

Capítulo 4

Cálculo de la f.d. compuesta de siniestros F

El cálculo numérico de la f.d. compuesta del importe acumulado de los siniestros F utilizando (3.1.2) con frecuencia es difícil y podría forzar los recursos de cálculo disponibles. Sin embargo, existen algunos casos importantes donde el denominado método de recurrencia puede hacer que los cómputos sean factibles, como se verá en la sección 4.1. Por otro lado se pueden utilizar métodos de aproximación para evaluar la f.d. F ; estos métodos se estudiarán en la sección 4.2.

4.1 Fórmula de recurrencia para F

(a) **Supuestos.** La familia más importante de distribuciones compuestas del importe acumulado de siniestros a la que se puede aplicar el método de recurrencia puede especificarse mediante las dos condiciones siguientes :

- (1) Las probabilidades del número de siniestros obedece a la fórmula de recurrencia

$$p_k = (a + b/k) \cdot p_{k-1} \text{ para } k = 1, 2, 3, \dots \quad (4.1.1)$$

donde a y b son constantes que especifican la distribución del número de siniestros.

- (2) La distribución de la intensidad de los siniestros es no-negativa, discreta y equidistante (es decir una distribución escalonada no negativa con valores equidistantes de la variable de intensidad del

siniestro). Más explícitamente, sólo son posibles los importes de siniestros

$$Z_i = i \cdot C \quad (i = 0, 1, 2, \dots, r) \quad (4.1.2)$$

donde C es una constante positiva (denominada **longitud del escalón**).

Como se especifica en (2.3.7) y (2.5.9), la condición (1) se cumple tanto en los casos de Poisson como en los de Pólya (es decir en la binomial negativa). La distribución binomial también cumple esta condición, pero no existe otra distribución del número de siniestros que lo cumpla. (Ejercicios 4.1.3 y 4.1.8). Véase Tabla 4.1.1.

Nótese que una distribución continua, y en realidad, cualquier otro tipo de distribución, puede aproximarse por medio de una distribución discreta equidistante. Por lo tanto, el supuesto (2) no resulta ser muy restrictivo. Sin embargo, una buena aproximación utilizando una distribución equidistante puede requerir un valor r grande, conduciendo a un cálculo que lleva demasiado tiempo, especialmente cuando el número esperado de siniestros es grande.

Denominaremos a las probabilidades de la intensidad de los siniestros mediante:

$$s_i = \text{Prob} \{Z = i \cdot C\}, \quad (4.1.3)$$

donde se permite que algunas s_j , $0 \leq j \leq r$ sean iguales a 0. Para $i < 0$ ó $i > r$, $s_i = 0$. Con frecuencia es conveniente elegir r como la i más grande para la cual $s_i > 0$.

Es importante permitir que s_0 sea mayor que cero, ya que la discretización (N.d.e T.: transformación de la variable continua en discreta) de una distribución positiva de la intensidad de un siniestro generalmente conduce a un s_0 positivo.

El método de recurrencia se puede generalizar más aún, flexibilizando, en cierta medida, las condiciones (1) y (2), como lo demostraron Panjer (1981), Jewell y Sundt (1981) y Willmot (1986). Jewell y Sundt, por ejemplo, extienden el método de recurrencia a un grupo de distribuciones del número de siniestros mucho más amplia que cumplen la condición de recurrencia (4.1.1) para $k \geq k'$, donde $k' \geq 2$. En Panjer y Willmot (1992) se podrá encontrar más información acerca de las fórmulas de recurrencia.

(b) Fórmula de recurrencia. Con las condiciones anteriores, la distribución del importe acumulado de siniestros también es discreta y equidistante, ya que sólo son posibles los valores que son múltiplos de la longitud de escalón C . Las probabilidades asociadas se indicarán mediante

$$f_i = \text{Prob} \{X = j \cdot C\} \quad (j = 0, 1, 2, \dots) \quad (4.1.4)$$

donde X señala la variable compuesta del importe de siniestros. Estas probabilidades pueden calcularse recurrentemente a partir de las ecuaciones

$$f_j = \frac{1}{1 - a \cdot s_0} \sum_{i=1}^{\min(j,r)} \left(a + \frac{ib}{j} \right) \cdot s_i \cdot f_{j-i}, \quad j = 1, 2, \dots, \quad (4.1.5)$$

con el valor inicial k

$$p_0, \quad \text{si } s_0 = 0, \\ f_0 = \\ \sum_{i=0}^{\infty} p_i \cdot s_0^i = M_k(\ln s_0), \quad \text{si } s_0 > 0, \quad (4.1.6)$$

o, más explícitamente,

$$f_0 = e^{n \cdot s_0 - n} \quad (\text{caso Poisson}) \\ f_0 = \left(1 + \frac{a \cdot n \cdot (1-s_0)}{a+b} \right)^{\frac{a+b}{a}} \quad (\text{caso Poisson}) \quad (4.1.7)$$

$$f_0 = \left(1 + \frac{a-1}{a \cdot s_0 - 1} \right)^{\frac{a+b}{a}} \quad (\text{caso binomial})$$

La demostración está dada en el Apéndice E.

El valor de la correspondiente f.d. compuesta F en cualquier punto dado $X = j \cdot C, j \geq 0$, se obtiene a partir de

$$F(X) = F(j \cdot C) = \sum_{i=0}^j f_i. \quad (4.1.8)$$

Y como F es una función escalonada, la fórmula (4.1.8) también se cumple para cada X entre $j \cdot C$ y $(j+1) \cdot C$.

OBSERVACION 1. Para planificar los cálculos, puede resultar útil saber que las operaciones aritméticas necesarias para calcular toda la distribución (4.1.4) son del orden de $5 \cdot \alpha \cdot r \cdot n \cdot m / C$, donde n es el número esperado de siniestros, m es la media de la intensidad de los siniestros, y α es un multiplicador, con frecuencia del orden 2 ... 10, indicando hasta cuántos múltiplos del valor esperado $m \cdot n$ de la distribución compuesta se debe extender el cálculo hasta que la masa de probabilidades restante pueda considerarse poco significativa.

OBSERVACION 2. Las observaciones en la sección 2.3(c) relacionadas con la programación de la computadora también se aplican a la recurrencia (4.1.5).

Para comodidad de los lectores, los valores de los parámetros a y b de las distribuciones del número de siniestros que cumplen la fórmula de recurrencia (4.1.1) aparecen en la Tabla 4.1.1.

Tabla 4.1.1 Parámetros para diferentes distribuciones del número de siniestros que cumplen con la fórmula de recurrencia (4.1.1).

Distribución del número de siniestros	a	b
Poisson(n)	0	n
Pólya(n, b)	$\frac{n}{n+h}$	$\frac{(h-1)n}{n+h}$
Binomial(p, N)	$\frac{p}{1-p}$	$\frac{(N+1)p}{1-p}$
Geométrica(p)	p	0

Como se sabe, las probabilidades binomiales y geométricas son

$$p_k = \binom{N}{k} p^k \cdot (1-p)^{N-k} \quad (\text{binomial})$$

$$(4.1.9) \quad p_k = (1-p) \cdot p^k \quad (\text{geométrica})$$

y la distribución Geométrica(p) es, en realidad, de Pólya($p/(1-p), 1$).

Panjer (1981) introdujo la fórmula de recurrencia (4.1.5) sin perjuicio de que el caso de Poisson compuesto había sido introducido por Adelson con anterioridad (1966).

(c) "Discretización" utilizando el método del punto medio. Si se tiene una distribución S de la intensidad de los siniestros continua, o si es una distribución discreta basada sobre los siniestros observados, entonces el primer paso en el cálculo de recurrencia es discretizarla en la forma (4.1.3). Un método directo es concentrar la masa de probabilidades de cada intervalo de longitud doble $(0, 2C], (2C, 4C], \dots$, en los puntos medios de los intervalos respectivos. Entonces se obtiene la distribución discreta como una aproximación a la distribución originaria, con toda la masa de probabilidades concentrada en los puntos

$$C, 3.C, 5.C, \dots, r.C \quad (r = 2.r'-1), \quad (4.1.10)$$

con probabilidades

$$\begin{aligned} s_1 &= S(2.C), \\ s_3 &= S(4.C) - S(2.C), \\ s_5 &= S(6.C) - S(4.C), \\ &\vdots \\ s_r &= 1 - S(2.(r'-1).C), \quad (r = 2.r'-1), \end{aligned} \quad (4.1.11)$$

y $s_i = 0$ en todo otro caso. El límite r' debería ser lo suficientemente grande como para que la masa de probabilidades a la derecha de $2 \cdot r'$ sea insignificante.

Una desventaja del método del punto medio, al igual que todos los métodos de discretización, es que las imprecisiones de redondeo surgen en $F(X)$ y en su valor esperado y en los momentos superiores. Como la densidad de siniestros usualmente es decreciente en la cola derecha de la f.d. S de la intensidad de los siniestros, habrá, en el área de la cola, una mayor masa de probabilidades en la primera mitad de cada intervalo que en la segunda. Entonces, típicamente surge un sesgo sistemático en la cola de la mano derecha que es la parte más significativa del rango, cuando se calculan los momentos. Una manera directa de probar la precisión es repetir el cálculo utilizando una r cada vez más grande hasta que cuando se la aumente más no afecte los resultados en forma significativa. Sin embargo, esto puede conducir hasta un valor r tan grande que los cálculos tomen demasiado tiempo. Un procedimiento más corto es comparar el valor esperado (y los momentos superiores) de la distribución discretizada con las características de la originaria. Una diferencia tolerable indica que el método es aceptable.

Otro método de discretización, que se introduce más adelante, proporciona una mejor solución al problema de las imprecisiones de redondeo.

(d) Discretización con esperanza no modificada. Una forma natural de aproximar una f.d. S dada de la intensidad de los siniestros por medio de una distribución equidistante y discreta es distribuir la masa de probabilidades relacionada con el intervalo entre dos puntos cercanos equidistantes de tal manera que el valor esperado de la distribución no se altere (**el método del valor esperado insesgado**). Nótese que cuando C es relativamente grande, la distribución de la masa de probabilidades en

el primer intervalo que comienza en cero puede conducir a que s_0 sea significativamente diferente a cero.

Definimos la distribución en escalera de la aproximación con probabilidades puntuales s_i y la longitud del escalón C con la fórmula

$$s_i = r_{i-1} \cdot l_i, \quad (4.1.12)$$

donde r_i es la masa de probabilidades que se movió al punto $i \cdot C$ del intervalo $(i \cdot C, (i+1) \cdot C]$ de la derecha, y l_i es la masa de probabilidades movida a $i \cdot C$ del intervalo $((i-1) \cdot C, i \cdot C]$ de la izquierda.

Con la condición de que la distribución discretizada $\{s_i\}_i$ y la distribución originaria S de la intensidad de los siniestros tenga el mismo valor esperado, se pueden obtener los siguientes reglas de cálculo (Apéndice E)

$$l_i = d_i - r_{i-1}, \quad r_{i-1} = d_i \cdot y \cdot e_i, \quad (4.1.13)$$

donde

$$d_i = \text{Prob}\{(i-1) \cdot C < Z \leq i \cdot C\} = S(i \cdot C) - S((i-1) \cdot C), \quad (4.1.14)$$

$$e_i = d_i \cdot \frac{E\{\mathbf{Z} | (i-1) \cdot C < \mathbf{Z} \leq i \cdot C\}}{C} = \frac{1}{C} \int_{(i-1)C}^{iC} Z dS(Z).$$

OBSERVACION. Como ejercicio (Ejercicio 4.17) se debe probar que la varianza de la distribución discretizada es siempre por lo menos igual a la varianza de la distribución originaria.

Cuando S es discreta, por ejemplo una f.d. muestral obtenida de las estadísticas de siniestros, entonces d_i y e_i se pueden calcular como

$$d_i = \sum_{k(i)} \text{Prob}\{\mathbf{Z} = Z_{k(i)}\}, \quad e_i = \frac{1}{C} \sum_{k(i)} \text{Prob}\{\mathbf{Z} = Z_{k(i)}\} Z_{k(i)} \quad (4.1.15)$$

donde las sumas están por sobre todos los puntos $Z_{k(i)}$ tales que

$$(i-1) \cdot C < Z_{k(i)} \leq i \cdot C.$$

EJEMPLO Sea S la f.d. de Pareto

$$S(Z) = 1 - \left(\frac{D+\beta}{Z+\beta}\right)^\alpha, \quad (Z \geq D),$$

con parámetros α, β y D ($\alpha > 0, \alpha \neq 1, D > 0, \beta > D$). Deducimos una fórmula de cálculo para e_i . Sustituyendo $S(M)$ en (3.3.15) la función L del valor esperado acotado de S , puede reformularse de la siguiente manera

$$L(M) = E(\min(Z, M)) = \frac{D+\beta}{\alpha-1} \cdot \left(1 - \left(\frac{D+\beta}{M+\beta}\right)^{\alpha-1}\right) + D$$

para $M > D$, y $L(M) = M$ para $M \leq D$. En lugar del cálculo directo de la integral en la segunda línea de (4.1.14), expresamos e_i en términos de las funciones L y S . Como

$$L(M) = \int_{-\infty}^M Z dS(Z) + M \cdot [1 - S(M)]$$

más generalmente tenemos

$$\int_{(i-1)C}^{iC} Z dS(Z) = L(i \cdot C) - L((i-1) \cdot C)$$

$$+ (i-1) \cdot C [1 - S((i-1) \cdot C)] - iC [1 - S(i \cdot C)]$$

$$= L(i \cdot C) - L((i-1) \cdot C) - C [1 - S((i-1) \cdot C)] + iC d_i.$$

Por lo tanto

$$e_i = i \cdot d_i + \frac{L(i \cdot C) - L((i-1) \cdot C)}{C} - [1 - S((i-1) \cdot C)].$$

Ejercicio 4.1.1 Calcular la f.d. $F(X)$ de Pólya compuesta para $X \leq 6$, cuando el número esperado de siniestros es $n=2$, $\sigma_q^2=0,1$ y la f.d. S de la intensidad de los siniestros es la distribución discreta de dos puntos $s_1=\text{Prob}\{Z=1\}=0,2$, $s_2=\text{Prob}\{Z=2\}=0,8$.

Ejercicio 4.1.2 Sea S la f.d. de Pareto($\alpha, 0, D$) con $\alpha=2$ y $D=0,5$, truncada en $Z_{\max}=10$. Discretizar S utilizando (a) el método del punto medio, (b) el método del valor esperado insesgado. La longitud del escalón es $C=1$. Comparar los valores esperados de las distribuciones discretizadas.

Ejercicio 4.1.3 Demostrar que la distribución binomial cumple la fórmula de recurrencia (4.1.1) con $a=-p/(1-p)$ y $b=-a/(N+1)$.

Ejercicio 4.1.4 Sea S la f.d. de la intensidad de los siniestros. Definir las f.d. L (L para la izquierda) equidistante y R (R para la derecha) con la longitud del escalón C que sigue a continuación:

$$R(Z) = S((i-1) \cdot C) \quad \text{para } (i-1) \cdot C \leq Z < i \cdot C$$

$$L(Z) = S(i \cdot C) \quad \text{para } (i-1) \cdot C \leq Z < i \cdot C$$

Demostrar que, si Z_L y Z_R son, respectivamente variables L y R distribuidas en forma aleatoria, entonces $E(Z_R) - E(Z_L) = C$.

Ejercicio 4.1.5 Sean S , L y R como en el ejercicio anterior. Sea T la f.d. equidistante con longitud de escalón $1/2 \cdot C$ obtenida utilizando el método del punto medio, y sea U la f.d. equidistante (4.1.12) que se obtiene utilizando el método del valor esperado insesgado. Sean F_S , F_L , F_R , F_T , F_U las f.d. compuestas correspondientes, cada una de las cuales tiene la misma distribución del número de siniestros.

(a) ¿En qué circunstancias $F_R = F_S$ es válida y cuándo $F_T = F_S$?

(b) Demostrar que $F_R \leq F_S \leq F_L$, $F_R \leq F_U \leq F_L$ y que si $S(0)=0$ y $S(r \cdot C)=1$, entonces $F_R \leq F_T \leq F_L$.

Ejercicio 4.1.6 Sean X e Y dos variables aleatorias tales que $E(X)=E(Y)$, $m \leq X \leq M$ y $\text{Prob}\{Y=m \text{ o } Y=M\}=1$. Probar que $\text{Var}(X) \leq \text{Var}(Y)$.

Ejercicio 4.1.7 Demostrar que la varianza de la distribución discretizada (4.1.12), que se obtiene utilizando el método del valor esperado insesgado, es mayor o igual a la varianza de la distribución originaria. (Clave: Utilizar el resultado del ejercicio 4.1.6)

***Ejercicio 4.1.8** Demostrar que las distribuciones de Poisson binomial negativa (es decir de Pólya) y la binomial son las únicas distribuciones del número de siniestros que cumplen la fórmula de recurrencia (4.1.1).

4.2 Fórmulas de aproximación para F

4.2.1 Introducción a las fórmulas de aproximación

(a) Necesidad de aproximación. En las secciones siguientes quedará claro que la f.d. F de Poisson mixta compuesta, que proporciona la distribución de importes acumulados de siniestros, es complicada para su cálculo.

El método de recurrencia descrito en la sección 4.1 es una contribución significativa para resolver este problema de cálculo. Sin embargo, aún se tiene la necesidad de obtener métodos aproximados rápidos y de precisión razonables. Una desventaja del método de recurrencia es que se deben seguir una gran cantidad de pasos para el cálculo cuando la cartera de riesgos es muy grande (como lo son la mayoría de las carteras de los aseguradores) o cuando la distribución de la intensidad de los siniestros tiene una cola larga. Esto provoca serios problemas, especialmente en la simulación de los procesos siniestrales, donde el cálculo de F se necesita reiteradamente, usualmente miles de veces. Por esta razón es necesario que el tiempo de cálculo sea del orden de los milisegundos más que segundos.

Los métodos de aproximación también pueden tener la virtud de sugerir relaciones analíticas entre las variables más importantes que controlan el proceso, y a veces aclararlas.

(b) Simetrización y métodos basados en los momentos. Un enfoque clásico es aproximar la f.d. F por la f.d. normal con el mismo valor medio y el mismo desvío estándar. Por medio del teorema del límite central del cálculo de probabilidades, la f.d. F es asintóticamente normal en el caso de Poisson (sección 2.3(e)) compuesto (no-mixto). Desafortunadamente,

el área de aplicación es algo limitada, como se verá en la sección 4.2.2, ya que la f.d. F , usualmente, es asimétrica mientras que la f.d. normal no lo es.

Una familia de f.d. de aproximación se obtiene aplicando una transformación de simetrización adecuada para la variable del importe acumulado de los siniestros con el objeto de eliminar la asimetría de la f.d. F , y luego aplicar la aproximación normal a esta distribución transformada. En realidad, esta misma idea se aplicó cuando la f.d. de Poisson se approximó por medio de las aproximaciones de Anscombe y de Peizer y Prass (secciones 2.3(f) y (g)). Cuando se aplican dichas aproximaciones para la f.d. F , entonces además del valor medio y del desvío estándar de F , también se utiliza la asimetría de F como un elemento clave. Sin embargo, estos tipos de aproximaciones son sólo aplicables en los casos en los que la asimetría no es grande, como se verá en secciones posteriores. El denominado método de la Potencia Normal (abreviadamente "PN"), el método de **Haldane** y la fórmula de **Wilson-Hilferty** pertenecen a este grupo de aproximaciones y nos ocuparemos de ellas en las secciones 4.2.4 y 4.2.5. Como la parametrización de las funciones de aproximación eventualmente se basa en igualar sus características básicas, es decir los momentos menores, con las características correspondientes de la distribución a ser aproximada, la simetrización es una variante especial del método de los momentos.

Otra variante del método de los momentos es aproximar directamente la f.d. F por medio de una f.d. seleccionada de alguna familia de distribuciones conocidas que concuerdan con las características más bajas, con frecuencia la media, el desvío estándar y la asimetría. Una condición necesaria es que la f.d. de aproximación tienda a la f.d. normal a medida que la asimetría se acerca a cero. Este método también es aplicable sólo para valores de asimetría relativamente pequeños. La distribución gamma trasladada que se introduce en la sección 3.3.5 es un ejemplo. Con este propósito se la adapta reemplazando las características de la variable de intensidad de siniestros

Z en la sección 3.3.5. por las características correspondientes de la variable \bar{X} de siniestros acumulados. Muchos autores utilizan la aproximación gamma, porque ofrece la posibilidad de trabajar con F en forma analítica y sus propiedades son bien conocidas y se encuentran fácilmente en libros de texto comunes. La conveniencia de la aproximación Gamma fue investigada por Bohman y Esscher (1964) y Seal (1977) la apoyó ampliamente, a pesar de que Pentikäinen (1977) la criticó moderadamente. Las pruebas han demostrado que es aplicable en circunstancias similares a la de la aproximación PN y a la de Haldane (secciones 4.2.4 y 4.2.5).

OBSERVACION. La mayoría de los métodos de aproximación se basan en los momentos inferiores (media, desvío estándar y asimetría) de la distribución objetivo. Entonces, en forma manifiesta, se obtendrá la misma aproximación para todo el grupo de las fd. con estas características iguales, pero con curtosis y con características superiores diferentes. Esto conduce a la pregunta de con qué precisión los tres momentos inferiores determinan las funciones que pertenecen a este grupo. El denominado **problema de los momentos** está bien estudiado en la bibliografía sobre el tema. Está comprobado que, si no se plantean condiciones restrictivas particulares, tales como las que se detallaron arriba, las distribuciones pueden estar muy lejos unas de otras, y las soluciones estándares del problema de los momentos no son útiles para nuestro propósito. (Goevaerts *et al.*, 1984)

(c) Otros métodos Vale la pena mencionar que la f.d. F también se puede calcular utilizando la **Transformación Rápida de Fourier** (Bohman y Esscher, 1964; Bertram, 1981; y Meyers y Schenker, 1983). Este método también se aplica a distribuciones más asimétricas.

4.2.2 Aproximación normal.

(a) Fórmula. Es habitual, tanto en la teoría de riesgo como en otros temas, aproximar la f.d. F del importe acumulado de siniestros por medio de la f.d. normal, como se hizo en el contexto de la distribución del número de siniestros (sección 2.3(e)).

Por lo tanto tenemos

$$F(X) \approx N(\nu(X)) \quad (4.2.1)$$

donde N es la f.d. normal (véase (2.3.9)) y

$$\nu(X) = (X - \mu_x)/\sigma_x \quad (4.2.2)$$

estandariza la variable para que tenga media cero y desvío estándar unitario. Las fórmulas para el evaluación numérica de N y su inversa aparecen en el Apéndice D.

OBSERVACION. La fórmula (4.2.1) está, para el objetivo de la uniformidad, formalmente delineada de acuerdo con las expresiones de Anscombe y de Peizer (2.3.12) y (2.3.13) así como las fórmulas de simetrización de las secciones siguientes, cada una de las cuales tiene a N como la función externa y una transformación diseñada adecuadamente como función interna.

(b) Análisis acerca de la aplicabilidad. La aproximación normal simplifica los cálculos y con frecuencia permite analizar los problemas que involucran numerosas variables e interrelaciones, de tal manera que de otra forma no sería factible o sólo podría hacerse con mucha dificultad. Por ejemplo la ecuación

$$\text{Prob}\{|X - \mu_x| \leq y_{c/2} \cdot \sigma_x\} = \varepsilon, \quad (4.2.3a)$$

proporciona una estimación de la probabilidad de que el importe acumulado de siniestros estandarizado se desvíe de su media como mucho $y_{c/2}$ veces el desvío estándar. $y_{c/2}$ es la raíz de la ecuación $\varepsilon = (1 - N(\gamma_{c/2})) + N(-\gamma_{c/2}) = 2N(-\gamma_{c/2})$. La ecuación limitada por un solo lado se utiliza con más frecuencia en las aplicaciones de la teoría del riesgo

$$\text{Prob}\{X \leq \mu_x + y_\epsilon \cdot \sigma_x\} = \epsilon, \quad (4.2.3b)$$

que evalúa la probabilidad de siniestros excesivos, siendo y_ϵ la raíz de $N(y_\epsilon) = 1-\epsilon$.

Los valores numéricos se pueden hallar en los libros de texto de cálculo de probabilidades. Para la conveniencia del lector, en la Tabla 4.2.1 aparecen algunos valores comunes.

Tabla 4.2.1. Soluciones aproximadas para y_ϵ y $y_{\epsilon/2}$.

ϵ	0.05	0.01	0.005	0.001
$y_\epsilon \cdot \sigma$	1.64	2.33	2.580	3.090
$y_{\epsilon/2} \cdot \sigma$	1.96	2.58	2.810	3.290

Desafortunadamente, la aproximación normal no es habitualmente lo suficientemente precisa. Es aceptable sólo si la asimetría de X es muy pequeña, como se verá en algunos ejemplos de la sección 4.2.6.

4.2.3 Simetrización

Una manera de mejorar la precisión de la aproximación normal es transformar la variable X normal en una variable auxiliar y ,

$$y = v(X), \quad (4.2.4)$$

por medio de una transformación v elegida adecuadamente que hace que la distribución sea aproximadamente simétrica. En la práctica frecuentemente esto se realiza imponiendo en la especificación para v la

condición de que la asimetría de la distribución transformada $y=v(X)$ sea igual a cero. Más aún, especificando otra v tal que estandarice la media de y a cero y el desvío estándar a la unidad, puede esperarse que la f.d. F de la variable simetrizada $y=v(X)$ pueda aproximarse más satisfactoriamente por la distribución normal.

$$F(X) = \bar{F}(y) \approx N(y). \quad (4.2.5)$$

Las diferentes funciones de transformación v dan lugar a una variedad de aproximaciones. En las próximas secciones se considerarán dos posibilidades. Primero, en la sección 4.2.4, la inversa de v se toma como un polinomio de segundo grado, dando lugar a la denominada fórmula de "PN". En segundo lugar, en la sección 4.2.5 se introduce una función potencial general, sugerida por Haldane.

Un análisis general del método de simetrización puede encontrarse en Box y Cox (1964).

4.2.4 Aproximación por Potencia Normal

(a) Fórmula PN Continuando con la idea descripta en la sección anterior, y eligiendo la inversa $v^{-1}(y)$ de la función de transformación v para que sea un polinomio de segundo grado, se obtiene la siguiente fórmula

$$x = \frac{X - \mu_x}{\sigma_x} = y + \frac{\gamma_x}{6} (y^2 - 1) (X \mu_x) \quad (4.2.5)$$

Resolviendo $y = v(X)$ y sustituyendo en (4.2.5) tenemos

$$F(X) \approx N(v(X)) = N\left(-\frac{3}{\gamma_X} + \sqrt{\frac{9}{\gamma_X^2} + 1 + \frac{6}{\gamma_X} \cdot \frac{X - \mu_X}{\sigma_X}}\right). \quad (4.2.7)$$

(b) Aplicabilidad. Las fórmulas (4.2.6) y (4.2.7) son válidas solamente para la cola derecha de la distribución y en tanto que la asimetría γ_X no supere la unidad. Lo anterior ofrece una extensión muy útil de los límites dados en (4.2.3) en relación a la aproximación normal, para el rango de variación de los siniestros. También se tiene en cuenta la asimetría de la distribución.

La fórmula (4.2.7) no tiene ninguna ventaja aparente con respecto a la fórmula de Wilson-Hilferty (que se introducirá en la sección 4.2.5), especialmente ya que la última cubre todo el rango de X , incluyendo los valores negativos.

(c) Referencias. La aproximación anterior fue hallada por Kauppi y Ojantakanen (1969) al comparar los resultados de los valores calculados simultáneamente por medio de diferentes métodos de aproximación. La fórmula es un caso especial de una transformación similar presentada por Cernish y Fisher (1937). Kendall y Stuart (1977) proporcionan una buena explicación de la idea, párrafos 6.25-27.

OBSERVACION. La idea de la simetrización puede ampliarse introduciendo más parámetros libres en la transformación v . Por ejemplo, es posible encontrar una función F transformada que tenga la misma curtosis que la f.d. normal. Entonces se puede proporcionar la fórmula con un término adicional controlado por la curtosis de la distribución originaria (Beard *et al.*, 1984, p.117). Las pruebas han demostrado que esto puede mejorar el ajuste, pero que funciona mejor en los casos donde la fórmula corta en sí misma da resultados satisfactorios.

Ejercicio 4.2.1 Obtener la fórmula aproximada (3.4.27) para la prima del reaseguro de exceso de siniestralidad.

4.2.5 Fórmula de Wilson-Hilferty

(a) Enfoque de Haldane. Haldane (1938) desarrolló la simetrización a la que nos referimos en la sección 4.2.3 introduciendo la expresión potencial

$$Y = (X/\mu_X)^b \quad (4.2.8)$$

donde b es un parámetro auxiliar que debe establecerse según la condición de que la asimetría de la variable transformada debe igualarse a cero. Además calculando la media μ_Y y el desvío estándar σ_Y para la variable Y poniendo

$$y = (Y - \mu_Y)/\sigma_Y = [(X/\mu_X)^b - \mu_Y]/\sigma_Y \quad (4.2.9)$$

se obtiene la transformación deseada $y = v(X)$ (4.2.4).

La deducción de las fórmulas de Haldane es elemental pero algo trabajosa (Pentikäinen, 1987).

(b) La fórmula de Wilson-Hilferty. Wilson y Hilferty (1931) utilizaron el mismo método que Haldane para deducir una fórmula para la aproximación numérica de la f.d. gamma. Se puede obtener como un caso especial de la fórmula general de Haldane suponiendo que el valor del parámetro b sea $1/3$. Como la f.d. gamma puede aproximar la f.d. F del importe acumulado de siniestros, como se indica en la sección 4.2.1(b), la fórmula de Wilson-Hilferty (abreviadamente WH) se indica como una aproximación para F .

En la sección 4.2.6, hay ejemplos de las aproximaciones de Haldane y de WH. En Pentikäinen (1987) se pueden encontrar pruebas. Gracias a su mayor flexibilidad, la fórmula general de Haldane es un poco

mejor, pero las diferencias no son significativas en el área de mayor interés práctico. Por lo tanto, considerando que la fórmula WH tiene una estructura más simple, la hemos adoptado en el resto de este libro como la norma que se utilizará en las aplicaciones.

La fórmula Wilson-Hilferty está dada por

$$F(X) \approx N(y) = N(W(x)), \quad (4.2.10)$$

donde x es la variable estandarizada $x = (X - \mu_x)/\sigma_x$, N es la f.d. normal estándar (1.4.22a) y la transformación WH es

$$\begin{aligned} W(x) &= c_1 + c_2 \cdot (x + c_3)^{1/3} \\ c_1 &= 1/(3g) - 3g, \quad c_2 = 3g^{2/3}, \quad c_3 = g, \quad g = 2/\gamma_x \end{aligned} \quad (4.2.11)$$

En muchos problemas se da el valor $F(X)$ y se pide la X correspondiente. Para resolver las ecuaciones anteriores, se debe cumplir con los pasos siguientes.

- (1) Resolver y para $N(y) = F(X)$ (ver Apéndice D)
- (2) Establecer $x = (1/c_2)^3 \cdot (y - c_1)^3 c_3$
- (3) Establecer $X = \mu_x + x \cdot \sigma_x$.

La derivada de (4.2.10) por la derecha es

$$\frac{dN(W(x))}{dx} = \frac{c_2}{\sigma_x \sqrt{18\pi}} \cdot (x + c_3)^{-2/3} \cdot \exp [c_1 + c_2(x + c_3)^{1/3}] \quad (4.2.13)$$

donde nuevamente x es la variable estandarizada.

(c) Aplicabilidad. Las comparaciones que se consideran en la sección 4.2.6 demuestran que, para las distribuciones levemente asimétricas, la aproximación WH así como las fórmulas de PN y la de Haldane, proporcionan resultados razonablemente precisos. Sin embargo, cuando la asimetría es grande la aplicabilidad se deteriora rápidamente. Si la asimetría γ_x supera la unidad o como mucho el valor 1,2, todos estos métodos se tornan poco confiables y no se deberían utilizar.

Otro método de prueba se obtiene calculando en forma numérica las características básicas μ , σ , γ y γ_2 , a partir de la f.d. de aproximación $N(v(X))$ y comparándolas con las correspondientes cantidades originarias μ_x , σ_x , γ_x y $\gamma_{2,x}$. Esto comprueba que las tres características inferiores son muy cercanas de las originarias en tanto que

$$0 \leq \gamma_x \leq 1.2, \quad (4.2.14)$$

mientras que para las distribuciones más asimétricas aún la media y los desvíos estándares no concuerdan, indicando que las fórmulas no son aceptables. Además, la curtosis γ_2 de la distribución Wilson-Hilferty satisface aproximadamente la ecuación

$$\gamma_2 \approx 1.6 \cdot \gamma^2 \quad (4.2.15)$$

donde se cumple la condición (4.2.14). Esto se puede utilizar para probar la aplicabilidad de esta aproximación. Si la curtosis de una f.d. F del importe acumulado de los siniestros F se desvía considerablemente de la curtosis (4.2.15) de la distribución Wilson-Hilferty $N(W(x))$ correspondiente, entonces no se puede esperar que el ajuste de la aproximación sea preciso. Si el desvío de (4.2.15) es menor que 0,2, el ajuste, generalmente, es muy bueno.

4.2.6 Comparación de aproximaciones

(a) Aplicabilidad. Para conocer bien la aptitud de los métodos de aproximación, se calcularon numerosas distribuciones utilizando, en paralelo, métodos exactos - especialmente de recurrencia - juntamente con las distintas aproximaciones analizadas. Pentikäinen publicó una amplia colección (1987) que consiste en 54 distribuciones de diferentes tipos. La tabla 4.2.2 muestra algunos resultados típicos.

Si se puede demostrar en forma fehaciente que con un método de aproximación se obtiene una precisión aceptable en diferentes áreas de aplicación, entonces se puede justificar su uso en cálculos prácticos. Un aspecto útil surgió de estas pruebas, como ya se señaló en la sección 4.2.5(c): la precisión es, como regla, buena, o por lo menos satisfactoria, en los casos en los que la asimetría de la distribución es pequeña. Otras condiciones son que el número de siniestros no debe ser muy pequeño y la distribución de la intensidad de un siniestro no debe ser heterogénea, es decir que no tenga una cola muy larga. La última condición, generalmente, se cumple cuando se reaseguran los picos de los riesgos.

OBSERVACION. Debe apreciarse que los datos básicos, con frecuencia, son bastante limitados, hacen espuria la exigencia de una precisión excesiva de la técnica del cálculo, especialmente si la consecuencia es un incremento significativo del tiempo de cálculo.

(b) Conclusiones. Las pruebas realizadas hasta el momento indican que no existen diferencias significativas entre los enfoques de PN, WH y el de Haldane. Por lo tanto, en los casos donde se den los criterios de aplicabilidad mencionados más arriba, la elección entre estas fórmulas debería realizarse sobre la base de la conveniencia práctica. Esto apunta a la fórmula WH. Sin embargo, la expresión analítica (4.2.6) que se obtuvo con el método de PN con frecuencia, es muy útil, mientras que los otros enfoques no tienen esta cualidad.

Como es previsible, la aproximación normal no es muy satisfactoria, excepto cuando la asimetría es pequeña.

Por último, señalemos que los métodos aproximados y los exactos (y la simulación directa, a la que nos referiremos en la sección 5.4(a)) con frecuencia se complementan entre ellos. Los primeros son más convenientes en los colectivos grandes que típicamente tienen asimetría moderada, mientras que los métodos exactos pueden utilizarse para colectivos pequeños, que con frecuencia son tan asimétricos que las aproximaciones son de valor limitado.

Ejercicio 4.2.2 Demostrar que la asimetría γ y la curtosis γ_2 de la f.d. gamma trasladada cumplen (comparar (4.2.15) la igualdad

$$\gamma_2 = (3/2)\gamma^2$$

Tabla 4.2.2. Ejemplo de aproximaciones según Penttiläinen (1987)

Distribución	Exacto			Aproximación			Error %			
	x	F o 1-F	N	PN	WH	HA	AN	ΔPN	ΔWH	ΔHA
Ejemplo 1 $\gamma_x = 0.08$	-2.0	0.0205	0.0228	—	0.0205	0.0205	11.0	—	0.1	-0.0
	-1.5	0.0646	0.0668	—	0.0645	0.0646	3.4	—	-0.0	-0.0
	-1.0	0.1586	0.1587	—	0.1586	0.1587	0.0	—	-0.0	0.0
	1.0	0.1586	0.1587	0.1587	0.1586	0.1587	0.0	0.0	-0.0	0.0
	2.0	0.0249	0.0228	0.0249	0.0249	0.0249	-8.6	0.1	0.0	-0.0
	3.0	0.0019	0.0013	0.0019	0.0019	0.0019	-28.9	0.4	0.5	-0.1
Ejemplo 2 $\gamma_x = 0.24$	4.0	0.0001	0.0000	0.0001	0.0001	0.0001	-68.3	1.0	1.7	-0.5
	-2.0	0.0162	0.0228	—	0.0160	0.0160	40.4	—	-1.1	-1.4
	-1.5	0.0595	0.0668	—	0.0594	0.0594	12.3	—	-0.2	-0.2
	-1.0	0.1578	0.1587	—	0.1580	0.1581	0.5	—	0.1	0.2
	1.0	0.1579	0.1587	0.1587	0.1581	0.1582	0.5	0.4	0.1	0.1
	2.0	0.0288	0.0228	0.0289	0.0288	0.0288	-21.0	0.4	-0.1	-0.1
Ejemplo 3 $\gamma_x = 0.59$	3.0	0.0031	0.0013	0.0031	0.0031	0.0031	-56.5	-2.1	-1.7	-2.3
	4.0	0.0002	0.0000	0.0002	0.0002	0.0002	-84.2	-8.2	-5.1	-7.2
	-2.0	0.0078	0.0228	—	0.0059	0.0067	191.7	—	-25.1	-14.4
	-1.5	0.0456	0.0668	—	0.0443	0.0447	46.5	—	-2.8	-2.0
	-1.0	0.1506	0.1587	—	0.1538	0.1522	5.3	—	2.1	1.1
	1.0	0.1529	0.1587	0.1587	0.1554	0.1541	3.8	3.8	1.6	0.8
Ejemplo 4 $\gamma_x = 1.08$	2.0	0.0362	0.0228	0.0372	0.0361	0.0359	-23.2	■ 2.8	-0.1	-0.6
	3.0	0.0068	0.0013	0.0064	0.0064	0.0066	-80.1	■ 4.9	-5.4	-2.1
	4.0	0.0011	0.0000	0.0009	0.0009	0.0010	-97.1	■ -17.4	-12.2	-0.8
	-2.0	0.0033	0.0228	—	0.0000	0.0006	589.4	—	-100.0	-83.1
	-1.5	0.0322	0.0668	—	0.0145	0.0209	107.5	—	-55.2	-35.1
	-1.0	0.1357	0.1587	—	0.1384	0.1307	16.9	—	2.0	-3.7
	1.0	0.1376	0.1587	0.1587	0.1491	0.1419	15.3	15.3	8.4	3.1
	2.0	0.0376	0.0228	0.0470	0.0431	0.0407	-39.5	25.1	14.7	8.3
	3.0	0.0125	0.0013	0.0119	0.0112	0.0116	-89.2	-4.9	-10.4	-7.3
	4.0	0.0042	0.0000	0.0027	0.0028	0.0035	-99.2	-35.5	-34.0	-16.5

Notas: x es la variable estandarizada ($(x-\bar{x})/\sigma$), y se calcula por medio de la fórmula de recurrencia. Las columnas N, PN, WH y HA proporcionan los valores aproximados obtenidos por las fórmulas Normal, PN, Wilson-Hilferty y Haldane. La distribución del número de siniestros ha sido tanto de Poisson o de Pólya y las distribuciones del importe de siniestros fueron de Pareto truncada o log-normal.