

CURVAS DE FRECUENCIA COMPUESTAS

POR PEDRO M. GONZÁLEZ QUIJANO
INGENIERO DE CAMINOS

II

En un artículo anterior, publicado bajo el mismo título que encabeza el presente, hemos explicado el método general que podría seguirse para el cálculo y construcción de las curvas de frecuencia compuestas, con especial aplicación al estudio de las series pluviométricas, aunque pueda ser aplicado a otras series de fenómenos análogos. Hoy hemos de estudiar el valor práctico del método y las rectificaciones que en él hubiera que introducir, cuando no se realizaran con suficiente aproximación las hipótesis fundamentales.

Una primera condición para que el método pueda ser de alguna utilidad, es que los valores medios de una y otra serie parcial sean comparables entre sí. Si se recurre a combinar las series parciales, es porque la serie total es muy poco numerosa y quedan entre sus elementos huecos demasiado grandes e irregulares, que dejan indeciso el trazado de la curva, y que hasta podrían sugerir la existencia de máximos o mínimos, debidos tan solo a una agrupación casual de las observaciones, y que llegarían a desaparecer cuando el número de éstas aumentara; pero si una de las series fuera de elementos tan pequeños que quedarán por debajo de las diferencias existentes entre los elementos consecutivos de la otra serie, el objeto no se habría logrado, porque los claros más importantes subsistirían los mismos, y las aglomeraciones no habrían hecho más que acentuarse.

Convendrá, pues, que las dos series en que la total se divide tengan medias aproximadamente iguales, porque es así como probablemente se logrará el máximo efecto. Y ello será, por lo general, fácil de conseguir cuando se trate de series pluviométricas en las que se conozcan los valores mensuales, pues no es forzoso que los meses de cada grupo sean consecutivos, ni tampoco que haya el mismo número de meses en cada grupo, bastando solo que para la formación de cada serie parcial se utilicen siempre los mismos meses en los distintos años.

Pero es también condición fundamental para la aplicación del método que las dos series en que se

divide la principal sean independientes. Si la lluvia del segundo semestre, por ejemplo, fuera siempre igual a la del primero o guardara con ella una relación fija, es claro que todas las combinaciones que no cumplieran con esta condición deberían ser desechadas y que, por consiguiente, la consideración de esa segunda serie sería inútil, pues nada añadiría a la primera.

Será, pues, preciso que de algún modo se pueda comprobar este extremo, y si hubiéramos de seguir una práctica bastante generalizada, cabría aplicar a este caso el llamado coeficiente de correlación de Galton, definido en nuestro caso por la fórmula

$$r = \frac{\sum (a_i - a) (b_i - b)}{\sqrt{\sum (a_i - a)^2} \cdot \sqrt{\sum (b_i - b)^2}}$$

en la que a_i y b_i representan los valores de las series parciales correspondientes al año i , y a y b , las medias correspondientes a cada serie.

Este coeficiente es igual a ± 1 cuando los b_i y los a_i están ligados por una relación lineal, e igual a cero cuando una de las series tuviera valores iguales, que serían en tal caso independientes de los de la otra serie. Entre estos valores extremos, el coeficiente puede tomar todos los valores posibles, y muchos autores consideran ese valor como una especie de medida de la dependencia mutua de las dos series.

Tal interpretación tiene, sin embargo, mucho de arbitraria, y no es admitida hoy por gran número de matemáticos de los más autorizados ¹. Puede, en efecto, existir una estrecha relación no lineal entre las dos series con valores del coeficiente de correlación muy próximos a cero, y esta sola consideración debería bastar para rechazar ese criterio de dependencia, al menos como criterio absoluto.

Por otra parte, la determinación del coeficiente de correlación, cuando las series son algo numero-

¹ Véase la comunicación e informe de Fréchet: *Sur l'usage du soi-disant coefficient de corrélation*, presentados en la XXII sesión del Institut International de Statistique. Londres, 1934.

sas, conduce a cálculos prolijos que no están justificados, si no han de dar en cambio sino una indicación dudosa. Para precisar ésta, se ha propuesto por algunos autores sustituir el nombre de coeficiente de correlación por el de coeficiente de *colinealidad*; pero el cambio de nombre no mejoraría en nada la aplicación del concepto, y aun si hubiéramos de aceptar la hipótesis de la relación lineal, cabría obtener un coeficiente más significativo con cálculos más sencillos.

Supongamos, en efecto, que ordenamos de menor a mayor la serie de los a y que debajo o al lado de cada a escribimos el b correspondiente. Si dividimos cada una de las dos series en dos partes que se correspondan, se tendrá

$$\begin{array}{l} A_1 = a_1 + a_2 + a_3 + \dots + a_k \quad a_{k+1} + a_{k+2} + \dots + a_m = A_2 \\ B_1 = b_a + b_b + b_c + \dots + b_n \quad b_{\lambda} + b_{\mu} + \dots + b_m = B_2 \end{array}$$

y si los a y los b están ligados por una relación lineal, el cociente

$$c = \frac{B_2 - B_1}{A_2 - A_1}$$

tendrá un valor constante, cualquiera que sea k . Pero este valor no será ya igual a la unidad, sino que determinará, con excepción de la constante, la fórmula lineal que liga las dos series.

Si la dependencia no fuera tan precisa, pero seguimos suponiendo que existe y que es de forma lineal, el coeficiente c no será ya independiente de k ; tomará valores más o menos variables, que, por lo general, lo serán tanto menos cuanto más se aproxime k al valor $\frac{m}{2}$, en el que probablemente llegará a alcanzar la máxima estabilidad.

Podremos admitir entonces, al menos como una primera aproximación, que los valores b son las sumas de otros dos: b' y b'' , de los cuales, el primero es exactamente proporcional al valor a correspondiente, del cual se deduce multiplicando por el coeficiente de proporcionalidad c , mientras que el segundo sería completamente independiente y vendría a constituir el elemento aleatorio aportado por la segunda serie.

Si, pues, representamos por a_i y b_j dos elementos correspondientes, se tendrá

$$b_j = b'_j + b''_j = c a_i + b''_j$$

donde b''_j sería ya completamente independiente de a_i ; y si se sustituyen a_i y b_j por

$$(1 + c) a_i \quad b''_j$$

y hacemos lo mismo con todos los términos correspondientes de una y otra serie, tendremos definidas dos series nuevas, que serían independientes una de otra, y a las que se podría ya aplicar el método propuesto, siempre bajo la hipótesis de que la dependencia, de existir, afecta una forma lineal.

Hagamos aplicación de las consideraciones expuestas a la discusión del ejemplo tratado en el artículo anterior. Las medias de las dos series son bastante próximas la una de la otra: la primera es 230,2; la segunda, 254,6. Aunque es aquella un poco inferior a ésta, no obtendríamos mayor aproximación a las condiciones óptimas incorporando el mes de julio al primer semestre, porque la media de julio es 21,9 y entonces las de las dos series pasarían a ser 252,1 y 232,7. Se habría invertido el sentido de la desigualdad, pero la diferencia habría permanecido sensiblemente la misma.

Examinemos ahora la cuestión de la independencia. Para calcular el coeficiente de correlación, podrán simplificarse un poco los cálculos, si se tiene en cuenta que

$$\begin{aligned} \Sigma (a_i - a)(b_j - b) &= \Sigma a_i a_j - b \Sigma a_i - a \Sigma b_j + nab = \\ &= \Sigma a_i a_j - nab \end{aligned}$$

$$\begin{aligned} \Sigma (a_i - a)^2 &= \Sigma a_i^2 - 2a \Sigma a_i + na^2 = \Sigma a_i^2 - na^2 \\ \Sigma (b_j - b)^2 &= \Sigma b_j^2 - nb^2. \end{aligned}$$

Hechos los cálculos, se obtiene

$$r = 0,285;$$

el coeficiente, como se ve, es bastante reducido. Los autores que conceden valor práctico al coeficiente de correlación suelen admitir que por bajo de 0,30 se puede concluir que no existe relación de dependencia entre las dos series; el resultado sería, pues, satisfactorio.

La consideración del coeficiente c nos haría ser más exigentes. Si ponemos en orden de magnitud creciente las observaciones del primer semestre, y en correspondencia con ellas las del segundo semestre de los mismos años, se tendría para las sumas más atrás consideradas:

$$\begin{array}{ll} A_1 = 2297,8 & B_1 = 3464,8 \\ A_2 = 4607,8 & B_2 = 4168,7 \\ A_2 - A_1 = 2310,0 & B_2 - B_1 = 703,9 \end{array}$$

de donde

$$c = \frac{B_2 - B_1}{A_2 - A_1} = 0,304.$$

El segundo semestre resulta así influido por el primero, aunque conserva todavía un elemento independiente considerable, que podrá aislarse restando de los valores observados los correspondientes del primer semestre, multiplicados por 0,304. Los valores que así se resten de la segunda serie se aumentarán a los términos correspondientes de la primera, y se tendrán así las dos series rectificadas, a las cuales habría que aplicar el método.

El cuadro 1.º contiene los elementos del cálculo y las series finales, debidamente ordenadas y dispuestas para su ulterior combinación.

Operando sobre estas últimas en la forma expli-

Cuadro número 1.

a_i		ca_i		b_j		b''_j		$(1+c)a_i$		b'_j
84,0	+	25,2	-	254,6	-	229,4	-	109,2		39,0
110,6	+	33,2	-	170,3	-	137,1	-	143,8		42,2
120,9	+	36,3	-	194,2	-	157,9	-	157,2		47,1
121,2	+	36,4	-	475,3	-	438,9	-	157,6		64,7
127,2	+	38,2	-	177,9	-	139,7	-	165,4		81,6
137,6	+	41,3	-	124,9	-	83,6	-	178,9		83,6
142,1	+	42,6	-	126,8	-	84,2	-	184,7		84,2
162,2	+	48,7	-	95,8	-	47,1	-	210,9		137,1
168,4	+	50,5	-	292,0	-	241,5	-	218,9		138,0
170,9	+	51,3	-	309,6	-	258,3	-	222,2		138,4
171,2	+	51,4	-	353,5	-	302,1	-	222,6		139,7
178,0	+	53,1	-	92,4	-	39,0	-	231,4		142,0
185,0	+	55,7	-	221,5	-	165,8	-	241,3		146,2
191,3	+	57,4	-	335,9	-	278,5	-	248,7		157,9
226,6	+	68,0	-	240,1	-	172,1	-	294,6		165,8
233,4	+	70,0	-	368,1	-	298,1	-	303,4		172,1
235,4	+	70,6	-	209,0	-	138,4	-	306,0		178,5
238,7	+	71,6	-	264,7	-	193,1	-	310,3		178,8
241,4	+	72,4	-	251,2	-	178,8	-	313,8		193,0
242,6	+	72,8	-	265,8	-	193,0	-	315,4		193,1
243,2	+	73,0	-	154,6	-	81,6	-	316,2		227,6
270,5	+	82,1	-	124,3	-	42,2	-	32,6		229,4
304,6	+	91,4	-	325,9	-	234,5	-	396,0		234,5
316,6	+	95,0	-	241,2	-	146,2	-	411,6		241,5
348,1	+	104,4	-	332,0	-	227,6	-	452,5		258,3
349,8	+	104,9	-	503,0	-	398,1	-	454,7		298,1
364,7	+	109,4	-	418,6	-	309,2	-	474,1		302,1
368,4	+	110,5	-	175,2	-	64,7	-	478,9		309,2
376,5	+	112,9	-	254,9	-	142,0	-	489,5		398,1
473,9	+	142,2	-	280,2	-	138,0	-	614,1		438,9

Las dos columnas señaladas con b''_j , contienen los mismos números; los de la primera, correspondiendo con la b_j ; los de la segunda, ordenados de menor a mayor.

Cuadro número 2.

	2	5	7	1	6	2	1	5	0	0	1
1	6	15	21	3	18	6	3	15	0	0	3
6	8	20	28	4	24	8	4	20	0	0	4
23	12	30	42	6	36	12	6	30	0	0	6
53	14	35	49	7	42	14	7	35	0	0	7
75	8	20	28	4	24	8	4	20	0	0	4
107	4	10	14	2	12	4	2	10	0	0	2
109	4	10	14	2	12	4	2	10	0	0	2
96	2	5	7	1	6	2	1	5	0	0	1
103	2	5	7	1	6	2	1	5	0	0	1
87	71	66	37	26	20	10	7	2	1	1	

6, 29, 76, 128, 182, 216, 205, 199, 190, 158, 137, 103, 63, 46, 30, 17, 9, 3, 2, 1
 2, 17, 52, 106, 158, 209, 246, 249, 212, 164, 139, 103, 64, 37, 21, 13, 5, 2, 1

cada en el artículo anterior, se deducen los números que han de servir para componer la curva de frecuencia, que son los que figuran en el cuadro 2.º al pie de los productos parciales y por encima de los valores obtenidos anteriormente, al operar sobre las series no corregidas. Las diferencias no son muy considerables; las más aparentes se traducen en un avance del máximo de frecuencia en el sentido de las lluvias crecientes; se encontraría entre los 400 y los 450 mm. para la curva construida con las series sin rectificar, mientras que con las series rectificadas vendría a caer entre los 500 y los 550.

No es, en realidad, de gran importancia práctica la determinación precisa de este máximo de frecuencia, puesto que, en las proximidades del máximo, la función varía poco; pero la acumulación de las diferencias en esos intervalos sugiere una comprobación que puede ser interesante. En la curva anteriormente construida, el número de combinaciones comprendidas entre 400 y 600 mm. son 810 para 1 800 combinaciones; para el mismo número de combinaciones totales son 916 las que corresponderían según la segunda curva entre los mismos límites. En igual proporción corresponderían para treinta años, como valores más probables en cada uno de los dos casos, 13,5 y 15,2 para el número de años en que la lluvia quedara comprendida entre 400 y 600 mm. Los observados han sido 13, cifra mucho más próxima a la primera que a la segunda de las dos calculadas. ¿Habrá que deducir de aquí que es la más aceptable

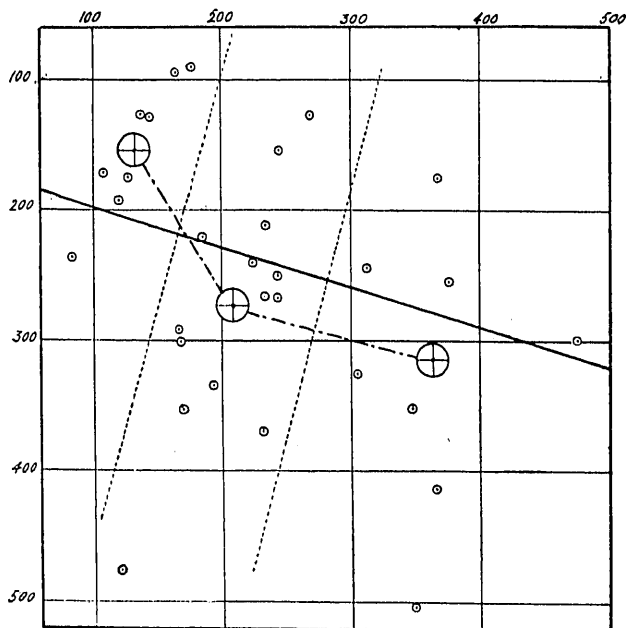


Figura 1.^a

la que resulta de operar con las series sin rectificar? Los datos no son bastante numerosos para poder llegar a esa conclusión; el 13 ó el 15 podrían ser los valores más probables, pero el valor más probable puede ser muy poco probable, y treinta años no es un plazo tan largo que excluya anomalías de esta importancia ni aun mayores.

Por otra parte, si entre los límites estrictos de 400 y 600 mm. no hay más que 13 observaciones, una hay de 397,8, y otra, de 601,5, valores tan próximos a los límites, que sus diferencias apenas si se pueden tomar en cuenta, dada la magnitud de los errores posibles. Las observaciones pasarían entonces de 13 a 15, y nos encontraríamos ya más cerca de la segunda solución.

Todo ello demuestra que nos encontramos en el límite de la precisión, y que en la práctica podría hacerse igualmente uso de una u otra curva, mien-

tras observaciones más prolongadas no dieran motivo en qué fundar una preferencia.

Porque tampoco hay que olvidar que la rectificación de las series se ha hecho en el supuesto de que la relación de dependencia fuese lineal, y esta hipótesis tampoco es más que aproximada. Para juzgar de su grado de aproximación, se podría acudir también a un procedimiento gráfico, con frecuencia empleado. Si se representa cada año por un punto, cuyas coordenadas sean las alturas de lluvia observadas en sus dos semestres, se tendrá el sistema de puntos representado en la figura 1.^a A primera vista, los puntos parecen repartidos al azar, lo que vendría a significar que la independencia entre las dos coordenadas era completa; pero si se examina la figura más atentamente, se observan entre los puntos numerosas alineaciones en la dirección aproximada de la recta señalada en la figura con trazo lleno, y cuyo coeficiente angular es 0,304. En las proximidades de esa recta, que divide el sistema en otros dos igualmente numerosos, se encuentran 13 de los puntos (la tercera parte, aproximadamente, del total), y a uno y a otro lado se observan, con más o menos regularidad, las otras alineaciones paralelas.

Pero, al mismo tiempo, otras alineaciones, aunque no tan numerosas, parecen observarse en dirección casi normal, entre las cuales el conjunto aparece dividido en tres grupos claramente distintos, sin que el escaso número de observaciones permita asegurar si esta distribución responde a la íntima esencia de los fenómenos o es sólo un resultado accidental del azar.

La figura 2.^a representa la curva de frecuencia rectificada, y la 3.^a reproduce la cuarta del artículo anterior, en la que convendrá advertir la equivocación padecida al numerar la escala, cuyas cotas deben ser todas aumentadas en 50. Si se comparan estas curvas con los tipos de Pearson, se observa

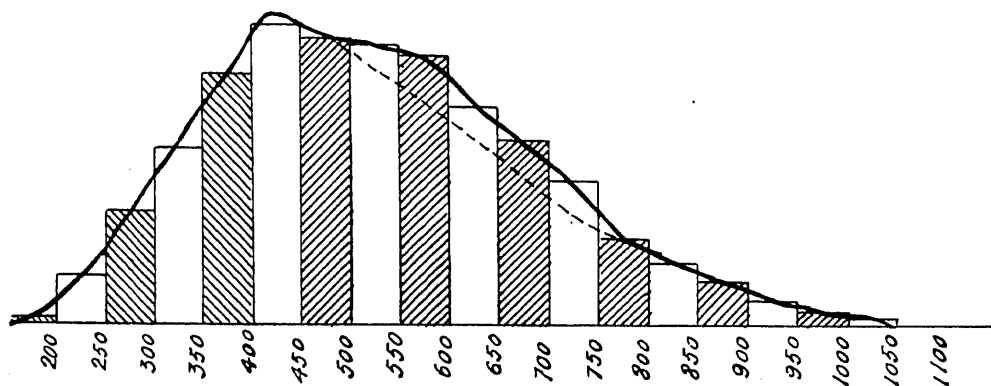


Figura 2.^a

una diferencia bastante marcada (aunque más en una que en otra), constituida por la protuberancia que aparece del lado allá del máximo de frecuencia, que viene a interrumpir la caída regular de la curva, que sería, aproximadamente, la que con líneas de trazos se señalan en una y otra figura.

¿Es esta circunstancia un resultado puramente accidental o procede de causa física? Aventurado

da, y la última se atenúa. Parece, pues, que aumenta la probabilidad de que la razón del hecho se encuentre en que no todas las lluvias puedan agruparse en un sistema único, obediente a una misma ley regular de probabilidad. La división en tres grupos que aparece marcada en la figura 1.^a añade todavía verosimilitud al supuesto; pero el desarrollo del tema exigiría aún estudios más detenidos,

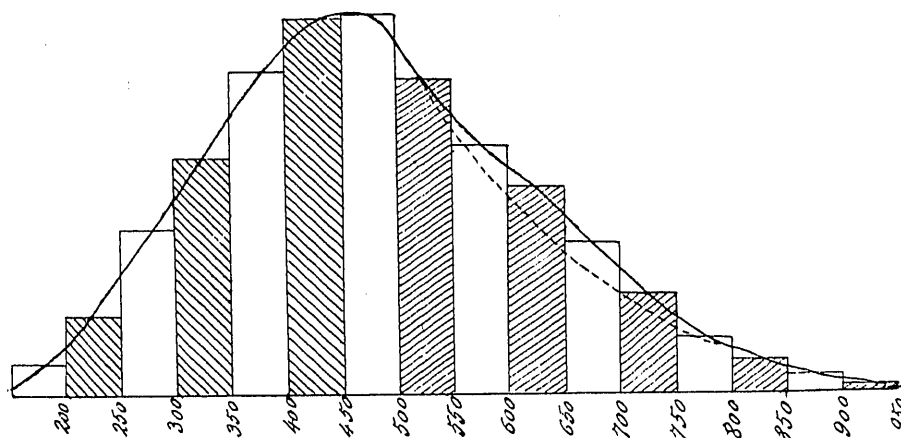


Figura 3.ª

sería decidir en un caso tan particular y contándose con tan pocas observaciones; pero si se tiene en cuenta que ya había llegado a un resultado parecido en 1932¹ al construir la curva de frecuencia media de 141 estaciones pluviométricas españolas observadas en el período decenal de 1916-25, la verosimilitud de la segunda hipótesis aparece mayor. Ya allí indicaba (pág. 35) que parecía como si a una ley regular de frecuencia viniera a superponerse, en esta rama descendente de la curva, otra ley secundaria con máximo distinto, y añadía: "Si esta ley secundaria corresponde a una distinción real en la naturaleza de los estados atmosféricos a que obedecen las precipitaciones, o es debida a la superposición de curvas demasiado heterogéneas procedentes de las distintas regiones de la Península, o es, por último, circunstancia accidental dependiente del período estudiado, cuestiones son que sólo podrían encontrar contestación por un estudio más detenido, para el que convendrá tener más datos que los que actualmente se poseen".

En el caso que ahora estudiamos se trata de una estación única, y el período de observación es de treinta años; una de las hipótesis queda descarta-

y bastará por hoy con dejarlo señalado a la investigación.

Es significativo, sin embargo, que si se determinan los centros de gravedad de los tres grupos, estos puntos, que en la figura están señalados con círculos mayores que los de los puntos ordinarios, se apartan bastante de la línea recta, lo que parece indicar que, si alguna correlación existe, ésta no podría expresarse, como media, por una fórmula lineal.

En cuanto al objeto propio del artículo, puede quedar resumido en las siguientes conclusiones:

1.ª Prácticamente, las curvas de frecuencia compuestas pueden dar solución al problema, con bastante aproximación, aun sin necesidad de rectificaciones, que nunca pueden ser completas y que, a veces, pueden resultar ilusorias.

2.ª Aparte de su valor práctico inmediato, el análisis de las probabilidades elementales y de la mayor o menor dependencia entre los fenómenos meteorológicos sucesivos tiene en sí mismo un interés científico, que puede llegar a ser fecundo en aplicaciones. Constituye un método utilizable en materias como ésta, en la que la complejidad de las causas hace tan difícil descubrir el íntimo mecanismo y la fórmula determinista que hubiera de representarlo.

¹ González Quijano: *Avance para una evaluación de la energía hidráulica de España*. Publicación del Consejo de la Energía, página 33.