $h-1$. Los estimadores de dicho factor de cambio de columna y de la cantidad total pagada hasta un año determinado propuestos por Chain-Ladder se obtienen fácilmente de este modelo de Mack.

Pero además, a partir de las hipótesis, se obtienen también estimadores para el error cuadrático medio de las provisiones estimadas, por año de origen y total. Por tanto obtenemos, no sólo el valor esperado de las provisiones, sino también el error correspondiente.

Existen otros modelos estocásticos, con otras hipótesis distintas a las del modelo de Mack que también incluyen, como estimadores puntuales, los proporcionados por Chain-Ladder; de forma que la versión estocástica del modelo de Chain-Ladder no es un tema cerrado desde el punto de vista de la investigación.

Como vemos a continuación, el modelo de Mack tiene la ventaja de ser un modelo de distribución libre, de forma que no hace hipótesis sobre la distribución de las variables aleatorias incluidas, sino tan sólo hipótesis generales, válidas para cualquier distribución.

Datos: C_{ij} en D triangular igual que en Chain-Ladder: $D = \{i, j \,| 0 \leq i \leq k, 0 \leq j \leq k, 0 \leq i + j \leq k\}$.

Hipótesis: Se asumen las siguientes hipótesis:

H1:

$$E\left[\frac{C_{i,h+1}}{C_{i,h}} \,|C_{i,h}\right] = m_h \text{ para } i = 0, 1, ..., k-1 \text{ y } h = 0, ..., k-1,$$

o, de forma equivalente,

$$E\left[C_{i,h+1} \,|C_{i,h}\right] = m_h \cdot C_{i,h} \text{ para } i = 0, 1, ..., k-1 \text{ y } h = 0, ..., k-1.$$

H2:

$$V\left[\frac{C_{i,h+1}}{C_{i,h}} \,|C_{i,h}\right] = \sigma_h^2 \text{ para } i = 0, 1, ..., k-1 \text{ y } h = 0, ..., k-1,$$

o, de forma equivalente,

$$V\left[C_{i,h+1} \,|C_{i,h}\right] = C_{i,h} \cdot \sigma_h^2 \text{ para } i = 0, 1, ..., k-1 \text{ y } h = 0, ..., k-1.$$

H3: C_{i_1,j_1} y C_{i_2,j_2} son variables aleatorias independientes para $i_1 \neq i_2$.

Estimadores: Los estimadores son los siguientes:

- $\hat{m}_h = \sum\limits_{i=0}^{k-h-1} \frac{C_{i,h}}{\sum\limits_{i=0}^{k-h-1} C_{i,h}} \cdot \frac{C_{i,h+1}}{C_{i,h}} = \frac{\sum\limits_{i=0}^{k-h-1} C_{i,h+1}}{\sum\limits_{i=0}^{k-h-1} C_{i,h}}.$

  El estimador con este modelo de Mack coincide con el del método de Chain-Ladder. Tal y como hemos comentado anteriormente, $\hat{m}_h$, es una media aritmética ponderada de los factores de desarrollo siendo las ponderaciones $C_{i,h}$.

- $\hat{\sigma}_h^2 = \frac{1}{k-h} \cdot \sum\limits_{i=0}^{k-h-1} C_{i,h} \cdot \left(\frac{C_{i,h+1}}{C_{i,h}} - \hat{m}_h \right)^2$ para $h = 0, ..., k-2$

  Falta un estimador para $h = k-1$. En [Mack, T. (1993)] se proponen diversos métodos para su estimación:

  1. Si $\hat{m}_{k-1} = 1$ y se espera que el desarrollo de los siniestros finalice tras $k-1$ años, entonces $\sigma_{k-1}^2 = 0$.
  2. Si lo anterior no ocurre, es necesario extrapolar el siguiente valor de la secuencia $\left\{ \hat{\sigma}_0^2, \hat{\sigma}_{01}^2, \ldots, \hat{\sigma}_{k-2}^2 \right\}$ que suele ser una sucesión de valores exponencialmente decrecientes.

- Suponiendo que las hipótesis **H1**, **H2** y **H3**, también sirven para los valores futuros $j = k-i+1, ..., k$,

$$\hat{C}_{i,j} = C_{i,k-i} \cdot \prod_{h=k-i}^{j-1} \hat{m}_h \text{ para } i, j \in F = \{i, j \,|0 \leq i \leq k, 0 \leq j \leq k, i+j > k\}.$$

  A partir de estos valores estimados y los datos originales se calculan las provisiones estimadas por año de origen, R_i para $i = 0, 1, ..., k$.

- Se propone el siguiente estimador para el Error Cuadrático Medio (ECM o MSE en inglés) de las provisiones por año de origen:

$$MSE(\hat{R}_i) = \hat{C}_{i,k}^2 \cdot \sum_{s=k-i}^{k-1} \frac{\hat{\sigma}_s^2}{\hat{m}_s^2} \cdot \left(\frac{1}{\hat{C}_{i,s}} + \frac{1}{\sum\limits_{q=0}^{k-s} \hat{C}_{q,s}} \right) \text{ para } i = 1, ..., k.$$

  Comentario: este estimador del error cuadrático medio de las provisiones por año de origen sólo tiene sentido, sólo puede utilizarse, cuando las provisiones se calculan como suma aritmética de los pagos futuros; es decir sin considerar rentabilidad sobre dichas provisiones.

- Se propone también un estimador para el error cuadrático medio de las provisiones totales, que de nuevo sólo tiene sentido cuando las provisiones se calculan sin considerar rentabilidad sobre dichas provisiones. El estimador es:

$$MSE\left(\hat{R}\right) = \sum_{i=1}^{k} \left\{ \left(s.e. \left(\hat{R}_i \right) \right)^2 + \hat{C}_{i,k} \left(\sum_{j=i+1}^{k} \hat{C}_{j,k} \right) \sum_{h=k-i}^{k-1} \frac{2\hat{\sigma}_h^2 / \hat{m}_h^2}{\sum\limits_{n=0}^{k-h} \hat{C}_{n,h}} \right\}$$

donde $s.e.\left(\hat{R}_i\right)$ significa error estándar de la provisión con año de origen i y se refiere a la raíz cuadrada de la estimación del error cuadrático medio anterior.

Ejemplo 9 (Estimación de pagos futuros en cálculo de provisiones (5)) *Cálculo de provisiones según el modelo de Mack*

Para calcular las provisiones con el modelo de Mack utilizaremos la función `MackChainLadder` *de la librería* `ChainLadder` *de* `R` *(ver [Gesmann et al. (2017)]). Los datos deben introducirse en una matriz, que contenga los pagos acumulados por año de origen en cada fila y con NA en el caso en que no se haya realizado aún el pago correspondiente. Los datos deben introducirse en una matriz transformada a un objeto de tipo triángulo mediante la función* `as.triangle` *del paquete* `ChainLadder`*. Veamos su ejecución con los datos del ejemplo:*

```
R> c0 <- c(88,43.6,51,54.15,15.6)
R> c1 <- c(93.2,45,64.2,54.8)
R> c2 <- c(109,69.2,57.4)
R> c3 <- c(122.4,63.4)
R> c4 <- 136.8
R> C0 <- cumsum(c0)
R> C1 <- c(cumsum(c1),NA)
R> C2 <- c(cumsum(c2),NA,NA)
R> C3 <- c(cumsum(c3),NA,NA,NA)
R> C4 <- c(c4,NA,NA,NA,NA)
R> C <- matrix(c(C0,C1,C2,C3,C4),ncol=5)
R> C <- t(C); C
      [,1]  [,2]  [,3]   [,4]   [,5]
[1,]  88.0 131.6 182.6 236.75 252.35
[2,]  93.2 138.2 202.4 257.20     NA
[3,] 109.0 178.2 235.6     NA     NA
[4,] 122.4 185.8    NA     NA     NA
[5,] 136.8    NA    NA     NA     NA

R> C <- as.triangle(C); C
      dev
origin     1     2     3      4      5
     1  88.0 131.6 182.6 236.75 252.35
     2  93.2 138.2 202.4 257.20     NA
     3 109.0 178.2 235.6     NA     NA
     4 122.4 185.8    NA     NA     NA
     5 136.8    NA    NA     NA     NA

R> mch <- MackChainLadder(C); mch
Warning message:
In Mack.S.E(CL[["Models"]], FullTriangle, est.sigma = est.sigma,  :
  'loglinear' model to estimate sigma_n doesn't appear appropriate.
p-value > 5.
 est.sigma will be overwritten to 'Mack'.
```

```
 Mack's estimation method will be used instead.
MackChainLadder(Triangle = C)

  Latest Dev.To.Date Ultimate  IBNR Mack.S.E CV(IBNR)
1    252       1.000      252   0.0     0.00      NaN
2    257       0.938      274  16.9     1.67   0.0983
3    236       0.731      322  86.6     5.58   0.0644
4    186       0.528      352 166.2    20.58   0.1239
5    137       0.344      398 261.3    28.77   0.1101

             Totals
Latest:   1,067.75
Dev:          0.67
Ultimate: 1,598.75
IBNR:       531.00
Mack.S.E     40.57
CV(IBNR):     0.08
```

4.6. Modelo lineal generalizado

En este apartado se resume el uso del modelo lineal generalizado en cálculo de provisiones. Existe una amplia literatura en la que se describen, para este modelo de regresión, los datos a utilizar y las hipótesis a aplicar. Nos referimos por ejemplo a las siguientes [England and Verrall (1999)], [England and Verrall (2002)], [England and Verrall (2006)], [Kaas et al. (2008)], [Boj et al. (2014)], [Boj y Costa (2014)] y [Charpentier (2014)] dentro de una amplia lista bibliográfica.

Si utilizamos MLG para estimar las provisiones técnicas, y realizamos los supuestos de distribución Poisson, $\zeta = 1$ en la fórmula (2.1), junto con la función de enlace logarítmica, obtenemos como estimación puntual de los pagos futuros la misma que con el modelo clásico Chain-Ladder determinista. Por lo tanto, se puede considerar que el MLG, aplicando la distribución Poisson junto con el enlace logarítmico, es una generalización estocástica de Chain-Ladder. Cabe notar que puesto que estimamos cuantías de pagos podría ser adecuado realizar otros supuestos si se adaptan mejor a los datos empíricos. Por ejemplo, suponer distribución Gamma o Gaussiana inversa, u otras funciones de enlace diferentes de la logarítmica.

A continuación se detallan las formulaciones para el caso en que se asume distribución Poisson (sobre-dispersa) y función de enlace logarítmica. Las fórmulas son fácilmente extendibles a otras distribuciones de la familia dada por la fórmula (2.1).

Datos: c_{ij} en D triangular: $D = \{i, j \,|0 \leq i \leq k, 0 \leq j \leq k, 0 \leq i + j \leq k\}$.

Hipótesis: Se asumen las siguientes hipótesis en el MLG

$$\mu_{ij} = E\,[c_{ij}],\, V\,(\mu_{ij}) = \mu_{ij},\, \phi > 1,\, w_{ij} = 1.$$

Con lo que

$$\mu_{ij} = E\left[c_{ij}\right] \text{ y } Var\left[c_{ij}\right] = \phi\mu_{ij}.$$

Y combinado con la función de enlace logarítmica

$$\log \mu_{ij} = \eta_{ij}.$$

Estimación y error de predicción: El predictor lineal en el MLG es de la forma

$$\eta_{ij} = c_0 + \alpha_i + \beta_j,$$

siendo α_i el factor correspondiente a los años de origen $i = 1, \ldots, k$, β_j el factor correspondiente a los años de desarrollo $j = 1, \ldots, k$, y c_0 el término que se correspondería al año de origen y desarrollo 0. Cabe notar que, en la construcción de las estimaciones, el término c_0 siempre debe ser añadido en la predicción, junto con el coeficiente correspondiente al año de origen y el correspondiente al año desarrollo. A partir de aquí podemos realizar las estimaciones de las cuantías del resto del rectángulo mediante la expresión:

$$\hat{c}_{ij} = \exp\left(c_0 + \alpha_i + \beta_j\right),$$

y con ellas calcular los pagos futuros por año de origen, $\hat{R}_i$, sumando las cuantías estimadas por filas para $i = 1, \ldots, k$, y los pagos futuros totales, $\hat{R}$, como la suma de los individuales.

Podemos calcular el error cometido en la predicción de los pagos futuros utilizando la fórmula analítica que se desprende del MLG y que tiene en cuenta la distribución supuesta en los datos, mediante las expresiones que a continuación detallamos. Al igual que con cualquiera de los métodos de cálculo de provisiones, es posible estimar las distribuciones predictivas haciendo uso de metodología *bootstrap*. En este apartado detallamos las expresiones analíticas del error de predicción en el caso de distribución Poisson y link logarítmico para cada estimación de los pagos futuros, para los pagos futuros por año de ocurrencia y para el pago futuro total.

- Error cuadrático medio para cada estimación de los pagos futuros:

$$MSE\left(\hat{c}_{ij}\right) \cong \phi\hat{\mu}_{ij} + \hat{\mu}_{ij}^2 Var\left[\hat{\eta}_{ij}\right].$$

- Error cuadrático medio para la estimación de los pagos futuros por año de origen, para $i = 1, \ldots, k$:

$$MSE\left(\hat{R}_i\right) \cong \sum_{j=k-i+1}^{k} \phi\hat{\mu}_{ij} + \sum_{j=k-i+1}^{k} \hat{\mu}_{ij}^2 Var\left[\hat{\eta}_{ij}\right] +$$
$$2 \sum_{\substack{j_1=k-i+1 \\ j_2=k-i+1 \\ j_2>j_1}}^{k} \hat{\mu}_{ij_1}\hat{\mu}_{ij_2} Cov\left[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}\right].$$

- Error cuadrático medio para la estimación de los pagos futuros totales:

$$MSE\left(\hat{R}\right) \cong \sum_{i=1}^{k} \sum_{j=k-i+1}^{k} \phi\hat{\mu}_{ij} + \sum_{i=1}^{k} \sum_{j=k-i+1}^{k} \hat{\mu}_{ij}^2 Var\left[\hat{\eta}_{ij}\right] +$$
$$2 \sum_{i=1}^{k} \sum_{\substack{j_1=k-i+1 \\ j_2=k-i+1 \\ j_2>j_1}}^{k} \hat{\mu}_{ij_1}\hat{\mu}_{ij_2} Cov\left[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}\right].$$

Ejemplo 10 (Estimación de pagos futuros en cálculo de provisiones (6)) *Cálculo de provisiones según el modelo lineal generalizado con distribución Poisson (sobredispersa) y función de enlace logarítmica.*

Para calcular las provisiones con modelo lineal generalizado utilizaremos la función `glmReserve` *de la librería* `ChainLadder` *de R (ver [Gesmann et al. (2017)]). Los datos deben introducirse en una matriz transformada a un objeto de tipo triángulo igual que ocurre con el modelo de Mack anterior mediante la función as.triangle. Veamos su ejecución con los datos del ejemplo:*

```
R> glmformula <- glmReserve(C, var.power = 1, link.power = 0,
 mse.method = "formula"); glmformula

Latest Dev.To.Date Ultimate IBNR       S.E         CV
2      257.2   0.9380015    274.2   17  5.572718 0.32780693
3      235.6   0.7303162    322.6   87 12.996502 0.14938508
4      185.8   0.5281410    351.8  166 20.196744 0.12166713
5      136.8   0.3438914    397.8  261 30.445739 0.11665034
total  815.4   0.6056150   1346.4  531 48.263824 0.09089232

R> names(glmformula)

[1] "call" "summary" "Triangle" "FullTriangle" "model"
[6] "sims.par" "sims.reserve.mean" "sims.reserve.pred"

R> glmformula$FullTriangle

      dev
origin     1     2     3      4      5
     1  88.0 131.6 182.6 236.75 252.35
     2  93.2 138.2 202.4 257.20 274.20
     3 109.0 178.2 235.6 302.60 322.60
     4 122.4 185.8 257.8 330.80 352.80
     5 136.8 209.8 290.8 372.80 397.80

R> glmformula$model

Call:  glm(formula = value ~ factor(origin) + factor(dev), family = fam,
    data = ldaFit, offset = offset)

Coefficients:
    (Intercept)  factor(origin)2  factor(origin)3  factor(origin)4
        4.46267          0.08285          0.24432          0.33275

 factor(origin)5     factor(dev)2     factor(dev)3
       0.45585         -0.62341         -0.52456

   factor(dev)4     factor(dev)5
       -0.50721         -1.71539
```

```
Degrees of Freedom: 14 Total (i.e. Null);  6 Residual
Null Deviance:     211.6
Residual Deviance: 4.872  AIC: NA
```

En este ejemplo los coeficientes del MLG con distribución de Poisson (sobre-dispersa) y función de enlace logarítmica que reproduce la estimación Chain-Ladder de los pagos futuros, toman los siguientes valores estimados:

$$\hat{c}_0 = 4.46267; \hat{\alpha}_1 = 0.08285; \hat{\alpha}_2 = 0.24432; \hat{\alpha}_3 = 0.33275; \hat{\alpha}_4 = 0.45585;$$

$$\hat{\beta}_1 = -0.62341; \hat{\beta}_2 = -0.52456; \hat{\beta}_3 = -0.50721; \hat{\beta}_4 = -1.71539.$$

4.7. Método de Bornhuetter-Ferguson

A continuación realizamos un resumen descriptivo del método de Bornhuetter-Ferguson extraido de [Schmidt, K. D. and Zocher, M. (2008)]. El método es originario de [Bornhuetter and Ferguson (1972)].

1. **Preliminares**

   Se definen las cuotas acumuladas individuales, $\gamma_{i,h}$,

   $$\gamma_{i,h} = \frac{E[C_{i,h}]}{E[C_{i,k}]} \text{ para } 0 \leq h \leq k.$$

   Un vector $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_k)$ de parámetros, con $\gamma_k = 1$, patrón de desarrollo de las cuotas acumuladas, si la identidad $\gamma_h = \frac{E[C_{i,h}]}{E[C_{i,k}]}$ se cumple $\forall h \in [0, 1, \ldots, k]$ y $\forall i \in [0, 1, \ldots, k]$. El patrón de desarrollo γ existe, si y sólo si, para cada año de desarrollo h, las cuotas acumuladas individuales $\gamma_{i,h}$ son idénticas para todos los años de origen.

   Recordemos que los factores de desarrollo de las variantes de Chain-Ladder eran:

   $$d_{i,h} = \frac{C_{i,h+1}}{C_{ih}} \text{ para } i = 0, 1, ..., k-1 \text{ y } h = 0, 1, ..., k-1$$

   con

   $$\hat{d}_h = \frac{\sum\limits_{i=0}^{k-h-1} w_{i,h} \cdot d_{i,h}}{\sum\limits_{i=0}^{k-h-1} w_{i,h}},$$

   dónde si los pesos eran $w_{i,h} = C_{i,h}$ se obtenía como caso particular el Método de Chain-Ladder clásico.

   Podemos definir $d = (d_0, d_1, \ldots, d_k)$ como el patrón de desarrollo de los factores de desarrollo. Veremos en el apartado 3 cómo γ y d están relacionados.

2. **Método de Bornhuetter-Ferguson extendido**

Datos: C_{ij} en D triangular.

Hipótesis: Se asumen las siguientes hipótesis:

H1: Existen los vectores $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_k)$ y $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_k)$ con $\gamma_k = 1$, tales que la identidad $E\left[C_{i,j}\right] = \gamma_j \cdot \alpha_i$ sirve $\forall j \in [0, 1, \ldots, k]$ y $\forall i \in [0, 1, \ldots, k]$. Entonces, $E\left[C_{i,j}\right] = \alpha_i$ y $\gamma_j = \frac{E[C_{i,j}]}{E[C_{i,k}]}$, lo que significa que γ es el patrón de desarrollo de las cuotas acumuladas. **H2**: Los vectores $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_k)$ de estimadores *a priori* de las cuotas acumuladas (con $\gamma_k = 1$) y el vector $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k)$ vienen dados.

Estimadores de las cuantías: Los estimadores de las cuantías vienen dados por las siguientes expresiones:

$$\hat{c}_{ij} = \hat{\alpha}_i \left(\hat{\gamma}_j - \hat{\gamma}_{j-1}\right) \text{ para } j = k - i + 1, \ldots, k.$$

$$\hat{C}_{ij} = C_{i,k-i} + \hat{\alpha}_i \left(\hat{\gamma}_j - \hat{\gamma}_{k-1}\right) \text{ para } j = k - i + 1, \ldots, k.$$

Comentario sobre dichos estimadores *a priori*: Pueden obtenerse a partir de información proporcionada por varias fuentes:

- Información interna: es aquella contenida en el triángulo de desarrollo de la cartera analizada.
- Información externa: es aquella información que no está contenida en el triángulo de desarrollo. Puede obtenerse a partir de estadísticas de los mercados, de otras carteras que se consideren con características similares a la analizada, o de las primas u otras medidas de volumen de la cartera analizada.

3. **Relación entre el patrón de desarrollo de las cuotas acumuladas y el patrón de desarrollo de los factores de desarrollo**

Cuotas acumuladas:

$$\gamma_{i,h} = \frac{E[C_{i,h}]}{E[C_{i,k}]} 0 \leq h \leq k.$$

Factores de desarrollo:

$$d_{i,h} = \frac{E[C_{i,h+1}]}{E[C_{i,h}]} 0 \leq h \leq k - 1.$$

Con lo que:

$$\gamma_{i,h} = \frac{E[C_{i,h}]}{E[C_{i,k}]} = \frac{E[C_{i,k-1}]}{E[C_{i,k}]} \frac{E[C_{i,k-2}]}{E[C_{i,k-1}]} \cdots \frac{E[C_{i,h+1}]}{E[C_{i,h+2}]} \frac{E[C_{i,h}]}{E[C_{i,h+1}]} = \prod_{j=h}^{k-1} \frac{1}{d_{i,j}}.$$

Por lo tanto, si utilizamos los patrones γ_h y d_h (y en concreto el estimador de Chain-Ladder para d_h) tenemos que:

$$\hat{\gamma}_h = \prod_{s=h}^{k-1} \frac{1}{\hat{m}_s}.$$

4. **Caso particular del modelo de Bornhuetter-Ferguson original**

   El modelo de Bornhuetter-Ferguson original propone utilizar información externa para el estimador *a priori* de las $\hat{\alpha}$, en concreto:

$$\hat{\alpha}_i = \text{Primas}_i \times \text{ Ratio de pérdidas esperado}_i,$$

   e información interna para el estimador *a priori* de las $\hat{\gamma}$, en concreto su equivalencia con el estimador de Chain-Ladder de los factores de desarrollo (como hemos visto en el apartado anterior), $\hat{\gamma}_h = \prod_{s=h}^{k-1} \frac{1}{\hat{m}_s}$.

Ejemplo 11 (Cálculo de provisiones con el método de Bornhuetter-Ferguson) *Este ejemplo numérico se ha realizado utilizando los datos analizados en [Schmidt, K. D. and Zocher, M. (2008)] y que se recogen en la siguiente tabla 4.4.*

Tabla 4.4: Triángulo con los pagos por siniestros de [Schmidt, K. D. and Zocher, M. (2008)].

	0	1	2	3	4	5
0	1001	854	568	565	347	148
1	1113	990	671	648	422	
2	1265	1168	800	744		
3	1490	1383	1007			
4	1725	2536				
5	1889					

A continuación se realizan los cálculos con `R` *para* $V11$ *con* $\hat{\alpha}^{external}$ *y* $\hat{\gamma}^{external}$*, y para* $V13$ *con* $\hat{\alpha}^{external}$ *y* $\hat{\gamma}^{CL}$*:*

```
R> # Datos acumulados:
R> C0 <- c(1001,1855,2423,2988,3335,3483)
R> C1 <- c(1113,2103,2774,3422,3844)
R> C2 <- c(1265,2433,3233,3977)
R> C3 <- c(1490,2873,3880)
R> C4 <- c(1725,4261)
R> C5 <- 1889
```

```
R> fd <- c(2.051107,1.328800,1.232147,1.119969,1.044378)
R> alphaext <- c(3520,3980,4620,5660,6210,6330)
R> gammaext <- c(0.28,0.53,0.71,0.86,0.95,1)
R> gammaest <- rep(0,6)
R> gammaest[1] <- 1/prod(fd)
R> gammaest[2] <- 1/prod(fd[2:5])
R> gammaest[3] <- 1/prod(fd[3:5])
R> gammaest[4] <- 1/prod(fd[4:5])
R> gammaest[5] <- 1/prod(fd[5:5])
R> gammaest[6] <- 1
R> gammaest
 [1] 0.2545809 0.5221727 0.6938630 0.8549413 0.9575077 1.0000000
R> C15 <- C1[5]+(gammaext[6]-gammaext[5])*alphaext[2];C15
 [1] 4043
R> C24 <- C2[4]+(gammaext[5]-gammaext[4])*alphaext[3];C24
 [1] 4392.8
R> C25 <- C2[4]+(gammaext[6]-gammaext[4])*alphaext[3];C25
 [1] 4623.8
R> C33 <- C3[3]+(gammaext[4]-gammaext[3])*alphaext[4];C33
 [1] 4729
R> C34 <- C3[3]+(gammaext[5]-gammaext[3])*alphaext[4];C34
 [1] 5238.4
R> C35 <- C3[3]+(gammaext[6]-gammaext[3])*alphaext[4];C35
 [1] 5521.4
R> C42 <- C4[2]+(gammaext[3]-gammaext[2])*alphaext[5];C42
 [1] 5378.8
R> C43 <- C4[2]+(gammaext[4]-gammaext[2])*alphaext[5];C43
 [1] 6310.3
R> C44 <- C4[2]+(gammaext[5]-gammaext[2])*alphaext[5];C44
 [1] 6869.2
R> C45 <- C4[2]+(gammaext[6]-gammaext[2])*alphaext[5];C45
 [1] 7179.7
R> C51 <- C5[1]+(gammaext[2]-gammaext[1])*alphaext[6];C51
 [1] 3471.5
R> C52 <- C5[1]+(gammaext[3]-gammaext[1])*alphaext[6];C52
 [1] 4610.9
R> C53 <- C5[1]+(gammaext[4]-gammaext[1])*alphaext[6];C53
 [1] 5560.4
R> C54 <- C5[1]+(gammaext[5]-gammaext[1])*alphaext[6];C54
 [1] 6130.1
R> C55 <- C5[1]+(gammaext[6]-gammaext[1])*alphaext[6];C55
 [1] 6446.6
R> Provextext <- C15+C25+C35+C45+C55-C1[5]-C2[4]-C3[3]-C4[2]-C5; Provextext
 [1] 9963.5

R> C15bis <- C1[5]+(gammaest[6]-gammaest[5])*alphaext[2];C15bis
 [1] 4013.119
R> C24bis <- C2[4]+(gammaest[5]-gammaest[4])*alphaext[3];C24bis
```

```
 [1] 4450.857
R> C25bis <- C2[4]+(gammaest[6]-gammaest[4])*alphaext[3];C25bis
 [1] 4647.171
R> C33bis <- C3[3]+(gammaest[4]-gammaest[3])*alphaext[4];C33bis
 [1] 4791.703
R> C34bis <- C3[3]+(gammaest[5]-gammaest[3])*alphaext[4];C34bis
 [1] 5372.229
R> C35bis <- C3[3]+(gammaest[6]-gammaest[3])*alphaext[4];C35bis
 [1] 5612.735
R> C42bis <- C4[2]+(gammaest[3]-gammaest[2])*alphaext[5];C42bis
 [1] 5327.197
R> C43bis <- C4[2]+(gammaest[4]-gammaest[2])*alphaext[5];C43bis
 [1] 6327.493
R> C44bis <- C4[2]+(gammaest[5]-gammaest[2])*alphaext[5];C44bis
 [1] 6964.431
R> C45bis <- C4[2]+(gammaest[6]-gammaest[2])*alphaext[5];C45bis
 [1] 7228.308
R> C51bis <- C5[1]+(gammaest[2]-gammaest[1])*alphaext[6];C51bis
 [1] 3582.856
R> C52bis <- C5[1]+(gammaest[3]-gammaest[1])*alphaext[6];C52bis
 [1] 4669.656
R> C53bis <- C5[1]+(gammaest[4]-gammaest[1])*alphaext[6];C53bis
 [1] 5689.281
R> C54bis <- C5[1]+(gammaest[5]-gammaest[1])*alphaext[6];C54bis
 [1] 6338.527
R> C55bis <- C5[1]+(gammaest[6]-gammaest[1])*alphaext[6];C55bis
 [1] 6607.503
R> Provestext <- C15bis+C25bis+C35bis+C45bis+C55bis-C1[5]-C2[4]-C3[3]-C4[2]-C5
R> Provestext
 [1] 10257.84
```

Hemos calculado $V11$ *con* $\hat{\alpha}^{external}$ *y* $\hat{\gamma}^{external}$*, obteniendo una provisión total de* 9964 *y también* $V13$ *con* $\hat{\alpha}^{external}$ *y* $\hat{\gamma}^{CL}$*, obteniendo una provisión de* 10258*. El resto de resultados para los diferentes escenarios se pueden encontran resumidos en la Tabla 5 de [Schmidt, K. D. and Zocher, M. (2008)] al igual que los calculados en este ejemplo.*

El rango de todas las provisiones es:

Máximo: 11987 *(con* $V43$*).*

Mínimo: 9872 *(con* $V14$*).*

Rango = *Máximo* − *Mínimo* = 2115*.*

Los rangos no son probabilísticos pero reflejan la incertidumbre de los diferentes medios de información según las versiones del método (ver página 108 de [Schmidt, K. D. and Zocher, M. (2008)]).

En las siguientes tablas 4.5 y 4.6 se muestran las estimaciones de los pagos para los dos casos estudiados en este ejemplo:

Tabla 4.5: Estimaciones del caso $V11$ de los pagos por siniestros de [Schmidt, K. D. and Zocher, M. (2008)].

	0	1	2	3	4	5
0	1001	854	568	565	347	148
1	1113	990	671	648	422	**199.0**
2	1265	1168	800	744	**415.8**	**231.0**
3	1490	1383	1007	**849.0**	**509.4**	**283.0**
4	1725	2536	**1117.8**	**931.5**	**558.9**	**310.5**
5	1889	**1582.5**	**1139.4**	**949.5**	**569.7**	**316.5**

Tabla 4.6: Estimaciones del caso $V13$ de los pagos por siniestros de [Schmidt, K. D. and Zocher, M. (2008)].

	0	1	2	3	4	5
0	1001	854	568	565	347	148
1	1113	990	671	648	422	**169.1**
2	1265	1168	800	744	**473.9**	**196.3**
3	1490	1383	1007	**911.7**	**580.5**	**240.5**
4	1725	2536	**1066.2**	**1000.3**	**636.9**	**263.9**
5	1889	**1693.9**	**1086.8**	**1019.6**	**649.2**	**268.9**

4.8. Supuestos de Provisiones

1. Los datos respecto a los pagos anuales por siniestros ocurridos en los años de origen 0, 1, 2, 3 y 4, vienen recogidos en el siguiente triángulo run-off recogido en la Tabla 4.7:

Tabla 4.7: Triángulo con pagos anuales por siniestros, siendo las filas los años de origen y las columnas los años de desarrollo.

	0	1	2	3	4
0	3487.2	3003.1	1442.1	803	330
1	2972.2	3000.8	1155.3	955.6	
2	3312.4	3059.6	1431.9		
3	3894.7	3991.9			
4	4483.6				

Se pide:

a) Utilizando el método de Chain-Ladder, completar los datos del triángulo para obtener el rectángulo. Suponiendo que todos los siniestros ocurridos en un año se pagan completamente durante ese año y los cuatro siguientes, calcular las provisiones para siniestros pendientes de declaración o pago a finales del año 4, en los siguientes casos:

1) Sin tener en cuenta la rentabilidad futura, es decir como suma únicamente de los pagos futuros.
2) Teniendo en cuenta la estructura de tipos de interés libre de riesgo con ajuste por volatilidad de la tabla 4.8 extraída de eiopa, `https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii-technical-information/risk-free-interest-rate-term-structures`, y publicada con fecha 31-1-2016, que se corresponde con tasas anuales de cupón cero, y tratando a los pagos futuros como una renta.

Tabla 4.8: Estructura de tipos de interés libre de riesgo con ajuste por volatilidad publicada por eiopa con fecha 31-1-2016, correspondiente a tasas anuales de cupón cero.

Plazo	Interés anual
1	0.036 %
2	0.015 %
3	0.067 %
4	0.155 %
5	0.267 %
6	0.389 %
7	0.514 %
8	0.641 %
9	0.760 %
10	0.870 %

3) Teniendo en cuenta la estructura de tipos de interés libre de riesgo con ajuste por volatilidad de la tabla 4.8 y tratando a los pagos futuros como un flujo.

b) Aplicando las variantes del método de Chain-Ladder, calcular las provisiones para siniestros pendientes de declaración o pago a finales del año 4, teniendo en cuenta los mismos casos que en el apartado anterior.

2. Se dispone de los siguientes datos referentes a los pagos anuales, en miles de euros, por los siniestros ocurridos durante los últimos 5 años que se recogen en la Tabla 4.9:

Tabla 4.9: Triángulo con pagos por siniestros, siendo las filas los años de origen y las columnas los años de desarrollo.

	0	1	2	3	4
0	288	132	240	40	50
1	398	102	198	102	
2	530	170	90		
3	610	190			
4	715				

Se pide:

a) Utilizando el método de Chain-Ladder:

1) Completar los datos del triángulo para obtener el rectángulo.
2) Calcular las provisiones para siniestros pendientes de declaración o pago a finales del año de origen 4 teniendo en cuenta la estructura de tipos de interés libre de riesgo con ajuste por volatilidad de la tabla 4.8 y tratando a los pagos futuros como una renta, en los siguientes supuestos:
   - Suponiendo que todos los siniestros ocurridos en un año se pagan completamente durante ese año y los cuatro siguientes.
   - Suponiendo que durante el año de ocurrencia y los cuatro siguientes se paga el 85 % de la cuantía total a la que ascienden los siniestros

b) Calcular las provisiones para siniestros pendientes de declaración o pago a finales del año de origen 4 teniendo en cuenta la estructura de tipos de interés libre de riesgo con ajuste por volatilidad de la tabla 4.8 y tratando a los pagos futuros como una renta, utilizando el método de mínimos cuadrados de De Vylder.

c) Teniendo en cuenta que el número de siniestros de cada año, estimado a final de dicho año se recogen en la Tabla 4.10,

Tabla 4.10: Número de siniestros en cada año de origen, estimado a final de año para los datos de la Tabla 4.9.

Año de origen	Número estimado de siniestros
0	651
1	752
2	825
3	900
4	1135

y que la inflación (incluyendo, en su caso, otros factores causantes de una variación exponencial) futura se estima en el 5 % anual, aplicando el método de separación aritmética de Taylor, construir el triángulo ajustado y completar el rectángulo. Calcular las provisiones para siniestros pendientes a finales del año de origen 4 teniendo en cuenta la estructura de tipos de interés libre de riesgo con ajuste por volatilidad de la tabla 4.8 y tratando a los pagos futuros como una renta.

Bibliografía

[ICEA (2017)] ALEGRE, A., BADÍA, C., BOJ, E., BOSCH, M., CASANOVAS, M., CASTAÑER, A., CLARAMUNT, M.M., COSTA, M.T., GALISTEO, M., GONZÁLEZ-VILA, L., MÁRMOL, M., MARTÍNEZ DE ALBÉNIZ, F.J., MORILLO, I., ORTÍ, F.J., PONS, M.A., PREIXENS, T., RIBAS, C., ROCH, O, SÁEZ, J., SARRASÍ, F.J. y VAREA, J., *Teoría general del seguro*. Asociación ICEA, Madrid 2017.

[Asmussen (2014)] ASMUSSEN, S., Modeling and performance of Bonus-Malus systems: stationary versus age-correction, *Risks*, **2**, 49-73, 2014.

[Baxter et al. (1980)] BAXTER, L.A., COUTTS, S.M. and ROSS, G.A.F., Applications of linear models in motor insurance, *AProceedings of the 21st International Congress of Actuaries*, Zurich, Society of Actuaries, 11-29, 1980.

[Berquist and Sherman (1977)] BERQUIST, J.R. and SHERMAN, R.E., Loss reserve adequacy testing: a comprehensive systematic approach, *Proceedings of the Casualty Actuarial Society*, **64**, 123-185, 1977.

[Boj et al. (2004)] BOJ, E., CLARAMUNT, M.M. y FORTIANA, J., *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación*. Cuadernos de la Fundación MAPFRE, **88**, Fundación MAPFRE Estudios, Madrid, 2004.

[Boj y Costa (2014)] BOJ, E. y COSTA, T., *Modelo lineal generalizado y cálculo de la provisión técnica*. Depósito digital de la Universidad de Barcelona, colección OMADO, 2014.
URL `http://hdl.handle.net/2445/49068`

[Boj et al. (2014)] BOJ, E., COSTA, T. y ESPEJO, J., Provisiones técnicas por años de calendario mediante modelo lineal generalizado. Una aplicación con RExcel, *Anales del Instituto de Actuarios Españoles*, **20**, 83-116, 2014.

[Bornhuetter and Ferguson (1972)] BORNHUETTER, R.L. and FERGUSON, R.E., The actuary and IBNR, *Proceedings of the Casualty Actuarial Society*, **59**, 181-195, 1972.

[Brockman and Wright(1992)] BROCKMAN, M.J. and WRIGHT, T.S., Statistical motor rating: making effective use of your data, *Journal of the Institute of Actuaries*, **119:3**, 457-543, 1992.

[Charpentier (2014)] CHARPENTIER, A., *Computational actuarial science with R*, Chapman and Hall, New York, 2014.

[Claramunt y Costa (2003)] CLARAMUNT, M.M. y COSTA, T., *Matemática actuarial no vida. Un enfoque práctico*. Colección de Publicaciones del Departamento de Matemática Económica, Financiera y Actuarial de la Universidad de Barcelona, **63**, 2003.

[Clarke and Harland (1974)] Clarke, T.G. and Harland, N., A practical statistical method of estimating claims liability and claims cash flow, *ASTIN Bulletin*, **8**, 26-37, 1974.

[Claramunt et al. (2017)] Claramunt, M.M., Costa, T. y Boj, E., *provisio: Área de trabajo en lenguaje R para el cálculo de provisiones técnicas en seguros no de vida con métodos deterministas.* Depósito digital de la Universidad de Barcelona, colección de investigación-software, 2017.
URL `http://hdl.handle.net/2445/106653`

[Denuit y Charpentier (2005)] Denuit, M. and A. Charpentier, *Mathématiques de l'assurance non-vie. Tome II: tarification et provisionnement*, Economica, 2005.

[De Vylder (1978)] De Vylder, F., Estimation of IBNR claims by least squares, *Mittelungen der Vereinigung Schewizerischer Versicherungsmathematike*, **78**, 249Ű254, 1978.

[De Vylder (1986)] De Vylder, F.E, *Advanced risk theory*, Editions de l'Université de Bruxelles, Swiss Association of Actuaries, 1986.

[England and Verrall (1999)] England, P.D. and Verrall, R.J., Analytic and bootstrap estimates of prediction errors in claims reserving, *Insurance: Mathematics and Economics*, **25**, 281-293, 1999.

[England and Verrall (2002)] England, P.D. and Verrall, R.J., Stochastic claims reserving in general insurance (with discussion), *British Actuarial Journal*, **8**, 443-544, 2002.

[England and Verrall (2006)] England, P.D. and Verrall, R.J., Predictive Distributions of Outstanding Liabilities in General Insurance, *Annals of Actuarial Science*, **1:II**, 221-270, 2006.

[Gesmann et al. (2017)] Gesmann, M., Murphy, D., Zhang, Y., Carrato, A., Crupi, G., Dutang, Ch., Lacoume, A., Charpentier, A., Wuthrich, M. and Concina, F., *ChainLadder*, Paquete de R en Cran, versión 0.2.4, 2017.
URL `http://cran.r-project.org/package=ChainLadder`

[Haberman and Renshaw A.E. (1996)] Haberman, S. and Renshaw, A.E., Generalized linear models and actuarial science, *The Statistician*, **45:4**, 407-436, 1996.

[Kaas et al. (2008)] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M., *Modern actuarial risk theory: using R. Second edition*, Springer-Verlag, Heidelberg, 2008.

[Konis et al. (2012)] Konis, K., Riani, M. and Scrucca, L., *Forward: Forward search*, Paquete de R en Cran, versión 1.03, 2012.
URL `http://CRAN.R-project.org/package=forward`

[Kramreiter and Straub (1973)] Kramreiter, H. and Straub, E., On the calculacation of IBNR reserves II, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **73**, 177-190, 1973.

[Lemaire (1995)] Lemaire, J., *Bonus-Malus systems in automobile insurance*, Kluwer Academic Publishers, Norwell, 1995.

[Mack, T. (1993)] Mack, T., Distribution-free calculation of the standard error of chain-ladder reserve estimaties, *Astin Bulletin*, **23:2**, 213-225, 1993.

[Mahmoudvand (2013)] Mahmoudvand, R., Edalati, A. and F. Shokoohi, Bonus-Malus system in Iran: an empirical evaluation, *Journal of Data Science*, **11**, 29-41, 2013.

[McCullagh and Nelder (1989)] McCullagh, P. and Nelder, J., *Generalized linear models (2nd edition)*, Chapman and Hall, London, 1989.

[Meyer (2000)] Meyer, U., *Third party motor insurance in Europe. A comparative study in the economical-statistical situation*, University of Bamberg, 2000.

[Partrat y Besson (2005)] Partrat, C. and J.-L. Besson, *Assurance non-vie. Modélisation, simulation*, Economica, 2005.

[R Development Core Team (2018)] R Development Core Team, *R: a language and environment for statistical computing*, Vienna, Austria, 2018.
URL `http://www.R-project.org/`

[Rolski et al. (1999)] Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J., *Stochastic processes for insurance and finance*, Wiley, Chichester, 1999.

[Schmidt, K. D. and Zocher, M. (2008)] Schmidt, K. D. and Zocher, M., The Bornhuetter-Ferguson principle, *Variance*, **2:1**, 85-110, 2008.

[Skurnick (1973)] Skurnick, D., A survey of loss reserving methods, *Proceedings of the Casualty Actuarial Society*, **60:113**, 16-58, 1973.

[Tarbell (1934)] Tarbell, T.F., Incurred but not reported claims reserves, *Proceedings of the Casualty Actuarial Society*, **20**, 275Ů280, 1934.

[Taylor (1979)] Taylor, G.C., Statistical testing of a non-life insurance model, Appears in: *de Vylder and Goovaerts (1979), Proceedings of the contact goup Actuarial Sciences. Mimeograph published by Instituut voor Actuariele Wetenschappen. Katholieke Universiteit te Leuven, Belgium, 37-64*, 1979.

[Taylor (1986)] Taylor, G.C., Taylor, G.C., *Claim reserving in non-life insurance*, North Holland, Insurance Series 1, 1986.

[Van Eeghen and De Vylder (1981)] Van Eeghen, J., De Vylder, F.E, *Loss reserving methods*, Surveys of Actuarial Studies, n. 1, National Nederlanden, 1981.

[Van Eeghen et al. (1983)] Van Eeghen, J., Greup, E.K. and Nijssen, J.A., *Rate making*, Surveys of Actuarial Studies, n. 2, National Nederlanden, 1983.

[Verbeek (1972)] Verbeek, H.G., An approach to the analysis of claims experience in motor liability excess loss reinsurance, *ASTIN Bulletin*, **6**, 195-202, 1972.