Aplicaciones al seguro de la distribución binomial negativa

Por

EUGENIO PRIETO PEREZ

Catedrático de la Universidad Autónoma de Madrid.
Director del Gabinete Técnico del Sindicato Nacional del Seguro

Introducción

Pretendemos con este trabajo, en primer lugar, presentar las características esenciales del modelo de la distribución binomial negativa, analizando algunos fenómenos del mundo empírico que ofrecen tales características, de entre los que le sirven de substrato empírico.

Muchos de estos fenómenos pertenecen al campo actuarial, de aquí la importancia que este asunto tiene para los interesados en el desarrollo teórico de la ciencia actuarial. Las aplicaciones que desarrollaremos y la mayor parte de la bibliografía consultada se refieren a este campo.

Consideramos con cierto detalle las distintas formas de generarse el modelo, por cuanto lo exige su racional utilización.

En este orden de ideas supongamos el fenómeno del acaecimiento de los accidentes de automóviles a las pólizas aseguradas en una cierta entidad aseguradora. Las pólizas dentro de la cartera de la entidad presentan diferentes grados de propensión al accidente, concretados en diferentes valores de λ, en una distribución de Poisson, pues bien, demostraremos que si λ se distribuye según una ley de Pearson de tipo III, la distribución del número de accidentes para el total de la cartera de la entidad sigue el modelo de la distribución binomial negativa.

Si con Arbous y Kerrich distinguiéramos entre *propensión al accidente* y *exposición al accidente,* caracterizando la primera "por ser un atributo del individuo que no presenta fluctuación alguna, y no muestra, por tanto, tendencia a aumentar o a disminuir durante un largo período de tiempo; incluyendo la segunda, esto es, la exposición al accidente, además de la propensión al mismo, los efectos de otros factores significativos como son la edad, la experiencia conseguida, fatiga, estados emotivos, etc., resulta que, al admitir una idéntica propensión inicial al accidente en cada individuo hasta que aquél se presente, pero cuyo acaecimiento comporta modificaciones en la propensión elevándola o reduciéndola según los casos y siempre que los accidentes o sucesos en general ocurran aleatoriamente, en ciertas condiciones, llegamos al esquema de la binomial negativa. Más adelante trataremos ampliamente esta cuestión, como un proceso estocástico.

M. G. Kendall y A. Stuart (1) llegan al modelo de la bnomial negativa de la siguiente forma:

Despues de n accidentes para la totalidad de la cartera de la entidad considerada, la probabilidad de que una póliza haya sufrido 0, 1, 2, ..., accidentes, seguirá, evidentemente, una distribución binomial de parámetros n, p, q, siendo p la probabilidad de que una póliza se siniestre. Supongamos, ahora, que la cantidad aseguradora aplica una política de selección tal que r siniestros en un tiempo menor al año, por ejemplo, supone la anulabilidad de la correspondiente póliza. Ocurre que la probabilidad de una anulación como consecuencia del acaecimiento del n-ésimo siniestro y de la aplicación de la citada política de selección, sigue la función densidad de la *distribución binomial negativa.*

Históricamente fueron Greenwood y Yule los que presentaron por vez primera (1920) un interesante estudio encaminado a la elaboración de un modelo matemático que describiera los problemas psicológicos y sociológicos que lleva aparejado el fenómeno de la propensión al accidente. Evidentemente, si personas distintas presentaran diferentes propensiones al accidente, sería conveniente detectarla y cambiarla. Este estudio fue magníficamente resumido por Luigi Vajani, en su libro *Saggi sui processi stocastici.* Examinaron Yule y Greenwood la distribución del número de accidentes sufridos por los operarios de una fábrica de proyectiles, y vieron que, a los datos por ellos recopilados, no era edecuado ajustarle ni el modelo de la distribución binomial ni la de Poisson, que, como es sabido, se caracterizan:

a) Distribución binomial:

$$m_1 = \text{media} = np \qquad \sigma_1^2 = \text{varianza} = npq$$

b) Distribución de Poisson:

$$m_2 = \text{media} = np \qquad \sigma_2^2 = \text{varianza} = np$$

Teniendo en cuenta que q es menor que 1, se verifica

$$\sigma_1^2 < m_1 = m_2 = \sigma_2^2$$

Los datos disponibles sugerían, por el contrario, un modelo de distribución que satisfaciera la desigualdad $m < \sigma^2$. Precisamente, la distribución binomial negativa cumple esta propiedad. A este modelo, como ya hemos señalado, llegaron Greenwood y Yule, suponiendo que la población considerada

(1) Véase M. G. KENDALL y A. STUARD, *The avanced theory of statistics*, volumen I, 1958.

estaba compuesta por individuos con diferentes grados de propensión al accidente, representados por diferentes valores de λ en una distribución de Poisson (2), y siempre que λ se distribuya siguiendo un modelo de distribución de Pearson-tipo III.

La primera aplicación a cuestiones actuariales del modelo de la binomial negativa, a juicio de Le Roy J. Simon, tiene lugar en un trabajo de Keffer en 1929, referente a un plan de tarificación del seguro de vida colectivo, basado en la experiencia (3). En los últimos años fue utilizada frecuentemente, entre otros, por Bichsel, Lundberg, Ammeter, Delaporte, Thyrion y otros muchos actuarios, como puede verse sin más que consultar el *The Astin Bulletin,* las Memorias de los Congresos Internacionales de Actuarios o cualquiera de las revistas especializadas en cuestiones actuariales.

El modelo de la binomial negativa nos es mostrado con frecuencia bajo diferentes formas: distribución de Polya-Eggenberger, distribución compuesta de Poisson y otras. De las relaciones entre estos modelos y de la estimación de los parámetros del modelo que estudiamos nos ocuparemos ampliamente en las páginas siguientes.

III. EL MODELO DE LA DISTRIBUCION BINOMIAL NEGATIVA Y SUS PROPIEDADES

$$P(\xi = x) = \binom{x+n-1}{x} p^n q^x \qquad \begin{matrix} x = 0, 1, 2 \ldots \\ p + q = 1 \end{matrix} \qquad [1]$$

decimos que sigue la distribución binomial negativa.

Propiedades

A) La función generatriz de momentos de [1] vendrá dada por:

$$G(\theta) = \sum_{x=0}^{\infty} e^{\theta x} \binom{n+x-1}{x} (1-q)^n q^x = (1-q)^n \sum_{x=0}^{\infty} \binom{n+x-1}{x} e^{\theta x} q^x =$$

$$= (1-q)^n \sum_{x=0}^{\infty} \binom{n+x-1}{x} (q\, e^{\theta})^x$$

---

(2) Esta distribución tiene por función de cuantía $P(\xi = X) = \frac{\lambda^x}{x!} e^{-\lambda}$ y es $E(\xi) = \lambda$, $\mathrm{Var}(\xi) = \lambda$.

(3) En su trabajo *An introduction to the Negative Binomial distribution and its applications,* Le Roy J. Simon, comentando el de Keffer, escribe: "Al replicar a la discusión por escrito que aquello desencadenó, desarrolló las ecuaciones para la media y para la varianza y comentó el hecho de que la varianza fuera de un valor superior al dado por la distribución de Poisson de una manera que puede interpertarse como indicativa de la heterogeneidad de los datos". Evidentemente, la propiedad puede deberse, asimismo, a otros efectos: el contagio, por ejemplo.

Dando valores a x resulta:

$$G(\theta) = (1-q)^n \left[1 + \binom{n}{1} q\, e^\theta + \binom{n+1}{2} (q\, e^\theta)^2 + \dots\right]$$

La serie contenida entre corchetes es el desarrollo de

$$(1 - q\, e^\theta)^{-n}$$

Así, pues,

$$G(\theta) = (1 - q)^n (1 - q\, e^\theta)^{-n} = \left(\frac{p}{1 - q\, e^\theta}\right)^n \qquad [2]$$

Recordando que la función generatriz de la distribución binomial es:

$$G(\theta) = (q + p\, e^\theta)^n = q^n \left(1 + \frac{p}{q} e^\theta\right)^n$$

la comparación con [2] parece justificar la denominación de *binomial negativa*, por el hecho de aparecer en [2] el segundo factor elevado al exponente (-n), mientras que en la binomial ordinaria el exponente es n.

El cálculo de los momentos basándonos en $G(\theta)$ se simplifica haciendo el cambio:

$$p = \frac{1}{1+\alpha} \qquad q = \frac{\alpha}{1+\alpha}$$

con lo que

$$G(\theta) = \left(\frac{\frac{1}{1+\alpha}}{1 - \frac{\alpha}{1+\alpha}}\right)^n = \left(\frac{1}{1+\alpha-\alpha\, e^\theta}\right)^n = (1+\alpha-\alpha\, e^\theta)^{-n}$$

B) La función cumulativa $\kappa(\theta) = \log_e G(\theta)$ en nuestro caso es:

$$\kappa(\theta) = -\, n \log_e (1 + \alpha - \alpha\, e^\theta) \qquad [3]$$

C) Los cuatro primeros momentos cumulantes de la distribución binomial negativa son:

$$\kappa_1 = \left[\frac{d\,\kappa(\theta)}{d\,\theta}\right]_{\theta=0} = \left[\frac{n\,\alpha\, e^\theta}{1+\alpha-\alpha\, e^\theta}\right]_{\theta=0} = n\,\alpha$$

$$\kappa_2 = \left[\frac{d^2\,\kappa(\theta)}{d\,\theta^2}\right]_{\theta=0} = \left[\frac{n\,\alpha\, e^\theta(1+\alpha-\alpha\, e^\theta) + \alpha\, e^\theta\, n\,\alpha\, e^\theta}{(1+\alpha-\alpha\, e^\theta)^2}\right]_{\theta=0} = n\,\alpha\,(1+\alpha)$$

$$\kappa_3 = \left[\frac{d^3\,\kappa(\theta)}{d\,\theta^3}\right]_{\theta=0} = n\,\alpha\,(1+\alpha)(1+2\,\alpha)$$

$$\kappa_4 = \left[\frac{d^4\,\kappa(\theta)}{d\,\theta^4}\right]_{\theta=0} = n\,\alpha\,(1+\alpha)(1+6\,\alpha+6\,\alpha^2)$$

Las expresiones de κ_1, κ_2, κ_3 y κ_4 en función de p y q aparecen con sólo considerar que:

$$\alpha = \frac{q}{p}$$

Por tanto, serán:

$$\kappa_1 = n\frac{q}{p}$$

$$\kappa_2 = n\frac{q}{p}\left(1+\frac{q}{p}\right) = \frac{nq}{p}\cdot\frac{p+q}{p} = \frac{nq}{p^2}$$

$$\kappa_3 = n\frac{q}{p}\left(1+\frac{q}{p}\right)\left(1+\frac{2q}{p}\right) = \frac{nq(1+q)}{p^2}$$

$$\kappa_4 = \frac{nq}{p}\left(1+\frac{q}{p}\right)\left(1+6\frac{q}{p}+6\frac{q^2}{p^2}\right) = \frac{nq(1+4q+q^2)}{p^4}$$

Recordando que

$$m = \text{media} = \kappa_1$$
$$\sigma^2 = \text{varianza} = \kappa_2$$
$$\mu_3 = \kappa_3$$
$$\mu_4 = \kappa_4 + 2(\sigma^2)^2$$

tenemos:

$$\text{media} = \frac{nq}{p}$$

$$\text{varianza} = \frac{nq}{p^2}$$

$$\mu_3 = \frac{nq}{p^2}\cdot\frac{1+q}{p}$$

$$\mu_4 = \frac{nq(1+4q+q^2)}{p^4} + 3\frac{n^2q^2}{p^4}$$

$$= \frac{nq}{p^2}\cdot\frac{1+4q+q^2+3nq}{p^2}$$

III. LA BINOMIAL NEGATIVA GENERADA POR UNA SUMA DE n VARIANTES INDEPENDIENTES DISTRIBUIDAS GEOMETRICAMENTE

Resumimos brevemente las principales características de la distribución geométrica:

Función de cuantía:

$$P(\xi = x) = pq^x \; (x = 0, 1, 2, \ldots, \infty) \qquad p + q = 1$$

Función generatriz:

$$G(\theta) = \sum_{x=0}^{\infty} e^{\theta x} pq^x = \frac{p}{1 - q e^\theta} = \frac{1}{\frac{1}{p} - \frac{q}{p} e^\theta} = \frac{1}{1 + \alpha - \alpha e^\theta}$$

La última expresión resulta del cambio

$$p = \frac{1}{1+\alpha} \quad \text{y} \quad q = \frac{\alpha}{1+\alpha}$$

Función cumulativa:

$$\kappa(\theta) = -\log_e(1 + \alpha - \alpha e^\theta)$$

Momentos: media$= \frac{q}{p}$

varianza$= \frac{q}{p^2}$

$$\kappa_3 = \frac{q}{p^2} \cdot \frac{1+q}{p^2}$$

$$\kappa_4 = \frac{q}{p^2} \cdot \frac{1 + 7q + q^2}{p^2}$$

Veamos, pues, que la distribución binomial negativa puede ser obtenida como suma de n distribuciones geométricas independientes e igualmente distribuidas. En efecto, sea

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n$$

en donde

$$P(\xi_i = x) = pq^x \begin{cases} x = 0, 1, \ldots \\ i = 1, 2, \ldots, n \end{cases}$$

La función generatriz de ξ es:

$$G_\xi(\theta) = E[e^{\theta\xi}] = E[e^{\theta(\xi_1+\xi_2+\ldots+\xi_n)}] = \prod_{i=1}^{n} E(e^{\theta\xi_i}) = \prod_{i=1}^{n} \varphi_{\xi_i}(\theta) = \left(\frac{1}{1 + \alpha - \alpha e^\theta}\right)^n$$

esto es, la función generatriz de ξ es la de una variable aleatoria con *distribución binomial negativa.* Como corolario podemos señalar que la distribución geométrica es el caso particular ($n=1$) de la binomial negativa.

IV. LA DISTRIBUCION DE POISSON COMO LIMITE DE LA BINOMIAL NEGATIVA

Partiendo de la función de cuantía

$$P(\xi = x) = \binom{n+x-1}{x} p^n q^x$$

al hacer

$$p = \frac{n}{n+\lambda t} \qquad q = 1 - p = \frac{\lambda t}{n+\lambda t} \qquad \lambda > 0$$

resulta:

$$P(\xi = x) = \binom{n+x-1}{x}\left(\frac{n}{n+\lambda t}\right)^n \cdot \left(\frac{\lambda t}{n+\lambda t}\right)^x =$$

$$= \binom{n+x-1}{x}\left(\frac{1}{1+\frac{\lambda t}{n}}\right)^n \frac{(\lambda t)^x}{\left(1+\frac{\lambda t}{n}\right)^x} \cdot \frac{1}{n^x} =$$

$$= \frac{n^x\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)\cdots\left(1+\frac{x-1}{n}\right)}{x!} \cdot$$

$$\cdot \frac{1}{\left(1+\frac{\lambda t}{n}\right)^n} \cdot \frac{(\lambda t)^x}{\left(1+\frac{\lambda t}{n}\right)^x} \xrightarrow[n\to\infty]{} e^{-\lambda t}\frac{(\lambda t)^x}{x!}$$

esto es, la función de cuantía de la distribución de Poisson de parámetro λt aparece como límite de la correspondiente de la distribución binomial negativa cuando

$$p = \frac{n}{n+\lambda t}$$

y n tiende a infinito.

V. DISTRIBUCIONES COMPUESTAS DE POISSON Y BINOMIAL NEGATIVA

Sin entrar en detalles nos limitaremos a recordar que llamamos distribución compuesta de la variable aleatoria ξ, respecto a otra η, a la que tiene por función de probabilidad

$$P(x) = \int_{-\infty}^{\infty} f(x, \lambda)\, d\, U(\lambda) \qquad [4]$$

donde $f(x, \lambda)$ representa la función de densidad de ξ dependiente del parámetro λ, que a su vez es una variable aleatoria η con función de distribución $U(\lambda)$.

Cuando en [4] sea

$$f(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{n!}$$

tenemos la denominada distribución compuesta de Poisson.

Si, como en la introducción, suponemos que

$$f(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

es la probabilidad de x siniestros para una póliza correspondiente a la cartera de una determinada entidad aseguradora, con propensión al siniestro de λ en un cierto período, y si fuera

$$d\,U(\lambda) = \frac{c^p}{\Gamma(n)} e^{-c\lambda} \lambda^{n-1} d\lambda \quad \text{con} \quad \begin{cases} 0 \leqslant \lambda < \infty \\ c > 0 \end{cases}$$

aparece que el número de accidentes o siniestros por póliza para el conjunto de la cartera de la entidad se distribuye según una binomial negativa.

En efecto, la probabilidad de x siniestros o accidentes será:

$$\int_0^\infty \frac{c^n}{\Gamma(n)} e^{-c\lambda} \lambda^{n-1} e^{-\lambda} \frac{\lambda^x}{x!} d\lambda \qquad [5]$$

La [5] es el coeficiente de t^x en el desarrollo de

$$\varphi(t) = \frac{c^n}{\Gamma(n)} \int_0^\infty e^{-c\lambda} \lambda^{n-1} e^{-\lambda+\lambda t} dt \qquad [6]$$

Haciendo en [6] $(c+1-t)\lambda = u$, tenemos:

$$\frac{c^n}{\Gamma(n)} \int_0^\infty e^{-(c+1-t)\lambda} \lambda^{n-1} d\lambda = \frac{c^n}{\Gamma(n)} \cdot \frac{\Gamma(n)}{(c+1-t)^n} = \left(\frac{c}{c+1}\right)^n \left(1 - \frac{t}{c+1}\right)^{-n}$$

La probabilidad de x éxitos es, por consiguiente:

$$\frac{n(n+1)\ldots(n+x-1)}{x!} \left(\frac{c}{c+1}\right)^n (c+1)^{-x}$$

Al hacer $p = \dfrac{c}{c+1}$ y $q = \dfrac{1}{c+1}$, resulta la función de cuantía de la distribución binomial negativa:

$$\binom{n+x-1}{x} p^n q^x \qquad x = 0, 1, 2, \ldots$$

VI. EL PROCESO DE POLYA-EGGENBERBER Y LA DISTRIBUCION BINOMIAL NEGATIVA

Denotemos por $P_n(t)$ la probabilidad de que en el intervalo $(0, t)$ una póliza haya tenido n siniestros. En el transcurso del intervalo $(t, t+dt)$, es posible:

a) El acaecimiento de un nuevo siniestro, esto es, en el lenguaje habitual de los procesos estocásticos, se produce el tránsito del estado E_n al E_{n+1}.

b) Que la póliza permanezca en $t+dt$, en el estado E_n.

Denotando por $\lambda_n(t)$ la probabilidad de transición de E_n a E_{n+1}, podemos escribir:

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

o sea:

$$\frac{P_n(t+dt)-P_n(t)}{dt} = P_{n-1}(t)\,\lambda_{n-1}(t) - P_n(t)\,\lambda_n(t)$$

Pasando al límite para $dt \rightarrow 0$ queda:

$$P_n'(t) = P_{n-1}(t)\,\lambda_{n-1}(t) - P_n(t)\,\lambda_n\,t) \qquad [7]$$

ecuación válida solamente para $h \geqslant 1$, pues para $n=0$ sólo es posible la transición E_0 a E_1, y la [7] se reduce a:

$$P'_0(t) = -\lambda P_0(t)$$

Al dar valores a n en [7] obtenemos un sistema de infinitas ecuaciones diferenciales (o en diferencias) que, junto con las condiciones iniciales, nos permite estudiar el comportamiento del proceso estocástico de riesgo.

Las condiciones iniciales, evidentemente, son:

$$P_n(0) = \begin{cases} 1 \text{ para } n=0 \\ 0 \text{ para } n \neq 0 \end{cases}$$

El proceso de Polya-Eggerberger se caracteriza porque para él $\lambda_n(t)$ es directamente proporcional al número n de siniestros acaecidos a la póliza en cuestión, en $(0, t)$,, e inversamente proporcional a la longitud del período considerado. Por consiguiente, tendrá una expresión tal como:

$$\lambda_n(t) = \frac{a+n}{b+t} \qquad [8]$$

Trataremos de obtener la llamada distribución de Polya-Eggenberger, caracterizada por ser la distribución marginal del proceso estocástico que acabamos de definir.

Veremos cómo esta distribución pertenece a la familia de las binomiales negativas.

Para obtener este resultado se requiere integrar el sistema

$$\left.\begin{aligned} P_n'(t) &= \lambda_{n-1}(t)\,P_{n-1}(t) - \lambda_n(t)\,P_n(t) \\ P_n'(t) &= -\,\lambda_0(t)\,P_0(t) \end{aligned}\right\} \qquad [9]$$

Satisfaciendo las condiciones

$$P_0(0) = 1 \quad \text{y} \quad P_n(0) = 0 \quad (n \geqslant 1)$$

y siendo

$$\lambda_n(t) = \frac{a+n}{b+t}$$

La solución aparece fácilmente si efectuamos en [8] la siguiente transformación:

$$\lambda_n(t) = \frac{a+n}{b+t} = \frac{a\left(1+\dfrac{n}{a}\right)}{b\left(1+\dfrac{t}{b}\right)} = \frac{1+\dfrac{n}{a}}{1+\dfrac{t}{b}}\cdot\frac{a}{b}$$

Haciendo $\dfrac{a}{b}=\lambda$ encontraremos:

$$b = \frac{a}{\lambda}$$

y, por tanto:

$$\lambda_n(t) = \frac{1+\dfrac{n}{a}}{1+\dfrac{\lambda}{a}t}\,\lambda \qquad [10]$$

Entonces, la solución de [9] podría obtenerse así:

Si en [8] hacemos $n=0$, resulta:

$$\lambda_0(t) = \frac{a\lambda}{a+\lambda t}$$

así que la ecuación diferencial

$$P_0(t) = \lambda_0(t)\,P_0(t)$$

se concreta en

$$P_0'(t) = -\frac{a\lambda}{a+\lambda t}\,P_0(t)$$

o sea,

$$\log P_0(t) = -a \log_e (a + \lambda t) + \log K$$

de donde

$$P_0(t) = k\left(\frac{1}{a+\lambda t}\right)^a$$

La constante K podemos determinarla haciendo uso de las condiciones iniciales. En efecto, para $t=0$ tenemos:

$$P_0(0) = k\left(\frac{1}{a}\right)^a = 1 \quad \text{o sea} \quad k = a^a$$

Resulta, pues,

$$P_0(t) = \left(\frac{a}{a+\lambda t}\right)^a$$

La primera ecuación de [9] para $n=1$ es:

$$P_1'(t) = \lambda_0(t)\,P_0(t) - \lambda_1(t)\,P_1(t)$$

Siendo:

$$\lambda_1(t) = \frac{a+1}{a+\lambda t}\,\lambda$$

$$\lambda_0(t) = \frac{a}{a+\lambda t}\,\lambda$$

por tanto,

$$P_1'(t) = \frac{a\lambda}{a+\lambda t}\left(\frac{a}{a+\lambda t}\right)^a - \frac{a+1}{a+\lambda t}\,\lambda P_1(t) \qquad [11]$$

La [11] es, a diferencia de la primera ecuación de [9], una ecuación diferencial, pero no en diferencias finitas como ésta. Es una ecuación diferencial lineal de primer orden.

La solución particular correspondiente a las condiciones iniciales propuestas es:

$$P_1(t) = a\left(\frac{a}{a+\lambda t}\right)^a \frac{\lambda t}{a+\lambda t}$$

Para obtener $P_2(t)$ se procede en forma análoga a la seguida para deducir $P_1(t)$, método que es general para cualquier valor de n entero y positivo. Resulta:

$$P_x(t) = \frac{a(a+1)(a+2)\ldots(a+x-1)}{x!}\cdot\left(\frac{a}{a+\lambda t}\right)^a\cdot\left(\frac{\lambda t}{a+\lambda t}\right)^x \qquad [12]$$

para $x=0,\ 1,\ 2, \ldots, \infty$.

La [12] se denomina comúnmente distribución de Polya, por cuanto proviene de este modelo, pero es fácil demostrar que pertenece a la familia de la distribución binomial negativa.

En efecto, la [12] puede escribirse así:

$$P_x(t) = \binom{x+a-1}{x}\left(\frac{a}{a+\lambda t}\right)^a \left(\frac{\lambda t}{a+\lambda t}\right)^x$$

Haciendo

$$P_t = \frac{a}{a+\lambda t} \qquad q_t = 1 - p_t = \frac{\lambda t}{a+\lambda t} \qquad \text{y} \qquad a = n$$

queda

$$P_x(t) = \binom{x+n-1}{x} p_t^n \cdot q_t^x$$

o sea, la distribución binomial negativa.

La distribución de Polya aparece en ocasiones bajo al forma

$$P_x(t) = \binom{x+a-1}{x}(1+\gamma t)^{-a}\left(\frac{\gamma t}{1+\gamma t}\right)^x$$

que surge de [12], sin más que dividir numerador y denominador por a en el segundo y tercer factor de [12] y hacer $\frac{\lambda}{a} = \gamma$. Tenemos:

$$P_x(t) = \binom{x+a-1}{x}\left(\frac{1}{1+\frac{\lambda}{a}t}\right)^a \cdot \left(\frac{\lambda/a \cdot t}{1+\frac{\lambda}{a}t}\right)^x =$$

$$= \binom{x+a-1}{x}(1+\gamma t)^{-a}\left(\frac{\gamma t}{1+\gamma t}\right)^x$$

VIII. EL PROCEDIMIENTO DEL MUESTREO SUCESIONAL DE M. G. KENDALL Y A. STUART

Este procedimiento de obtención de la binomial negativa se encuentra en el libro *The avanced theory of statistic* (4) y es en extremo simple. En efecto, después de N siniestros, para el total de pólizas de la cartera de una entidad aseguradora, la proporción de las siniestradas 0, 1, ..., veces vendrá dada por el correspondiente término del desarrollo, $(p+q)^N$, en donde p es la probabilidad de que una póliza se siniestre en el período considerado y $p+q=1$. Tal y como hemos señalado en la introducción, si la

(4) Véase pág. 130, vol. I, 1958.

entidad anulara toda póliza al acaecerle el n-ésimo siniestro, la probabilidad de anulación para una póliza de tal cartera con el n-ésimo siniestro es:

$$\binom{N-1}{n-1} q^{N-n} p^{n-1} q = \binom{N-1}{n-1} q^{N-1} p^{n}$$

Ahora bien, las anulaciones no comienzan hasta el n-ésimo siniestro; por consiguiente, la probabilidad de una anulación con el acaecimiento del n-ésimo, $(n+1)$, ésimo, ..., siniestro, son:

$$p^n \left\{ 1, nq, \frac{n(n+1)}{2!} q^2, \ldots \right\}$$

esto es:

$$\binom{n+x-1}{x} p^x q^x \qquad x = 0, 1, 2, \ldots$$

Kendall y Stuart, en su libro, deducen la binomial negativa, contemplando el caso de una población sujeta a los efectos de los sucesivos ataques de una cierta enfermedad, cuando las personas afectadas lo son al azar, y en una proporción p en el supuesto de que n ataques a un mismo individuo de la población le fueran fatales.

VIII. LA DISTRIBUCION BINOMIAL NEGATIVA Y LA LEY DE DUBOURDIEU

Denotemos con $P_x(t)$ la probabilidad de que en el intervalo $(0, t)$ ocurra x veces un cierto suceso S.

Definición

Llamamos ley de Dubourdieu (5) la definida así:

$$P_x(t) = (-1)^x \frac{t^x}{x!} \frac{d^x P_0(x)}{dt^x}$$

con

$$P_0(0) = 1 \qquad \begin{cases} x \geqslant 0 \\ t > 0 \end{cases}$$

donde $P_0(t)$ es una función analítica absolutamente monótona, tal que

$$(-1)^x \frac{d^x P_0(t)}{dt^x} > 0$$

(5) Véase J. Dubourdieu, *Remarques relatives a la théorie de l'assurance accidents*, 1938. *Les fonctions absolutement monotones et la théorie methemátique de l'assurance accidents*, 1938.

Demostraremos que las distribuciones de Poisson y binominal negativa cumplen las anteriores condiciones de Dubourdieu.

La probabilidad de que el suceso S no ocurra en $(0, t)$, en la hipótesis de Poisson, es:

$$P_0(t) = e^{-\lambda t} \qquad \lambda > 0$$

que es una función absolutamente monótona. Además se verifica

$$\frac{d^x P_0(t)}{dt^x} = (-1)^x e^{-\lambda t} \lambda^x \tag{13}$$

Sustituyendo [13] en

$$P_x(t) = (-1)^x \frac{t^x}{x!} \cdot \frac{d^x P_0(t)}{dt^x}$$

resulta

$$= (-1)^x \frac{t^x}{x!} (-1)^x e^{-\lambda t} \lambda^x = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

es decir, la ley de Poisson de parámetro λt.

En cuanto a la distribución binomial negativa tenemos:

Partiendo de la expresión

$$P_x(t) = \binom{x + a - 1}{x} (1 + \lambda t)^{-a} \left(\frac{\gamma t}{1 + \gamma t}\right)^x$$

encontramos

$$P_0(t) = (1 + \gamma t)^{-a}$$

y

$$\frac{d^x P_0(t)}{dt^x} = (-1)^x a (a + 1)(a + 2) \ldots (a + x - 1) \cdot \gamma^x (1 + \gamma t)^{-a-x}$$

Llevando esta expresión a [12] resulta:

$$P_x(t) = (-1)^x \frac{t^x}{x!} (-1)^x a (a + 1)(a + 2) \ldots (a + x - 1) \cdot \gamma^x (1 + \gamma t)^{-a-x}$$

$$= \binom{a + x - 1}{x} (1 + \gamma t)^{-a} \left(\frac{t\gamma}{1 + \gamma t}\right)^x$$

es decir, la distribución de Polya. Basta remitirnos al epígrafe VI de este trabajo para que la proposición enunciada quede probada.

Entonces, la relación entre la ley de reclamaciones de Dubourdieu y las distribuciones de Poisson y binomial negativa es que aquélla es una generalización de éstas. Esta generalización supone, como muy bien expuso Guido

Santacroce (6), que si en la ley general expresada por [12] la probabilidad de que, dado un intervalo $(0, T)$, al que dividimos en dos contiguos $(0, t)$ y (t, T), y conocido que en $(0, t)$ han ocurrido x sucesos, se verifiquen m sucesos en (t, T), depende, en general de X y de t —así ocurre en la binomial negativa— mientras que resulta independiente en el modelo de Poisson.

IX. ESTIMACION DE LOS PARAMETROS

La determinación del modelo de la binomial negativa exige la estimación de los parámetros que en él figuran. De los diferentes métodos estadísticos para la estimación de parámetros consideraremos dos, por resultar especialmente simples, a saber:

a) El método de los momentos, y
b) El método de máxima verosimilitud.

Sea $\bar{X}$ la media de la distribución empírica del número de siniestros por póliza, acaecidos en un determinado período a una cierta cartera. El método de Karl Pearson consiste en igualar los momentos del modelo con los correspondientes de la muestra. En nuestro caso haremos:

$$\frac{n(1-p)}{p} = \bar{x}$$

o sea,

$$n(1-p) = \bar{x}p$$

$$\hat{p} = \frac{n}{\bar{x}+n}$$

El método de la máxima verosimilitud considera el máximo de

$$L = \binom{x+n-1}{x} p^n q^x$$

o bien de

$$\log L = \log \binom{x+n-1}{x} + n \log p + x \log(1-p) \qquad [14]$$

En el último término de [14] se sustituyó q por $1\text{-}p$, para así manejar un solo parámetro. Entonces:

$$\frac{d \log L}{dp} = \frac{n}{p} - \frac{x}{1-p} = 0$$

de donde

$$\hat{p} = \frac{n}{x+n}$$

(6) Véase G. SANTACROCE, *Intorno alli distribuzioni binomiali ad esponente negativo*, 1959.

Por tanto, la aplicación de los métodos *a*) y *b*) conducen a un mismo estimador para p, la *frecuencia relativa.*

Como es sabido:

$$E(\hat{p}) = p = \frac{nq}{m}$$

$$\sigma^2 = \operatorname{var}(\hat{p}) = -\frac{1}{E\left[\dfrac{d^2 \log L}{dp^2}\right]}$$

En nuestro caso es:

$$\frac{d^2 \log L}{dp^2} = \frac{d}{dp}\left[\frac{n}{p} - \frac{x}{1-p}\right] = -\frac{n}{p^2} - \frac{x}{(1-p)^2}$$

luego

$$\sigma^2 = \operatorname{var}(\hat{p}) = \frac{1}{E\left[\dfrac{n}{p^2} + \dfrac{x}{(1-p)^2}\right]} = \frac{1}{\dfrac{n}{p^2} + \dfrac{\frac{n \cdot q}{p}}{(1-p)^2}} = \frac{p^2 q}{n} = \sigma^2$$

Podemos concluir: el estimador de p así obtenido es eficiente y centrado.

Resumen

El autor estudia con cierto detalle la distribución binomial negativa desde el punto de vista de las aplicaciones actuariales, analizando las diversas formas de generarse este modelo, a saber:

a) Distribución compuesta de Poisson, cuando el parámetro λ sigue una distribución de Pearson-tipo III.

b) Como un proceso de Polya.

c) Según el método del muestreo sucesional de M. C. Kendall y A Stuart.

Se prueba que cumple las condiciones de J. Dubourdieu y se discute el ajuste del modelo por los métodos de K. Pearson y máxima verosimilitud.

BIBLIOGRAFIA

M. S. BARTLETT: *An introduction to stochastic process*, 1955.

W. FELLER: *An introduction to probability theory, and its application*, vol. I, 1957.

M. B. KENDALL, y A. STUART: *The avanced theory of statistics*, vol. I, 1958.

LUIGI VAJANI: *Saggi sui processi stocastici*, 1960.

GONZALO ARNÁIZ: *Introducción a la estadística teórica*, 1965.

S. DUBOURDIEU: *Remarques relatives a la théorie de l'assurance accidents*, 1938.

— *Les fonctions absolutement monotones et la théorie methématique de l'assurance accidents*, 1938.

A. G. ARBOUS, y J. E. KERRICH: *Accident statistics and the concept of accident propeness*, 1951.

J. L. SIMON: *Fitting negative binomial distribution by the method of maximum likelibood*, 1961.

— *An introduction to the Negative Binomial Distribution and its applications*, 1962.

THOMAS O. CARLSON: *Negative Binomial Rationale*, 1962.

UBALDO NIETO DE ALBA: *La distribución del número de siniestros en la matemática del seguro*, 1964.