Una vez que ya conocemos la frecuencia y severidad por separado hay que comenzar a unir estos dos conceptos. Para eso vamos a construir una v.a. que no solo combine los temas anteriores sino que los generalice.

Comencemos con un caso particular.

Modelo individual

Suponga que tiene un portafolio de n pólizas individuales de seguros válidas por un año como se muestra en la Figura 1.1.

Figura 1.1

Sea p_j la probabilidad de que el j-ésimo asegurado no efectúe ninguna reclamación durant el tiempo de vigencia del seguro, y sea q_j la probabilidad de que se observe exactamente una reclamación. Suponga que la igualdad $p_j + q_j = 1$ se cumple, ello significa que no puede haber más de una reclamación por cada asegurado. Tal situación puede corresponder, por ejemplo, a los seguros de vida. Defina una variable aleatoria

$$\boxed{\mathrm{D}_j = \begin{cases} 1 & \text{si hay reclamación en la póliza } j, \\ 0 & \text{si no hay reclamación en la póliza } j. \end{cases}}$$

Claramente D_j tiene una distribución Bernoulli con parámetro q_j. El uso de la letra D viene del término en inglés *Death*. Observe que el número total de reclamaciones está dado por la variable aleatoria $N = D_1 + \cdots + D_n$. Ahora suponga artificialmente que cada póliza efectúa una reclamación, y sea la variable aleatoria $C_j > 0$ el monto de la reclamación efectuada por la póliza j. Debido a que los siniestros pueden presentarse con características distintas y ello puede derivar en distintos montos de reclamación, consideraremos de manera general a C_j no como una constante sino como una variable aleatoria. La letra C proviene del término en inglés *Claim*, que se traduce en español como *reclamación*. La verdadera reclamación de la póliza j está dada por el producto

$$\boxed{\mathrm{D}_j C_j = \begin{cases} C_j & \text{si } D_j = 1, \\ 0 & \text{si } D_j = 0. \end{cases}}$$

Observe que esta variable aleatoria puede ser mixta, es decir, no ser continua ni discreta. véase la Figura 1.2 en donde se muestran posibles gráficas de la función de distribución de esta variable aleatoria. De esta forma se considera como datos en este modelo la colección de vectores aleatorios $(D_1, C_1), ..., (D_n, C_n)$, que supondremos independientes. Consideraremos además que las variables D_j y C_j también son independientes entre sí.

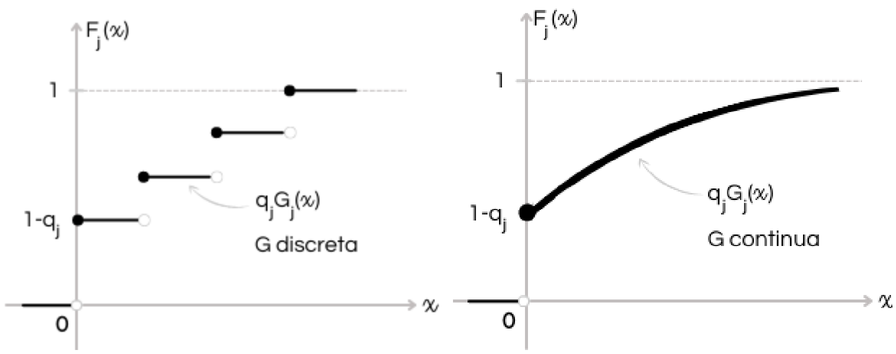


Figura 1.2

El monto de reclamaciones agregadas, o también llamado agregado de reclamaciones, en el modelo individual, es la variable aleatoria

$$S = \sum_{j=1}^{n} D_j C_j. \tag{1.1}$$

Esta variable es el monto que afronta una compañía aseguradora por concepto de reclamaciones durante el periodo completo del seguro. La ecuación (1.1) representa el *modelo individual* para una póliza de seguros de las características señaladas. El modelo tiene este nombre pues supone conocer las probabilidades de reclamación y posible monto de reclamación de todos y cada uno de los asegurados de manera individual. Una posible desventaja de este modelo es que presupone que el número de asegurados en la cartera se mantiene constante durante todo el tiempo de vigencia del seguro. Desde el punto de vista matemático, y también desde la perspectiva del negocio del seguro, nuestro objetivo es conocer las características de la variable S, a quien llamaremos *riesgo*. Si $F_j(x)$ denota la función de distribución del producto D_jC_j, entonces la función de distribución $F(x)$ del riesgo S adquiere la siguiente expresión en términos de convoluciones:

$$F(x) = (F_1 * \cdots * F_n)(x)$$

Esta fórmula general y compacta es, sin embargo, un tanto difícil de calcular y no la utilizaremos con frecuencia. Como primeros resultados generales se presentan a continuación algunas características de S. Denotaremos por $G_j(x)$ la función de distribución de C_j, y como es costumbre, cuando exista, $M_X(t)$ denota la función generadora de momentos de una variable X cualquiera

Proposición *Bajo la notación e hipótesis del modelo individual se tienen los siguientes resultados.*

1. $\mathbb{E}(S) = \sum_{j=1}^{n} q_j \mathbb{E}(C_j)$.

2. $Var(S) = \sum_{j=1}^{n} \left[q_j Var(C_j) + q_j p_j \mathbb{E}^2(C_j) \right]$.

3. $F_j(x) = \begin{cases} 1 + q_j(G_j(x) - 1) & si \quad x \geq 0, \\ 0 & si \quad x < 0. \end{cases}$

4. $M_{D_jC_j}(t) = 1 + q_j(M_{C_j}(t) - 1)$.

5. $M_S(t) = \prod_{j=1}^{n} [1 + q_j(M_{C_j}(t) - 1)]$.

Nota: $p_j = 1 - q_j$

Demostración.

1. Por la hipótesis de independencia,

$$\mathbb{E}(S) = \sum_{j=1}^{n} \mathbb{E}(D_jC_j) = \sum_{j=1}^{n} \mathbb{E}(D_j)\mathbb{E}(C_j) = \sum_{j=1}^{n} q_j \mathbb{E}(C_j).$$

2. Primeramente tenemos que

$$\begin{aligned} Var(D_jC_j) &= \mathbb{E}(D_j^2 C_j^2) - \mathbb{E}^2(D_jC_j) \\ &= q_j \mathbb{E}(C_j^2) - q_j^2 \mathbb{E}^2(C_j) \\ &= q_j [Var(C_j) + \mathbb{E}^2(C_j)] - q_j^2 \mathbb{E}^2(C_j) \\ &= q_j Var(C_j) + q_j p_j \mathbb{E}^2(C_j). \end{aligned}$$

Por lo tanto

$$Var(S) = \sum_{j=1}^{n} Var(D_jC_j) = \sum [q_j Var(C_j) + q_j p_j \mathbb{E}^2(C_j)].$$

3. Para cualquier número real $x \geq 0$,

$$
\begin{aligned}
F_j(x) &= P(D_jC_j \leq x) \\
&= P(D_jC_j \leq x | D_j = 0)P(D_j = 0) + P(D_jC_j \leq x | D_j = 1)P(D_j = 1) \\
&= P(0 \leq x | D_j = 0)p_j + P(C_j \leq x | D_j = 1)q_j \\
&= p_j + q_j G_j(x) \\
&= 1 + q_j(G_j(x) - 1).
\end{aligned}
$$

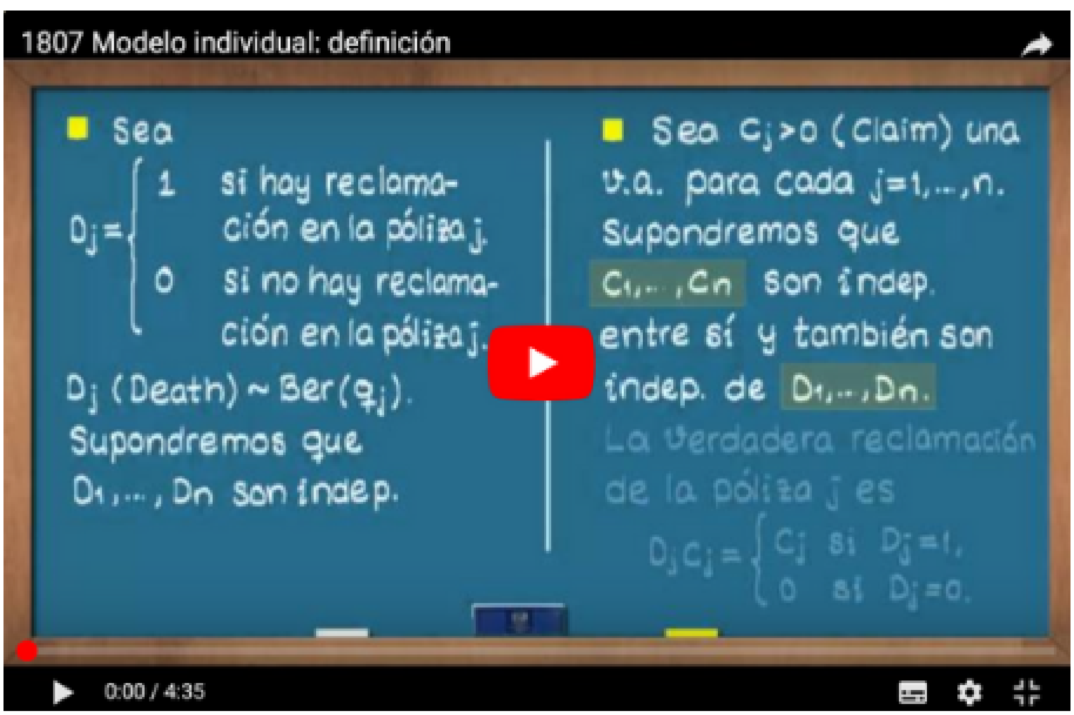
4. Nuevamente condicionando sobre el valor de D_j,

$$
\begin{aligned}
M_{D_jC_J}(t) &= \mathbb{E}(e^{tD_jC_j}) \\
&= \mathbb{E}(e^{tD_jC_J} | D_j = 0)P(D_j = 0) + \mathbb{E}(e^{tD_jC_J} | D_j = 1)P(D_j = 1) \\
&= p_j + q_j M_{C_j}(t) \\
&= 1 + q_j(M_{C_j}(t) - 1).
\end{aligned}
$$

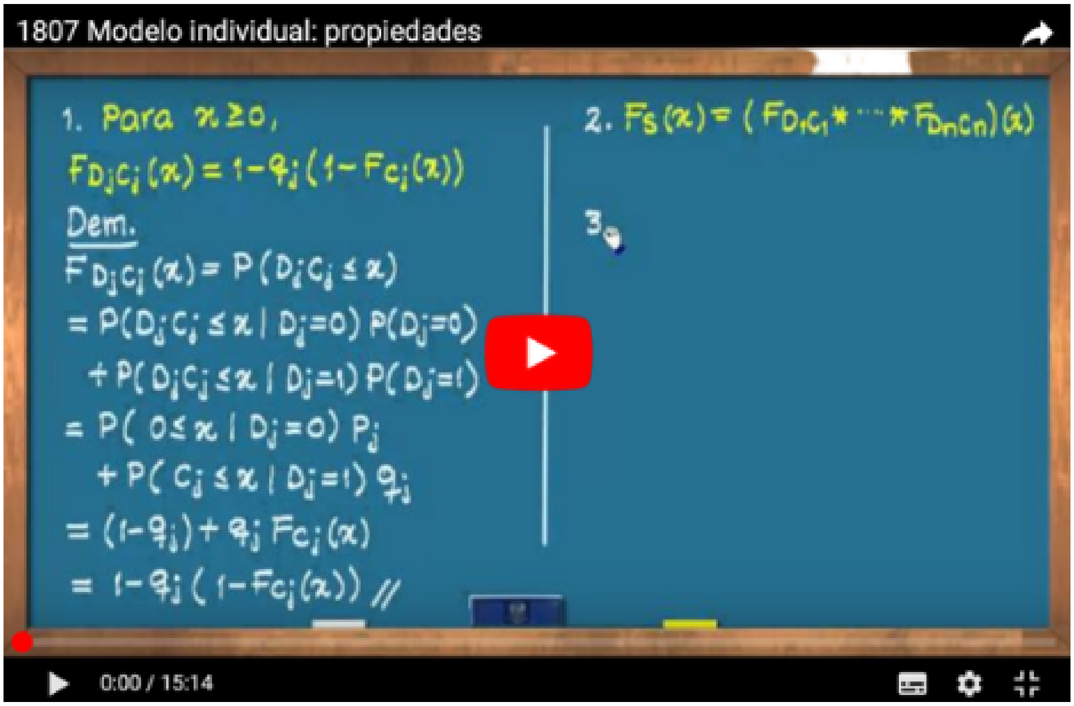
5. Esta igualdad se sigue directamente de la anterior usando la hipótesis de independencia.■

Nota: $S \geq 0$ bajo el contexto y definiciones que se mencionaron en esta sección.

A continuación se muestran vídeos con la explicación de lo mencionado anteriormente:



https://youtu.be/rekGEr6sGoQ



https://youtu.be/uGhUhhZh30k

Ejemplo

Sean $\{C_i\} \sim Exp\left(\frac{1}{200}\right)$ los montos de reclamación para $n = 100$ asegurados, de los cuales la probabilidad de reclamación está dada por la siguiente tabla:

# de asegurados	45	37	18
q_j	0.02	0.04	0.07
	Grupo 1	Grupo 2	Grupo 3

- $\mathbb{E}[S] = \sum_{j=1}^{100} q_j \mathbb{E}[C_j] = 200(\underbrace{45(0.02)}_{\text{Grupo 1}} + \underbrace{37(0.04)}_{\text{Grupo 2}} + \underbrace{18(0.07)}_{\text{Grupo 3}}) = 728$

- $Var(S) = \sum_{j=1}^{100} q_j Var(C_j) + q_j \overbrace{(1-q_j)}^{p_j} \mathbb{E}^2[C_j]$

$$= 200^2 \sum_{j=1}^{100} [q_j + q_j(1-q_j)] \quad \left\{ Var(C_j) = \mathbb{E}^2[C_j] = 200^2 \right.$$

$$=200^2[\underbrace{(0.02+0.02(0.98))(45)}_{\text{Grupo 1}} + \underbrace{(0.04+0.04(0.96))(37)}_{\text{Grupo 2}} + \underbrace{(0.07+0.07(0.93))(18)}_{\text{Grupo 3}}] = 284,584$$

Observamos que a pesar de contar con ciertas propiedades del riesgo S ninguna de ellas nos dice cómo obtener probabilidades de esta variable aleatoria. Lo que sucede es que en general resulta complicado calcular probabilidades de sumas de variables aleatorias el cual está ligado con una expresión matemática conocida como convoluciones. Para entender más a fondo esto primero veremos algunos resultados.

Derivación bajo el signo de integral

Teorema de Leibniz *Sea* $f : [a, b] \times [c, d] \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ *una función continua y sean* $\alpha, \beta : [c, d] \rightarrow \mathbb{R}$ *funciones derivables tales que*

$$a \leq \alpha(y) \leq x \leq \beta(y) \leq b \quad \forall y \in [c, d]$$

supongamos que $\frac{\partial f}{\partial y}$ *existe y es continua en el conjunto*

$$T = \{(x, y)\mathbb{R}^2 | \alpha(y) \leq x \leq \beta(y), \quad y \in [c, d]\}$$

entonces

$$F(y) = \int_{\alpha(y)}^{\beta(y)} f(x, y)dx$$

existe es derivable $\forall y \in [c, d]$ *y*

$$F'(y) = \int_{\alpha(y)}^{\beta(y)} \frac{\partial f(x, y)}{\partial y} dx + f(\beta(y), y)\beta'(y) + f(\alpha(y), y)\alpha'(y)$$

Nota: Si α y β son funciones constantes, entonces:

$$\frac{\partial}{\partial y} \int_{\alpha}^{\beta} f(x, y)dx = \int_{\alpha}^{\beta} \frac{\partial}{\partial y} f(x, y)dx$$

Teorema de probabilidad total en el caso continuo

Sea X y Y v.a. continuas con $f_X(t)$ y $f_Y(t)$ sus respectivas funciones de densidad. Entonces:

$$F_Y(t) \doteq \mathbb{P}[Y \leq t] = \int_{\mathbb{R}} \mathbb{P}[Y \leq t | X = x] f_X(x)dx \doteq \int_{\mathbb{R}} F_{Y|X=x}(t) f_X(x)dx$$

Dem

Llamemos $A \doteq Y \leq t$ el evento " 'Y' fue menor o igual a 't' ". Entonces:

$$\mathbb{I}_A(t) \sim Ber(\mathbb{P}[A] \doteq F_Y(t))$$

Recuerden que $\mathbb{E}[Y|X]$ es una v.a que depende del valor que toma X. Así:

$$\begin{aligned}
\underline{F_Y(t)} &= \mathbb{E}[\mathbb{I}_A(t)] = \mathbb{E}[\mathbb{E}[\mathbb{I}_A(t)|X]] && \left\{ \text{Aplicando esperanza iterada.} \right. \\
&= \int_{\mathbb{R}} E[\mathbb{I}_A(t)|X = x] f_X(x) dx && \left\{ \text{Aplicando el teorema del estadístico inconsciente.} \right. \\
\mathbb{E}[\mathbb{I}_A(t)|X = x] &= 0 \cdot \mathbb{P}[\neg A|X = x] + 1 \cdot \mathbb{P}[A|X = x] = \mathbb{P}[A|X = x] \\
&= \int_{\mathbb{R}} \mathbb{P}[A|X = x] f_X(x) dx && \left\{ \mathbb{I}_A(t)|X = x \sim Ber(\mathbb{P}[A|X = x]) \text{ esto pues } \mathbb{I}_A(t) \in \{0, 1\} \text{ siempre.} \right. \\
&= \int_{\mathbb{R}} \mathbb{P}[Y \leq t|X = x] f_X(x) dx && \left\{ \text{Por definición de A.} \right. \\
&\doteq \underline{\int_{\mathbb{R}} F_{Y|X=x}(t) f_X(x) dx}
\end{aligned}$$

Una vez teniendo el resultado anterior con v.a. continuas, recordemos el Teorema de probabilidad total en el caso discreto:

Teorema de probabilidad total en el caso discreto:

Si X, Y son v.a. con soporte en los $\mathbb{N} \cup \{0\}$ entonces:

$$f_Y(t) = \mathbb{P}[Y = t] = \sum_{x=0}^{\infty} \mathbb{P}[Y = t|X = x]\mathbb{P}[X = x] = \sum_{x=0}^{\infty} f_{Y|X=x}(t) f_X(x)$$

Pero esta notación con "f" no debe confundirse, sería "muy cómodo" simplemente cambiar la suma por una integral para decir que es válido en el caso continuo, pero estaríamos abusando de la notación y reaccionando de una manera heurística al confundir f con $\mathbb{P}$.

https://youtu.be/pJI8LfnPPB8

Sin embargo, y como ha pasado antes en la historia, la idea anterior es correcta pero no directamente del caso discreto.

En el pasado, por ejemplo, la gente afirmaba lo siguiente sin haber realizado una prueba formal de que esto ocurría:

$$\mathbb{E}[g(x)] = \int_{\mathbb{R}} g(x) dF_X = \begin{cases} \int_{\mathbb{R}} g(x) f_X(x) dx & \text{si } X \text{ es continua} \\ \sum_{\forall x} g(x) f_X(x) & \text{si } X \text{ es discreta} \end{cases}$$

Es la fórmula del estadístico inconsciente. De ahí recibe su nombre y la llamada ”suerte del estadístico”. El teorema de probabilidad total para el caso continuo se da porque:
Gracias al último resultado, tenemos que: $F_y(t) = \int_{\mathbb{R}} F_{y|x=x}(t) f_X(x) dx$

$$\implies f_Y(t) \doteq \frac{d}{dt} F_Y(t) = \frac{d}{dt} \int_R F_{Y|X=x}(t) f_X(x) dx$$

$$= \int_{\mathbb{R}} \frac{d}{dt} F_{Y|X=x}(x) f_X(x) dx \left\{ \begin{array}{l} \textit{aplicando la regla de derivacion bajo} \\ \\ \textit{el signo integral} \end{array} \right.$$

$$\boxed{\therefore\ f_Y(t) = \int_{\mathbb{R}} f_{Y|X=x}(t) f_X(x) dx}$$

Teniendo en mente estos resultados anteriores, podemos seguir con el tema de obtener probabilidad de el riesgo S.

Consideremos X, Y v.a.i. discretas y tomemos $S \doteqdot X + Y$, notemos que S es también discreta y además:

$$\begin{aligned} f_S(t) &= \mathbb{P}[S = t] = \mathbb{P}[X + Y = t] = \mathbb{P}[X = t - Y] \\ &= \sum_{\forall y} \mathbb{P}[X = t - Y | Y = y][Y = y] \left\{ \textit{Probabilidad total} \right. \\ &= \sum_{\forall y} \mathbb{P}[X = t - y][Y = y] \left\{ \begin{array}{l} \textit{Pues } X \perp Y \textit{ y} \\ \textit{además } Y = y \end{array} \right. \\ &= \sum_{\forall y} f_X(t - y) f_Y(y) = \mathbb{E}[f_X(t - Y)] \end{aligned}$$

Por lo tanto, para cualesquiera par de v.a.i. X, Y discretas y tomando $S = X + Y$ entonces:

$$f_S(t) = \underbrace{\sum_{\forall y} f_X(t - y) f_Y(y)}_{\text{Convolución}} = \mathbb{E}[f_X(t - Y)]$$

o bien:

$$f_S(t) = \underbrace{\sum_{\forall x} f_Y(t - x) f_X(x)}_{\text{Convolución}} = \mathbb{E}[f_Y(t - X)]$$

Un resultado similar sucede en el caso continuo.

Consideremos X, Y v.a.i. continuas y tomemos $S \doteqdot X + Y$, notemos que S es también continua y además:

$$
\begin{aligned}
F_S(t) &\ddot{=} \mathbb{P}[S \leq t] = \mathbb{P}[X+Y \leq t] = \mathbb{P}[X \leq t-Y] \\
&= \int_{\mathbb{R}} \mathbb{P}[X \leq t-y|Y=y] f_Y(y)dy \left\{ \begin{array}{l} Probabilidad\ Total\ en\ el \\ caso\ continuo \end{array} \right. \\
&= \int_{\mathbb{R}} \mathbb{P}[X \leq t-y] f_Y(y)dy \left\{ \begin{array}{l} X \perp Y \\ \implies \mathbb{P}[X \leq t-y|Y=y] = P[X \leq t-y] \end{array} \right. \\
&\doteq \int_{\mathbb{R}} F_X(t-y) f_Y(y)dy = \mathbb{E}[F_X(t-Y)]
\end{aligned}
$$

Por lo tanto, para cualesquiera par de v.a.i. X, Y continuas y tomando $S = X+Y$ entonces:

$$
F_S(t) = \underbrace{\int_{\mathbb{R}} F_X(t-y) f_Y(y)dy}_{\text{Convolución}} = \mathbb{E}[F_X(t-Y)]
$$

o bien:

$$
F_S(t) = \underbrace{\int_{\mathbb{R}} F_Y(t-x) f_X(x)dx}_{\text{Convolución}} = \mathbb{E}[F_Y(t-X)]
$$

Del resultado anterior se sigue que:

$$
\begin{aligned}
f_S(t) &= \frac{d}{dt} F_S(t) = \frac{d}{dt} \int_{\mathbb{R}} F_X(t-y) f_Y(y)dy \\
&= \int_{\mathbb{R}} \frac{d}{dt} F_X(t-y) f_Y(y)dy \left\{ \begin{array}{l} Aplicando\ la\ regla\ de\ Leibniz \\ de\ derivacion\ bajo\ el\ signo\ integral \end{array} \right. \\
&= \int_{\mathbb{R}} f_X(t-y) f_Y(y)dy = \mathbb{E}[f_X(t-Y)]
\end{aligned}
$$

Por lo tanto, para culesquiera par de v.a.i. X, Y continuas y tomando $S = X+Y$ entonces:

$$F_S(t) = \underbrace{\int_{\mathbb{R}} f_X(t - y) f_Y(y) dy}_{\text{Convolución}} = \mathbb{E}[F_X(t - Y)]$$

o bien:

$$F_S(t) = \underbrace{\int_{\mathbb{R}} f_Y(t - x) f_X(x) dx}_{\text{Convolución}} = \mathbb{E}[F_Y(t - X)]$$

Que es básicamente el mismo resultado que en el caso discreto.

En resumen:

Sumas y Convoluciones

Si X e Y son variables aleatorias independientes con f.d. F_X y F_Y, respectivamente, entonces la f.d. de la suma $Z = X + Y$ es la convolución de F_x y F_Y

$$F_Z = \int_{-\infty}^{\infty} F_X(z-t)dF_Y(t) = \int_{-\infty}^{\infty} F_Y(z-t)dF_X(t)$$

Si X e Y toman valores en los enteros no-negativos con funciones de probabilidad despectivas p_X y p_Y entonces

$$p_Z(n) = P(Z=n) = \sum_{i=0}^{n} P(X=i)P(Y=n-i) = \sum_{i=0}^{n} p_X(i)p_Y(n-i) = \sum_{i=0}^{n} p_X(n-i)p_Y(i)$$

Si consideramos la situación en la cual X tienen Y densidades f_X y f_Y, respectivamente, la densidad f_Z de la suma es la convolución de las densidades f_X y f_Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-t)f_Y(t)dt = \int_{-\infty}^{\infty} f_Y(z-t)f_X(t)dt$$

Nota: $\frac{dF_Y(t)}{dt} = f_Y(t) \Longleftrightarrow dF_Y(t) = f_Y(t)dt$

Ahora, pensando en el modelo individual, definamos $X_j \doteqdot C_j D_j$ entonces $\{X_j\}_{j=1}^{n}$ son v.a.i. $\forall j \in \{1, ..., n\}$.

$$S_n \doteqdot \sum_{j=1}^{n} C_j D_j \doteqdot \sum_{j=1}^{n} X_j$$

Luego, notemos que $S_{n-1} \perp X_n$ pues $\{X_j\}_{j=1}^{n-1} \perp X_n$. De hecho $S_{n-k} \perp X_m$ $\forall m \in \{n-k+1, ..., n\}$ y variando $k \in \{1, ..., n-1\}$. Entonces aplicando los resultados obtenidos anteriormente tendríamos en particular para el caso continuo con la función de distribución:

$$F_{S_n}(t) = \int_{\mathbb{R}} F_{S_{n-1}}(t-x_n) f_{X_n}(x_n) dx_n$$

$$= \int_{\mathbb{R}} (\int_{\mathbb{R}} F_{S_{n-2}}(t-x_n-x_{n-1}) f_{X_{n-1}}(x_{n-1}) dx_{n-1}) f_{X_n}(x_n) dx_n$$

$$= \int_{\mathbb{R}} (\int_{\mathbb{R}} (\int_{\mathbb{R}} F_{S_{n-3}}(t - \textstyle\sum_{j=0}^{2} x_{n-j}) f_{X_{n-2}}(x_{n-2}) dx_{n-2}) dx_{n-1}) f_{X_n}(x_n) dx_n$$

$$= ... = \underbrace{\int_{\mathbb{R}} (\int_{\mathbb{R}} (...(\int_{\mathbb{R}}}_{\text{n-1 veces}} F_{S_{n-3}}(t - \textstyle\sum_{j=0}^{n-2} x_{n-j}) f_{X_2}(x_2) dx_2)...) f_{X_{n-1}}(x_{n-1}) dx_{n-1}) f_{X_n}(x_n) dx_n$$

Estos cálculos obtienen probabilidades exactas del riesgo y así tendríamos una expresión de la probabilidad de S en términos de sus sumandos $\{X_j\}_{j=1}^n$. Como hacer cálculos como el anterior es "demasiado fácil", se ha optado por alternativas que nos ayudan a calcular probabilidades mediante ciertas aproximaciones o simplemente asumiendo casos particulares un tanto más manejables.

Aproximación Normal

Recordemos el siguiente teorema:

TEOREMA CENTRAL DEL LÍMITE.

Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que para cada natural n, $\mathbb{E}(X_n) = \mu$ y $Var(X_n) = \sigma^2 < \infty$. Entonces

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$$

Cuando n es grande y el portafolio de asegurados es homogéneo en el sentido de que las variables $D_j C_j$ son idénticamente distribuidas con segundo momento finito, puede usarse el teorema central del límite para aproximar la distribución de S mediante la distribución normal, es decir,

$$P(S \leq x) = P\left(\frac{S - \mathbb{E}(S)}{\sqrt{Var(S)}} \leq \frac{x - \mathbb{E}(S)}{\sqrt{Var(S)}}\right) \approx \Phi\left(\frac{x - \mathbb{E}(S)}{\sqrt{Var(S)}}\right)$$

Esta aproximación puede ser adecuada para ciertos riesgos pero tiene la desventaja de que asigna una probabilidad positiva al intervalo $(-\infty, 0)$, lo cual no es consistente con el hecho de que $S \geq 0$. Sin embargo, dado que la distribución $N(\mu, \sigma^2)$ se concentra principalmente en el intervalo $(\mu - 4\sigma, \mu + 4\sigma)$, cuando la esperanza y la varianza de S son tales que $\mathbb{E}(S) - 4\sqrt{Var(S)} \geq 0$, la probabilidad asignada a la parte negativa del eje es realmente pequeña, ello alivia un poco el hecho de que esta distribución no tenga soporte en el intervalo $[0, \infty)$. Tal vez la situación más comprometedora sea que la función de densidad normal decae muy rápidamente pues existen riesgos cuyas funciones de densidad no cumplen con tal característica.

Nota: Si $X_j \doteqdot C_j D_j$ y $\{X_j\}_{j=1}^{n}$ v.a.i.i.d. tomando $S \doteqdot \sum_{j=1}^{n} X_j$ entonces:

$$\text{Ejercicio al lector} \begin{cases} \circ\mathbb{E}[S] = n\mathbb{E}[X] \\ \\ \circ Var(S) = nVar(X) \end{cases}$$

Ejemplo

Supongamos un grupo homogéneo de asegurados tales que $D_j \sim Ber(0.1)$ y $C_j \sim \chi^2_{(3)}$. Tomando $X_j \doteqdot D_j \bullet C_j$ entonces:

$$\begin{aligned} \mathbb{E}[X_j] &= \mathbb{E}[D_j]\mathbb{E}[C_j]\{\ D_j \perp C_j \quad \forall j \\ &=(0.1)(3) && false \\ &=0.3 \quad \forall j && false \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X_j^2] &= \mathbb{E}[(D_j C_j)^2] \\ &= \mathbb{E}[D_j^2]\mathbb{E}[C_j^2]\{\ D_j \perp C_j \\ & && false \\ &=1.5 \quad \forall j && false \end{aligned}$$

$$\Longrightarrow Var(X_j) = \mathbb{E}[X_j^2] - \mathbb{E}^2[X_j] = 1.5 - 0.3^2 = 1.41 \quad \forall j$$

Considerando un portafolio con $n = 150$ pólizas, calculamos la media y varianza del riesgo S.

$$\circ\mathbb{E}[S] = n\mathbb{E}[X] = 150(0.3) = 45$$

$$\circ Var(S) = nVar(X) = 150(1.41) = 211.5$$

Bajo aproximación normal tenemos:

$$
\begin{aligned}
F_S(t) &\doteq \mathbb{P}[S \leq t] \\
&= \mathbb{P}\left[\frac{S - \mathbb{E}[S]}{\sqrt{Var(S)}} \leq \frac{t - \mathbb{E}[S]}{\sqrt{Var(S)}}\right] \\
&\approx \Phi\left(\frac{t - \mathbb{E}[S]}{\sqrt{Var(S)}}\right)
\end{aligned}
$$

Entonces

$$
F_S(t) \approx \Phi\left(\frac{t - 45}{\sqrt{211.5}}\right) \quad \forall t \geq 0
$$

Con lo cual podemos aproximar probabilidades del riesgo S y misma herramienta para cuantiles.

Por ejemplo si quisiéramos saber cuál es la probabilidad de aproximada de que S sea menor a 65 entonces:

$$
\begin{aligned}
F_S(65) &\approx \Phi\left(\frac{65 - 45}{\sqrt{211.5)}}\right) \\
&= 0.9154697
\end{aligned}
$$

También si quisiéramos obtener un cuantil de probabilidad p, podríamos estimarlo de forma general:

$$
\Phi\left(\frac{t_0 - \mathbb{E}[S]}{\sqrt{Var(S)}}\right) = p \Leftrightarrow q_S(p) \approx t_0 = \sqrt{Var(S)}\Phi^{-1}(p) + \mathbb{E}(S)
$$

Que en nuestro ejemplo sería para $p = 97.5\,\%$

$$
q_S(97.5\,\%) \approx t_0 = \sqrt{211.5}(1.96) + 45 \approx 73.50383
$$

Un video con otra aplicación relacionada con el tema lo pueden ver a continuación

Figura 1: `https://www.youtube.com/watch?v=TAjhROWBkiE`

Ahora, intentemos obtener probabilidades exactas de un riesgo, pero buscando simplificar el problema con un ejemplo:

Supongamos que $D_j \sim Ber(0,1)$ y que C_j tiene la siguiente f.m.p

$$\mathbb{P}[C_j = c] \quad = \quad \begin{cases} 0.8 & \text{si} \quad c = 1 \quad \forall j \\ 0.2 & \text{si} \quad c = 2 \\ 0 & \quad\quad\; e.o.c \end{cases}$$

Tomando un portafolio con $n = 3$ pólizas bajo el modelo individual, busquemos probabilidades de S.

Notemos primero que $S \ddot{=} \sum_{i=1}^{n} x_i \in \{0,1,2,3,4,5,6\}$.

- Cuando $S = 0$ fue porque:

$$\overbrace{\underbrace{C_1 D_1}_{\text{asegurado 1}} = \underbrace{C_2 D_2}_{\text{asegurado 2}} = \underbrace{C_3 D_3}_{\text{asegurado 3}}}^{\text{Reclamaciones por asegurado}} = 0 \quad \xrightarrow{\text{la denotamos}} \quad \overbrace{(0,0,0)}^{\text{suman cero}}$$

Entonces S=0.

¡Solo hay un caso posible!

- Cuando S=1 fue porque:

$$\underbrace{\overbrace{\{(1,0,0)\}}^{\text{evento \#1}} \bigcup \overbrace{\{(0,1,0)\}}^{\text{evento \#2}} \bigcup \overbrace{\{(0,0,1)\}}^{\text{evento \#3}}}_{\text{Posibles eventos}} \quad \longrightarrow \quad \text{Hay 3 casos} \tag{1}$$

Recordemos un poco de combinatoria antes de continuar.

¿Cuantas "palabras"diferentes podemos formar con las letras de la palabra missíssippi?

$\frac{m}{1}\ \frac{i}{2}\ \frac{s}{3}\ \frac{s}{4}\ \frac{i}{5}\ \frac{s}{6}\ \frac{s}{7}\ \frac{i}{8}\ \frac{p}{9}\ \frac{p}{10}\ \frac{i}{11} \longrightarrow$ Hay 11 letras.

$\frac{m}{1}\ \frac{i}{2}\ \frac{s}{3}\ \frac{p}{4} \longrightarrow$ pero solamente hay 4 letras diferentes.

Pero contamos con una "m", cuatro "i", cuatro "s", y dos "p". Como buscamos "palabras"diferentes, buscamos hacer permutaciones. Utilizando el coeficiente multinomial tenemos:

$\frac{11!}{1!4!4!2!} = 34,650$ palabras diferentes/permutaciones.

Una vez visto esto, seguimos sin contar tanto. .

- Cuando S=2 fue porque:

a. (2,0,0) y este evento puede suceder de $\frac{3!}{1!2!} = 3$ formas diferentes.

O bien puede pasar que:

b. (1,1,0) y este evento puede suceder de $\frac{3!}{2!1!} = 3$ formas diferentes.

Para un total de <u>6 casos</u> distintos.

- Cuando S=3 fue porque:

$$\left.\begin{array}{lcr} (2,1,0); & 3! = 6 & \text{casos} \\ (1,1,1); & 1 & \text{caso} \end{array}\right\} \text{7 casos}$$

- Cuando S=4 fue porque:

$$\left.\begin{array}{ll} (2,2,0); & \dfrac{3!}{2!} = 3 \quad \text{casos} \\ \\ (1,1,2); & \dfrac{3!}{2!} = 3 \quad \text{casos} \end{array}\right\} 6 \text{ casos}$$

- Cuando S=5 fue porque:

$$(2,2,1); \quad \frac{3!}{2!} = 3 \quad \text{casos}$$

- Cuando S=6 fue porque:

$$(2,2,2); \quad \text{solo hay un caso}$$

Ahora que tenemos la cantidad de casos, vamos a calcular probabilidades del riesgo S. Denotemos como $X_i = D_iC_i \quad \forall i \in \{1,2,3\}$. Entonces:

$$\begin{aligned} \mathbb{P}[S=t] &= \mathbb{P}[D_1C_1 + D_2C_2 + D_3C_3 = t] \\ &= \mathbb{P}[X_1 + X_2 + X_3 = t] \quad \forall t \in \ \{0,1,...6\} \end{aligned}$$

1. $t = 0$

$$\left.\begin{aligned} \mathbb{P}[S=0] &= \mathbb{P}[X_1 + X_2 + X_3 = 0] \\ &= \mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 0] \\ &= \mathbb{P}[X_1 = 0]\mathbb{P}[X_2 = 0]\mathbb{P}[X_3 = 0] \\ &= \mathbb{P}[X = 0] \\ &= \mathbb{P}[\text{No hubo reclamación}] \\ &= 0.9^3 \end{aligned}\right\} \text{Todo pues } X_i \geq, X_i \perp X_j \text{ si } i \leq j, X_i \text{ son i.d}$$

2. t=1

$$\begin{aligned} \mathbb{P}[S=1] &= \mathbb{P}[X_1 + X_2 + X_3 = 1] \\ &= \mathbb{P}[\{X_1 = 0, X_2 = 0, X_3 = 0\} \cup \{X_1 = 0, X_2 = 1, X_3 = 0\} \cup \{X_1 = 0, X_2 = 0, X_3 = 1\}] \\ &= \mathbb{P}[\{X_1 = 0, X_2 = 0, X_3 = 0\}] + \mathbb{P}[\{X_1 = 0, X_2 = 1, X_3 = 0\}] + \mathbb{P}[\{X_1 = 0, X_2 = 0, X_3 = 1\}] \\ &= 3\mathbb{P}[(1,0,0)] \\ &= 3[\underbrace{(0.1)(0.8)}_{\text{1 reclamo \$1}}][\underbrace{0.9^2}_{\text{2 no reclamaron}}] \quad \text{Pues cada evento es ajeno y las v.a son i.i.d} \end{aligned}$$

Nota: Si $a \in \{1, 2\} \Rightarrow \mathbb{P}[CD = a] = \mathbb{P}[D = 1, C = a] = \mathbb{P}[D = 1]\mathbb{P}[C = a]$

3. t=2

$$\begin{aligned}
\mathbb{P}[S=2] &= [\overbrace{3}^{\#\text{ De veces que ocurre el evento}}] \cdot \underbrace{\mathbb{P}[\overbrace{(2,0,0)}^{\#\text{ De veces que ocurre el evento}}]}_{\text{Evento en probabilidad}} \cdot 3\mathbb{P}[(1,1,0)] \\
&= \underbrace{3 \cdot [(0.1)(0.2)][0.9]^2}_{\text{Evento en probabilidad}} + 3 \cdot [(0.1)(0.8)]^2[0.9]
\end{aligned}$$

4. t=3

$$\begin{aligned}
\mathbb{P}[S=3] &= 6 \cdot \mathbb{P}[(2,1,0)] + \mathbb{P}[(1,1,1)] \\
&= 6 \cdot [(0.1)(0.2)][(0.1)(0.8)][0.9] + [(0.1)(0.8)]^3
\end{aligned}$$

5. t=4

$$\begin{aligned}
\mathbb{P}[S=4] &= 3 \cdot \mathbb{P}[(2,2,0)] + 3\mathbb{P}[(1,1,2)] \\
&= 3 \cdot [(0.1)(0.2)]^2[0.9] + 3 \cdot [(0.1)(0.8)]^2[(0.1)(0.2)]
\end{aligned}$$

6. t=5

$$\begin{aligned}
\mathbb{P}[S=5] &= 3 \cdot \mathbb{P}[(2,2,1)] \\
&= 3 \cdot (0.1)^3[(0.2)^2(0.8)]
\end{aligned}$$

7. t=6

$$\begin{aligned}
\mathbb{P}[S=6] &= \mathbb{P}[(2,2,2)] \\
&= [(0.1)(0.2)]^3
\end{aligned}$$

Lo cual nos permite contar con las probabilidades exactas de S. Así, su f.m.p es:

$$
\begin{aligned}
f_S(t) &= \mathbb{P}[S = t] \\
&= \begin{cases}
72.9\,\% & \text{si} \quad t = 0 \\
19.44\,\% & \text{si} \quad t = 1 \\
6.588\,\% & \text{si} \quad t = 2 \\
0.9152\,\% & \text{si} \quad t = 3 \\
0.1464\,\% & \text{si} \quad t = 7 \\
0.0096\,\% & \text{si} \quad t = 5 \\
0.0008\,\% & \text{si} \quad t = 6 \\
0 & \qquad\; e.o.c
\end{cases}
\end{aligned}
$$