Introducción a la Tarificación en Seguros No Vida

Antonio Heras
Universidad Complutense de Madrid

Diciembre de 2015

El Seguro

- Una operación de Seguro consiste en la transferencia del riesgo de una pérdida a cambio de una cierta cantidad de dinero (la ***prima***).
- Una operación de Seguro consiste en el intercambio de una pérdida de cuantía incierta y posiblemente grande por una pérdida cierta de cuantía pequeña (la ***prima***).
- Se denomina ***Tarificación*** al proceso de cálculo de una prima suficiente y justa.

Características de los riesgos asegurables

- No todos los riesgos son asegurables. En teoría, los riesgos asegurables deberían tener ciertas características:
- Debería haber un gran número de riesgos asegurados (los riesgos deberían estar **agrupados**, “pooled”, en **carteras grandes**).
- Los riesgos deberían ser **homogéneos**.
- Los riesgos deberían ser **independientes**.

La importancia de las carteras grandes

- En una situación ideal, deberíamos tener un gran número de asegurados, cada uno de los cuales paga una prima pequeña. El dinero del fondo así creado se usará para compensar a los pocos asegurados que hayan sufrido pérdidas.
- Agrupando los riesgos individuales, el riesgo global de la aseguradora es menor que la suma de los riesgos de los asegurados considerados independientemente. Cuanto más grande es el grupo (la cartera), más predecible resulta su comportamiento.

Importancia de la Independencia

- Las pérdidas deberían ser independientes: El que un asegurado sufra una pérdida no debería afectar a las pérdidas de los demás.
- Por ejemplo, una compañía aseguradora no debería asegurar todos los edificios de una misma zona geográfica contra incendios o terremotos, debido a la posibilidad de sufrir pérdidas catastróficas.

Importancia de la Homogeneidad

- Si los riesgos son heterogéneos, la compañía aseguradora está expuesta al problema de la ***Selección Adversa***: los asegurados con poco riesgo abandonarán la cartera, y los asegurados con mayor riesgo se quedarán.
- En el caso de que la cartera sea heterogénea, la compañía debería subdividirla en subcarteras homogéneas.

La Prima Pura

- El componente más importante de la prima es la llamada ***prima pura***, definida como la esperanza matemática de la siniestralidad: P=E(S).
- Si se cumplen las condiciones, la ***ley de los grandes números*** implica que la siniestralidad media estará cerca de la prima pura:

$$P = E\left(S\right) \simeq \overline{S}_n = \frac{S_1 + ... + S_n}{n}$$

La ley de los Grandes Números

Bajo las condiciones anteriores, la Ley de los Grandes Números establece que:

$$\lim_{n\to\infty} \Pr\left[\left|\overline{S}_n - E(S)\right| > \varepsilon\right] = 0, \forall \varepsilon > 0$$

Si S es la siniestralidad total y P=E(S) es la prima pura, la Ley de los Grandes Números asegura que la prima pura estará cerca del valor observado de la siniestralidad media.

La Prima Recargada

- Pero que la prima pura esté cerca de la siniestralidad media no implica que los ingresos totales por dichas primas sean también cercanos a la siniestralidad total.
- Para conseguir lo anterior, la prima pura se recarga con un Recargo de Seguridad, dando lugar a la ***Prima Recargada***.
- Aquí nos vamos a ocupar únicamente de la prima pura.

Un ejemplo de heterogeneidad

- Consideremos una cartera de 10 asegurados, y supongamos que cada uno de ellos solo puede tener, como mucho, un siniestro al año, de cuantía unidad. La “prima colectiva” para esta cartera se estima en 0.2, y esta será la prima pagada por cada asegurado el primer año.

Siniestros de la cartera en 1 año (siniestralidad media = 0.1)

Insured	1	2	3	4	5	6	7	8	9	10
Year 1	0	0	0	0	0	0	0	0	1	0

Siniestros en 2 años (siniestralidad media = 0.2)

Insured	1	2	3	4	5	6	7	8	9	10
Year 1	0	0	0	0	0	0	0	0	1	0
Year 2	1	1	0	0	0	0	0	0	1	0

Insured	1	2	3	4	5	6	7	8	9	10
Year 1									1	
Year 2	1	1							1	
Year 3	1		1						1	
Year 4			1						1	
Year 5									1	
Year 6		1								
Year 7	1	1		1	1					
Year 8	1			1		1			1	
Year 9	1				1					
Year 10	1								1	
Av claim amount	0.6	0.3	0.2	0.2	0.2	0.1	0	0	0.7	0

- Aunque la prima colectiva en este ejemplo puede ser adecuada globalmente, claramente no es una prima justa: algunos asegurados (1 y 9) deberían pagar una prima mayor, mientras que otros (7,8 y 10) deberían pagar menos.

- La heterogeneidad de la cartera puede llevar a una selección negativa de los asegurados (7,8 y 10 podrían abandonar la compañía aseguradora), lo que puede dar lugar a problemas de solvencia.

Los Factores de Riesgo

- Para obtener subcarteras homogéneas, es importante reconocer las diferentes características de los asegurados que tienen influencia en su nivel de riesgo. Estas características se denominan **Factores de Riesgo** (Risk Factors).
- En el Seguro de Vida los Factores de Riesgo son la edad, el sexo (que ya no se puede tener en cuenta), el estilo de vida (fumador / no fumador, etc.), la predisposición genética a las enfermedades (que se podrá medir en el futuro), estado de salud, etc.

Problemas de los Factores de Riesgo

En algunos seguros No Vida como el Seguro de Autos, los Factores de Riesgo pueden ser…

- Difíciles de cuantificar: ¿cómo se mide el temperamento de un conductor?
- Difíciles de comprobar: ¿cómo se mide el grado de consumo de alcohol o drogas al volante? ¿y el número de kilómetros conducidos?
- Políticamente delicados: pensemos en el sexo o nacionalidad de un conductor.

- A veces se distingue entre **Factores de Riesgo** (que a menudo son difícilmente observables) y **Factores de Tarificación** (Rating Factors), que sí son observables y están correlacionados con los anteriores.

  Por ejemplo, en el Seguro de Autos, hay tres categorías de factores de tarificación:
- Relacionados con los asegurados: edad del conductor, sexo, profesión, estado civil,…
- Relacionados con el vehículo: edad y modelo de coche, potencia del motor,…
- Relacionados con la región geográfica: provincia, tipo de área residencial (ciudad grande o pequeña, zona rural, etc.)…

Más problemas de los Factores de Riesgo

- En muchas ramas del seguro existe un gran número de factores de riesgo o tarificación que tienen influencia en la siniestralidad (el seguro de Autos es el ejemplo típico). Esto puede tener como consecuencia un excesivo número de clases de tarificación.

- Por ejemplo, 4 factores con 10 valores posibles cada uno darían lugar a 10000 clases de tarificación. Muchas de estas clases estarán vacías o casi vacías, y por tanto proporcionarán muy poca información sobre la siniestralidad.
- Por otra parte, si reducimos el número de clases, ya no podremos mantener la hipótesis de homogeneidad.

Métodos de Tarificación

Dos grandes familias de métodos:

- ***Métodos de Tarificación a priori***, que tienen en cuenta los factores de riesgo para segmentar la cartera en clases y para calcular la prima pura en cada clase.
- ***Métodos de Tarificación a posteriori*** o basados en la experiencia, que toman en cuenta la siniestralidad pasada de cada póliza para calcular la prima: Bonus-Malus, Credibilidad,… (no hablaremos de ellos)

Modelo del Riesgo Colectivo

- Supongamos que ya hemos conseguido una segmentación satisfactoria en clases homogéneas. Para calcular la prima pura de cada clase necesitamos ciertas hipótesis.
- Supondremos que la Siniestralidad Total S de cada póliza se puede calcular a partir del Número de Siniestros N y la cuantía individual X. Esto se conoce como el ***Modelo del Riesgo Colectivo***:

$$S = X_1 + \ldots + X_N$$

- Asumimos que las Xs son i.i.d. e independientes de N.

Modelo del Riesgo Colectivo

- En este modelo se tiene que

Prima Pura = E(S) = E(N) x E(X)

Prima Pura = Frecuencia media x Cuantía media

Pure Premium = Claim Frequency x Claim Severity

Los métodos de tarificación a priori suelen estimar E(N) y E(X) independientemente, a partir de los factores de riesgo.

Modelos Aditivos y Multiplicativos

- Supongamos solo dos factores, por ejemplo edad del conductor y uso del vehículo. Si el efecto de cada nivel de factor sobre la prima pura se calcula mediante una suma, tenemos un ***Modelo Aditivo***. Si se calcula mediante un producto, tenemos un ***Modelo Multiplicativo*** (el que realmente se usa en la práctica)

$$P\_Pura_{ij} = \mu_0 + \alpha_i + \beta_j$$

$$P\_Pura_{ij} = \eta_0 . \gamma_i . \delta_j$$

Datos Coste medio y exposure (Mildenhall, 1999)

	Business	DriveLong	DriveShort	Pleasure
Age 17-20	797,8 (5)	244,52 (23)	274,78 (40)	250,48 (21)
Age 21-24	362,23 (44)	298,13 (92)	298,6 (171)	213,71 (63)
Age 25-29	342,31 (129)	297,9 (318)	248,56 (343)	250,57 (140)
Age 30-34	367,46 (169)	293,87 (361)	228,48 (448)	229,09 (123)
Age 35-39	256,21 (166)	238,21 (381)	201,67 (479)	153,62 (151)
Age 40-49	352,49 (304)	236,06 (719)	202,80 (970)	208,59 (245)
Age 50-59	340,56 (162)	253,63 (504)	202,67 (859)	207,57 (266)
Age >60	342,58 (96)	259,79 (312)	196,33 (578)	192 (260)

Los mismos datos en otro formato

	Age	***Use***	***Severity***	***Claim Count***
1	17-20	Pleasure	250,48	21
2	17-20	D_Short	274,78	40
3	17-20	D_Long	244,52	23
4	17-20	Business	797,8	5
5	21-24	Pleasure	213,71	63
6	21-24	D_Short	298,13	171
…	…	…	…	…
…	…	…	…	…
32	>60	Business	342,58	96

Modelos Aditivos

Se basan en la Regresión Lineal Simple

$$Y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$$

Donde ε está distribuido como Normal de media 0, etc. Equivalentemente, Y_i es normal, de media

$$\mu_i = \beta_0 + \sum_k \beta_k x_{ik}$$

Resultados de la Regresión Lineal

Inter cept	17 20	21 24	25 29	30 34	35 39	40 49	50 59	>60	Buss	Long	Short	Pleas
194,82	70,48	63,58	43,89	34,94	-19,48	0,53	4,04	---	132,28	53,96	8,75	---
sí	sí	sí	sí	sí	¿?	no	no	---	sí	sí	no	---

Cálculo de primas en el Modelo Aditivo

- Por ejemplo, la prima pura asociada a la clase (Business, 50-59) sería:

  194,82 (Intercept) + 132,28 (efecto Business) + 4,04 (efecto edad 50-59) = 331,14

Ajuste del Modelo Aditivo

	Business	DriveLong	DriveShort	Pleasure
Age 17-20	397,58	319,26	274,05	265,30
Age 21-24	390,68	312,36	267,16	258,40
Age 25-29	370,99	292,67	247,46	238,71
Age 30-34	362,04	283,72	238,52	229,76
Age 35-39	307,62	229,30	184,09	175,34
Age 40-49	327,63	249,32	204,11	195,35
Age 50-59	331,14	252,82	207,62	198,86
Age >60	327,10	248,78	203,57	194,82

Modelos basados en GLMs

- La respuesta Y_i puede ser cualquier distribución de la familia exponencial, no necesariamente normal.
- La relación entre las medias μ_i y los predictores viene dada por una función LINK g:

$$g(\mu_i) = \beta_0 + \sum_k \beta_k x_{ik}$$

Modelos Multiplicativos

- Cuando la función LINK es logarítmica, tenemos un Modelo Multiplicativo:

$$Log(\mu_i) = \beta_0 + \sum_k \beta_k x_{ik}$$

$$\mu_i = \exp(\beta_0).\prod_k \exp(\beta_k x_{ik})$$

Resultados del GLM (LINK = Log, respuesta Gamma)

Inter cept	17 20	21 24	25 29	30 34	35 39	40 49	50 59	>60	Buss	Long	Short	Pleas
5,273	0,268	0,263	0,187	0,145	-0,072	0,007	0,022	---	0,497	0,234	0,041	---
sí	sí	sí	sí	sí	¿?	no	no	---	sí	sí	no	---

Cálculo de primas en el Modelo Multiplicativo

- Por ejemplo, la prima pura asociada a la clase (Business, 50-59) sería:

  Exp(5,273 (Intercept) + 0,4972 (efecto Business) + 0,022 (efecto edad 50-59)) =
  = Exp(5,273).Exp(0,4972).Exp(0,022) =
  (195).(1,6441).(1,02224) = 327,729

Ajuste del Modelo Multiplicativo

	Business	DriveLong	DriveShort	Pleasure
Age 17-20	419,05	322,14	265,55	254,88
Age 21-24	417,09	320,63	264,30	253,69
Age 25-29	386,68	297,26	245,03	235,19
Age 30-34	370,52	284,83	234,79	225,36
Age 35-39	298,36	229,36	189,07	181,47
Age 40-49	322,79	248,14	204,55	196,33
Age 50-59	327,73	251,94	207,68	199,34
Age >60	320,60	246,46	203,16	195,00

GLM vs Regresión Lineal

- Los GLMs generalizan al modelo clásico de Regresión Lineal Múltiple porque:
- Las respuestas pueden pertenecer a cualquier distribución de la familia exponencial, no necesariamente normal. Pueden, por tanto, ser asimétricas.
- La relación entre medias y predictores no es directa, sino que viene dada por una cierta función LINK.
- Si la respuesta es normal y LINK = Identidad, entonces GLM = Regresión Lineal Múltiple.

En resumen, los GLMs…

- Son muy flexibles y engloban una gran cantidad de modelos estadísticos, entre ellos el modelo clásico de regresión.
- Son una herramienta muy útil para la tarificación, ya que…
- Aconsejan sobre la partición en clases, y asocian una tarifa a cada clase, incluso a las clases con pocos datos, suavizando la distribución global.
- Aunque las primas de mercado pueden diferir de las teóricas, por razones comerciales o regulatorias, siempre resulta útil la comparación entre ambos tipos de primas.

Programación R

- model1 = lm(Severity~factor(Age)+factor(Vehicle_Use), data=coste_siniestros, weights=Claim_Count)

- model2 = glm(Severity~factor(Age)+factor(Vehicle_Use), data=coste_siniestros, weights=Claim_Count, family=Gamma(link="log"))

- model3 = glm(Severity~factor(Age)+factor(Vehicle_Use), data=coste_siniestros, weights=Claim_Count, family=gaussian (link="identity"))

Programación SAS

```
proc genmod data = coste_siniestros;
 class age vehicle_use;
 model severity = age vehicle_use / dist=normal  link=identity;
 weight claim_count;
 run;

 proc genmod data = coste_siniestros;
 class age vehicle_use;
 model severity = age vehicle_use / dist=gamma  link=log;
 weight claim_count;
 run;
```

Gracias por su atención