

EL METODO DE SIMULACION Y LA RESISTENCIA CARACTERISTICA DEL HORMIGON (*)

Por VALENTIN MARTIN JADRAQUE

Doctor Ingeniero de Caminos, Canales y Puertos

El estudio que se describe, y que se ha propuesto como base para una futura Instrucción para el Hormigón, aborda el cálculo y estudio de un estimador centrado de la resistencia característica, por medio del método de simulación estadística.

En una primera parte se analizan algunos aspectos negativos del estimador propuesto por el Comité Europeo del Hormigón y, en consecuencia, por la "Instrucción 1968 para el proyecto y la ejecución de obras de hormigón en masa o armado" del Ministerio de Obras Públicas.

En una segunda parte se describe un estimador de la resistencia característica que elimina en buena parte los inconvenientes antes apuntados, estudiándose algunos aspectos, como son su distribución, su normalidad y, por último, las curvas de eficacia, es decir, las probabilidades de aceptar o rechazar hormigones rechazables y hormigones aceptables, definidos éstos según que la resistencia característica sea menor o mayor o igual que la resistencia característica nominal figurada en proyecto.

Por último, se estudia un estimador de la resistencia característica en función del menor de los valores de una muestra de extensión n , y de forma que la mediana del estimador sea precisamente la resistencia característica.

Por su extensión, dividimos el trabajo en dos partes, dejando la segunda para el número próximo, que comprende los apartados 4 y 5.

0. INTRODUCCION

La contribución del ordenador a la resolución de multitud de problemas de difícil planteamiento teórico-matemático ha sido en los últimos tiempos decisiva. En el caso concreto que nos ocupa de calcular estimadores de la resistencia característica del hormigón por medio del método de simulación, la aportación del ordenador ha sido fundamental, pues algunos de estos estimadores de los que se habla en el estudio tienen unas leyes de distribución difíciles de abordar teóricamente, y gracias al ordenador ha quedado reducido el problema a fijar el grado de precisión con que queríamos obtener los resultados. En esta aportación fundamental de ordenador se deben citar al Gabinete de Cálculo y a la Oficina de Mecanización del Ministerio de Obras Públicas, que con sus ordenadores 1130 de IBM y 4004 de Siemens han resuelto todos los problemas que lleva consigo la simulación, y en particular deben citarse a título personal a los Ingenieros de Caminos Jesús Martínez, del Gabinete de Cálculo, y a Celso Nores, de la Oficina de Mecanización, que con la preparación de programas en paralelo han hecho posible la comparación de los resultados obtenidos y, a la vista de lo acorde de éstos, el poder garantizar la fiabilidad de los mismos. (Por supuesto que los programas y, sobre todo, la generación de números aleatorios han sido diferentes en los centros de cálculo antes citados).

Por último, hemos de decir que el estudio que se desarrolla a continuación viene a ser la contribución que ha hecho el que suscribe al grupo de trabajo que ha abordado —con posterioridad a la primera publicación de la Instrucción 1968— el problema más general de la resistencia característica del hormigón y el control de calidad del mismo.

(*) Se admiten comentarios sobre el presente artículo, que pueden remitirse a la Redacción de esta Revista hasta el 31 de agosto de 1971.

1. ANTECEDENTES

El Comité Europeo del Hormigón y, por consecuencia, la "Instrucción para el proyecto y la ejecución de obras de hormigón en masa o armado" del Ministerio de Obras Públicas aprobado por Decreto de la Presidencia del Gobierno 2987/1968, de 20 de septiembre, establecía al tratar de la resistencia característica del hormigón, que ésta se obtuviera a partir de una serie de ensayos de resistencia sobre probetas, al multiplicar por 2 la media aritmética de los $n/2$ resultados más bajos y restar después la media aritmética del conjunto de los n resultados.

Este procedimiento suponía unas hipótesis y adolecía de unos inconvenientes que pasamos a analizar:

La hipótesis que se admite es la distribución normal $N(\mu, \sigma)$ de los resultados x de la rotura a compresión de probetas, es decir, que si la media de la distribución es μ y la desviación típica es σ se verifica:

$$\text{Prob.}[a \leq \rho \leq b] = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_a^b e^{-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma}\right)^2} \cdot dx \quad (a \text{ y } b \text{ positivos})$$

siendo ρ la variable aleatoria, valor de la rotura por compresión de una probeta, tomada al azar.

En cuanto a los inconvenientes de este procedimiento, en primer lugar y desde un punto de vista teórico, estaba la propia definición, ya que la resistencia característica es un valor único (desconocido generalmente) de la población, definido por unas condiciones de probabilidad y, por tanto, no puede confundirse con el valor obtenido de una muestra de extensión n utilizando un cierto algoritmo o fórmula de aplicación.

En realidad, lo que se definía en la Instrucción era un procedimiento para obtener una estimación de la resistencia característica del hormigón.

Por tanto, conviene distinguir que resistencia característica de un hormigón —cuyo valor de rotura de probetas es una variable aleatoria— es aquel valor —fijo para una población— que tiene una probabilidad del 95 por 100 de ser superado al romper una probeta al azar. Si a esta resistencia característica la llamamos R'_{bk} se verificará que: $\text{Prob}[\rho > R'_{bk}] = 0,95$, siendo ρ la variable aleatoria, valor obtenido de la rotura de una probeta elegida al azar.

Según lo anterior y suponiendo que la distribución de ρ , valores de rotura por compresión de probetas, es una normal $N(\mu, \sigma)$ de media μ y desviación típica σ , el valor de la resistencia característica R'_{bk} es —según puede comprobarse en cualquier tabla en que viene tabulada la normal—:

$$R'_{bk} = \mu - 1,645 \cdot \sigma$$

El procedimiento que propugnaba la Instrucción para calcular una estimación de la resistencia característica era tomar como estimador la variable:

$$\alpha = 2 \cdot \left[\frac{x_1 + x_2 + \dots + x_{n/2}}{n/2} \right] - \left[\frac{x_1 + x_2 + \dots + x_{n/2} + \dots + x_n}{n} \right]$$

en donde n es la extensión de la muestra y se han ordenado previamente según

valores crecientes las x_i —valores de rotura—, de forma que:

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_{n/2} \leq \dots \leq x_{n-1} \leq x_n$$

Inmediatamente surge la pregunta de ¿por qué esta fórmula de α es un estimador aceptable (relativamente centrado) de la resistencia característica?

Nosotros hemos investigado ese ¿por qué? y hemos llegado a las siguientes conclusiones:

a) Para el caso de extensión de la muestra n muy grande, y, por tanto, desde un punto de vista teórico $n = \infty$, se demuestra fácilmente cuál es el valor a que tiende la media del estimador α .

En efecto, si la extensión de muestra n es infinita, y suponemos por hipótesis que los valores de rotura x_i se distribuyen según una ley normal $N(\mu, \sigma)$ de media μ y desviación típica σ , se tendrá una distribución de frecuencias relativas como la indicada en la figura 1.

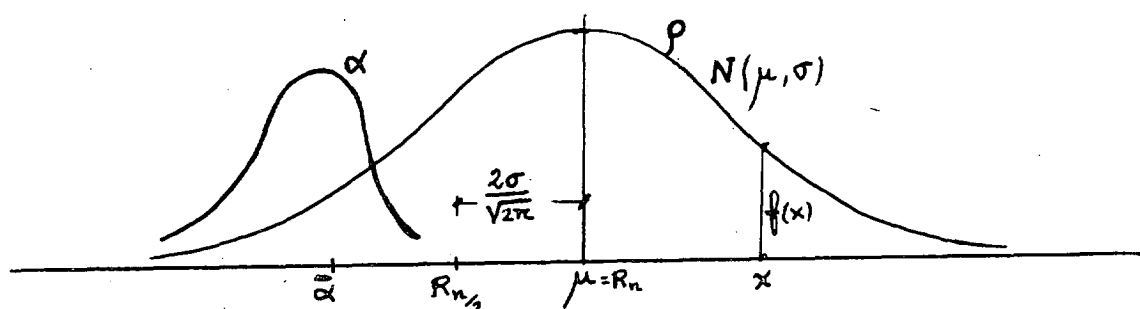


Figura 1.

Tenemos en este caso:

$$\alpha = 2 \cdot \left[\frac{x_1 + x_2 + \dots + x_{n/2}}{n/2} \right] - \left[\frac{x_1 + x_2 + \dots + x_{n/2} + \dots + x_n}{n} \right] = 2 \cdot R_{n/2} - R_n$$

siendo $R_{n/2}$ el valor medio de los $\frac{n}{2}$ valores más bajos y R_n la media de los n valores:

$$R_{n/2} = \frac{\int_{-\infty}^{\mu} x \cdot f(x) \cdot dx}{\int_{-\infty}^{\mu} f(x) \cdot dx} \quad \text{siendo} \quad f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

Haciendo:

$$\left. \begin{aligned} \frac{x-\mu}{\sigma} &= u \\ \frac{dx}{\sigma} &= du \end{aligned} \right\} R_{n/2} = \frac{\int_{-\infty}^0 (\mu + \sigma \cdot u) \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \cdot du}{\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \cdot du} = \mu + \frac{2 \cdot \sigma}{\sqrt{2\pi}} \cdot \int_{-\infty}^0 u \cdot e^{-\frac{u^2}{2}} \cdot du$$

$$R_{n/2} = \mu - \frac{2 \cdot \sigma}{\sqrt{2\pi}}$$

Dada la simetría de la curva $f(x)$ respecto al valor μ , la media $R_n = \mu$ y por tanto:

$$\bar{\alpha} = 2 \cdot R_{n/2} - R_n = 2\mu - \frac{4 \cdot \sigma}{\sqrt{2\pi}} - \mu = \mu - \frac{4 \cdot \sigma}{\sqrt{2\pi}}$$

$$\bar{\alpha} \approx \mu - 1,60 \cdot \sigma$$

Como puede verse, este valor $(\mu - 1,60 \cdot \sigma)$ está muy próximo al valor teórico de resistencia característica, y que era, según vimos, $R'_{bk} = \mu - 1,645 \cdot \sigma$.

En términos de probabilidad, si R'_{bk} tiene una probabilidad del 95 por 100 de ser superado, el valor obtenido $\bar{\alpha} = \mu - 1,60 \cdot \sigma$ tiene una probabilidad del 94,5 por 100 de ser superado.

b) Hasta ahora se ha supuesto $n = \infty$, pero el problema se complica cuando n es finito. En este caso, hemos utilizado el método de simulación para ver en cuánto se desviaba la media del estimador α respecto de la resistencia característica teórica $R'_{bk} = \mu - 1,645 \cdot \sigma$.

Para ello observamos que si en una distribución normal $N(\mu, \sigma)$ tipificamos la variable de la forma $z_i = \frac{x_i - \mu}{\sigma}$, la resistencia característica $R'_{bk} = \mu - 1,645 \cdot \sigma$ se convierte en $-1,645$, es decir, la resistencia característica de la normal $N(0,1)$.

Primeramente se ha generado una variable aleatoria normal $N(0,1)$. El procedimiento seguido, y según el teorema central del límite, ha sido generar números y_i aleatorios del intervalo $[0,1]$. Esta variable rectangular tiene de media 0,5, y de va-

rianza, $1/12$. La variable $z_j = \frac{\left(\sum_{i=1}^{27} y_i\right) - 13,5}{1,5}$ es muy aproximadamente normal $N(0,1)$. Por este procedimiento y para una extensión de muestra dada, se han generado con ordenador los valores y_i , y con ellos los z_j . Estos z_j se han ordenado de menor a mayor y se ha calculado la estimación:

$$\alpha^* = 2 \cdot \left[\frac{z_1 + z_2 + \dots + z_{n/2}}{n/2} \right] - \left[\frac{z_1 + z_2 + \dots + z_n}{n} \right]$$

Por comodidad de cálculo, la extensión de la muestra se la ha llamado $2 \cdot N$, y, por tanto:

$$\alpha^* = 2 \cdot \left[\frac{z_1 + z_2 + \dots + z_N}{N} \right] - \left[\frac{z_1 + z_2 + \dots + z_N + \dots + z_{2 \cdot N}}{2 \cdot N} \right]$$

con $z_1 \leq z_2 \leq \dots \leq z_N \leq \dots \leq z_{2N}$

Los resultados obtenidos para la media, varianza y desviación típica del estimador α^* han sido los que se indican en el cuadro 1.

CUADRO 1.

Tamaño de la muestra $2.N =$	Número de muestras de extensión $2.N$: 2.000			Número de muestras de extensión $2.N$: 1.500		
	ORDENADOR 4004 de Siemens			ORDENADOR 1130 de I.B.M.		
	Media α^*	Varianza α^*	Desv. Típ. α^*	Media α^*	Varianza α^*	Desv. Típ. α^*
6 (*)	—1,411	0,410	0,640	—1,393	0,400	0,632
8	—1,459	0,316	0,562	—1,471	0,310	0,557
10	—1,478	0,231	0,481	—1,484	0,248	0,498
12	—1,498	0,206	0,454	—1,507	0,202	0,450
14	—1,522	0,172	0,414	—1,516	0,160	0,400
16	—1,512	0,153	0,391	—1,522	0,151	0,388
18	—1,532	0,136	0,368	—1,542	0,130	0,361
20	—1,553	0,122	0,349	—1,528	0,114	0,338
22	—1,544	0,110	0,332	—1,547	0,107	0,328
24	—1,548	0,096	0,310	—1,550	0,100	0,317
26	—1,561	0,091	0,302	—1,543	0,089	0,299
28	—1,547	0,085	0,292	—1,560	0,084	0,290
30	—1,558	0,085	0,292	—1,560	0,078	0,280
∞	—1,596	0	0	—1,596	0	0

(*) Número de muestras: 50000.

Las diferencias observadas por ambos procedimientos con ordenador provienen, por un lado, por los diferentes números de muestras utilizadas — 2000 y 1500— y, por otro lado, por haber sido diferentes los modos de generar números aleatorios y_i del intervalo $[0,1]$ en cada uno de los ordenadores. De todas maneras, esto nos ha servido para que uno de ellos sirva de comprobación del otro y viceversa, y podamos, por tanto, afirmar como buenos cualesquiera de los resultados obtenidos con errores de algunas centésimas y que hubiesen sido aminoradas en el caso de haber utilizado un número de muestras superior.

Como puede observarse, el valor medio de α^* para extensión $2.N$ de la muestra pequeña difiere significativamente del valor a estimar que era —1,645.

En términos de probabilidad, si —1,645 era superado en el 95 por 100 de las veces, en el caso de la $2.N = 6$ con media $\bar{\alpha}^* = -1,41$, solamente es superado en el 92 por 100 de las veces y, por tanto, podría ser aceptado el estimador propuesto como un buen estimador de la resistencia característica.

De lo anterior se deduce que el inconveniente que este estimador α pueda presentar no proviene fundamentalmente de consideraciones teóricas, sino más bien por consideraciones de tipo práctico.

Veámoslo con un ejemplo:

Supongamos un hormigón cuyas probetas han roto por compresión a: 165, 172, 175, 180, 182 y 188 Kg/cm².

Según la fórmula propuesta, una estimación de la resistencia característica sería:

$$\hat{\alpha} = 2 \cdot \left[\frac{165 + 172 + 175}{3} \right] - \left[\frac{165 + 172 + 175 + 180 + 182 + 188}{6} \right]$$

$$\hat{\alpha} = 341,333 - 177,000 = 164,333 \text{ Kg/cm}^2$$

Supongamos que el hormigón anterior se ha mezclado con otro de calidad superior (hecho que puede ocurrir), o las últimas probetas han tenido errores de medición de laboratorio y que las probetas obtenidas han roto a: 165, 172, 175, 180, 182, 188, 225, 232, 238 y 242 Kg/cm².

Según la fórmula:

$$\hat{\alpha}_1 = 2 \cdot \left[\frac{165 + 172 + 175 + 180 + 182}{5} \right] - \left[\frac{165 + 172 + 175 + 180 + 182 + 188 + 225 + 232 + 238 + 242}{10} \right]$$

$$\hat{\alpha}_1 = 349,600 - 199,900 = 149,700 \text{ Kg/cm}^2$$

Puede observarse, por tanto, una clara contradicción con el fin buscado, ya que por el hecho de a un hormigón considerado como aceptable (resistencia característica 164,3 Kg/cm²) añadirle un hormigón de mejor calidad se ha conseguido rebajar extraordinariamente su resistencia característica a 149,7 Kg/cm², con el consiguiente riesgo de rechazar la partida conjunta cuando no se rechazaría cada una de las partes.

En realidad, lo que ha fallado no ha sido el procedimiento, sino la hipótesis de normalidad del hormigón, pues en este ejemplo claramente se trataba de una mezcla de dos hormigones diferentes, con dos leyes normales diferentes también. Como por otro lado el ejemplo citado puede ocurrir con alguna frecuencia en las obras, ha de tratarse en lo posible que el hecho de aumentar la calidad del hormigón añadiendo hormigones más resistentes *no penalice*, disminuyendo, la resistencia característica, con el consiguiente peligro de rechazar todo el conjunto.

Para ello hay que tratar de encontrar otro estimador de la resistencia característica que en lo posible elimine los dos inconvenientes antes citados: *desviación significativa del valor — 1,645 para muestras pequeñas y, sobre todo, que las probetas de mayor resistencia no penalicen, disminuyendo, la resistencia característica.*

2. ESTIMADOR PROPUESTO R

Después de diferentes ensayos de estimadores, hemos llegado a un estimador que en nuestra opinión elimina de manera muy importante los inconvenientes antes apuntados. Este estimador que llamaremos *R* se define para una muestra de extensión $2 \cdot N$:

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_{N-1} \leq x_N \leq x_{N+1} \leq \dots \leq x_{2N}$$

$$R = 2 \cdot \left[\frac{x_1 + x_2 + \dots + x_{N-1}}{N-1} \right] - x_N = 2 \cdot R_{N-1} - x_N$$

Es decir, *multiplicando por 2 la media aritmética de los (N — 1) —mitad menos uno—, valores más bajos, y restando después el valor x_N de rotura de la probeta de lugar mitad.*

Este estimador *R* al no considerar en la muestra de extensión $2 \cdot N$ los valores correspondientes a la mitad más alta ($x_{N+1}, x_{N+2}, \dots, x_{2N}$) elimina o al menos aminora en gran parte, el inconveniente antes citado de penalizar, disminuyendo, la resistencia característica cuando se mezclan hormigones con diferentes leyes de normalidad, o cuando los valores más altos de rotura se deben fundamentalmente a errores de medición del laboratorio.

Así, en el ejemplo que se citaba anteriormente con roturas de probetas de 165, 172, 175, 180, 182 y 188 Kg/cm² y que aplicando el estimador α se obtenía como estimación de la resistencia característica 164,333 Kg/cm², con el estimador R se obtendría:

$$\hat{R} = 2 \cdot \left[\frac{165 + 172}{2} \right] - 175 = 162 \text{ Kg/cm}^2$$

Al añadir a la serie anterior las roturas de probetas de 225, 232, 238 y 242 kilogramos por centímetro cuadrado vimos cómo la estimación de la resistencia característica bajaba a $\hat{\alpha}_1 = 149,7 \text{ Kg/cm}^2$, lo que encontrábamos contradictorio. Con el estimador R propuesto, la estimación de la resistencia característica sería ahora:

$$\hat{R}_1 = 2 \cdot \left[\frac{165 + 172 + 175 + 180 + 182}{5} \right] - 188 = 349,6 - 188 = 161,6 \text{ Kg/cm}^2$$

Aparte de estos casos —que es de desear que no proliferen— y ya en el caso de que la muestra de extensión $2 \cdot N$ provenga ciertamente de un hormigón con distribución normal $N(\mu, \sigma)$ vamos a comprobar por *simulación* cómo se distribuye el estimador R antes definido, calculando para diferentes extensiones de la muestra el valor medio, la varianza y la desviación típica.

Primeramente comparemos el comportamiento y las propiedades de las funciones de densidad y distribución del estimador $R = 2 \cdot R_{N-1} - x_N$ proveniente de una población normal $N(\mu, \sigma)$ con las de la variable $R^* = 2 \cdot R_{N-1}^* - z_N$ proveniente de una población tipificada z normal $N(0,1)$:

$$\left(z_i = \frac{x_i - \mu}{\sigma} \right)$$

En primer lugar si:

$$x_1 \leq x_2 \leq \dots \leq x_N \leq x_{N+1} \leq \dots \leq x_{2N}$$

se verificará:

$$\frac{x_1 - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma} \leq \dots \leq \frac{x_N - \mu}{\sigma} \leq \frac{x_{N+1} - \mu}{\sigma} \leq \dots \leq \frac{x_{2N} - \mu}{\sigma}$$

es decir:

$$z_1 \leq z_2 \leq \dots \leq z_N \leq z_{N+1} \leq \dots \leq z_{2N}$$

En segundo lugar al tipificar la variable x de la forma $z = \frac{x - \mu}{\sigma}$ la estimación $\hat{R} = 2 \cdot R_{N-1} - x_N$ se convierte en:

$$\begin{aligned} \hat{R} &= 2 \cdot \left[\frac{\sigma \cdot z_1 + \mu + \sigma \cdot z_2 + \mu + \dots + \sigma \cdot z_{N-1} + \mu}{N-1} \right] - (\sigma \cdot z_N + \mu) = \\ &= \sigma \cdot (2 \cdot R_{N-1}^* - z_N) + \mu = \sigma \cdot \hat{R}^* + \mu \end{aligned}$$

siendo:

$$\begin{aligned} R_{N-1}^* &= \frac{z_1 + z_2 + \dots + z_{N-1}}{N-1} \\ \hat{R}^* &= 2 R_{N-1}^* - z_N = \frac{\hat{R} - \mu}{\sigma} \end{aligned}$$

En tercer lugar:

$$\text{Prob}(R \leq x) = F(x) = \text{Prob}\left(R^* \leq \frac{x - \mu}{\sigma}\right) = \text{Prob}(R^* \leq z) = F^*(z)$$

En cuarto lugar:

$$\text{Prob}(x_i \leq R \leq x_j) = \text{Prob}\left(\frac{x_i - \mu}{\sigma} \leq R^* \leq \frac{x_j - \mu}{\sigma}\right) = \text{Prob}(z_i \leq R^* \leq z_j)$$

Generados con ordenador valores y_i de la distribución rectangular [0,1] y mediante la fórmula:

$$z_j = \frac{\left(\sum_{i=1}^{27} y_i\right) - 13,5}{1,5}$$

que obtiene una distribución muy, aproximadamente, normal $N(0,1)$; se ordenan z_j de menor a mayor:

$$z_1 \leq z_2 \leq \dots \leq z_{N-1} \leq z_N \leq \dots \leq z_{2N}$$

y se calcula:

$$R^* = 2 \cdot \left[\frac{z_1 + z_2 + \dots + z_{N-1}}{N-1} \right] - z_N = 2 \cdot R^*_{N-1} - z_N$$

Los valores obtenidos de la media, varianza y desviación típica de R^* repitiendo este ensayo 3000 y 2000 veces, se expresan en el cuadro 2.

CUADRO 2.

Tamaño de la muestra 2 . N =	Número de muestras de extensión 2 . N: 3.000			Número de muestras de extensión 2 . N: 2.000		
	ORDENADOR 4004 de Siemens			ORDENADOR 1130 de I.B.M.		
	Media R^*	Varianza R^*	Desv. Típ. R^*	Media R^*	Varianza R^*	Desv. Típ. R^*
6	— 1,691	0,700	0,836	— 1,700	0,714	0,845
8	— 1,703	0,486	0,697	—	—	—
10	— 1,656	0,352	0,594	—	—	—
12	— 1,660	0,290	0,539	— 1,661	0,284	0,533
14	— 1,632	0,243	0,493	—	—	—
16	— 1,656	0,205	0,452	—	—	—
18	— 1,641	0,191	0,437	— 1,647	0,184	0,429
20	— 1,631	0,153	0,391	—	—	—
22	— 1,630	0,143	0,378	—	—	—
24	— 1,627	0,133	0,365	—	—	—
26	— 1,617	0,124	0,353	—	—	—
28	— 1,637	0,121	0,348	—	—	—
30	— 1,623	0,097	0,312	—	—	—

Puede observarse cómo las diferencias entre los valores medios obtenidos y el teórico —1,645 de la resistencia característica, son del orden de unas centésimas y que en términos de probabilidad tienen poca trascendencia. Ya sabemos que —1,645 corresponde a la resistencia característica y, por tanto, con una probabilidad del 95 por 100 de ser superada. Los valores —1,69 y —1,62 se corresponden con probabilidades de ser superados del 95,4 y 94,7 por 100, respectivamente, es decir, *prácticamente* podemos afirmar que el estimador R propuesto está centrado en la resistencia característica, que era otro de los objetivos que debía cumplir el estimador.

3. DISTRIBUCION DEL ESTIMADOR R .

Considerando el estimador $R = 2 \cdot R_{N-1} - x_N$ como un buen estimador (centrado) de la resistencia característica, con extensión de la muestra $2 \cdot N$:

$$(x_1 \leq x_2 \leq x_3 \leq \dots \leq x_{N-1} \leq x_N \leq x_{N+1} \leq \dots \leq x_{2N})$$

y siendo:

$$R_{N-1} = \frac{x_1 + x_2 + \dots + x_{N-1}}{N-1}$$

hemos extendido la simulación por medio de ordenador hasta 15 000 muestras de extensión $2 \cdot N$, habiéndose obtenido los resultados que se indican en el cuadro 3, para la media, varianza y desviación típica del estimador:

$$R^* = 2 \cdot \left[\frac{z_1 + z_2 + \dots + z_{N-1}}{N-1} \right] - z_N$$

con:

$$z_1 \leq z_2 \leq \dots \leq z_{N-1} \leq z_N \leq \dots \leq z_{2N}$$

CUADRO 3.

Extensión de la muestra $2 \cdot N =$	15.000 muestras de extensión $2 \cdot N$					
	ORDENADOR 4004 de Siemens			ORDENADOR 1130 de I.B.M.		
	Media R^*	Varianza R^*	Desv. Típ. R^*	Media R^*	Varianza R^*	Desv. Típ. R^*
6	—1,711	0,699	0,836	—1,710	0,695	0,834
12	—1,657	0,291	0,540	—1,652	0,290	0,539
18	—1,637	0,183	0,428	—1,640	0,184	0,429

Se observa cómo al haber aumentado el número de muestras a 15 000, las diferencias entre los resultados de ambos ordenadores se han reducido a milésimas, lo cual es una garantía de la verosimilitud del procedimiento y de los resultados obtenidos.

Finalmente, el problema que se ha resuelto por simulación ha sido el encontrar la distribución del estimador $R^* = 2 \cdot R_{N-1}^* - z_N$. En primer lugar se contrasta la hipótesis de que este estimador R^* se distribuye según una ley normal. Para ello se ha dividido el intervalo $(-\infty, +\infty)$ en 40 intervalos (Ordenador Siemens 4004), de forma que, si efectivamente este estimador R^* se distribuyese, según una ley normal $N(\mu^*, \sigma^*)$, el área delimitada por cada intervalo, en la función de densidad (frecuencia relativa), sería la misma e igual a 0,025.

CUADRO 4.

(Estimador $R^* = 2 \cdot R_{N-1}^* - z_N$)

$N(0,1)$ t	Media $R^* = -1,7110$ Desv. Típ. $R^* = 0,8364$			Media $R^* = -1,6575$ Desv. Típ. $R^* = 0,5400$			Media $R^* = -1,6368$ Desv. Típ. $R^* = 0,4283$		
	$2N = 6$			$2N = 12$			$2N = 18$		
	z	Frecuenc. absoluta f_i	$F_n^*(z)$	z	Frecuenc. absoluta f_i	$F_n^*(z)$	z	Frecuenc. absoluta f_i	$F_n^*(z)$
— 1,9600	— 3,3503	592	0,0395	— 2,7160	458	0,0305	— 2,4762	467	0,0311
— 1,6449	— 3,0868	336	0,0619	— 2,5458	370	0,0552	— 2,3412	358	0,0550
— 1,4395	— 2,9150	342	0,0847	— 2,4349	351	0,0786	— 2,2533	387	0,0808
— 1,2816	— 2,7830	307	0,1051	— 2,3496	326	0,1003	— 2,1856	354	0,1044
— 1,1503	— 2,6732	295	0,1248	— 2,2787	369	0,1249	— 2,1294	331	0,1265
— 1,0364	— 2,5779	347	0,1479	— 2,2172	332	0,1471	— 2,0806	337	0,1489
— 0,9346	— 2,4927	304	0,1682	— 2,1622	343	0,1699	— 2,0370	343	0,1718
— 0,8416	— 2,4150	305	0,1885	— 2,1120	348	0,1931	— 1,9972	347	0,1949
— 0,7554	— 2,3429	323	0,2101	— 2,0655	361	0,2172	— 1,9603	373	0,2198
— 0,6745	— 2,2752	334	0,2323	— 2,0218	345	0,2402	— 1,9256	356	0,2435
— 0,5978	— 2,2110	341	0,2551	— 1,9804	335	0,2625	— 1,8928	332	0,2657
— 0,5244	— 2,1496	303	0,2753	— 1,9407	377	0,2877	— 1,8614	320	0,2870
— 0,4538	— 2,0906	334	0,2975	— 1,9026	353	0,3112	— 1,8311	356	0,3107
— 0,3853	— 2,0333	350	0,3209	— 1,8656	377	0,3363	— 1,8018	341	0,3335
— 0,3186	— 1,9775	372	0,3457	— 1,8296	371	0,3611	— 1,7732	344	0,3564
— 0,2533	— 1,9229	341	0,3684	— 1,7943	359	0,3850	— 1,7453	409	0,3837
— 0,1891	— 1,8692	363	0,3926	— 1,7596	366	0,4094	— 1,7178	409	0,4109
— 0,1257	— 1,8162	378	0,4178	— 1,7254	341	0,4321	— 1,6906	366	0,4353
— 0,0627	— 1,7635	349	0,4411	— 1,6914	380	0,4575	— 1,6636	356	0,4591
± 0,0000	— 1,7110	380	0,4664	— 1,6575	368	0,4820	— 1,6368	385	0,4847
0,0627	— 1,6586	361	0,4905	— 1,6236	378	0,5072	— 1,6099	395	0,5111
0,1257	— 1,6059	388	0,5163	— 1,5896	360	0,5312	— 1,5829	373	0,5359
0,1891	— 1,5529	404	0,5433	— 1,5554	371	0,5559	— 1,5558	386	0,5617
0,2533	— 1,4992	409	0,5705	— 1,5207	435	0,5849	— 1,5283	386	0,5874
0,3186	— 1,4445	410	0,5979	— 1,4854	427	0,6134	— 1,5003	399	0,6140
0,3853	— 1,3887	385	0,6235	— 1,4494	380	0,6387	— 1,4718	428	0,6425
0,4538	— 1,3314	402	0,6503	— 1,4124	375	0,6637	— 1,4424	402	0,6693
0,5244	— 1,2724	432	0,6791	— 1,3743	377	0,6889	— 1,4122	400	0,6960
0,5978	— 1,2110	468	0,7103	— 1,3347	446	0,7186	— 1,3807	386	0,7217
0,6745	— 1,1468	443	0,7399	— 1,2932	366	0,7430	— 1,3479	378	0,7469
0,7554	— 1,0792	459	0,7705	— 1,2496	430	0,7717	— 1,3133	357	0,7707
0,8416	— 1,0071	465	0,8015	— 1,2030	386	0,7974	— 1,2763	401	0,7975
0,9346	— 0,9293	410	0,8288	— 1,1528	408	0,8246	— 1,2365	401	0,8242
1,0364	— 0,8441	443	0,8583	— 1,0978	394	0,8509	— 1,1929	411	0,8516
1,1503	— 0,7489	423	0,8865	— 1,0363	415	0,8785	— 1,1441	391	0,8777
1,2816	— 0,6391	414	0,9141	— 0,9654	403	0,9054	— 1,0879	404	0,9046
1,4395	— 0,5070	453	0,9443	— 0,8801	443	0,9349	— 1,0203	434	0,9335
1,6449	— 0,3352	346	0,9674	— 0,7692	359	0,9589	— 0,9323	347	0,9567
1,9600	— 0,0716	334	0,9897	— 0,5990	362	0,9830	— 0,7974	362	0,9808
∞	∞	155	1,0000	∞	255	1,0000	∞	288	1,0000
	$z = 0,8364 \cdot t - 1,7110$			$z = 0,54 \cdot t - 1,6575$			$z = 0,4283 \cdot t - 1,6368$		

En el cuadro 4 se especifican los extremos t de estos intervalos que cumplen la mencionada condición en la $N(0,1)$, así como las correspondientes z para los casos de extensión de la muestra igual a 6, 12 y 18; la frecuencia absoluta f_i obtenida en cada intervalo y la frecuencia relativa acumulada $F^*_n(z)$.

El test de normalidad de la χ^2 calcula:

$$\chi^2 = \left(\sum_{i=1}^r \frac{f_i^2}{n \cdot p_i} \right) - n$$

En nuestro caso:

$$n = 15\,000$$

$$p_i = 0,025$$

$$r = 40$$

Con lo cual:

$$\chi^2 = \frac{1}{375} \cdot \left(\sum_{i=1}^{40} f_i^2 \right) - 15\,000$$

El número de grados de libertad, por haberse estimado los parámetros, media y desviación típica, son $\nu = r - 3 = 37$.

Este valor de χ^2 para los casos $2N = 6, 12$ y 18 resulta ser:

$2N =$	6	12	18
$\chi^2 =$	504,72	150,49	124,49

Por ser los grados de libertad $\nu = 37 > 30$ puede ponerse que:

$$\text{Prob}(\chi^2 > \chi^2_p) \cong 1 - \Phi \left(\frac{\chi^2_p - \nu}{\sqrt{2\nu}} \right)$$

Para un nivel de significación de 0,001 se tiene:

$$\text{Prob}(\chi^2 > \chi^2_p) = 0,001 \cong 1 - \Phi \left(\frac{\chi^2_p - \nu}{\sqrt{2\nu}} \right)$$

siendo $\Phi(x)$ la función de distribución de la $N(0,1)$.

En nuestro caso:

$$\Phi \left(\frac{\chi^2_p - 37}{\sqrt{74}} \right) \cong 0,999$$

con lo cual:

$$\frac{\chi^2_p - 37}{8,60233} \cong 3,09$$

$$\chi^2_p \cong 63,58$$

Por ser:

$$\begin{aligned} 504,72 &> 63,58 \\ 150,49 &> 63,58 \\ 124,49 &> 63,58 \end{aligned}$$

hay que *rechazar la hipótesis de normalidad del estimador R^* muy significativamente* (nivel de significación del 0,001).

La función de distribución $F^*(z)$ del estimador R^* viene dada muy, aproximadamente, por la columna $F^*_n(z)$ del cuadro 4 para los casos $2 \cdot N = 6, 12$ y 18 y su representación se adjunta en figura aparte.

En el Ordenador 1130 de IBM, y siguiendo el paralelismo que venimos realizando, también se ha simulado la rotura de 15000 muestras de extensión $2 \cdot N = 6, 12$ y 18 —procedentes de una distribución normal, $N(0,1)$ —, analizándose la distribución del estimador:

$$R^* = 2 \cdot R^*_{N-1} - z_N = 2 \cdot \left(\frac{z_1 + z_2 + \dots + z_{N-1}}{N-1} \right) - z_N$$

Para ello se ha distribuido el intervalo $(-\infty, +\infty)$ en 22 intervalos.

En los cuadros 5 y 6 se detallan los extremos de los intervalos y la distribución por frecuencias del estimador, así como los valores tipificados de esos extremos.

Contrastamos, al igual que en el caso anterior, la hipótesis de normalidad del estimador R^* con el *test de Kolmogorof* para un nivel de significación del 0,001. Para ello hallamos los valores que toma la expresión:

$$Z = \sqrt{n} \cdot D_n = \sqrt{n} \cdot \sup |F^*(z) - F^*_n(z)| = \sqrt{15000} \cdot \sup |F^*(z) - F^*_n(z)|$$

en los casos de extensión de muestra $2N = 6, 12$ y 18 , siendo $F^*(z)$ y $F^*_n(z)$ las funciones de distribución de la variable normal y la variable estadística de las 15000 muestras.

En el cuadro 7 se especifican para estos casos los valores de $F^*(z)$ y $F^*_n(z)$.

Los valores que se obtienen para:

$$Z = \sqrt{15000} \cdot \sup |F^*(z) - F^*_n(z)| = 122,47 \cdot \sup |F^*(z) - F^*_n(z)|$$

son los siguientes:

$$\begin{aligned} \text{Para } 2N = 6, \quad Z &\geq 122,47 \times 0,0368 = 4,507 \\ \text{Para } 2N = 12, \quad Z &\geq 122,47 \times 0,0229 = 2,805 \\ \text{Para } 2N = 18, \quad Z &\geq 122,47 \times 0,0162 = 1,984 \end{aligned}$$

Para un nivel de significación de 0,001 la Z_1 correspondiente a la distribución de Kolmogorof vale $Z_1 = 1,95$.

Por ser:

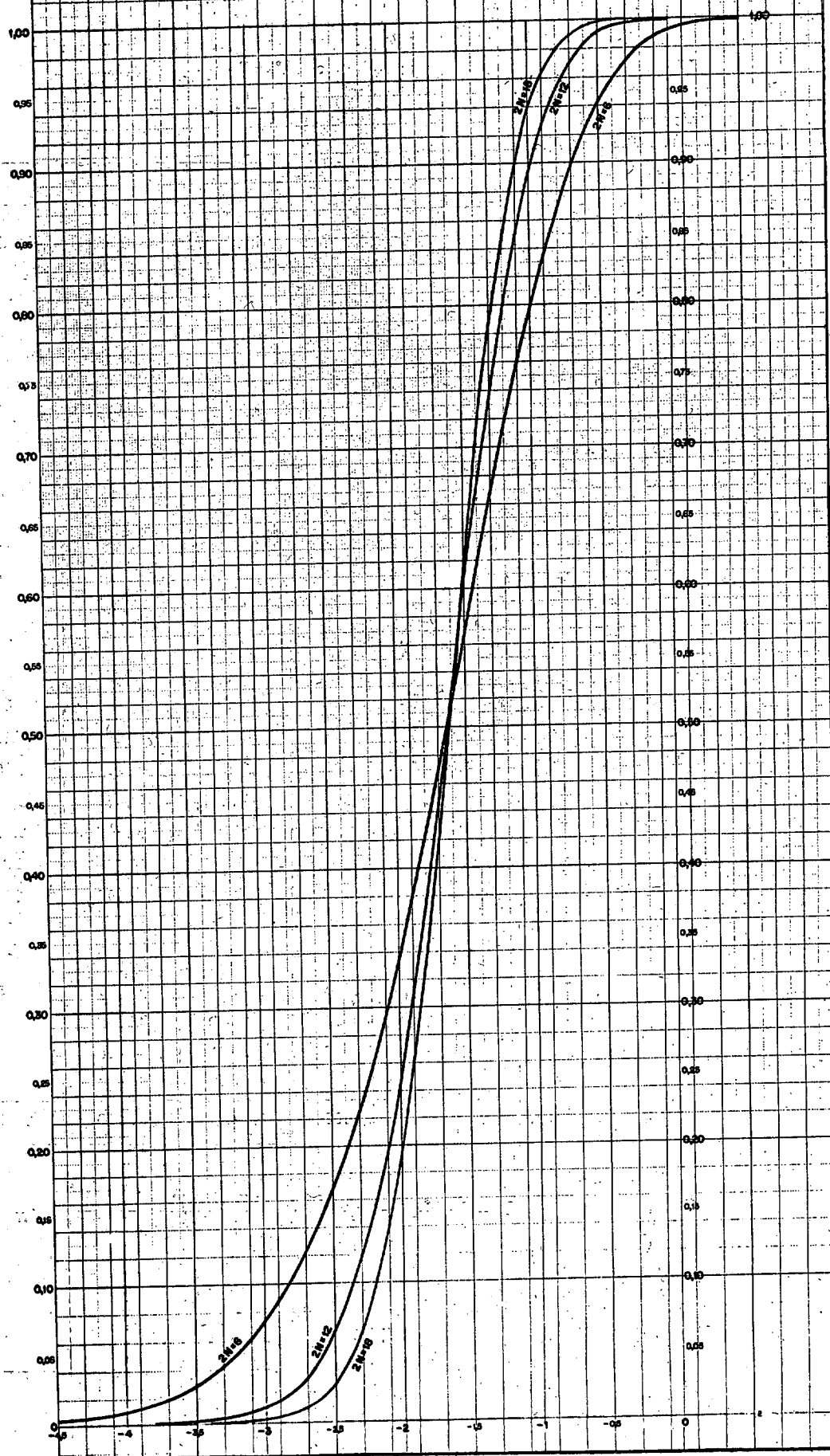
$$\begin{aligned} 4,507 &> 1,95 \\ 2,805 &> 1,95 \\ 1,984 &> 1,95 \end{aligned}$$

hay que *rechazar la hipótesis de normalidad del estimador R^* muy significativamente* (nivel de significación del 0,001).

FUNCION DE DISTRIBUCION $F^*(t)$ DEL ESTIMADOR $R^* = 2A \frac{t^2}{N-1} - t^2$

(EXTENSION DE LA MUESTRA = 2N)

$F^*(t) = \text{PROB}(F^* \leq t)$



CUADRO 5.

Estimador $R^* = 2 \cdot R_{N-1}^* - z_N$

Intervalos	TAMAÑO DE LA MUESTRA 2 N					
	2 . N = 6		2 . N = 12		2 . N = 18	
	Media $R^* = -1,710$ Desv. típ. $R^* = 0,834$		Media $R^* = -1,652$ Desv. típ. $R^* = 0,539$		Media $R^* = -1,640$ Desv. típ. $R^* = 0,429$	
	Frecuencia		Frecuencia		Frecuencia	
	Absoluta	Relativa	Absoluta	Relativa	Absoluta	Relativa
$-\infty; -4,0$	132	0,0088	2	0,0001	0	0,0000
$-4,0; -3,8$	89	0,0059	2	0,0001	0	0,0000
$-3,8; -3,6$	120	0,0080	8	0,0005	0	0,0000
$-3,6; -3,4$	189	0,0126	28	0,0019	2	0,0001
$-3,4; -3,2$	226	0,0151	45	0,0030	13	0,0009
$-3,2; -3,0$	336	0,0224	88	0,0059	33	0,0022
$-3,0; -2,8$	466	0,0311	183	0,0122	57	0,0038
$-2,8; -2,6$	607	0,0405	356	0,0237	173	0,0115
$-2,6; -2,4$	773	0,0515	630	0,0420	386	0,0257
$-2,4; -2,2$	954	0,0636	994	0,0663	808	0,0539
$-2,2; -2,0$	1127	0,0751	1365	0,0910	1479	0,0986
$-2,0; -1,8$	1309	0,0873	1881	0,1254	2168	0,1445
$-1,8; -1,6$	1408	0,0939	2148	0,1432	2694	0,1796
$-1,6; -1,4$	1566	0,1044	2287	0,1525	2715	0,1810
$-1,4; -1,2$	1462	0,0975	1919	0,1279	2226	0,1484
$-1,2; -1,0$	1256	0,0837	1493	0,0995	1342	0,0895
$-1,0; -0,8$	1010	0,0673	877	0,0585	627	0,0418
$-0,8; -0,6$	813	0,0542	458	0,0305	220	0,0147
$-0,6; -0,4$	559	0,0373	178	0,0119	50	0,0033
$-0,4; -0,2$	327	0,0218	45	0,0030	5	0,0003
$-0,2; 0,0$	173	0,0115	7	0,0005	1	0,0001
$0; +\infty$	98	0,0065	6	0,0004	1	0,0001
Suma	15000	1,0000	15000	1,0000	15000	1,0000

$$R^* = 2 \cdot R_{N-1}^* - z_N = 2 \cdot \left(\frac{z_1 + z_2 + \dots + z_{N-1}}{N-1} \right) - z_N$$

CUADRO 6.

Estimador $R^* = 2 \cdot R_{N-1}^* - z_N$

Extremos de intervalo tipificados:

z	2 N = 6	2 N = 12	2 N = 18
	$\mu^* = -1,710$ $\sigma^* = 0,834$	$\mu^* = -1,652$ $\sigma^* = 0,539$	$\mu^* = -1,640$ $\sigma^* = 0,429$
	$\frac{z - \mu^*}{\sigma^*}$	$\frac{z - \mu^*}{\sigma^*}$	$\frac{z - \mu^*}{\sigma^*}$
-4,0	-2,746	-4,356	-5,501
-3,8	-2,506	-3,985	-5,035
-3,6	-2,266	-3,614	-4,569
-3,4	-2,026	-3,243	-4,103
-3,2	-1,787	-2,872	-3,636
-3,0	-1,547	-2,501	-3,170
-2,8	-1,307	-2,130	-2,704
-2,6	-1,067	-1,759	-2,238
-2,4	-0,827	-1,388	-1,772
-2,2	-0,587	-1,017	-1,305
-2,0	-0,348	-0,646	-0,839
-1,8	-0,108	-0,275	-0,373
-1,6	+ 0,132	+ 0,096	+ 0,093
-1,4	0,372	0,467	0,559
-1,2	0,611	0,839	1,026
-1,0	0,851	1,210	1,492
-0,8	1,091	1,581	1,958
-0,6	1,331	1,952	2,424
-0,4	1,571	2,323	2,890
-0,2	1,811	2,694	3,357
-0	2,050	3,065	3,823

μ^* = media de R^* .

σ^* = desviación típica de R^* .

CUADRO 7.

Estimador $R^* = 2 \cdot R_{N-1}^* - z_N$

z	2 N = 6		2 N = 12		2 N = 18	
	$\mu^* = -1,710$ $\sigma^* = 0,384$		$\mu^* = -1,652$ $\sigma^* = 0,539$		$\mu^* = -1,640$ $\sigma^* = 0,429$	
	$F_n^*(z)$	$F^*(z)$	$F_n^*(z)$	$F^*(z)$	$F_n^*(z)$	$F^*(z)$
— 4,0	0,0088	0,0030	0,0001	0,0000	0,0000	0,0000
— 3,8	0,0147	0,0061	0,0002	0,0000	0,0000	0,0000
— 3,6	0,0227	0,0117	0,0007	0,0001	0,0000	0,0000
— 3,4	0,0353	0,0214	0,0026	0,0006	0,0001	0,0000
— 3,2	0,0504	0,0370	0,0056	0,0020	0,0010	0,0001
— 3,0	0,0728	0,0610	0,0115	0,0062	0,0032	0,0008
— 2,8	0,1039	0,0956	0,0237	0,0166	0,0070	0,0034
— 2,6	0,1443	0,1430	0,0474	0,0393	0,0185	0,0126
— 2,4	0,1959	0,2041	0,0894	0,0826	0,0443	0,0382
— 2,2	0,2595	0,2786	0,1557	0,1546	0,0981	0,0959
— 2,0	0,3346	0,3639	0,2467	0,2591	0,1967	0,2007
— 1,8	0,4219	0,4570	0,3721	0,3917	0,3413	0,3546
— 1,6	0,5157	0,5525	0,5153	0,5382	0,5209	0,5371
— 1,4	0,6201	0,6450	0,6678	0,6797	0,7019	0,7119
— 1,2	0,7176	0,7294	0,7957	0,7992	0,8503	0,8475
— 1,0	0,8013	0,8026	0,8952	0,8869	0,9397	0,9322
— 0,8	0,8687	0,8623	0,9537	0,9430	0,9815	0,9749
— 0,6	0,9229	0,9084	0,9842	0,9745	0,9962	0,9923
— 0,4	0,9601	0,9419	0,9961	0,9899	0,9995	0,9981
— 0,2	0,9819	0,9650	0,9991	0,9965	0,9999	0,9996
0	0,9935	0,9798	0,9996	0,9989	0,9999	0,9999

μ^* = media de R^* .

σ^* = desviación típica de R^* .