Maximización de la función de Verosimilitud de Distribuciones de Probabilidad usando Algoritmos Genéticos

Maximization of the Likelihood Function of Probability Distributions using Genetic Algorithms

Fuentes-Mariles, O. A.[a1], Arganis-Juárez, M. L.[a2], Domínguez-Mora, R.[a3], Fuentes-Mariles, G. E. [a4], Rodríguez-Vázquez, K. [b]

[a] Instituto de Ingeniería, UNAM, Edificio 5 Cub. 403. Circuito Escolar, Ciudad Universitaria, Coyoacán, México D.F. C.P. 04510 (México).
E-mail: [a1] ofm@pumas.iingen.unam.mx, [a2] marganisj@iingen.unam.mx, [a3] rdm@pumas.iingen.unam.mx, [a4] gfm@pumas.iingen.unam.mx

[b] Instituto de Investigaciones en Matemáticas Aplicada y en Sistemas. E-mail: [b] katya@uxdea4.iimas.unam.mx

Recibido: 31/07/2014 Aceptado: 10/12/2014 Publicado: 29/01/2014

RESUMEN

Tradicionalmente, para obtener los parámetros de una función de distribución con el método de máxima verosimilitud se acostumbra igualar a cero la derivada del logaritmo de la función de verosimilitud y resolver el sistema de ecuaciones no lineales que resulta. La popularidad del procedimiento se debe a su sencillez; sin embargo, cuando la función de verosimilitud no es suficientemente regular, puede llevar a obtener un valor muy alejado del máximo Por ese motivo, en este documento se presenta el uso de un algoritmo genético que permite encontrar los parámetros de la función de distribución (con lo que se maximiza directamente la función de verosimilitud, o su logaritmo, sin recurrir a la derivada de los logaritmos de dicha función). Se halló buena concordancia de los resultados respecto a los obtenidos usando un software de uso frecuente en México, para el caso las funciones Gumbel y Gumbel de dos poblaciones.

Palabras clave | Algoritmo genético; máxima verosimilitud; Función Gumbel; Gumbel de dos poblaciones; optimización.

ABSTRACT

Traditionally, to get the parameters of a distribution function with the maximum likelihood method is usually equaled to zero the derivative of the logarithm of the likelihood function and then the resulting non-linear system of equations is solved. The popularity of the procedure is due to its simplicity; however, when the likelihood function is not regular enough, can lead to obtain a value very far away from the maximum sought. This document presents the use of a genetic algorithm that allows to find the parameters of the distribution function by directly maximizing the likelihood function, or its logarithm, without need to resort to the derivative of the logarithms of the function. The results are compared with those obtained using a software frequently used in Mexico, for the case functions Gumbel and Gumbel of two populations.

***Key words** | Genetic Algorithm; máximum likelihood; Gumbel Function; Gumbel of two populations; optimization.*

INTRODUCCIÓN

Las funciones de distribución de algunas variables hidrológicas tienen gran importancia en los estudios de Hidrología. Para estimar sus parámetros se pueden utilizar métodos como los de momentos y de máxima verosimilitud, aprovechando diversos programas o paqueterías de cómputo que en su mayoría resuelven las ecuaciones correspondientes a dichos métodos con procedimientos de tipo determinístico. Ejemplos de ellos son el programa de ajuste de funciones de distribución de probabilidad AX (Jiménez, 1996), el 60-Matrix (6O-Matrix Statistical Time Series Analysis, 2005), el paquete interactivo desarrollado en la Universidad de las Américas-Puebla (Mazariegos y Raynal, 2002).

En particular, con respecto al método de máxima verosimilitud, existen algunos interrogantes, ya que los algoritmos que usan muchos de estos paquetes recurren a igualar a cero las derivadas parciales de la función de verosimilitud, lo que puede llevar a encontrar valores de los parámetros que corresponden a un valor de la función de verosimilitud muy alejado del máximo (Smith, 1988, Horbelt *et al.*, 2002, Myiung, 2003).

En este documento se analiza el comportamiento de algoritmos de tipo aleatorio, como los algoritmos genéticos (AG), (Goldberg, 1989), al usarlos para maximizar directamente la función de verosimilitud o su logaritmo. En los últimos años este tipo de algoritmos se ha utilizado ampliamente en varios campos de la ingeniería hidrológica e hidráulica (Nicklow *et al*., 2003, Liu *et al.,* 2004, Jha *et al.,* 2004, Clark y Whu, 2006).

Antecedentes

Debido a la situación geográfica de la República Mexicana, las distribuciones de probabilidad de los gastos máximos anuales de muchas corrientes se apegan al comportamiento de las funciones Gumbel (Domínguez *et al.*, 2000, Domínguez *et al.*, 2006, Arganis *et al.*, 2009). Sin embargo, en muchos otros casos en los que los gastos máximos anuales son resultado tanto de eventos de tipo ciclónico como no ciclónico, la función doble Gumbel o Gumbel de dos poblaciones constituye una mejor alternativa (González, 1970). Por este motivo en este trabajo se hace especial énfasis en estas dos funciones de probabilidad.

Domínguez, *et al.,* 2004 aplicaron un algoritmo genético simple (AG) en la determinación de los parámetros de una función Gumbel de dos poblaciones; en dicho estudio la función objetivo consideraba minimizar el error medio cuadrático entre los valores de la probabilidad de no excedencia estimados con la fórmula de Weibull (Kite, 1988) y los calculados con la función de distribución ajustada. Utilizando la misma función objetivo, Fuentes *et al.,* 2005, hicieron ajustes a otras funciones de distribución aplicables a series hidrológicas que comúnmente se presentan en la República Mexicana. Un estudio antecedente al aquí presentado planteó la maximización directa de la función de verosimilitud para algunas distribuciones, entre ellas la distribución Gumbel (Fuentes, *et al.,* 2006).

MATERIAL Y MÉTODOS

Método de máxima verosimilitud

Este método establece que para una variable aleatoria x que tiene una función de densidad de probabilidad $p(x;\alpha,\beta,\ldots)$ donde $\alpha, \beta,\ldots$ son los parámetros de dicha función, la probabilidad de obtener un valor dado, x_i, es proporcional a $p(x_i;\alpha,\beta,\ldots)$, y la probabilidad conjunta de obtener una muestra aleatoria de n valores independientes $x_1, x_2,\ldots x_n$ es proporcional al producto:

$$L = \prod_{i=1}^{n} p(x_i;,\alpha,\beta,\ldots) \tag{1}$$

A la función anterior se le llama verosimilitud (Kite, 1988).

El método de máxima verosimilitud consiste en estimar los parámetros $\alpha, \beta,\ldots$ tales que L sea máxima.

Para hacer la estimación, diversos autores (Jiménez, 1996, Escalante y Reyes, 2002, Rao y Hamed, 2000,) proponen usar el logaritmo de la función L:

$$\log L = \log p(x_i; , \alpha, \beta, \ldots) + \log p(x_2; , \alpha, \beta, \ldots) + \log p(xn; , \alpha, \beta, \ldots) \tag{2}$$

Obtener las derivadas respecto a cada uno de sus parámetros:

$$e_1 = \frac{\partial \log L}{\partial \alpha}$$
$$e_2 = \frac{\partial \log L}{\partial \beta}$$

Para finalmente estimar los parámetros $\alpha, \beta, \ldots$ al igualar estas derivadas a cero.

En ocasiones, este procedimiento puede conducir a valores de la función de verosimilitud L alejados del máximo (Smith, 1988, Horbelt *et al.,* 2002, Myiung, 2003). Por este motivo, en este trabajo se propone aprovechar la capacidad de los algoritmos genéticos para obtener directamente los parámetros que hacen máxima la función de verosimilitud (o su derivada) sin necesidad de igualar a cero las derivadas del logaritmo de dicha función respecto a cada uno de sus parámetros.

Algoritmos determinísticos. Programa AX

Diversos software, como los ya mencionados: AX, el paquete interactivo de la Universidad de las Américas-Puebla, utilizan el logaritmo de la función de verosimilitud y el concepto de derivada) para estimar los parámetros que maximizan dicha función.

En particular, el programa AX desarrollado por el Centro Nacional de Desastres (CENAPRED) (Jiménez, 1996) es un software de uso frecuente en México, elaborado en ambiente Visual Basic, que ayuda a efectuar el análisis estadístico de series de valores máximos anuales. El programa lee un archivo que contiene los datos anuales históricamente registrados de la variable considerada, ordena los datos de mayor a menor, les calcula su periodo de retorno (Tr) con la fórmula de Weibull y estima los parámetros de distintas funciones de distribución con los métodos de momentos y de máxima verosimilitud con ayuda de algoritmos determinísticos; para el método de máxima verosimilitud utiliza el logaritmo de la función de verosimilitud y el concepto de derivada. Adicionalmente, hace el dibujo de las funciones ajustadas con respecto a los datos históricos y permite estimar eventos de diseño para varios periodos de retorno (Figura 1).

Algoritmos genéticos

Los Algoritmos Genéticos (AG) imitan de manera parcial los mecanismos de evolución biológica. Esta técnica fue desarrollada en la universidad de Michigan a principios de la década de 1960 por John H. Holland y un grupo de estudiantes (Holland, 1975). Las características que distinguen a esta técnica son las siguientes:

1. Comúnmente utilizan codificación binaria (cadenas de 0´s y 1´s) para las posibles soluciones del problema a resolver, aunque también se han utilizado el código gray, representación entera y real.
2. El operador principal es el de cruza (en el caso binario, cruza en un solo punto, esto es, intercambio de subcadenas).
3. La selección suele realizarse de manera probabilística.

A continuación se presenta la estructura general del AG simple.

i. Generar la población inicial.

ii. Evaluar la población, a cada individuo de la población se le asigna un valor de aptitud (*fitness*).

iii. Selección de la población basada en la aptitud (*fitness*) de cada individuo.

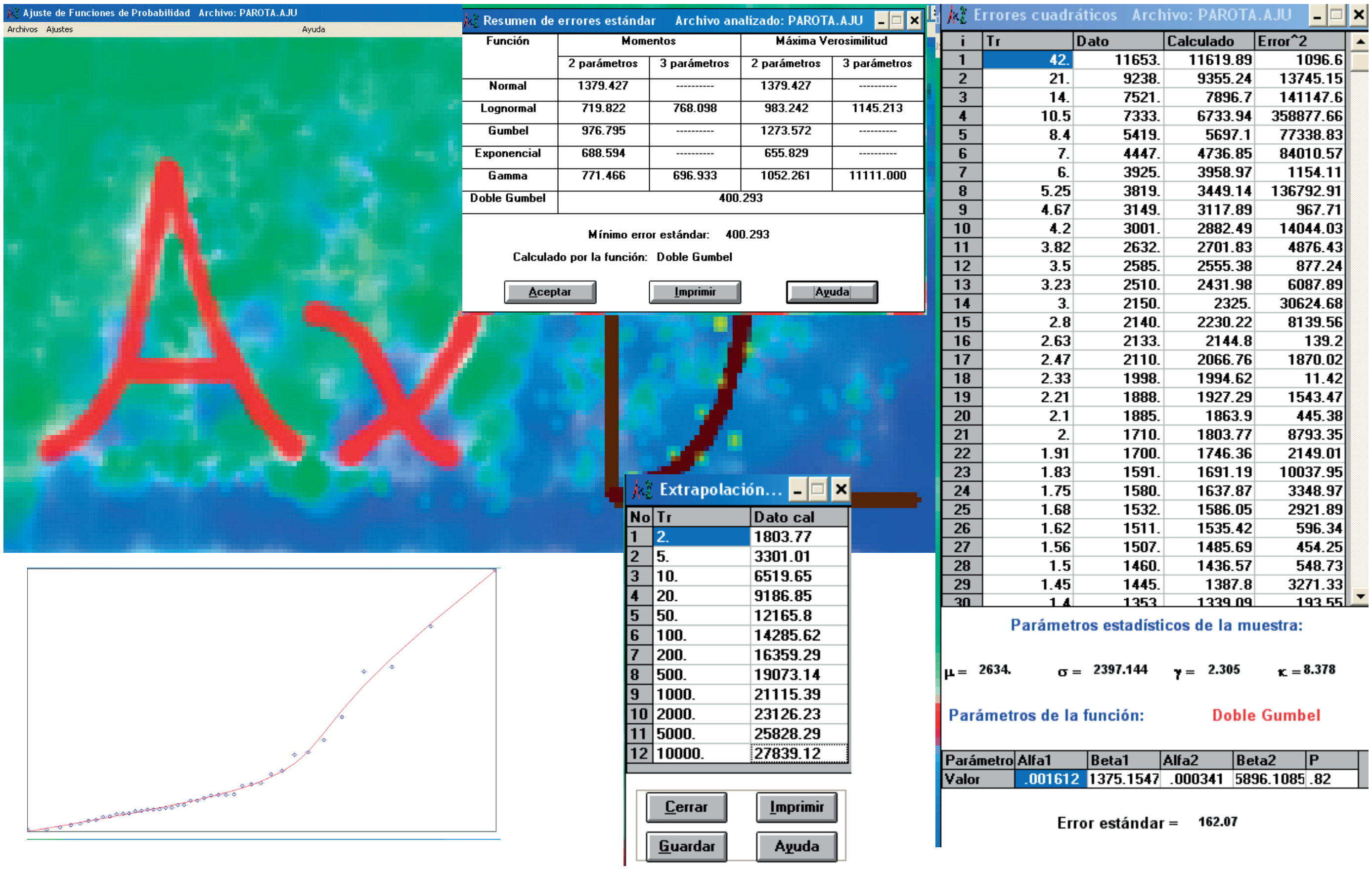


Figura 1 | Programa AX.

iv. Cruza de los individuos seleccionados (padres) para obtener una nueva población (hijos) cada pareja de padres genera una pareja de hijos.

v. Mutación en un porcentaje muy bajo de los individuos hijos.

vi. Los individuos hijos forman una nueva población sustituyendo por completo a la generación anterior (padres). Puede existir elitismo, esto ocurre cuando el mejor individuo de la generación de padres pasa intacto a la nueva generación (hijos).

vii. Evaluar la población.

viii. Se repiten los pasos de iii al vii hasta alcanzar un criterio de fin (comúnmente se trata de un determinando número de generaciones).

Estos algoritmos han sido empleados en diversas áreas, y se han obtenido resultados satisfactorios en problemas de optimización, búsqueda y aprendizaje de máquina. Actualmente los algoritmos genéticos tienen también amplia aplicación en hidrología; son algoritmos aleatorios que permiten estimar parámetros de un modelo dentro de un problema de optimización. El algoritmo requiere como datos de entrada el número de variables a determinar, en este caso, el número de parámetros de la función de distribución. Se define una población inicial o primera generación, con cierto número de individuos, donde cada individuo será un conjunto de valores de los parámetros de la función de distribución considerada; se evalúa el desempeño (*fitness*) de cada individuo con la función objetivo (en este caso la maximización de la función de verosimilitud L); se seleccionan los individuos con mejor desempeño usando el método de la ruleta o el método estocástico universal (Goldberg,1989). A los individuos seleccionados se les aplican los operadores de cruza y mutación generando una nueva población con la que se repite el proceso hasta llegar al número de generaciones previamente establecido. En esa última generación se selecciona el individuo de mejor desempeño (que será el conjunto de parámetros de la función de distribución analizada).

Dentro de las cajas de herramientas *Matlab* (The Mathworks, Inc. (1992)), se encuentra el *toolbox* sobre algoritmos genéticos, el cual contiene un programa correspondiente al algoritmo genético simple *sga.m* y las funciones de los operadores asociadas al mismo. Este programa se puede adaptar de manera sencilla a cualquier problema de optimización, con la creación de otro archivo que contenga a la función objetivo del problema analizado.

Función de densidad de probabilidad de la Distribución Gumbel

En el trabajo de Koutsoyiannis (Koutsoyiannis, 2003) se detalla que Gumbel (1958) desarrolló una teoría completa sobre distribuciones de valores extremos; de acuerdo con ella a medida que el número de datos de una serie de máximos tiende a infinito, la función de distribución General de Valores Extremos converge a una de tres posibles asíntotas, dependiendo de la forma matemática de la función de distribución *F(x)* que es la función de distribución de cada valor de la serie de máximos. Las tres asíntotas se pueden representar con la expresión matemática introducida por Jenkinson (1955,1969), conocida como la Distribución General de Valores Extremos (GEV): $F(x) = exp\left\{-\left[1+\frac{(x-\varepsilon)}{\lambda}k\right]^{-1/k}\right\}$, donde ε, λ y k son los parámetros de ubicación, escala y forma.

Cuando $k=0$, se tiene la distribución Tipo I, conocida como Gumbel:

$$F(x) = exp\{-exp[-(x-\varepsilon)/\lambda]\}$$

O

$$F(x;\alpha,\beta) = exp\{-exp[\alpha(x-\beta)]\} \tag{3}$$

Con $\alpha=-1/\lambda$ y $\beta=\varepsilon$; donde α y β son los parámetros de la función

Debido a su comportamiento, la función de distribución Gumbel es muy utilizada para modelar el comportamiento de los escurrimientos o las precipitaciones máximas anuales en muchas regiones.

Su función de densidad de probabilidades es de la forma:

$$f(x;a,b) = \alpha\, exp\{-\alpha[x-\beta] - exp[-\alpha(x-\beta)]\} \tag{4}$$

donde α y β son los parámetros de forma y escala de la función, respectivamente.

Distribución Gumbel de dos poblaciones

Al analizar el comportamiento de los escurrimientos en diversos ríos de la República Mexicana, se encontró que los gastos máximos anuales ocurrían generalmente en la época de verano, pero ocasionalmente se presentaban valores mucho mayores, ya sea en la época de invierno o por la incidencia directa de huracanes en la cuenca. Para modelar este fenómeno se desarrolló la función de distribución de probabilidades Gumbel de dos poblaciones (González, 1970) que desde entonces ha sido usada con éxito en diversos casos (Domínguez *et al.*, 2000, Domínguez *et al.*, 2006, Arganis *et al.*, 2009). Más adelante se observó un fenómeno semejante en otras partes, lo que dio lugar al desarrollo de las denominadas funciones de distribución mixtas (Rossi *et al.*, 1984).

En el caso de la función Gumbel de dos poblaciones, la función de densidad es de la forma:

$$f(x) = pf_1 + (1-p)f_2 \tag{5}$$

donde: $f_1 = \frac{df_1}{dx}$; $f_2 = \frac{df_2}{dx}$; siendo $F_1(x;\alpha_1,\beta_1) = exp\{-exp[-\alpha_1(x-\beta_1)]\}$ y $F_2(x;\alpha_2,\beta_2) = exp\{-exp[-\alpha_2(x-\beta_2)]\}$

En las expresiones anteriores es una función de distribución Gumbel con parámetros α_1 y β_1 a la que se ajustarían los valores máximos de los años en que no se presentan eventos ciclónicos; F_2 es la función de distribución tipo Gumbel con parámetros α_2 y β_2 a la que se ajustarían los valores máximos de los años en que se presentan eventos ciclónicos; f_1 y f_2 son las funciones de densidad de probabilidad correspondientes a F_1 y F_2; p es un parámetro que indica la proporción de datos que corresponden a la primera población (de eventos no ciclónicos), donde ($0 < p < 1$), por lo que $1-p$ corresponde a la proporción de los datos de la población de eventos ciclónicos.

APLICACIÓN

Funciones consideradas en el algoritmo genético y en el método determinístico

Distribución Gumbel

Al derivar la ecuación (1) con respecto a cada parámetro e igualar a cero, resultan expresiones como las indicadas a continuación:

$$e_1 = \sum_{i=1}^{n} x_i e^{-\widehat{\alpha} x_i} - \left(\overline{x} - \frac{1}{\widehat{\alpha}}\right)\sum_{i=1}^{n} e^{-\widehat{\alpha} x_i} \tag{6}$$

$$e_2 = \widehat{\beta} - \frac{1}{\widehat{\alpha}} ln\left[\frac{n}{\sum_{i=1}^{n} e^{-\alpha x}}\right] \tag{7}$$

donde n es el número de datos de la muestra; $\overline{x}$ es la media de la muestra; e_1 y e_2 son funciones que representan el error que puede cometer al tratar de igualar a cero las derivadas del logaritmo natural de L.

La función objetivo planteada en este caso fue:

$$FO_{1AG} = min\left(\sum_{j=1}^{np} e_j^2\right) \tag{8}$$

donde np es el número de parámetros de la función de densidad de probabilidad.

Las ecuaciones anteriores fueron resueltas con el algoritmo determinístico del programa AX que utiliza un procedimiento por aproximaciones sucesivas en el que se toman como valores iniciales a los resultados obtenidos por el método de momentos. Adicionalmente, ante la incertidumbre de que el uso de la primera derivada pueda llevar a un resultado que corresponda a un valor alejado del máximo global, las mismas ecuaciones fueron resueltas usando el algoritmo genético y considerando 2 casos: con un intervalo de búsqueda amplio o limitando dicho intervalo en el entorno de los resultados que se obtuvieron con el AX.

Por otra parte, se utilizó el algoritmo genético para determinar directamente los parámetros que hacen máxima de la función logaritmo natural de la función de verosimilitud (sin necesidad de usar derivadas), es decir:

$$FO_{2AG} = Máx\ [ln(L)] \tag{9}$$

o bien, los que hacen máxima la función de verosimilitud L directamente, sin tomar el logaritmo ni la derivada (ecuación 10). Las ventajas en el uso de esta última opción se destacan en este documento.

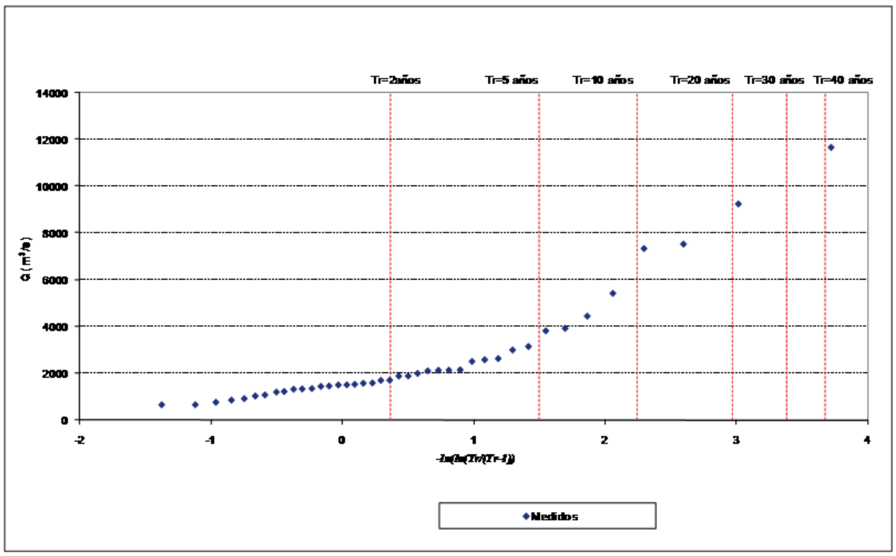
$$FO_{3AG} = Máx\ (L) \tag{10}$$

Función Gumbel de dos poblaciones

En el caso de la Función Gumbel de dos poblaciones el algoritmo determinístico AX emplea un método de optimización que toma como valor inicial al obtenido para dicha función con el método de momentos bajo la suposición de que el 80% de los datos caen en la primera población (es decir, para $p=0.8$). A partir de este valor inicial, se obtienen los parámetros de la distribución de probabilidades con un método de optimización con el que se minimiza el error estándar de ajuste (González, 1970).
Por otra parte, en el algoritmo genético se consideró como función objetivo la que maximiza directamente a la función de verosimilitud L (ecuación (10)), evitando con ello el uso de logaritmos y de la derivada.

Datos de entrada

En la Figura 2 se presentan los 41 datos de gastos máximos anuales utilizados, los cuales fueron obtenidos a partir de registros de la estación hidrométrica La Parota, ubicada en el río Papagayo del estado de Guerrero, México (Figura 3).



Año	Qm^3/s	Año	Qm^3/s
1962	1201	1983	1336
1963	2632	1984	9238
1964	1888	1985	1227
1965	921	1986	1579
1966	659	1987	1353
1967	11653	1988	2133
1968	767	1989	2585
1969	2150	1990	1445
1970	2510	1991	1885
1971	1034	1992	1507
1972	1591	1993	3149
1973	7333	1994	662
1974	7521	1995	1079
1975	1998	1996	3001
1976	5419	1997	3925
1977	1532	1998	1710
1978	1460	1999	1511
1979	2140	2000	2110
1980	1760	2001	859
1981	4447	2002	3819
1982	1324		

Figura 2 | Gastos máximos anuales. Estación hidrométrica La Parota, Gro. México.

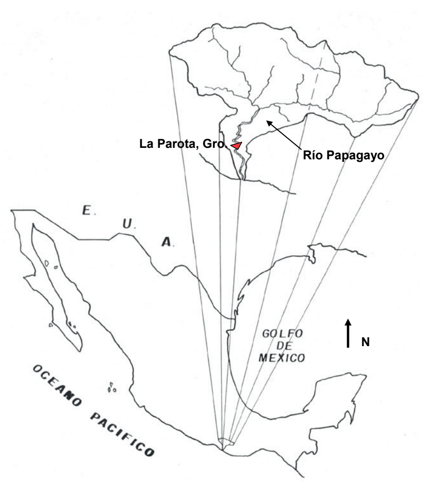


Figura 3 | Gastos Cuenca del río Papagayo, Gro. México.

Uso del programa AX y del AG

Se aplicó el software AX (Jiménez, 1996) y adicionalmente un algoritmo genético (AG) programado en Matlab (The MathWorks, 1992) en el que se codificaron las distintas funciones objetivos propuestas y un archivo de texto que contiene los datos registrados. En las aplicaciones del algoritmo genético se utilizaron poblaciones de 200 individuos y 5000 generaciones, además de considerar una probabilidad de mutación de *0.7/longitud del cromosoma* y en el operador selección el método estocástico universal (Baker, 1985).
Una vez estimados los parámetros para las distintas funciones de distribución, tanto con el algoritmo genético como con el AX, se graficaron los valores medidos y calculados para diferentes periodos de retorno.

RESULTADOS

Función Gumbel

Para el caso de la función Gumbel se estimaron sus parámetros usando distintos procedimientos:

Con el algoritmo determinístico AX, resolviendo la ecuación 8 partiendo de condiciones iniciales dadas por el método de momentos.

Mediante el algoritmo genético (AG), para resolver la misma ecuación 8 usando un intervalo amplio de búsqueda.

Usando el (AG) para resolver la misma ecuación 8, limitando el intervalo de búsqueda.

Usando el (AG) para resolver la ecuación 9 con un intervalo de búsqueda amplio.

Usando el (AG) para resolver la ecuación 10 con un intervalo de búsqueda amplio.

Los resultados se presentan en la Tabla 1, en la que se muestran los valores de los parámetros α y β, los valores de e_1 y e_2 (ecuaciones (6) y (7)), la suma de sus cuadrados (ecuación (8)), el valor de la función de verosimilitud L (ecuación (1)) y el error estándar de ajuste (EEA) definido como: $EEA = \sqrt{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 / (np - 1)}$; donde: x_i dato medido, $\hat{x}_i$ valor calculado, np número de parámetros de la función de distribución de probabilidades.

Tabla 1 | Resultados de la aplicación del algoritmo determinístico AX y el algoritmo genético (AG).Función Gumbel.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Método	α	β	e_1	e_2	$e_1^2+e_2^2$	L	EEA
AX	0.0008	1751.9343	343.68	-11.719	118254.5871	6.63533465E-158	14.8980
AG Ec. 8 a)	0.883311	663.12686	-3.3209E-250	3.854E-07	1.48524E-13	0	44.9390
AG Ec. 8 b)	0.0008177	1756.09670	0.00017591	3.311E-05	3.20417E-08	6.73015327E-158	14.7405
AG Ec. 9 c)	0.00081770	1756.10010	0.15016	2.376E-05	0.022548118	6.7301533E-158	14.7402
AG Ec. 10 d)	0.00081770	1756.10010	0.15016	2.376E-05	0.022548118	6.73015325E-158	14.7402

Notas: a) Se utilizó un intervalo de búsqueda amplio, b) Se disminuyó el intervalo de búsqueda de los parámetros, al considerar los resultados previos del AX, c) y d) Se utilizó un intervalo de búsqueda amplio.

En la Figura 4 se ilustra la comparación entre los valores históricos y calculados con la función Gumbel.

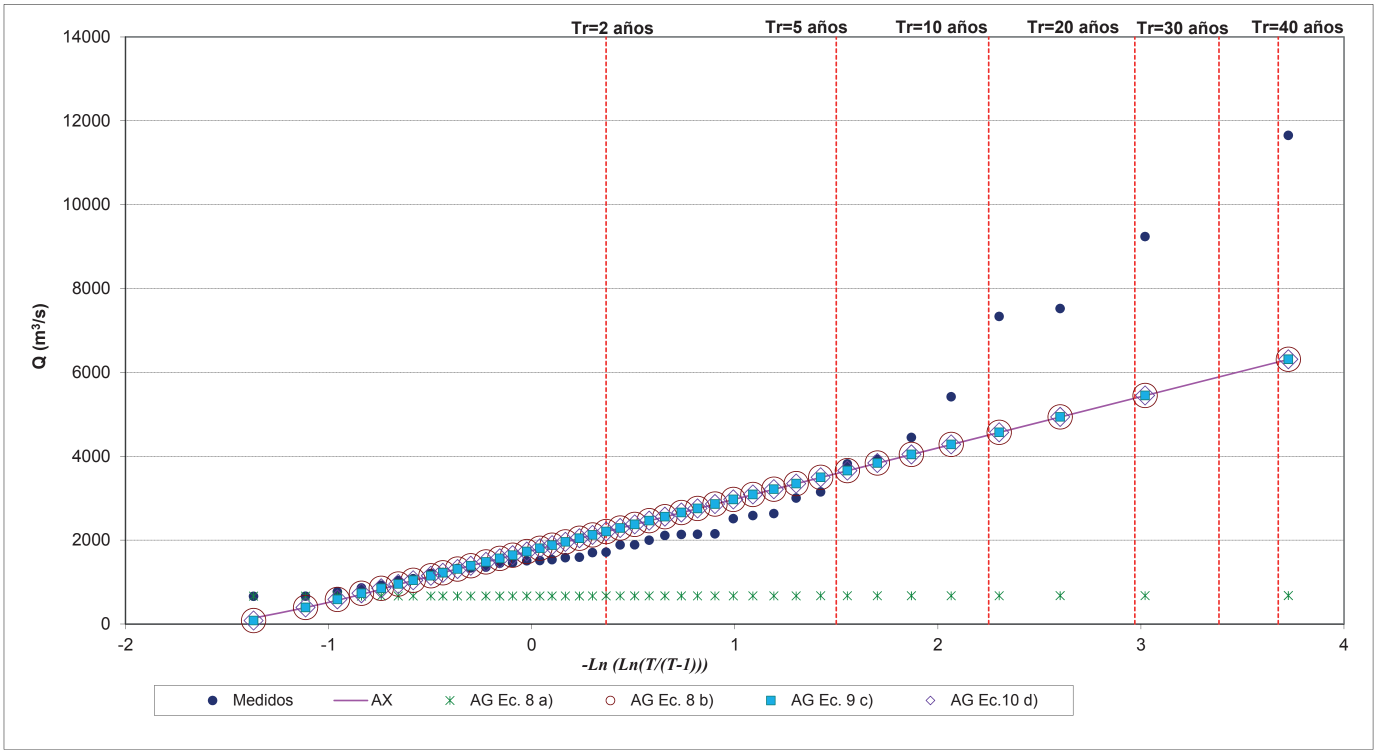


Figura 4 | Comparación entre valores históricos y calculados. Función Gumbel.

Función Gumbel de dos poblaciones

En la Tabla 2 se presentan los resultados obtenidos al estimar los parámetros de la función Gumbel de dos poblaciones, usando el AX y utilizando el AG. También se presentan los valores del error estándar de ajuste (EEA) y de la función de verosimilitud L (ecuación (1)).

Tabla 2 | Resultados de la aplicación del algoritmo determinístico AX y el algoritmo genético (AG). Función Gumbel de dos poblaciones.

Método	β_1	α_1	β_2	α_2	p	EEA	L
AX	1375.1547	0.001612	5896.1085	0.000341	0.82	162.0703	3.4E-153
AG Ec. 10 a)	1355.7867	0.001825	4668.9819	0.0003925	0.7853608	394.0569	1.17E-151
AG Ec. 10 b)	1355.9679	0.0018245	4672.2922	0.0003926	0.7855266	396.2322	1.24E-151

Notas: a) Se usó un intervalo de búsqueda amplio, b) Se disminuyó el intervalo de búsqueda para todos los parámetros, respecto al caso a).

En la Figura 5 se presenta la comparación entre los valores históricos y calculados con la función Gumbel de dos poblaciones.

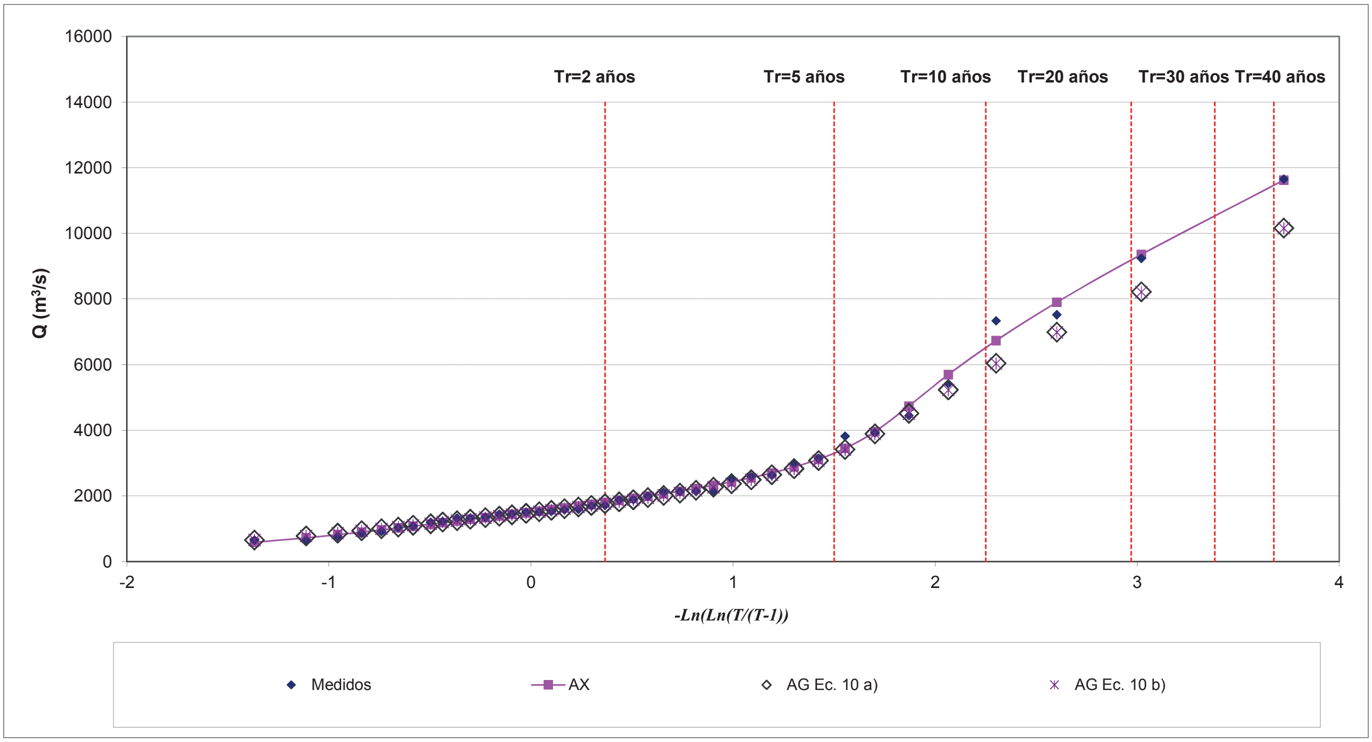


Figura 5 | Comparación entre valores históricos y calculados. Función Gumbel de dos poblaciones.

DISCUSIÓN

En la Tabla 1 se aprecia que:

1. Cuando en el AG se usa la ecuación 8 con un intervalo amplio de búsqueda (caso a), se pueden obtener valores de los parámetros α y β que hacen que las derivadas prácticamente sean cero (columnas 4 y 5), pero para los que la función L tiende a un valor alejado del máximo global y el error estándar de ajuste aumenta sensiblemente, de tal forma que los parámetros obtenidos llevan a una forma de la función de distribución Gumbel que no reproduce adecuadamente el comportamiento de los datos históricos, como se muestra en la Figura 4.

2. Cuando se reduce el intervalo de búsqueda (caso b) los valores que se obtienen son muy similares a los que da el software AX. Los valores de la derivada de la función se aproximan a cero y se logra obtener un valor de la función L similar al que reporta el AX. También se corrobora que con los parámetros obtenidos se consigue un error estándar de ajuste ligeramente menor al proporcionado por el algoritmo del software AX, por lo que la distribución Gumbel con estos parámetros se apega mejor a los datos históricos.

3. Finalmente, al emplear el AG para resolver la ecuación (9) que usa el logaritmo natural de L, (caso c) o bien la ecuación (10) para maximizar directamente la función de verosimilitud L (caso d), se logra obtener tanto el máximo global para L, como el menor error estándar de ajuste.

Por otro lado, al comparar los valores que toma la función de verosimilitud *L* al usar las derivadas del logaritmo de dicha función, se concluye que el algoritmo determinístico que utiliza el AX sí logró obtener el máximo valor de *L*, pero pudo haber llegado a un valor mínimo, dependiendo de los valores iniciales de búsqueda que se le proporcionen.

La Tabla 2 muestra que, para la serie analizada, en lo que se refiere al error estándar de ajuste, el AX permite estimar mejores valores de los parámetros que los obtenidos por el algoritmo genético. Lo anterior se corrobora en la Figura 5, en la que se observa la correspondencia entre los datos históricos y los estimados con las dos funciones de distribución. En contraposición, al emplear el algoritmo genético para optimizar la función objetivo dada por la ecuación (10), se obtienen valores de la función *L* mayores que los proporcionados por el AX; es decir, se logra maximizar a *L*. Esta aparente contradicción se debe esencialmente a que el valor del error estándar de ajuste depende de la forma en que se estiman las probabilidades de no excedencia correspondientes a los datos ordenados de mayor a menor (en este caso se utilizó la fórmula de Weibull).

Por otra parte, cuando se disminuye el intervalo de búsqueda de todos los parámetros (caso b) se obtienen resultados idénticos que los del caso 10.

CONCLUSIONES

El ajuste adecuado de los parámetros de una función de distribución de probabilidad es indispensable para poder hacer inferencias probabilísticas a partir de los datos de una muestra. En particular, en el caso de los gastos máximos anuales registrados en un sitio, conduce a la estimación de los valores asociados a distintos periodos de retorno necesarios para el diseño de las obras destinadas al control de las crecientes.

La estimación de los parámetros mediante el método de máxima verosimilitud conduce a encontrar la función de distribución de probabilidad de la que más probablemente provienen los datos de la muestra; sin embargo, es común que en las aplicaciones prácticas se recurra a igualar a cero la derivada de los logaritmos de dicha función para estimar su máximo, lo que, como se muestra en este trabajo puede llevar a resultados incorrectos.

En este trabajo se presenta un estudio comparativo entre los resultados obtenidos con un software que emplea algoritmos determinísticos basados en el uso de la primera derivada de sus logaritmos para la estimación del valor máximo de la función de verosimilitud y los que se obtienen usando los algoritmos genéticos para encontrar los parámetros que corresponden al óptimo de distintas funciones objetivo. Se utilizó para ello la muestra de 41 gastos máximos anuales registrados en la estación La Parota, en el estado de Guerrero, México a la que se ajustaron las funciones Gumbel y Gumbel de dos poblaciones.

Con el análisis realizado se pudo corroborar que la sola derivación e igualación a cero de una función para la búsqueda de su valor máximo puede llevar a resultados incorrectos, y que este problema puede subsanarse maximizando directamente la función de verosimilitud o su logaritmo, con ayuda de los algoritmos genéticos.

En el caso de la distribución de probabilidades Gumbel, al maximizar directamente a la función de verosimilitud L con un algoritmo genético simple, en lugar de los métodos determinísticos, se consiguió un mejor ajuste de los parámetros, como se muestra en la Tabla 1 y en la Figura 4.

En el caso de la función Gumbel de dos poblaciones el software AX condujo a errores estándar menores que los obtenidos con el algoritmo genético, pero el empleo de este algoritmo condujo a valores mayores de la función de verosimilitud.

Los resultados obtenidos mostraron que al aplicar un algoritmo genético es posible maximizar directamente la función de verosimilitud a partir de la función de densidad de probabilidades de una variable aleatoria, sin necesidad de usar las derivadas de su logaritmo, y correr el riesgo de obtener valores que no corresponden al máximo de la función de verosimilitud.
Los algoritmos genéticos pueden ser empleados para maximizar la función de verosimilitud o su logaritmo considerando que, aunque en teoría en los dos casos debe llegarse a los mismos parámetros de la función de distribución, la precisión numérica puede hacer que alguno de ellos dé mejores resultados.

AGRADECIMIENTOS

Se agradece al Dr. Martín Jiménez Espinoza por facilitar el uso del software AX

REFERENCIAS

Arganis-Juárez, M.L., Domínguez-Mora, R., González-Villarreal, F., Carrizosa-Elizondo, E., Esquivel-Garduño, G., Hollands, A.J., Ramírez-Salazar, L.E. (2009). *Estudio Integral de la Cuenca Alta del Río Grjialva. Actualización de Avenidas de Diseño*. Para CFE. Informe Final.

Baker, J.E. (1985). Adaptive Search Selection Methods for Genetic Algorithms, *in Proceedings of the First International Conference on Genetic Algorithms (Grefenstette, ed)*, Lawrence Erlbaum, 101-111.

Clark, C., Whu, Y.Z. (2006). Integrated hydraulic model and genetic algorithm optimization for informed analysis of a real water system. *Asce 8th Annual International Symposium On Water Distribution System Analysis,* Cincinnati, August 27-30, Ohio.

Domínguez-Mora, R., Carrizosa-Elizondo, E., Fuentes-Mariles, G.E., Arganis-Juárez, M.L. (2000). *Estudio de diferentes aspectos sobre el funcionamiento de la obra de excedencias del Proyecto Hidroeléctrico, la Angostura, Chiapas y actualización de la hidrología para el sistema de presas del Río Grijalva. "Estudio Hidrológico de la Cuenca alta del Río Grijalva"*. Para CFE. Informe final.

Domínguez-Mora, R., Fuentes-Mariles, G.E., Arganis-Juárez, M.L. (2004). Optimación de los parámetros de la función de distribución doble gumbel usando algoritmos genéticos en una serie de gastos máximos anuales. *XXI Congreso Latinoamericano de Hidráulica*, Sao Paulo, Brasil.

Domínguez-Mora, R., Arganis-Juárez, M.L., Carrizosa-Elizondo, E., Fuentes-Mariles, G.E., Echeverri, C.A. (2006). *Determinación de Avenidas de Diseño y Ajuste de los Parámetros del Modelo de Optimización de las Políticas de Operación del Sistema de Presas del Río Grijalva*. Para CFE. Informe Final.

Escalante-Sandoval, C., Reyes-Chávez, L. (2002). *Técnicas Estadísticas en Hidrología*. Facultad de Ingeniería. Universidad Nacional Autónoma de México.

Fuentes-Mariles, O.A., Fuentes-Mariles, G.E., Domínguez-Mora, R. (2005). Optimación de los parámetros de algunas funciones de distribución de probabilidad de gastos máximos anuales usando un algoritmo genético simple. *4a. Conferencia Iberoamericana en Sistemas Cibernética e Informática*, Cicsi, Orlando, Flo., Usa, Vol. 2, 156-159.

Fuentes-Mariles, O.A. Domínguez-Mora, R., Fuentes-Mariles, G.E., Arganis-Juárez, M.L., Rodríguez-Vázquez, K. (2006). Estimación de los parámetros de funciones de distribución empleadas en hidrología usando ecuaciones de máxima verosimilitud y algoritmos genéticos. *XXII Congreso Latinoamericano De Hidráulica*, Ciudad Guayana, Venezuela.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, USA.

González-Villarreal, F. (1970). *Contribución al análisis de frecuencias de valores extremos de los gastos máximos en un río*. Serie Azul, Instituto de Ingeniería, UNAM.

Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York. (citado por Koutsoyiannis, D., 2003)

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. The University of Michigan Press.

Horbelt, W., Timmer, J., Voss, H.U. (2002). Parameter estimation in nonlinear delayed feedback systems from noisy data. *Physics Letters A*. 299(5-6): 513–521. doi:10.1016/S0375-9601(02)00748-X

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) value of meteorological elements, *Quarterly Journal of the Royal Meteorological Society* 81, 158-171. (citado por Koutsoyiannis, D.,2003)

Jenkinson, A.F. (1969). *Estimation of maximum floods*, World Meteorological Organization, Technical Note No. 98, ch. 5, 183-257. (citado por Koutsoyiannis, D.,2003)

Jha, M.K., Nanda G., Samuel, M.P. (2004). Determining hydraulic characteristics of production wells using genetic algorithm. *Water Resources Management*, 18(4): 353–377. doi:10.1023/B:WARM.0000048485.62254.1c

Jiménez-Espinoza. M. (1996). *Programa Ax*. Área De Riesgos Hidrometeorológicos. Centro Nacional de Prevención de Desastres. México.

Kite, G.W. (1988). *Frequency And Risk Analyses In Hidrology*. Littletown, Colorado.USA.

Koutsoyiannis, D. (2003). On the appropriateness of the gumbel distribution in modelling extreme rainfall. Hydrological Risk: recent advances in peak river flow modelling, prediction and real-time forecasting. Assessment of the impacts of land-use and climate changes. *Proceedings of the ESF LESC Exploratory Workshop held at Bologna*, Italy, October 24-25, 303-319.

Liu, Y., Khu, S.T., Savic D. (2004). A Hybrid Optimization Method Of Multi-Objective Genetic Algorithm (Moga) And K-Nearest Neighbor (KNN) Classifier for Hydrological Model Calibration. *Lecture Notes In Computer Sciences*, Volume 3177, 546-551. doi:10.1007/978-3-540-28651-6_80

Mazariegos, B.R., Raynal-V., J.A. (2002). Paquete Interactivo Para La Estimación De Parámetros De La Distribución Weibull, B14. *Memorias Del XX Congreso Latinoamericano De Hidráulica*, La Habana, Cuba.

Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology,* 47(1): 90–100, doi:10.1016/S0022-2496(02)00028-7.

Nicklow, J.W., Ozkurt O., Bringer Jr, J.A. (2003). Control of Channel Bed Morphology in Large-Scale River Networks using a Genetic Algorithm*, Water Resources Management,* 17(2): 113–132. doi:10.1023/A:1023609806431

6O-Matrix Statistical Time Series Analysis. Stsa Toolbox Version 2. (2005). *The Time Series Analysis Toolbox For O-Matrix*, http://www.omatrix.com/Stsav2.html

Rao, A.R., Hamed, K.H. (2000). *Flood Frequency Analysis*. Crc Press, USA, Web Site: Google.Books.Com

Rossi, F., Florentino, M., Versace, P. (1984). Two-Component Extreme Value Distribution for Flood Frequency Analysis, *Water Resources Research* 20(7), 847-856. doi:10.1029/WR020i007p00847

Smith, R.L. (1988). Forecasting Records By Maximum Likelihood. *Journal Of The American Statistical Association*, 83(402): 331-338. doi:10.2307/2288847.

The Mathworks, Inc. (1992). *The Mathworks Matlab Reference Guide*.