

CALCULO DE VALORES EXTREMOS DE UNA VARIABLE. APLICACION AL ESTUDIO DE CAUDALES MENSUALES(*)

Por Dr. JOSE LLAMAS
Quebec (Canadá), 1969

En este trabajo se presentan ejemplos de la aplicación de la Estadística Matemática, a la determinación de frecuencias de valores extremos, en variables hidrológicas.

Partiendo de la teoría estadística de distribuciones conjuntas y del criterio general de independencia, se deducen las funciones de distribución de los valores extremos.

Por último, las distribuciones obtenidas se aplican al análisis frecuencial de los caudales máximos y mínimos en el río Chaudière, afluente del San Lorenzo (Quebec, Canadá), determinando previamente las funciones de densidad por el método de los polinomios de Laguerre, con la ayuda de un ordenador electrónico.

1. INTRODUCCION

Durante mucho tiempo, las ciencias físicas en general y la hidrología en particular se han apoyado para su progreso en la experiencia de un cierto número de investigadores que supieron obtener generalidades de un conjunto de observaciones. El ingeniero puso su técnica, y lo que es igualmente importante, su sentido común al servicio de una aplicación práctica e inmediata, siguiendo los cauces abiertos por el investigador.

Desde hace unos años, el ingeniero es testigo de un impacto que, en ocasiones, hace vibrar los cimientos de las ciencias clásicas. En hidrología este impacto se presenta de tres formas más o menos diferentes:

a) La cantidad de datos acumulados tras unos decenios de pacientes medidas: precipitación, caudales, nebulosidad, etc.

b) El avance tecnológico producido como consecuencia de un desarrollo industrial acelerado y de nuevas necesidades en recursos hidráulicos.

Una de las facetas más interesantes de esta nueva tecnología es la colaboración, hasta entonces inexistente, entre diversas disciplinas. La hidrología moderna se sirve, en su desarrollo, de la experiencia adquirida en otros sectores, y como ejemplo podría citarse el análisis armónico, análisis espectral, los conceptos de señal y ruido, etc., materias todas que fueron utilizadas anteriormente por el investigador en cuestiones electrónicas. La estadística, la teoría de probabilidades y de conjuntos son otras herramientas de las que el hidrólogo moderno se sirve continuamente. En un siglo de especialización se da el caso paradójico que la mayor parte de las especialidades son, en realidad, multi-disciplinarias.

c) La fantástica potencia de cálculo disponible con los computadores modernos. La facilidad operacional, desconocida hasta hace unos años, abre al hidrólogo nuevos horizontes y le obliga a mirar, desde un ángulo distinto, problemas que, hasta la fecha, fueron del dominio clásico. Hoy día un análisis de variables pluridimensio-

(*) Se admiten comentarios sobre el presente artículo, que pueden remitirse a la Redacción de esta Revista, hasta el 3^o de enero de 1971.

nales, la inversión de una matriz de covarianzas, etc., son operaciones medidas con la escala del segundo. Una cantidad asombrosa de datos puede ser utilizada simultáneamente para obtener, y este es el fin primordial, la mayor información posible.

Esta nueva metodología tiende a eliminar más y más las "recetas" de carácter casi-universal que el técnico utilizaba para compensar, con una generalización arriesgada, la insuficiencia de sus instrumentos de cálculo.

Un ejemplo: muy a menudo vemos en libros antiguos que una muestra hidrológica (caudal, precipitación, etc.) puede ser considerada como perteneciente a una familia cualquiera de probabilidades (normal, uniforme, etc.) de manera que con dos o tres parámetros como máximo la aproximación está terminada. El hidrólogo sabe que trabajar con más de tres parámetros, utilizando calculadoras no electrónicas es un

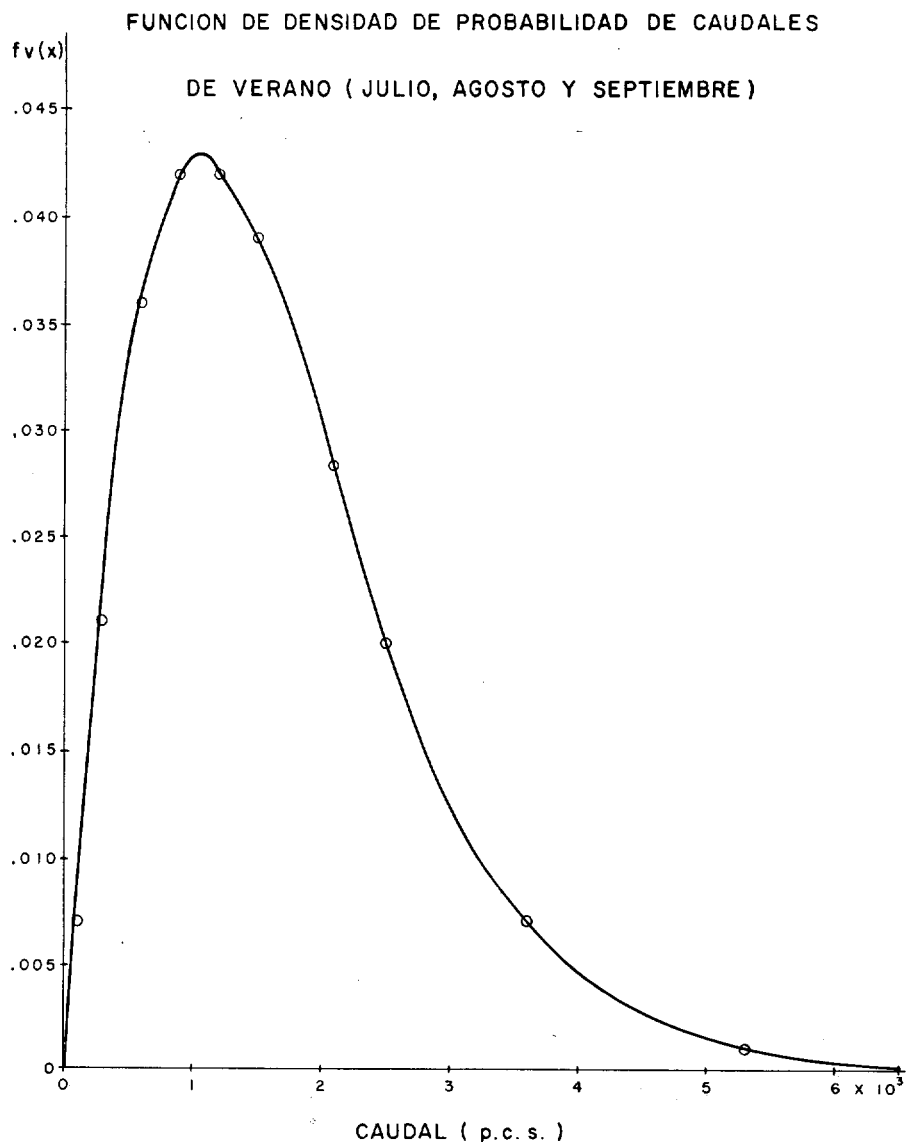


Figura 1.

$F_v(x)$ FUNCION DE DISTRIBUCION DE CAUDALES DE VERANO.

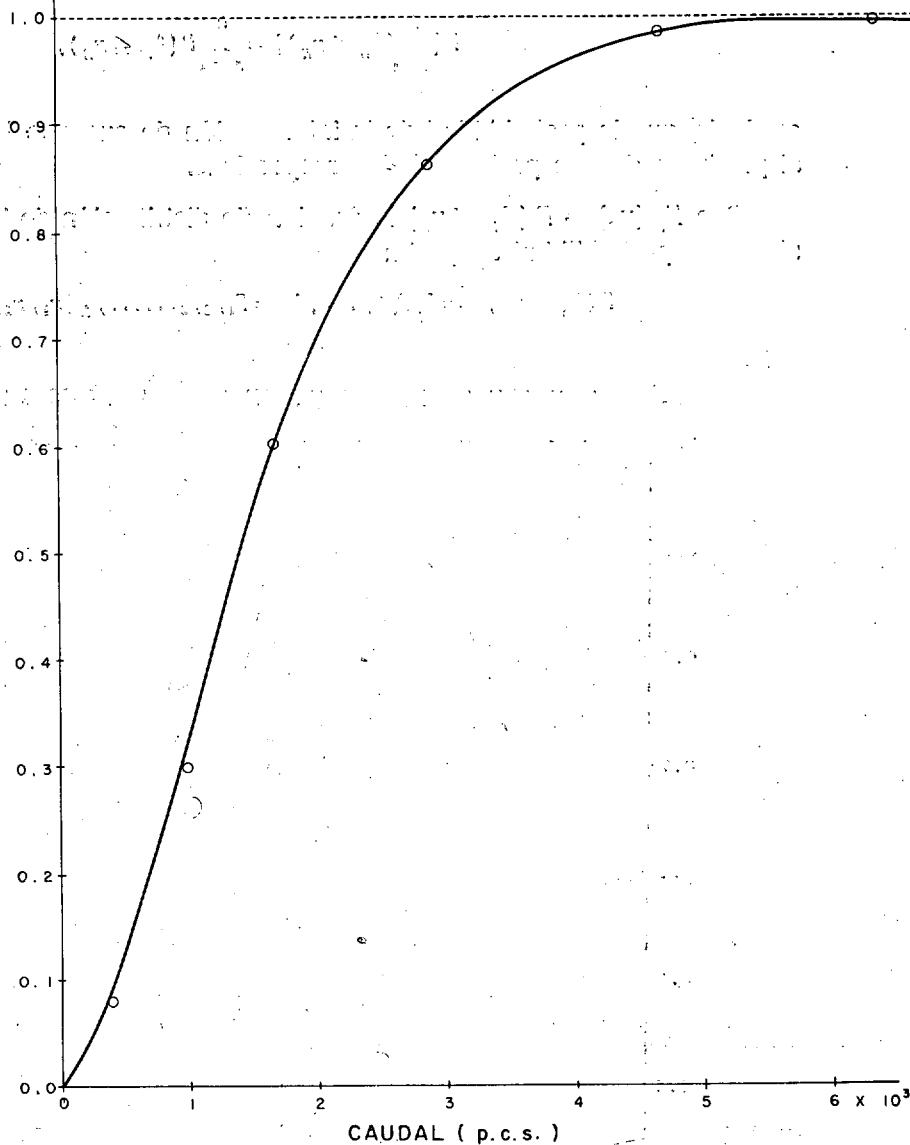


Figura 2.

asunto de días y a veces de meses, si la muestra tiene dimensiones aceptables. Por consiguiente era lógico sacrificar una gran parte de la información contenida en la muestra a la rapidez del proyecto.

Siguiendo estas nuevas tendencias, examinemos uno de los problemas más importantes en hidrología: el análisis de máximos y mínimos de una variable aleatoria. Dos ejemplos, uno teórico y otro práctico, completarán el estudio.

2. CALCULO DE LOS VALORES EXTREMOS DE UNA VARIABLE ALEATORIA

Consideremos una serie de k variables aleatorias independientes:

$$A_n, n = 1, 2, 3, \dots, k$$

De acuerdo con el "criterio general de independenciam" (C.G.I.):

$$P \left\{ \bigcap_{n=1}^k (A_n \leq x_n) \right\} = \prod_{n=1}^k P (A_n \leq x_n),$$

es decir, que la probabilidad de la intersección de sucesos independientes es igual al producto de sus probabilidades respectivas.

Sea $F_n(x_n) = P(A_n \leq x_n)$ la función de distribución de la variable A_n . Por definición de probabilidad conjunta:

$$P(A_n \leq x_n, n = 1, 2, 3, \dots, k) = F_{1, 2, 3, \dots, k}(x_1, x_2, \dots, x_k)$$

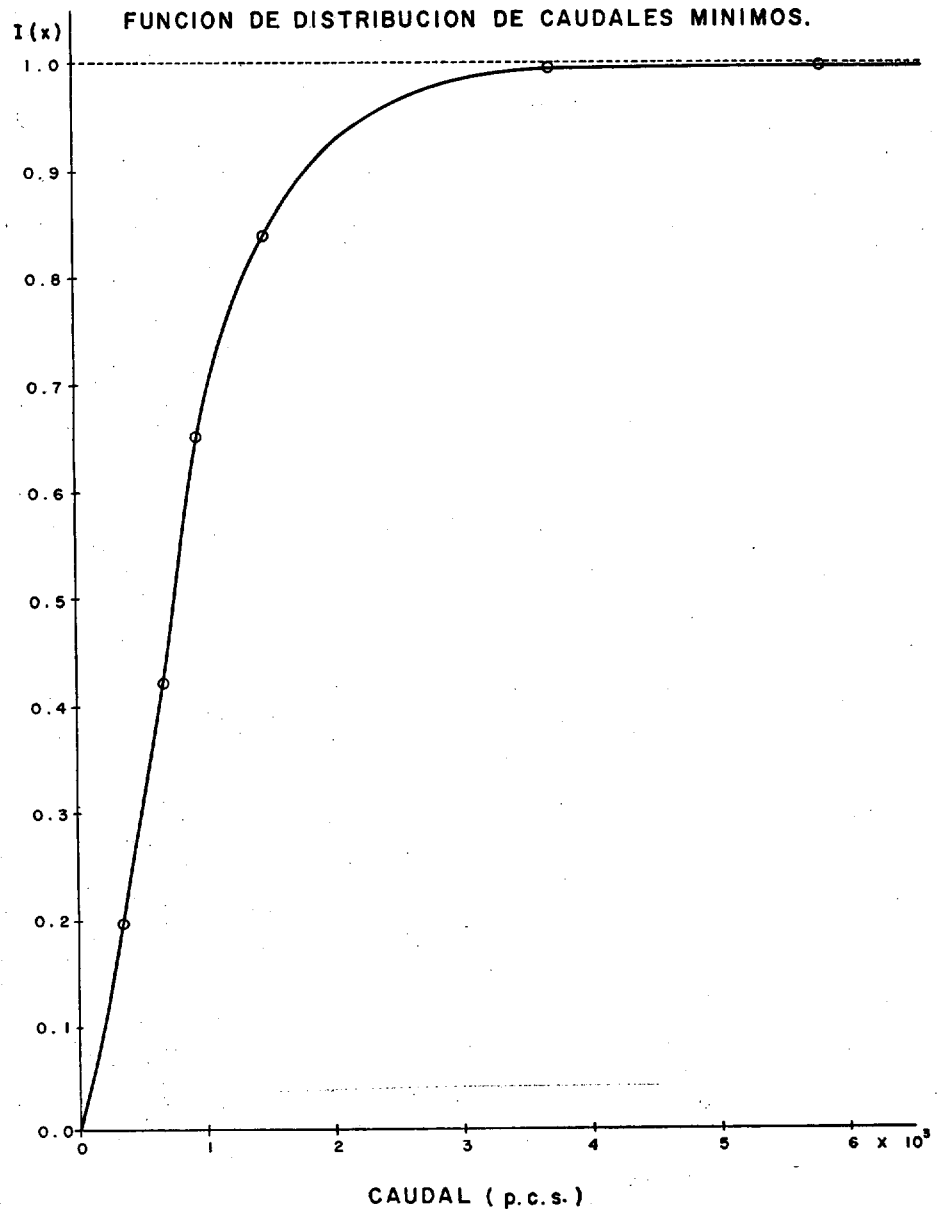


Figura 3.

FUNCION DE DENSIDAD DE PROBABILIDAD DE CAUDALES

DE PRIMAVERA (ABRIL Y MAYO)

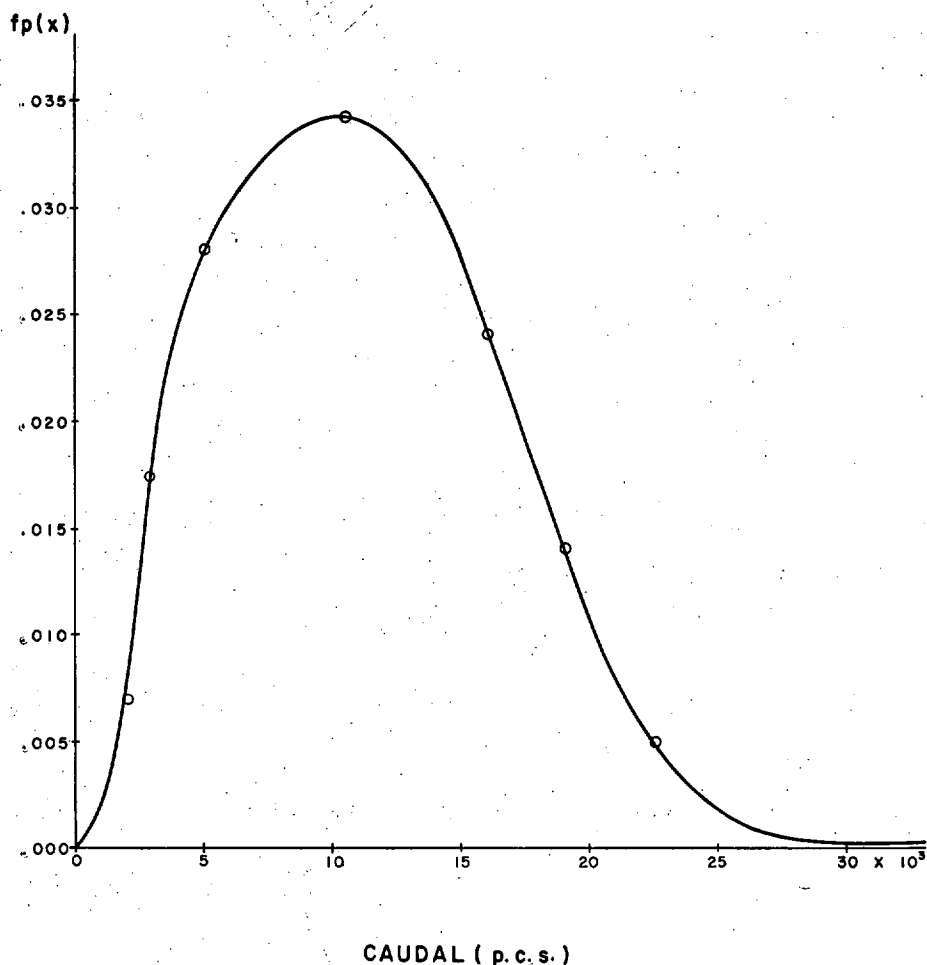


Figura 4.

y en virtud del C.G.I.:

$$P(A_n \leq x_n, n = 1, 2, 3, \dots, k) = F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_k(x_k)$$

Si todas las variables tienen una función de distribución común $F(x)$, entonces:

$$P(A_n \leq x_n, n = 1, 2, 3, \dots, k) = [F(x)]^k$$

Consideremos ahora la subserie $(A_p, p = 1, 2, 3, \dots, n)$ y las variables aleatorias siguientes:

$$\underline{A} = \inf. A_p = \text{variable de valor inferior en } [A_p] \quad (1 \leq p \leq n)$$

$$\overline{A} = \sup. A_p = \text{variable de valor superior en } [A_p] \quad (1 \leq p \leq n)$$

$\inf. A_p$ y $\sup. A_p$ representan los valores extremos de la subserie. El problema con-

siste en calcular las funciones de distribución o las funciones de densidad de probabilidades de inf. A y sup. A.

Sean $I_n(x)$ y $S_n(x)$ las funciones de distribución de inf. A_p y sup. A_p , respectivamente. Es decir:

$$I_n(x) = P(\text{inf. } A_p \leq x) \quad \text{y}$$

$$S_n(x) = P(\text{sup. } A_p \leq x)$$

Es evidente que si inf. A_p es superior a un valor determinado, todas las variables de la subserie lo serán igualmente:

$$(\text{inf. } A_p > x) = \prod_{p=1}^n (A_p > x)$$

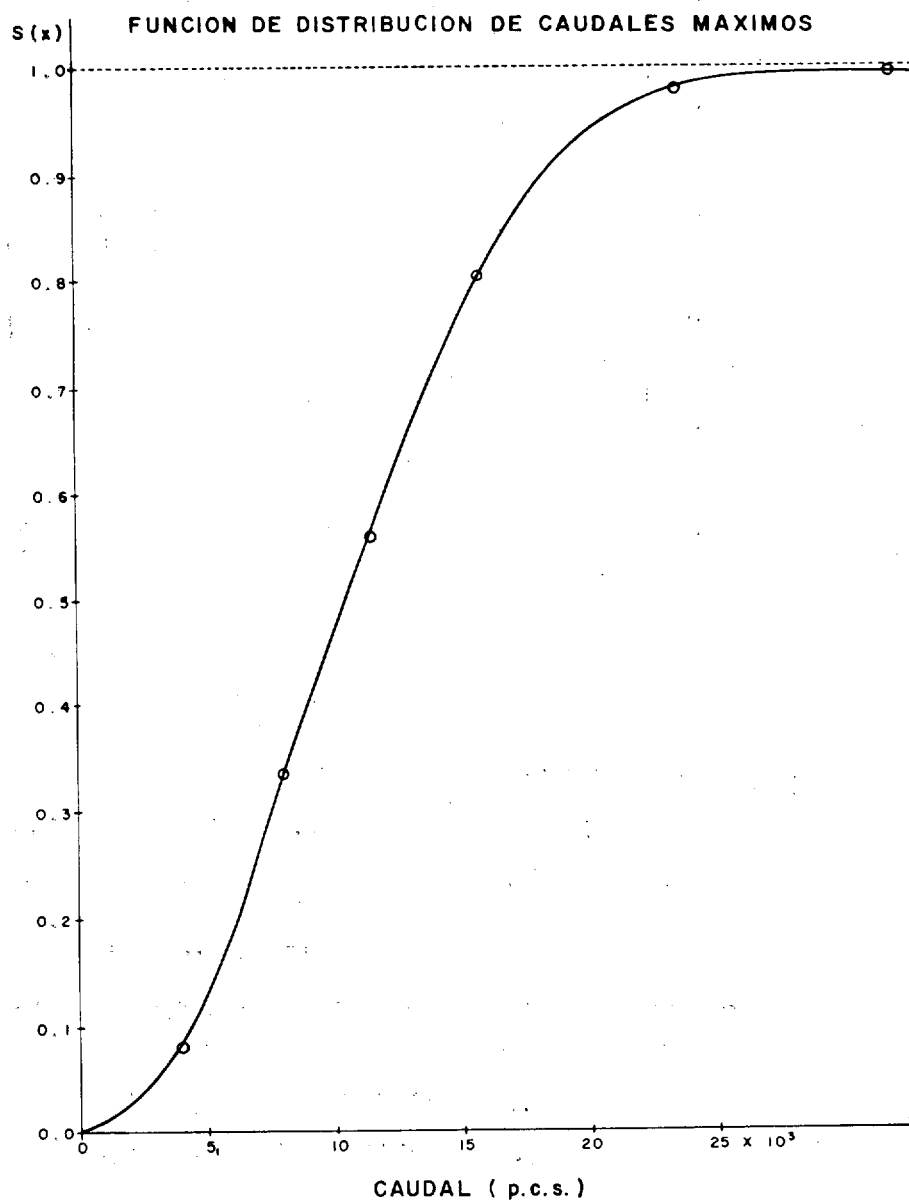


Figura 5.

Y en términos de probabilidad:

$$P(\inf. A_p > x) = P\left\{\prod_{p=1}^n (A_p > x)\right\} = \prod_{p=1}^n P(A_p > x) = \prod_{p=1}^n [1 - P(A_p \leq x)]$$

De donde:

$$1 - P(\inf. A_p \leq x) = \prod_{p=1}^n [1 - P(A_p \leq x)]$$

$$I_n(x) = 1 - \prod_{p=1}^n [1 - P(A_p \leq x)] \quad (1)$$

Igualmente:

$$P(\sup. A_p \leq x) = \prod_{p=1}^n P(A_p \leq x),$$

por consiguiente:

$$S_n(x) = \prod_{p=1}^n P(A_p \leq x) \quad (2)$$

Sean $F_p(x)$ y $f_p(x)$ la función de distribución y la función de densidad de probabilidad de la variable A_p . Es decir:

$$F_p(x) = P(A_p \leq x), \text{ y } f_p(x) = \frac{d F_p(x)}{d x}$$

Las ecuaciones (1) y (2) son:

$$I_n(x) = 1 - \prod_{p=1}^n [1 - F_p(x)] \quad (3)$$

$$S_n(x) = \prod_{p=1}^n F_p(x) \quad (4)$$

y si todas las variables de tipo A tienen la misma función de distribución, $F(x)$, entonces:

$$I(x) = 1 - [1 - F(x)]^n \quad (5)$$

$$S(x) = [F(x)]^n \quad (6)$$

Estas dos últimas expresiones son las funciones de distribución de los valores extremos de la variable aleatoria A . Las funciones de densidad de probabilidad correspondientes son:

$$i(x) = \frac{d I(x)}{d x} = n \cdot f(x) \cdot [1 - F(x)]^{n-1} \quad (7)$$

$$s(x) = \frac{d S(x)}{d x} = n \cdot f(x) \cdot [F(x)]^{n-1} \quad (8)$$

3. EJEMPLO TEORICO

Se puede demostrar matemáticamente que la función de densidad de probabili-

dad de la intensidad cumulativa de una serie de m aguaceros obedece a una ley gamma de tipo:

$$f(x) = \frac{p^m}{\Gamma(m)} e^{-px} x^{m-1} \quad \text{para } x > 0$$

El problema consiste en calcular, con la ayuda de una serie de datos experimentales, la función de densidad de probabilidades de los valores extremos de la intensidad de un aguacero.

En este caso, $m = 1$:

$$f(x) = p \cdot e^{-px}; x > 0$$

$$F(x) = p \cdot \int_0^x e^{-pt} dt = 1 - e^{-px}$$

La f.d.p. de la intensidad mínima es:

$$i(x) = n \cdot p \cdot e^{-px} \cdot e^{-px(n-1)} = n \cdot p \cdot e^{-npx} \quad (9)$$

Igualmente:

$$s(x) = n \cdot p \cdot e^{-px} (1 - e^{-px})^{n-1} = n \cdot p \cdot e^{-px} \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} e^{-kpx}$$

$$s(x) = n \cdot p \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} e^{-(k+1)px} \quad (10)$$

4. APLICACION AL CALCULO DE CAUDALES EXTREMOS

Un análisis de los caudales medios mensuales, durante un período de cuarenta y cuatro años, en el río Chaudière (afluente del San Lorenzo, en la provincia de Quebec, Canadá), ha permitido obtener la información siguiente:

a) El caudal anual mínimo absoluto se produce, sin excepción, durante los meses de julio, agosto o septiembre.

b) Los caudales medios mensuales, durante estos tres meses, son mutuamente independientes y poseen una distribución común. Con objeto de verificar estas afirmaciones, un test de independencia (test de Markov) y otro de distribución (test de Smirnov-Kolmogorov), fueron realizados. Ambos dieron resultados afirmativos a un nivel de confianza de 95 por 100. Esto permite considerar los caudales medios mensuales de estos tres meses, como pertenecientes a la misma población.

c) El caudal anual máximo absoluto se produce en los meses de abril o mayo (afectado por los deshielos). Los caudales medios mensuales, de estos dos meses, tienen las mismas propiedades que los caudales mínimos (independencia y distribución común).

La lista I muestra los caudales registrados durante el período 1923-66.

LISTA I

AÑO	PRIMAVERA		VERANO		
	Abril	Mayo	Julio	Agosto	Septiembre
1923	17 000	10 800	1 520	1 630	1 070
1924	8 370	12 300	821	817	2 170
1925	12 200	4 040	1 770	2 670	2 790
1926	4 300	18 200	1 060	653	947
1927	7 420	5 380	896	546	487
1928	10 400	13 400	1 750	2 650	3 120
1929	14 300	10 100	2 220	1 890	1 180
1930	15 000	10 900	2 480	1 020	756
1931	11 000	3 810	945	936	2 620
1932	16 000	4 930	1 760	1 680	2 220
1933	15 600	15 300	759	671	318
1934	22 700	5 420	854	719	597
1935	9 270	7 270	1 970	768	1 170
1936	9 150	11 900	822	1 020	4 090
1937	14 700	10 100	2 000	4 870	2 580
1938	15 800	6 980	1 770	3 590	2 960
1939	13 100	14 500	2 430	3 650	3 060
1940	10 400	10 100	1 990	740	504
1941	15 700	2 610	1 400	491	1 200
1942	20 700	8 570	1 220	1 210	853
1943	14 700	15 300	2 530	1 930	682
1944	12 000	10 200	1 650	1 460	1 740
1945	14 500	8 050	2 880	726	3 360
1946	8 570	6 320	595	2 080	574
1947	18 000	18 900	5 230	1 070	563
1948	11 100	10 000	1 490	835	337
1949	11 400	3 060	524	446	694
1950	14 000	3 490	2 070	2 540	2 110
1951	22 600	3 560	2 200	1 790	2 570
1952	16 300	3 630	421	398	656
1953	12 000	5 850	1 890	546	725
1954	14 800	5 590	2 430	4 040	8 100
1955	23 300	7 400	754	1 720	681
1956	12 300	11 800	2 140	2 920	2 540
1957	5 680	3 840	1 450	2 550	1 590
1958	15 400	4 600	2 900	3 960	2 080
1959	16 400	4 060	2 150	3 630	1 470
1960	22 400	7 140	1 920	780	1 510
1961	15 700	19 300	1 500	1 420	3 010
1962	19 700	7 840	1 340	1 660	1 300
1963	17 800	9 870	2 560	4 610	4 190
1964	14 500	5 260	2 410	778	641
1965	6 480	4 850	551	1 920	4 490
1966	12 600	5 720	1 630	1 010	2 130

Sean $f_v(x)$ y $f_p(x)$ las funciones de densidad de probabilidad de verano y primavera, respectivamente. Las funciones de distribución correspondientes son:

$$F_v(x) = P(X_v \leq x) \text{ y } F_p(x) = P(X_p \leq x).$$

Las funciones $f_v(x)$ y $f_p(x)$ pueden ser calculadas con ayuda de una serie convergente de polinomios ortogonales. Como en los dos casos, el coeficiente de asimetría de la muestra es mayor que cero y, por otra parte, las variables son definidas únicamente en el dominio positivo, la serie de polinomios de tipo Laguerre es la más indicada:

$$f_v(x) = f(x; g, h) \sum_{m=0}^k \frac{m! \Gamma(h/2)}{\Gamma(m+h/2)} \cdot \frac{d_m}{(2g)^m} \cdot L_m^{(h/2-1)}\left(\frac{x}{2g}\right); x > 0$$

donde:

$$f(x; g, h) = \left[e^{-\frac{x}{2g}} \left(\frac{x}{2g}\right)^{\frac{h}{2}-1} \right] / [(2g)^2 \Gamma(h/2)]; x > 0$$

$$d_m = \sum_{i=0}^m \binom{m-1+h/2}{m-i} (-1)^i (2g)^{m-i} \cdot \frac{\gamma_j}{i!}$$

$\gamma_j = E(x_v)^j$ es el momento central, de orden j , de la muestra.

$L_m^{(c)}(y)$ es el polinomio de Laguerre de orden m :

$$L_m^{(c)}(y) = \sum_{j=0}^m \binom{m+c}{m-j} \frac{(-y)^j}{j!}$$

El parámetro de escala, g , y el número de grados de libertad, h , se calculan por el método de los momentos, es decir:

$$E(x_v) = \int_0^{\infty} x f(x; g, h) dx = g h$$

$$E(x_v)^2 = \int_0^{\infty} x^2 f(x; g, h) dx = g^2 h (h+2)$$

En este estudio, $k = 5$, es decir, se han tenido en cuenta los 5 primeros momentos de la muestra.

Finalmente:

$$f_v(x) = 4.83 \cdot 10^{-7} \cdot \exp(-0.001205 x) \cdot x^{1.15316 \cdot 10^{-7}} (0.78256 + 0.000374 x - 1.4004 \cdot 10^{-7} x^2 + 8.19842 \cdot 10^{-12} x^3 + 1.68471 \cdot 10^{-15} x^4 - 1.24470 \cdot 10^{-19} x^5)$$

La función de distribución $F_v(x)$ se obtiene fácilmente integrando $f_v(x)$ en el dominio positivo.

Una vez calculadas estas funciones, las expresiones (5) y (6) permiten calcular las funciones de distribución de los caudales extremos. En el caso de caudales mí-

nímos (I), $n = 3$ y en el de caudales máximos (S), $n = 2$. Estas funciones de distribución están representadas en las figuras 3 y 5, respectivamente.

Los cálculos numéricos han sido realizados en lenguaje A.P.L. sobre ordenador IBM-360-40.

REFERENCIAS

1. LLAMAS, J.: "Deficit and Surplus in Precipitation Series". Tesis de Doctorado. Colorado State University, Fort Collins, Colorado, 1968.
2. LLAMAS, J.: "Runs of Precipitación". Colorado State University, Hydrology Paper, núm. 33, 1969.
3. SIDDIQUI, M. M.: "Test for Regression Coefficients when Errors are Correlated". An. Math. Statistics, vol. 31, 1960.
4. SZEGO, G.: "Orthogonal Polynomials". Am. Math. Society. Colloquium Publications, vol. XXIII, 1939.

© 1970 by American Statistical Association