

La distribución logarítmico-normal. Aplicación a la distribución del coste de siniestros

Por

ANGEL MARIN COBO

DEFINICION DEL MODELO

En diferentes campos de aplicación se hace uso de distribuciones asimétricas con fines de análisis, ajuste, inferencia y predicción. En la etapa de elección del modelo representativo o especificación de un proceso estocástico cuyos resultados aparecen más o menos asimétricamente distribuidos, suelen considerarse como funciones de densidad idóneas, la distribución Gamma, Beta, de Pareto y Logaritimico-normal, entre otras.

De las distribuciones citadas merece un especial interés la logaritmico-normal por sus aplicaciones en Economía como modelo teórico aplicable en la distribución personal de la renta y en las distribuciones de ventas y, en el campo actuarial, para el ajuste de poblaciones demográficas y para la distribución de la cuantía de siniestros en los seguros No-Vida de Accidentes, Responsabilidad Civil del automóvil, Incendios y Pérdida de beneficios debidos al fuego.

Esta distribución denominada también de Galton Mac-Alister puede interpretarse como aquella que resulta de una determinada causa y a la cual se superponen un gran número de impulsos o causas aleatorias independientes, cuyos efectos se caracterizan por determinados factores multiplicativos.

Así, a la *aditividad* de los efectos que conduce a la distribución normal, corresponde, en el caso de la ley logaritmico-normal, la *proporcionalidad* de los efectos. Es por ello que la distribución logaritmico normal se llama también *ley del efecto proporcional* dado por Gibrat.

Siguiendo el razonamiento que hace H. Cramer (1), supongamos que la cuantía del daño ocasionado por un siniestro depende de n causas o impulsos que expresamos por las variantes $\xi_1, \xi_2, \dots, \xi_n$, cuyo efecto depende de la fuerza del impulso y de la cuantía del siniestro potencial en el instante en que actúa el impulso en cuestión. Por x_v , indicaremos la cuantía del daño producido por las v primeras causas o impulsos.

Formulemos las siguientes hipótesis:

1.^a Las variantes $\xi_1, \xi_2, \dots, \xi_n$ son estocásticamente independientes.

2.^a El incremento del daño producido por el impulso ξ_{v+1} es proporcional a dicho impulso y al tamaño del siniestro producido por las v primeras causas, es decir: $\Delta x_v = \xi_{v+1} x_v$

Según esto,

$$\xi_{v+1} = \frac{\Delta x_v}{x_v} = \frac{x_{v+1} - x_v}{x_v}$$

y, por tanto,

$$S_n = \xi_1 + \xi_2 + \dots + \xi_n = \sum_{v=0}^{n-1} \xi_{v+1} = \sum_{v=0}^{n-1} \frac{x_{v+1} - x_v}{x_v}$$

donde $\sum_{v=0}^{n-1} \xi_{v+1} =$ Suma de las v primeras causas.

Si el aumento producido por cada impulso es pequeño, tenemos pasando al límite:

$$\sum_{v=0}^{n-1} \frac{\Delta x_v}{x_v} \sim \int_0^{\infty} \frac{dx}{x} = \log x$$

$$\Delta \rightarrow 0$$

$$n \rightarrow \infty$$

Al sumar v variables independientes y al hacer tender $n \rightarrow \infty$ damos entrada al Teorema Central del límite, con lo que la suma

$$S_n = \sum_{v=0}^{n-1} \xi_{v+1}$$

(1) Fundamentos de la Teoría de las Probabilidades. Editorial Aguilar.

tiende a una distribución normal, es decir:

$$S_n \rightarrow N(\alpha, \sigma)$$

Por lo tanto, si S_n es $N(\alpha, \sigma)$ y $S_n = \log x$ según la expresión de paso al límite, entonces se dice que x es una variante logaritmico-normal.

Según esto podemos decir que una variante ξ representativa de la cuantía del daño de un siniestro sigue una distribución logaritmico-normal, si su logaritmo sigue una distribución normal, $N(\alpha, \sigma)$, es decir,

$$\log \xi = \sigma \eta + \alpha \Leftrightarrow \xi = e^{\sigma \eta + \alpha}$$

siendo $\eta \rightarrow N(0, 1)$

LEYES DE PROBABILIDAD Y CARACTERISTICAS DE LA DISTRIBUCION

Función de densidad

$$g(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\alpha)^2}{2\sigma^2}} = N(\alpha, \sigma)$$

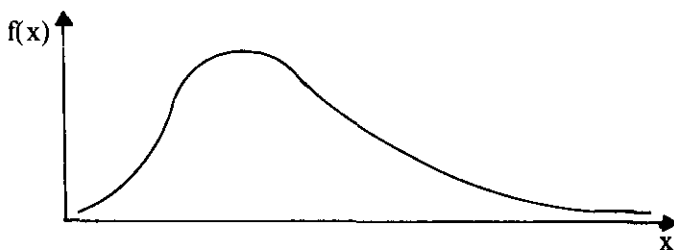
Haciendo el cambio de variable $y = \log x$;

$$dy = \left| \frac{1}{x} \right| dx$$

la función de densidad de x será:

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\log x - \alpha)^2}{2\sigma^2}} \quad 0 < x < \infty$$

Y su representación gráfica:



Función de distribución

$$\begin{aligned}
 F(x) = P(\xi \leq x) &= \int_0^x \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(lx - \alpha)^2}{2\sigma^2}} dx = \\
 &= \frac{1}{\sigma \sqrt{2\pi}} \int_0^x \frac{1}{x} e^{-\frac{(lx - \alpha)^2}{2\sigma^2}} dx
 \end{aligned}$$

Momento ordinario de orden r

Por definición

$$\alpha_r = E(\xi^r) = \int_0^{\infty} x^r f(x) dx = \int_0^{\infty} x^r \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(lx - \alpha)^2}{2\sigma^2}} dx$$

integral de inmediata solución haciendo el cambio de variable

$$t = \frac{lx - \alpha}{\sigma}$$

cuyo resultado es:

$$\alpha_r = e^{r\alpha + \frac{r^2 \sigma^2}{2}}$$

que para $r = 1$

$$\alpha_1 = e^{\alpha + \frac{\sigma^2}{2}}$$

y para $r = 2$

$$\alpha_2 = e^{2(\alpha + \sigma^2)}$$

Haciendo $r = it$ en α_r nos da la expresión de la función característica:

$$\varphi(t) = e^{it\alpha + \frac{(it)^2 \sigma^2}{2}}$$

Características

Esperanza matemática:

$$E(\xi) = \alpha_1 = e^{\alpha + \frac{\sigma^2}{2}}$$

Varianza:

$$D^2(\xi) = \alpha_2 - \alpha_1^2 = e^{2\alpha + \sigma^2} (e^{\sigma^2} - 1)$$

Otras características:

Mediana: $M_e = e^\alpha$

Moda: $M_o = e^{\alpha - \sigma^2}$

Coefficiente de variación: $C_v = \frac{\sigma}{\alpha_1} = \sqrt{e^{\sigma^2} - 1}$

INFERENCIA DEL MODELO

1. Estimación de los parámetros α y σ^2

La especificación del modelo exige la estimación de los parámetros α y σ^2 que en él figuran. De los diferentes métodos estadísticos para la estimación de parámetros consideraremos:

a) El método de la máxima verosimilitud

Sea $L(x_1, x_2, \dots, x_n, \alpha, \sigma) = \prod_{i=1}^n f(x_i; \alpha, \sigma)$ la función de verosimilitud de la muestra, es decir:

$$L = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \alpha)^2} = \left(\frac{1}{\sigma^2 \sqrt{2\pi}} \right)^{n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \alpha)^2}$$

$$\ln L = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \alpha)^2$$

la segunda función de verosimilitud.

Maximizando lL tenemos:

$$\frac{\partial lL}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^n (lx_i - \alpha) = 0 \quad ; \quad \sum_{i=1}^n (lx_i - \alpha) = 0$$

$$\frac{\partial lL}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2} \frac{\sum_{i=1}^n (lx_i - \alpha)^2}{\sigma^4} = 0 \quad ; \quad n\sigma^2 = \sum_{i=1}^n (lx_i - \alpha)^2$$

De donde las estimaciones máximo-verosímiles de α y σ^2 , serían:

$$\hat{\alpha} = \frac{\sum_{i=1}^n lx_i}{n} \quad y \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (lx_i - \alpha)^2}{n} = S^2$$

esto es, la media y varianza muestrales de los logaritmos neperianos de la variable x_i , respectivamente.

b) Método de los momentos de Karl Pearson

Al igualar los momentos del modelo con los correspondientes de la muestra, tendríamos el sistema:

$$\begin{aligned} \bar{x} &= e^{\alpha + \frac{\sigma^2}{2}} \\ S^2 &= e^{2(\alpha + \sigma^2)} - e^{2\alpha + \sigma^2} \end{aligned}$$

que resuelto nos da los siguientes valores:

$$\hat{\alpha} = 1 \left(\frac{\bar{x}^2}{\sqrt{\bar{x}^2 + S^2}} \right) \quad y \quad \hat{\sigma}^2 = 1 \left(1 + \frac{S^2}{\bar{x}^2} \right)$$

c) A través de la relación existente entre los momentos α_2 y α_1

De la relación existente entre los momentos α_2 y α_1 anteriores, se llega a:

$$\hat{\sigma}^2 = 1 \left(\frac{\alpha_2}{\alpha_1^2} \right) = 1\alpha_2 - 21\alpha_1$$

$$\hat{\alpha} = 1 \left(\frac{\alpha_1^2}{\sqrt{\alpha_2}} \right) = 21\alpha_1 - \frac{1}{2}1\alpha_2$$

en donde α_1 y α_2 se estiman a través de los correspondientes momentos de la distribución empírica.

2. Bondad del ajuste. Test χ^2

Para efectuar el contraste, elaboraremos el estadístico Id:

$$Id = \sum_{i=1}^n \frac{(f_i - Np_i)^2}{Np_i} \rightarrow \chi_{(n-1)-k}^2$$

Siendo n el número de clases y k los parámetros a estimar.

El criterio de contrastación será:

Si $Id < \chi_{(n-1)-k}^2(\alpha)$ se acepta la hipótesis de ajuste.

Si $Id \geq \chi_{(n-1)-k}^2(\alpha)$ se rechaza la hipótesis de ajuste.

Siendo α el nivel de significación marcado.

DISTRIBUCION LOGARITMICO-NORMAL GENERALIZADA

Se llama variante logarítmico-normal generalizada a una variante que, mediante un cambio de origen, sigue la ley logarítmico-normal, es decir:

Si

$$\xi' = \xi + x_0 \Leftrightarrow \xi' = x_0 + e^{\sigma \eta + \alpha}$$

o también,

$$\frac{\log(\xi' - x_0) - \alpha}{\sigma} = \eta \rightarrow N(0, 1)$$

Esta variante logarítmico-normal generalizada depende de tres parámetros: x_0 , α , σ .

Leyes de probabilidad y características

Las leyes de probabilidad y características de la variante logarítmico-normal generalizada se deducen fácilmente de las leyes y características correspondientes de la variante logarítmico-normal, teniendo en cuenta la transformación: $\xi' = \xi + x_0$

Función de distribución

$$F_{\xi'}(x) = P(\xi' \leq x) = P(\xi + x_0 \leq x) = P(\xi \leq x - x_0) = F_{\xi}(x - x_0)$$

esto es, la función de distribución de la variante ξ en el punto $(x - x_0)$.

Función de densidad

$$f_{\xi'}(x) = F'_{\xi'}(x) = F'_{\xi}(x - x_0) = f_{\xi}(x - x_0)$$

o sea,

$$f_{\xi'}(x) = \frac{1}{\sigma(x - x_0)\sqrt{2\pi}} e^{-\frac{|\ln(x - x_0) - \alpha|^2}{2\sigma^2}} \quad x_0 < x < \infty$$

Características

Esperanza matemática:

$$E(\xi') = E(\xi + x_0) = x_0 + E(\xi) = x_0 + e^{\alpha + \sigma^2/2}$$

Varianza:

$$D^2[\xi'] = D^2[\xi + x_0] = D^2[\xi] = e^{2\alpha + \sigma^2} (e^{\sigma^2} - 1)$$

Otras características:

$$\begin{aligned} \text{Mediana: } & M_e = x_0 + e^{\alpha} \\ \text{Moda: } & M_o = x_0 + e^{\alpha - \sigma^2} \end{aligned}$$

APLICACION A LA OBTENCION DE LA DISTRIBUCION BASICA DE LA CUANTIA DE SINIESTROS EN EL SEGURO DE RESPONSABILIDAD CIVIL DE AUTOMOVILES

De la distribución muestral del coste del siniestro para la modalidad de Responsabilidad Civil Ilimitada, publicada por UNESPA en septiembre de 1985 en la Estadística del Seguro Voluntario del Automóvil, referida a datos de

1983, que figura en la Tabla I, se ha hecho una estimación de los parámetros α y σ^2 por el Método de la máxima-verosimilitud, habiéndose obtenido para éstos los valores de:

$$\alpha = 9,56 \quad \text{y} \quad \sigma^2 = 1,064$$

Según cálculos que figuran en la Tabla II.

Asimismo, se ha determinado el estadístico $\chi^2 = \sum \frac{(f_i - p_i)^2}{p_i}$ según consta en la Tabla III, cuyo test de contraste para un nivel de significación $\alpha = 0,05$ y $(n - 1) - k = 26$ grados de libertad, nos ha permitido aceptar la hipótesis de ajuste logaritmico-normal, ya que:

$$Id = 25,59 < \chi_{26}^2(0,05) = 38,88$$

Especificación de la función

1. Función de densidad

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \alpha)^2}{2\sigma^2}} = \frac{1}{1,03 \times \sqrt{2\pi}} e^{-\frac{(\ln x - 9,56)^2}{2 \cdot 1,064}}$$

2. Parámetros

Esperanza: $E(\xi) = \alpha_1 = e^{\alpha + \sigma^2/2} = e^{10,092} = 24.149$

Varianza: $D^2(\xi) = \alpha_2 - \alpha_1^2 = e^{2\alpha + \sigma^2} (e^{\sigma^2} - 1) = e^{20,184} (e^{1,064} - 1)$

Desviación típica: $(D(\xi)) = e^{\alpha + \sigma^2/2} \sqrt{e^{\sigma^2} - 1} = 33.269$

3. Otras características:

Mediana: $M_c = e^\alpha = e^{9,56} = 14.186$

Moda: $M_o = e^{\alpha - \sigma^2} = e^{8,496} = 4.895$

Coefficiente de variación: $C_v = \frac{\sigma}{\alpha_1} = \sqrt{e^{\sigma^2} - 1} = \sqrt{e^{1,064} - 1} = 1,37766$

TABLA I
DISTRIBUCION MUESTRAL DEL COSTE DEL SINIESTRO
PARA LA MODALIDAD DE RESPONSABILIDAD
CIVIL ILIMITADA

Categoría: Turismos Datos: 1983

Importe del siniestro		Coste medio del intervalo	Frecuencia relativa (en porcentaje)
Intervalos del coste			
	Menos de 1.000	607	1,6756
1.000 -	2.000	1.649	1,3025
2.000 -	3.000	2.690	1,9612
3.000 -	4.000	3.654	2,9078
4.000 -	5.000	4.727	4,9075
5.000 -	6.000	5.642	4,2665
6.000 -	7.000	6.606	4,1229
7.000 -	8.000	7.650	4,3319
8.000 -	9.000	8.590	3,4877
9.000 -	10.000	9.803	6,4737
10.000 -	12.000	11.148	6,0885
12.000 -	14.000	13.171	13,3439
14.000 -	16.000	15.035	7,2944
16.000 -	18.000	17.384	5,4760
18.000 -	20.000	19.494	4,6026
20.000 -	25.000	23.024	5,6877
25.000 -	30.000	28.093	3,9851
30.000 -	35.000	32.786	2,4423
35.000 -	40.000	38.103	2,2093
40.000 -	50.000	46.610	3,7698
50.000 -	75.000	61.671	3,8273
75.000 -	100.000	88.035	2,1481
100.000 -	150.000	124.975	1,6258
150.000 -	200.000	177.466	0,7980
200.000 -	300.000	246.668	0,6274
300.000 -	400.000	349.193	0,2124
400.000 -	500.000	458.484	0,1229
500.000 -	750.000	594.293	0,1051
750.000 -	1.000.000	907.302	0,0497
1.000.000 -	1.500.000	1.259.003	0,0326
1.500.000 -	2.000.000	1.860.999	0,0525
2.000.000 -	2.500.000	2.308.695	0,0213
2.500.000 -	5.000.000	3.545.609	0,0284
5.000.000 -	7.500.000	6.839.426	0,0049
	Mas de 7.500.000	9.923.098	0,0049

100

Fuente: Estadística de Seguro Voluntario del Automóvil. UNESPA. Septiembre 1985.

TABLA II
ESTIMACION DE LOS PARAMETROS

x_i	$\ln x_i$	f_i	$f_i(\ln x_i)$	$f_i(\ln x_i)^2$
607	6,41	1,6756	10,740596	68,84722
1649	7,41	1,3025	9,651525	71,51780
2690	7,90	1,9612	15,493480	122,39849
3654	8,20	2,9078	23,843960	195,52047
4727	8,46	4,9075	41,517450	351,23763
5642	8,64	4,2665	36,862560	318,49252
6606	8,80	4,1229	36,281520	319,27738
7650	8,94	4,3319	38,727186	346,22104
8590	9,06	3,4877	31,598562	286,28297
9803	9,19	6,4737	59,493303	546,74345
11148	9,32	6,0885	56,744820	528,86172
13171	9,49	13,3439	126,633610	1201,75300
15035	9,62	7,2944	70,172128	675,05587
17384	9,76	5,4760	53,445760	521,63062
19498	9,88	4,6026	45,473688	449,28004
23024	10,04	5,6877	57,104508	573,32926
28093	10,24	3,9851	40,807424	417,86802
32786	10,40	2,4423	25,399920	264,15917
38103	10,55	2,2093	23,308115	245,90061
46610	10,75	3,7698	40,525350	435,64751
61671	11,03	3,8273	42,215119	465,63276
88035	11,39	2,1481	24,466859	278,67752
124975	11,74	1,6258	19,086892	224,08011
177466	12,09	0,7980	9,647820	116,64214
246668	12,42	0,6274	7,792308	96,78047
349193	12,76	0,2124	2,710224	34,58246
458484	13,04	0,1229	1,602616	20,89811
594293	13,30	0,1051	1,397830	18,59114
907302	13,72	0,0497	0,681884	9,35545
1259003	14,05	0,0326	0,458030	6,43532
1860999	14,44	0,0525	0,758100	10,94696
2308695	14,65	0,0213	0,312045	4,57146
3545609	15,08	0,0284	0,428272	6,45834
6389426	15,67	0,0049	0,076783	1,20319
9923098	16,11	0,0049	0,078939	1,27171
		100	955,539186	9236,15193

$a_1 = \bar{x} = 9,56$
 $a_2 = 92,362$

$m_2 = S^2 = a_2 - a_1^2 = 1,064$
 $S = 1,03$

TABLA III
DETERMINACION DEL ESTADISTICO χ^2

Intervalos $I_n x_i$	$P(Z_{i-1} < \eta < Z_i) = P_i$	f_i	$Id = \frac{(f_i - p_i)^2}{p_i}$
0 - 6,90	$P(-9,28 < \eta < -2,58) = 0,00494$	0,494	2,826
6,90 - 7,60	$P(-2,58 < \eta < -1,90) = 0,02378$	2,378	0,486
7,60 - 8	$P(-1,90 < \eta < -1,51) = 0,03680$	3,680	0,803
8 - 8,29	$P(-1,51 < \eta < -1,23) = 0,04383$	4,383	0,497
8,29 - 8,52	$P(-1,23 < \eta < -1) = 0,04931$	4,931	1,120
8,52 - 8,70	$P(-1 < \eta < -0,83) = 0,04461$	4,461	0,008
8,70 - 8,85	$P(-0,83 < \eta < -0,69) = 0,04183$	4,183	—
8,85 - 8,99	$P(-0,69 < \eta < -0,55) = 0,04606$	4,606	0,016
8,99 - 9,10	$P(-0,55 < \eta < -0,45) = 0,03519$	3,519	—
9,10 - 9,21	$P(-0,45 < \eta < -0,34) = 0,04058$	4,058	1,438
9,21 - 9,39	$P(-0,34 < \eta < -0,17) = 0,06557$	6,557	0,033
9,39 - 9,55	$P(-0,17 < \eta < 0) = 0,06750$	6,750	6,441
9,55 - 9,68	$P(0 < \eta < 0,12) = 0,04776$	4,776	1,328
9,68 - 9,80	$P(0,12 < \eta < 0,23) = 0,04319$	4,319	0,310
9,80 - 9,90	$P(0,23 < \eta < 0,33) = 0,03835$	3,835	0,154
9,90 - 10,13	$P(0,33 < \eta < 0,55) = 0,07954$	7,954	0,646
10,13 - 10,31	$P(0,55 < \eta < 0,73) = 0,05847$	5,847	0,593
10,31 - 10,46	$P(0,73 < \eta < 0,87) = 0,04054$	4,054	0,641
10,46 - 10,60	$P(0,87 < \eta < 1) = 0,03349$	3,349	0,388
10,60 - 10,82	$P(1 < \eta < 1,22) = 0,04743$	4,743	0,200
10,82 - 11,23	$P(1,22 < \eta < 1,62) = 0,05861$	5,861	0,706
11,23 - 11,51	$P(1,62 < \eta < 1,89) = 0,02324$	2,324	0,013
11,51 - 11,92	$P(1,89 < \eta < 2,29) = 0,01837$	1,837	0,024
11,92 - 12,21	$P(2,29 < \eta < 2,57) = 0,00593$	0,593	0,071
12,21 - 12,61	$P(2,57 < \eta < 2,96) = 0,00354$	0,354	0,211
12,61 - 12,90	$P(2,96 < \eta < 3,24) = 0,00094$	0,094	0,149
12,90 - 13,12	$P(3,24 < \eta < 3,46) = 0,00033$	0,033	0,245
13,12 - 13,53	$P(3,46 < \eta < 3,85) = 0,00021$	0,021	0,337
13,53 - 13,82	$P(3,85 < \eta < 4,14) = 0,00006$	0,006	5,91
		100,—	$\Sigma = 25,594$

BIBLIOGRAFIA

- AZORÍN, F. y PALACIOS, M. J.: «Exposición general e indicaciones para el uso de la distribución logarítmal». *Estadística Española*, n.º 68 y 69. Julio-Diciembre, 1975
- BANCKET, L. B.: «The Lognormal Model for the Distribution of One Claim». *The Astin Bulletin*. Vol. II. Part. 1. Enero, 1962.
- H. CRAMER: «Teoría de probabilidades». Editorial Aguilar.
- NIETO DE ALBA, U.: Apuntes de la Cátedra de Teoría Matemática del Seguro. Facultad de Económicas.
- LOPEZ CACHERO, M.: «Fundamentos y Métodos de Estadística». *Pirámide*.
- PRIETO PEREZ, E.: Anales del Instituto de Actuarios Españoles n.º 20. Modelos probabilísticos de la distribución de la renta.
- SIXTO RIOS: «Métodos Estadísticos». Ediciones Castillo.
- VELEZ CATALAN, A.: «Distribución de los stocks en almacén». *Estadística Española* n.º 23. Abril-junio, 1964.