Modelo estadístico aplicable a colectivos abiertos

Juan Antonio Pérez López
Actuario de Seguros

INTRODUCCION

La obtención del modelo estadístico de un colectivo abierto supone la determinación de la distribución de todos los elementos del colectivo, que tienen una edad fijada, en torno a su valor medio. Es claro, que la afirmación anterior equivale a decir que ha de determinarse cuál es la probabilidad de que existan n elementos de edad x dentro del colectivo.

Este es el problema abordado en el presente estudio y vamos a exponer en unas líneas muy generales cuál ha sido el método para su resolución, con el fin de que se posea en todo momento la visión de conjunto necesaria para la perfecta interpretación de los resultados parciales que van obteniéndose.

Diremos, en primer término, que por colectivo abierto entendemos aquel en que el número de elementos a él pertenecientes, fluctúa impulsado por dos fuerzas de signo contrario: una la entrada de nuevos elementos, y otra la salida de algunos de los que estaban dentro de él.

Nos apoyaremos en el proceso genético de dicho colectivo, es decir, consideraremos la entrada de un número de elementos de edad x (edad primera) y, en la edad siguiente $x + 1$, nos encontramos con que el total de elementos es la suma de los supervivientes de la edad anterior, más todos los que entran en aquél momento, y que han de ser de edad $x + 1$ precisamente; de igual modo sucedería en las edades sucesivas. De este modo, el tiempo en el colectivo vendrá medido en el sentido biológico, es decir, por las edades de sus componentes. Es evidente que la imagen total del colectivo obtenida de este modo, será idéntica a la que proporcione cualquier colectivo real que se considere; no hemos hecho más que descomponer su propia génesis. Es algo paralelo a lo que en

Biometría representa el estudio de una generación ideal (en la que se observan los cambios que el tiempo introduce en la cuantía de elementos que la componen, partiendo de un número fijo, y no admitiendo más que salidas del Grupo) cuando se establecen los modelos estadísticos de los colectivos cerrados.

La solución del problema que nos ocupa la abordaremos mediante el estudio del proceso estocástico que se engendra al considerar el curso del colectivo en el tiempo (tiempo biológico); dicho proceso estocástico es la imagen estadística de todas las posibles formas que puede adoptar el colectivo; cada una de estas formas lleva aparejada una probabilidad de existencia, y la curva modelo del colectivo (que será la que una los valores medios de las distribuciones correspondientes al número de individuos pertenecientes al colectivo para cada edad) será, precisamente, la de máxima probabilidad de existencia.

Al final del trabajo trataremos de modo sucinto algunas cuestiones, referentes a la aplicación práctica de la teoría desarrollada.

Observaremos, por último, que aunque a lo largo del presente estudio nos limitamos a considerar colectivos abiertos con una sóla causa de entrada y otra de salida, sus resultados son inmediatamente generalizables siguiendo un método idéntico o paralelo al que utilizaremos.

1.—IMAGEN ESTADÍSTICA DEL COLECTIVO ABIERTO

Al observar el desenvolvimiento de este colectivo en el tiempo, estudiando la incorporación y eliminación de sus elementos en función de dicho tiempo (biológico), obtenemos los siguientes resultados:

En el instante x (o para la edad x), al menos teóricamente, puede haber dentro del colectivo un número de elementos (miembros de edad x) que oscila entre 0 e ∞; cada uno de los números de este campo tiene una probabilidad de existencia, por lo cual, para cada edad x, tendremos una variante, cuya función de probabilidad nos dará la de que en aquel momento haya precisamente n elementos dentro del colectivo. El campo de variación de esta variante será teóricamente $(0, \infty)$. Cada instante (o edad) x tiene su correspondiente variante η_x; el estudio de la forma que va adoptando el colectivo en el transcurso de este tiempo (es decir, al recorrer las diversas edades que pueden tener sus miembros) es, sencillamente, el estudio de esta variante dependiente de x, lo que nos conduce, en definitiva, a la consideración de un *Proceso estocástico* (que será de parámetro discreto si se toman en cuenta solamente edades enteras, tal como se realiza en la práctica).

Sólo nos interesarán las variaciones que la edad introduce en el colectivo, por lo que el proceso estocástico será *estacionario*. Con este convenio lo que hacemos es suponer que el número de elementos de edad x dentro del colectivo depende de dicha edad, pero no del *momento histórico* (o lugar del eje del *tiempo físico o absoluto*) en que estemos situados.

II.—DISTRIBUCION DE LA VARIANTE η_x

Determinación por un método estadístico de la distribución de η_x

Vamos a estudiar primero la distribución del número de elementos de edad x que entran a formar parte del colectivo en el intervalo de tiempo (0,t), siendo $\lambda_x\,dt$ la probabilidad de que entre un individuo de edad x en el intervalo (0,dt); λ_x depende solamente de x y suponemos que la edad x comprende individuos de edad entre $(x-\frac{1}{2}, x+\frac{1}{2})$ tal como se realiza en la práctica; utilizamos, pues, sólo edades enteras. Es de resaltar que no ofrecería ninguna dificultad el dejar de considerarlo así (siendo entonces variable de la ecuación diferencial siguiente la misma x), y si lo efectuamos de este modo, se debe tan sólo a que, si determinamos la distribución que nos ocupa por otro método que indicaremos someramente, las fórmulas que obtendríamos serían idénticas a las que pudiésemos conseguir ahora sin dicha suposición. Por otra parte, el razonamiento estadístico quedará así más claro.

De acuerdo con lo anterior y siendo $P_n(x,t)$ la probabilidad de que entren n elementos de edad x en el intervalo (0,t) ($t \leqq 1$ y haciendo coincidir el intervalo (0,1) con el $(x-\frac{1}{2}, x+\frac{1}{2})$), tendremos la siguiente ecuación diferencial.

$$P_n(x,t+dt) = P_{n-1}(x,t)\,\lambda_x\,dt + P_n(x,t)\,[1-\lambda_x\,dt]$$

$$\frac{d\,P_n(x,t)}{dt} = P_{n-1}(x,t)\,\lambda_x - P_n(x,t)\,\lambda_x = P'_n(x,t)$$

cuya solución es:

$$P_n(x,t) = e^{-\int_0^t \lambda_x\,dt}\left[c + \int_0^t P_{n-1}(x,t)\,\lambda_x\,e^{\int_0^t \lambda_x\,dt}\,dt\right]$$

Como para $t = 0$ es ($P_n(x,0) = 0$) resulta ser $c = 0$.

Vamos a determinar $P_n(x,t)$ por inducción completa; de la ecuación diferencial tenemos:

$$P'_0(x,t) = -P_0(x,t)\lambda_x \quad ; \quad P_0(x,t) = e^{-\lambda_x t}$$

$$P_1(x,t) = e^{-\lambda_x t}\int_0^t e^{-\lambda_x t}\lambda_x e^{\lambda_x t}\,dt = e^{-\lambda_x t}\lambda_x t$$

$$P_2(x,t) = e^{-\lambda_x t}\frac{(\lambda_x t)^2}{2}$$

Suponiendo cierto que $P_{n-1}(x,t) = e^{-\lambda_x t}\dfrac{(\lambda_x t)^{n-1}}{\underline{|n-1}}$ veamos si se cumple $P_n(x,t) = e^{-\lambda_x t}\dfrac{(\lambda_x t)^n}{\underline{|n}}$

$$P_n(x,t) = e^{-\lambda_x t}\int_0^t e^{-\lambda_x t}\frac{(\lambda_x t)^{n-1}}{\underline{|n-1}}e^{\lambda_x t}\,dt = e^{-\lambda_x t}\frac{(\lambda_x t)^n}{\underline{|n}}$$

Si $t=1$, $P_n(x,t) = P_n(x)$ pues x recoge el intervalo $(x-\frac{1}{2}\,,\ x+\frac{1}{2})$ que coincide con el $(0,1)$ para t; es decir, $P_n(x,1) = P_n(x)$ es la probabilidad de que ingresen en el colectivo n elementos de edad entre $(x-\frac{1}{2}\,,\ x+\frac{1}{2})$ en el intervalo $(x-\frac{1}{2}\,,\ x+\frac{1}{2})$ de la vida del mismo colectivo.

$$P_n(x) = e^{-\lambda_x}\frac{\lambda_x^n}{\underline{|n}}. \tag{1}$$

Tenemos, pues, que el número de elementos de edad x que ingresan en el colectivo sigue la ley de Poisson.

Sabemos, por otra parte, que el número n de individuos (supervivientes de un grupo de N) que alcanzan la edad x, sigue la ley:

$$\binom{N}{n}{}_{x-x_0}p_{x_0}^n\,(1-{}_{x-x_0}p_{x_0})^{N-n} \tag{2}$$

en donde ${}_{x-x_0}p_{x_0}$ es la probabilidad de que un individuo de edad x_0 alcance la edad x.

La misma ley será válida (pues la demostración sería totalmente idéntica a la que en Biometría se sigue para (2)) para recoger la probabilidad de que estén aún n individuos dentro del colectivo, al cabo de $x-x_0$ años, de un grupo de N en x_0; ${}_{x-x_0}p_{x_0}$ será, en este caso, la probabilidad de *permanencia* en el colectivo.

Vamos ahora, partiendo de estos elementos, a estudiar la distribución de una variante τ_x. Esta variante es bidimensional y está compuesta de dos variantes unidimensionales (ξ_x, η_x). ξ_x (cuyo campo de variación es teóricamente $(0, \infty)$, tomando sólo valores enteros) representa el número total de elementos ingresados hasta aquel momento en el colectivo; η_x (con idéntico campo de variación) es la encargada de representar el número total de individuos que permanecen dentro del colectivo en dicho momento (o para dicha edad x). Por consiguiente, la variante τ_x al tomar cada par de valores (k_x, n_x) con una probabilidad, lo que nos indica dicha probabilidad es là de que ocurra el suceso de la permanencia de n individuos de edad x dentro del colectivo cuando eran en número de k los entrados en edades anteriores.

Estas variantes ξ_x, η_x son, evidentemente, dependientes. Ahora bien, a nuestros efectos, nos interesa exclusivamente la función de densidad marginal de η_x que será la que nos dé la probabilidad de tener en el colectivo n individuos de edad x, cualquiera que fuese el número de ellos que entrase en edades anteriores. A pesar de ello obtendremos la función característica de τ_x (de la que pasaremos a la función de densidad que constituye nuestro objeto de un modo muy sencillo) por el interés que pudiera presentar.

Nos encontramos, en primer lugar, con que la variante τ_x es una suma de x variantes bidimensionales τ_0^x, $\tau_1^x \ldots$; $[\tau_i^x = (\xi_i^x, \eta_i^x)]$ en las cuales la probabilidad de cada par de valores (k_i^x, n_i^x) que toma τ_i^x representa la probabilidad de que de k_i individuos de edad i que entraron en el colectivo, n_i permanezcan en el mismo a la edad x. Según esto, la función de densidad marginal de η_i^x, nos daría la probabilidad de que hubiese en el colectivo a la edad x, exactamente n individuos de todos los que entraron en él en la edad i.

Es, por otra parte, evidente que las τ_i^x son independientes, motivo por el cual, conocidas sus distribuciones, nos será fácil hallar la de τ_x.

De (1) y (2) obtenemos de un modo inmediato la función de densidad de η_i^x (número de individuos entrados a la edad i y que permanecen en el colectivo a la edad x) condicionada a ξ_i^x (número de individuos entrados a la edad i). Tenemos, en principio:

$$P\left(\eta_i^x = n_i \,\middle|\, \xi_i^x = k_i\right) = \binom{k_i}{n_i} {}_{x-i}p_i^{\,n_i} \left(1 - {}_{x-i}p_i\right)^{k_i - n_i}$$

y, de aquí, la función de densidad condicionada de η_i^x a ξ_i^x será:

$$f\left(\gamma_i^x \,\middle|\, z_i^x\right) = \binom{z_i^x}{\gamma_i^x} {}_{x-i}p_i^{\,\gamma_i^x} \left(1 - {}_{x-i}p_i\right)^{z_i^x - \gamma_i^x}$$

y, puesto que $f(z,\gamma) = f_1(z)\, f(\gamma/z)$, la función de densidad de la variante τ_i^x es:

$$f(t_i^x) = f(z_i^x, \gamma_i^x) = e^{-\lambda_i} \frac{\lambda_i^{z_i^x}}{\lfloor z_i^x} \binom{z_i^x}{\gamma_i^x}\, {}_{x-i}p_i^{\gamma_i^x} (1 - {}_{x-i}p_i)^{z_i^x - \gamma_i^x} \tag{3}$$

que tendrá como función característica: (vemos que para $\gamma_i^x > z_i^x$ se reduce todo a cero)

$$\varphi_{\tau_i^x}(u,v) =$$

$$= \sum_{z_i^x=0}^{\infty} e^{iuz_i^x} e^{-\lambda_i} \frac{\lambda_i^{z_i^x}}{\lfloor z_i^x} \sum_{\gamma_i^x=0}^{z_i^x} e^{iv\gamma_i^x} \binom{z_i^x}{\gamma_i^x}\, {}_{x-i}p_i^{\gamma_i^x} (1 - {}_{x-i}p_i)^{z_i^x - \gamma_i^x} =$$

$$= \sum_{z_i^x=0}^{\infty} e^{-\lambda_i} \frac{\left(e^{iu} \lambda_i \left[{}_{x-i}p_i\, e^{iv} + (1 - {}_{x-i}p_i)\right]\right)^{z_i^x}}{\lfloor z_i^x}$$

$$= e^{\lambda_i \left[e^{iu}\left[{}_{x-i}p_i\, e^{iv} + (1 - {}_{x-i}p_i)\right] - 1\right]} \tag{4}$$

Como las τ_i^x son independientes, la función característica de la variante suma, será igual al producto de las funciones características; así pues, la función característica de la variante que nos ocupa τ_x será:

$$\varphi_{\tau_x}(u,v) = \varphi_{\sum_{i=0}^{x} \tau_i^x}(u,v) = \prod_{i=0}^{x} \varphi_{\tau_i^x}(u,v) =$$

$$= e^{\sum_{i=0}^{x} \lambda_i \left[e^{iu}\left({}_{x-i}p_i\, e^{iv} + (1 - {}_{x-i}p_i)\right) - 1\right]} \tag{5}$$

de la cual obtenemos la función característica de la distribución marginal correspondiente a η_x dando sencillamente el valor cero al parámetro *u*.

$$\varphi_{\tau_x}(0,v) = \varphi_{\eta_x}(v) = e^{\sum_{i=0}^{x} \lambda_i \left[\left({}_{x-i}p_i\, e^{iv} + (1 - {}_{x-i}p_i)\right) - 1\right]} =$$

$$= e^{\sum_{i=0}^{x} \lambda_i \left[{}_{x-i}p_i\, (e^{iv} - 1)\right]} = e^{(e^{iv} - 1) \sum_{i=0}^{x} \lambda_i\, {}_{x-i}p_i} \tag{6}$$

que, como podemos comprobar, se trata de la función característica correspondiente a una distribución de Poisson de media

$$\sum_{i=0}^{x} \lambda_i\, {}_{x-i}p_i$$

En consecuencia, podemos afirmar que el número posible de elementos que estén dentro del colectivo a una edad x se distribuye según fa función de Poisson $e^{-\lambda} \frac{\lambda^n}{\underline{|n}}$ con media $\lambda = \sum_{i=0}^{x} \lambda_{x-i}\,{}_ip_i$, media que es, sencillamente, el número esperado de individuos que permanecerán dentro del colectivo de todos los entrados en edades anteriores (es, pues, una esperanza matemática total).

Determinación por otro procedimiento de la distribución problema

Vamos a dar unas indicaciones resumidas sobre la forma de efectuar la determinación de la distribución que nos ocupa por otro procedimiento distinto al anterior; este nuevo procedimiento tiene sobre el otro la ventaja de tener un planteo sencillísimo, y las únicas dificultades que presenta en su desarrollo son de tipo matemático. No obstante preferimos el que hemos expuesto, debido a que, a nuestro parecer, analiza mucho más profundamente la esencia del colectivo abierto, para servirse de sus características en la búsqueda de la solución. Con este segundo procedimiento los conceptos quedan relegados más bien a un segundo término, y todo el problema queda reducido a la simple solución de una ecuación diferencial.

Sea $P_n(x)$ la probabilidad de que nos encontremos en el colectivo con n individuos de edad x y hagamos las siguientes hipótesis:

δ_x dx es la probabilidad de salida, en el intervalo de tiempo dx, para un individuo del colectivo de edad x; se supone que la probabilidad de salida de dos individuos es infinitésimo con respecto a dx (es decir, es un infinitésimo de segundo orden para dx $\to 0$) por lo que, en lo sucesivo, no la tendremos en cuenta.

λ_x dx es la probabilidad de que entre a formar parte del colectivo un individuo de edad x, en el intervalo de tiempo dx. La probabilidad de dos entradas simultáneas no la tenemos en cuenta tampoco, por suponerla infinitésimo de segundo orden.

Observemos que λ_x dx es la probabilidad *con respecto al colectivo* de que entre un individuo en el intervalo dx, mientras que δ_x dx, es la probabilidad *para cada individuo* de salida del colectivo; así pues, la probabilidad *con respecto al colectivo* de salida de un individuo de edad x en el tiempo dx será (suponiendo n individuos de tal edad) igual a $n\,\delta_x\,dx$ pues, al no considerar los infinitésimos de segundo orden, que son las probabilidades de dos salidas simultáneas, nos resultan independientes las salidas para cada individuo.

De acuerdo con lo anterior podemos plantear la siguiente ecuación diferencial:

$$P_n(x+dx) = P_{n-1}(x)\,\lambda_x\,dx\,(1-(n-\tfrac{1}{2})\,\delta_x\,dx) +$$
$$+ P_n(x)\,(1-\lambda_x\,dx)\,(1-n\,\delta_x\,dx) +$$
$$+ P_{n+1}(x)\,(1-\lambda_x\,dx)\,(n+1)\,\delta_x\,dx + P_n(x)\,\lambda_x\,dx\,(n+\tfrac{1}{2})\,\delta_x\,dx$$

Eliminaremos infinitésimos de orden superior al primero, quedando

$$P'_n(x) = \frac{d\,P_n(x)}{dx} = P_{n-1}(x)\,\lambda_x + P_{n+1}(x)\,(n+1)\,\delta_x - P_n(x)\,\lambda_x -$$
$$- P_n(x)\,n\,\delta_x \qquad (7)$$

Abordaremos su integración utilizando la función característica (donde ponemos $e^{iv} = s$).

$$z = \varphi(s,x) = \sum_{n=0}^{\infty} s^n\,P_n(x) \quad ; \quad \frac{\partial z}{\partial x} = \sum_{n=0}^{\infty} s^n\,P'_n(x)$$

donde sustituyendo $P'_n(x)$ por su valor (7) obtendremos, tras algunas transformaciones, la siguiente ecuación diferencial de primer orden, en derivadas parciales, no homogénea:

$$\frac{\partial z}{\partial x} + \frac{\partial z}{\partial s}(s-1)\,\delta_x = \lambda_x\,(s-1)\,z \qquad (8)$$

Se integra a través de su sistema adjunto imponiendo como condición para la solución que se verifique

$$(s-1)\,e^{-\int_0^x \delta_x\,d_x} = K_o \qquad (9)$$

puesto que es condición cumplida por la ecuación correspondiente a ésta en los colectivos cerrados, teniendo que cumplirla también la que nos ocupa por ser más general e incluir a aquélla como caso particular.

Así resulta

$$z = e^{(e^{iv}-1)\int_0^x \lambda_t\,\frac{{}_xp_o}{{}_tp_o}\,dt} = e^{[e^{iv}-1]\int_0^x \lambda_t\,{}_{x-t}p_t\,dt} \qquad (10)$$

que vemos es la función característica correspondiente a la distribución de Poisson con media igual a $\int_0^x \lambda_t \, {}_{x-t}p_t \, dt$ (número esperado de individuos que alcanza la edad x dentro del colectivo, de todos los ingresados en edades anteriores).

Como puede comprobarse, las fórmulas (6) y (10) son iguales, salvo en lo que se refiere, naturalmente, al hecho de haber partido, para la demostración de (6), de considerar sólo edades enteras (por lo que la variable x tomó sólo valores enteros en dicha fórmula) con lo que la integral que aparece en (10) es un sumatorio en (6).

En el caso de que la salida del colectivo se deba tan sólo al fallecimiento de sus miembros, nos encontramos que $\delta_x = \mu_x$ (tanto instantáneo de mortalidad), siempre que éste μ_x esté calculado sobre los valores medios de individuos vivos a la edad de que se trate.

III.—ESTIMACION DE LOS PARÁMETROS

Tanto λ_r como ${}_{r-i}p_i$ pueden ser estimados por máxima verosimilitud a partir de una muestra, por sernos conocidas las leyes de distribución de la población original.

Supongamos, por ejemplo, que hemos observado la entrada de n_r elementos de edad r (esto para varios valores de r). La función de verosimilitud será (el muestreo es simplemente aleatorio):

$$L = \prod_{r=1}^{n} e^{-\overline{\lambda}_r} \frac{\overline{\lambda}_r^{\,n_r}}{\underline{|n_r}}$$

$$\log_e L = \sum_{r=1}^{n} \left[-\overline{\lambda}_r + n_r \log_e \overline{\lambda}_r - \log_e \underline{|n_r} \right]$$

$$\frac{\partial \log_e L}{\partial \overline{\lambda}_r} = 0 = -1 + \frac{n_r}{\overline{\lambda}_r} \quad ; \quad \overline{\lambda}_r = n_r \sim \lambda_r$$

Por lo tanto, el número medio de elementos que entran a cada edad se estima por el número de entradas observado para esa edad.

Si, en vez de tener observaciones sobre distintas edades, tenemos n observaciones para cada edad i, la estimación será (n_1, n_2, ... n_n son el número de entradas observadas para esa edad)

$$L = \prod_{r=1}^{n} e^{-\overline{\lambda}_i} \frac{\overline{\lambda}_i^{\,n_r}}{\underline{|n_r}}$$

y, siguiendo el camino anterior, obtendremos:

$$\overline{\lambda}_i = \frac{\sum_1^n n_r}{n} = \overline{x} \sim \lambda_i$$

es decir, se toma como estimador (que es centrado por ser la media muestral) el número medio de individuos entrados a esa edad.

Una aplicación exactamente igual cabe hacer con respecto a ${}_{r-i}p_i$. El resultado será, para el caso similar (en cuanto a las observaciones) al primero que planteábamos al estimar λ_r :

${}_{r-i}\overline{p}_i = \frac{n_r}{N_i} \sim {}_{r-i}p_i$ siendo n_r los individuos que permanecen en el colectivo a la edad r de un grupo de N_i ingresados a la edad i.

Y, para el caso en que tenemos en K observaciones, que de N_r^i elementos de edad i quedan n_r dentro del colectivo a la edad x ($r = 1, 2, 3 \ldots, K$) obtenemos:

$${}_{x-i}\overline{p}_i = \frac{\sum_{r=1}^{K} n_r}{\sum_{r=1}^{K} N_r^i} \sim {}_{x-i}p_i$$

IV.—NOTAS SOBRE APLICACIONES DE LA TEORIA

Al aplicarse esta teoría a colectivos abiertos concretos, habrá de darse a las probabilidades y valores medios que intervienen en las fórmulas, una interpretación, mediante el establecimiento de unas hipótesis que, expresadas matemáticamente, nos determinen la forma de dichos parámetros. El establecimiento de dichas hipótesis dependerá naturalmente del colectivo de que se trate, en el cual habrá que analizar las causas de entrada y salida de elementos, para poder expresarlas de un modo analítico.

Tomemos, por ejemplo, el caso de un colectivo laboral. Las salidas del mismo serán solamente: por muerte o por salida del trabajo; por lo tanto, puesto que la salida de la profesión podemos considerarla sin grave error independiente de la edad (pues los jubilados siguen perteneciendo al colectivo) la forma del tanto de salida puede ser idéntica a la de Makeham o Lázarus para el tanto instantáneo de mortalidad, sin más que una pequeña modificación (que el ajuste nos dará) de la constante que en ellas figura para recoger la mortalidad por causas fortuitas. En cuanto a las entradas en el colectivo, son, ciertamente, dependientes de la edad, pero, no obstante, dependen también, muy directamente, del colectivo en particular de que se trate; así pues, las hipótesis sobre la entrada habrán de ser hechas tras un estudio de dicho colectivo real (profesión de los que agrupa, etc.) y a la vista de los datos que se posean; habrán de ser, pues, unas hipótesis referentes a la *biología especial del colectivo,* pues no gravitan (como en el caso de salida) sobre todos y cada uno de los individuos de un grupo, sino, solamente, sobre el *colectivo como tal ente.* Bien es cierto que algunas de estas hipótesis serán válidas para una gran mayoría de colectivos laborales.