

## Algunos estudios estadísticos útiles en hidrología

### ALGUNOS ESTUDIOS ESTADISTICOS EN HIDROLOGIA

Son brevemente señalados algunos de los problemas asociados con la determinación de la distribución para modelar caudales máximos, entre ellos se señalan: Los límites de los estadísticos, el sesgo de éstos, así como la condición de separación.

#### ABSTRACT

Some of the problems related to the choice of the flood flow distribution are briefly outlined, such as the boundness and bias of the statistics, as well as the condition of separation.

**CARLOS A. GONZALEZ M.**  
Ing. Agrícola – M.Sc. en Hidrología  
Instructor Asociado  
U.N.

### 1. GENERAL

El análisis de caudales máximos o mínimos trata de escoger un modelo el cual tiene como objetivo final, la representación de una serie de extremos o picos del hidrógrafo.

Usualmente en hidrología se dispone de una muestra de datos (Ya sea caudales máximos o mínimos) la cual se supone es una muestra aleatoria e idénticamente distribuida generada de una distribución acerca de la cual se desean hacer inferencias.

Las muestras de observaciones son supuestamente generadas de una distribución cuya descripción depende de los valores de ciertos parámetros; entre mayor sea el número de ellos, más flexible se torna la distribución.

Las distribuciones de dos parámetros por ejemplo no poseen el parámetro de forma y por lo tanto éstas sólo tienen una conformación específica. La distribución Pearson tipo 3 (PT3), Weibull, La distribución general de Valores Extremos y otras muy comunes en hidrología adoptan diferentes conformaciones de acuerdo a su tercer parámetro de forma.

Las distribuciones más comunes usadas en hidrología son las de dos y tres parámetros, ello es debido a los diferentes problemas que se presentan en la estimación de estos cuando su número se incrementa. Algunas de las dificultades se enumeran más adelante cuando el método de estimación de parámetros más utilizado, (Método de momentos), es usado para su determinación.

Esta es una razón bastante poderosa del por qué los hidrólogos a pesar de percibir la necesidad de incluir distribuciones de más de tres parámetros, reconocen el hecho de la introducción de mayores errores en el uso de momentos mayores que el tercero.

## 2. COMPORTAMIENTO DE LAS MUESTRAS CON RESPECTO A LOS ESTADISTICOS

Wallis et al., [1974], condujeron un estudio de simulación tendiente a examinar el comportamiento aleatorio de las muestras con respecto a los estadísticos. Ellos obtuvieron la función de distribución para tres estadísticos, como lo son el promedio aritmético ( $\bar{X}$ ), Desviación estándar (Sd) y coeficiente de asimetría (Cs) para muestras de tamaño pequeño (De 10 en 10 hasta 90), las cuales fueron derivadas de las distribuciones Normal, Gumbel, Logarítmica Normal, Pearson tipo 3, Weibull y Pareto tipo 1 (también conocida como distribución Pearson tipo 4).

Este estudio mostró un coeficiente de asimetría fuertemente marcado de los estadísticos así también como sesgamiento y límites en la función de distribución de éstos.

De acuerdo al citado estudio, la función de distribu-

ción de los estadísticos depende del tamaño de la muestra, del coeficiente de asimetría de la población ( $\delta$ ) y de la naturaleza de la distribución de la población que genera el proceso.

Fue observado que  $\bar{X}$  es un estimador insesgado de ( $\mu$ ), mientras que la Desviación estándar (Sd) y el coeficiente de asimetría (Cs) son estimadores sesgados de ( $\sigma$ ) y ( $\delta$ ) respectivamente. ( $\sigma$ : Desviación estándar de la Población y  $\delta$ : Coeficiente de asimetría de la población).

Kirby [1974], ha mostrado que la función de distribución para valores positivos del coeficiente de Variación (Como lo son los valores utilizados en hidrología; ej: Caudales y Precipitaciones), el coeficiente de asimetría, la Desviación máxima estandarizada  $\{(X \text{ máximo} - \bar{X}) / Sd\}$  y el rango  $\{X_{\text{máximo}} - X_{\text{mínimo}}\}$ , son limitados. La tabla 1 muestra los límites citados.

### LIMITES ALGEBRAICOS PARA LOS ESTADISTICOS EXTRAIDOS DE MUESTRAS DE TAMAÑO N. (Tomado de Kirby (1974)).

Estadístico	Fórmula Empleada	Límite superior	Límite inferior
Coeficiente de Variación (Cv).	$\frac{Sd}{\bar{X}}$	$\frac{1/2}{(N-1)}$	0
Coeficiente de Asimetría (Cs)	$\frac{1/N \sum_{i=1}^N (X_i - \bar{X})^3}{(1/N \sum_{i=1}^N (X_i - \bar{X})^2)^{3/2}}$	$\frac{(N-2)}{(N-1)^{1/2}}$	$\frac{-(N-2)}{(N-1)^{1/2}}$
Desviación Max. estandarizada	$(X_{\text{max}} - \bar{X}) / Sd$	$(N-1)^{1/2}$	$(N-1)^{-1/2}$
Rango	$(X_{\text{max}} - X_{\text{min}}) / Sd$	$(2N)^{1/2}$	2 (N es par)

Estos límites son llamados límites algebraicos para enfatizar el hecho que ellos son inherentes a las fórmulas algebraicas que definen los estadísticos, y son totalmente independientes de cualquier suposición acerca del proceso que genera la muestra.

Como puede observarse en la columna 2 de la tabla las fórmulas usadas para computar los estadísticos no tuvieron corrección por sesgamiento.

### 3. CONDICION DE SEPARACION

Matalas et al., [1975], determinaron para diferentes regiones de los Estados Unidos, los valores de la Desviación estándar y el promedio del  $C_s$ , de series hidrológicas; graficándolos en los ejes Y y X respectivamente. Estos valores graficados caen por encima de la misma relación derivada de un estudio de simulación llevado a cabo con datos generados por las siguientes distribuciones: Normal, Uniforme, Gumbel, Logarítmica Normal, Pearson tipo 3, Weibull y Pareto.

Ellos llamaron este fenómeno la **condición de separación** y de acuerdo a sus resultados esto podría sugerir que ninguna de las distribuciones hipotetizadas es una buena aproximación a la distribución de caudales máximos.

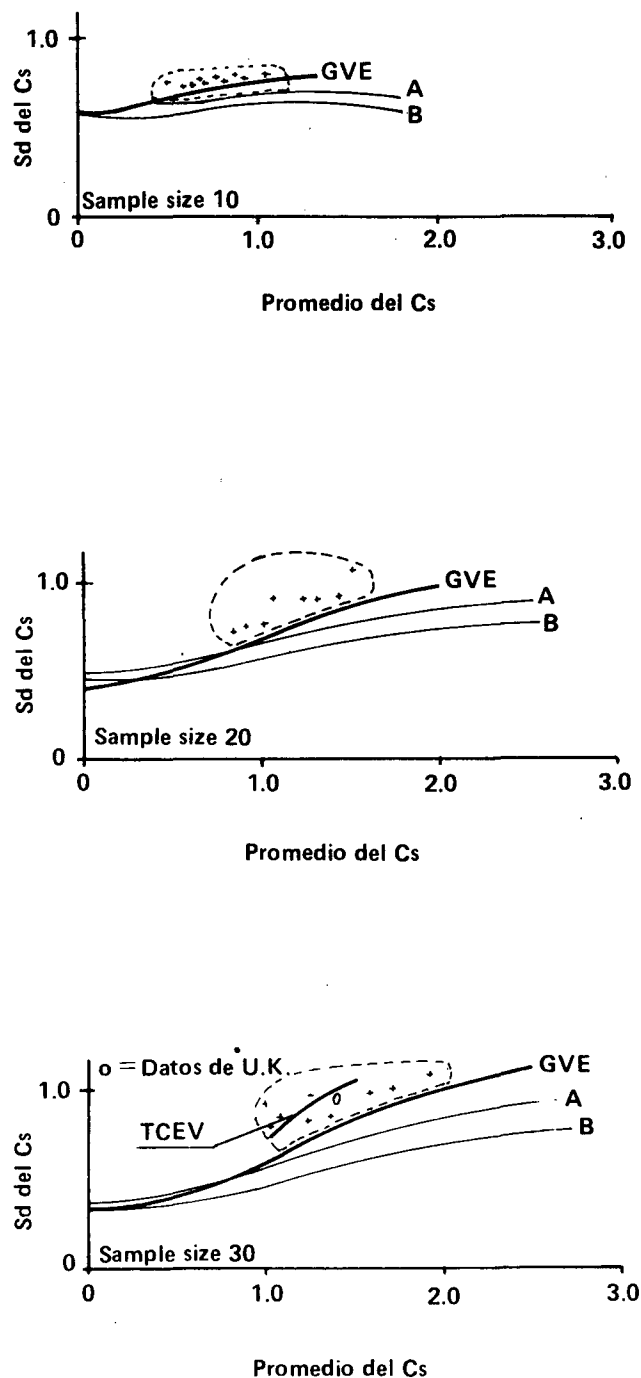
La figura 1 de Cunnane [1986], muestra estos resultados.

Subsecuentemente, otras distribuciones que no fueron incluidas en este estudio fueron examinadas tales como la distribución Logarítmica Pearson tipo 3, (Landwehr et al., [1978]), la distribución General de Valores Extremos (GVE), (Cunnane [1984]); fallando también en explicar la condición de separación.

Las únicas distribuciones que parecen satisfacer la condición de separación son las distribuciones Wakeby pero solo para ciertos valores de sus parámetros, (Houghton, [1978]) y la distribución de Valores Extremos de 2 componentes (TCEV) probada para series hidrológicas de caudales máximos de Italia (Rossi et al. [1984]) y Gran Bretaña (Beran et al. [1986]).

En el contexto de caudales mínimos, King [1985], trabajando con datos de Irlanda encontró que ninguna de las distribuciones usadas por Matalas et al., [1975], explican la condición de separación para caudales mínimos, este estudio también mostró que la Desviación estándar del coeficiente de asimetría graficados contra el promedio del coeficiente de asimetría caen por debajo del límite inferior encontrado en el estudio de Matalas et al., [1975]; estos resultados están en contraste con los datos de caudales máximos que en el gráfico se observan por encima del límite superior encontrado en el estudio.

De nuevo esto podría sugerir que estas distribuciones son inapropiadas para modelar caudales mínimos.



**Figura 1:**

Ilustración de la condición de separación. Desviación estándar del  $C_s$  para muestras de tamaño 10, 20 y 30 generadas de la distribución GVE (Cunnane, 1984). Curvas A y B encierran los resultados de todas las distribuciones estudiadas por Matalas et al., [1975]. Los resultados de la distribución TCEV y datos del Reino Unido (U.K) son tomados de Beran et al., [1986]. (Tomado de Cunnane, 1986).

#### 4. CORRECCION POR SESGAMIENTO EN EL COEFICIENTE DE ASIMETRIA

No hay duda que un problema que todavía subsiste es encontrar un adecuado estimador del coeficiente de asimetría para las series comúnmente encontradas en hidrología. El estimador de ( $\delta$ ) es muy variable debido primordialmente a la longitud de la muestra que usualmente se tiene a disposición.

En la aplicación de análisis de frecuencia el uso de un estimador insesgado de ( $\delta$ ), se torna muy importante debido al papel que desempeña en la estimación de los parámetros de la distribución escogida como la que genera las series hidrológicas a analizar; sin embargo, ha sido encontrado que la mayoría de las correcciones recomendadas hasta ahora producen un estimador sesgado de ( $\delta$ ).

El estimador sesgado de ( $\delta$ ) para una muestra de tamaño N es definido como:

$$C_s = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^3}{\left\{ \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 \right\}^{3/2}} \quad 1.1$$

El estimador del coeficiente de asimetría dado por la ecuación 1.1 resulta sesgado inferiormente en la estimación de  $\delta$  y este sesgamiento depende del tamaño de la muestra, la distribución que genera las muestras aleatorias y su coeficiente de asimetría, Wallis et al., [1974].

El USA Water Resources Council (WRC) en 1967 recomendó que el estimador del coeficiente de asimetría debería ser corregido por el factor Cf dado como:

$$C_f = \frac{[N(N-1)]^{1/2}}{N-1} \quad 1.2$$

Donde N = Tamaño de la muestra

Este factor fue inicialmente propuesto por Fisher [1929], para remover el sesgo en la estimación del coeficiente de asimetría cuando las muestras fueran provenientes de una distribución Normal. Cf es una combinación de correcciones para el segundo y tercer momento central como se demuestra a continuación:

$$C_{f1} = \frac{N}{N-1} \quad 1.3$$

$$C_{f2} = \frac{N^2}{(N-2)(N-1)} \quad 1.4$$

$$C_f = \frac{C_{f2}}{C_{f1}^{3/2}} \quad 1.5$$

Donde Cf1 = Corrección para el segundo momento central

Cf2 = Corrección para el tercer momento central

Cf = Factor recomendado por el WRC.

El USA Inter-Agency committee on Water Resources (IWRC) en 1966, sugirió que el factor arriba señalado debiera ser multiplicado por el factor dado por Hazen [1930]:

$$H_f = \left[ 1 + \frac{8.5}{N} \right] \quad 1.6$$

Donde Hf = Factor de Hazen

El factor (Hf x Cf) fue probado en el estudio de Wallis et al., [1974], y sus resultados mostraron que el uso de este factor resulta en un estimador aproximadamente insesgado de  $\delta$ , para un pequeño rango del coeficiente de asimetría de la población ( $1/2 < \delta < 2.0$ ); cuando la distribución que genera las muestras es la Logarítmica Normal. Para otras distribuciones y valores de los resultados mostraron que el uso de este factor resulta en un estimador sesgado inferiormente de  $\delta$ .

Bobee and Robitaille [1975], derivaron un factor de corrección para la distribución Pearson tipo 3 (PT3), usando los resultados obtenidos por Wallis et al., [1974], para valores del coeficiente de asimetría entre 0.25 y 5.0 y un tamaño de la muestra entre 25 y 90. El factor dado es el siguiente.

$$C_s = \bar{C}_s \left[ \left( 1 + \frac{6.51}{N} + \frac{20.2}{N^2} \right) + \left( \frac{1.48}{N} + \frac{6.77}{N^2} \right) \bar{C}_s^2 \right]$$

Donde  $C_s$  = Coeficiente de asimetría de la muestra  
 $\bar{C}_s$  = Coeficiente de asimetría de la muestra sin corrección por sesgamiento.

Esta corrección mejoró significativamente las correcciones recomendadas por el WRC e IWRC.

En el mismo estudio otras fórmulas fueron derivadas para corregir por sesgamiento el  $C_s$ , cuando las distribuciones son la Logarítmica Normal y la Weibull (ambas de tres parámetros), para el mismo rango del coeficiente de asimetría y longitud de la muestra. Estas fórmulas son:

Distribución Logarítmica Normal

$$C_s = \bar{C}_s \left[ \left( 1.01 + \frac{7.01}{N} + \frac{14.66}{N^2} \right) + \left( \frac{1.69}{N} + \frac{74.66}{N^2} \right) \bar{C}_s^3 \right]$$

## Distribución Weibull

$$Cs = \overline{Cs} \left[ \left( 1.01 + \frac{5.05}{N} + \frac{20.13}{N^2} \right) + \left( \frac{0.69}{N} + \frac{27.15}{N^2} \right) \overline{Cs}^3 \right]$$

1.9

Yevjevich and Obeysekera [1984], compararon las correcciones desarrolladas por Bobee and Robitaille [1975] y Hazen [1930], con un nuevo estimador el cual emplea la medida de dependencia entre el promedio y la varianza de las muestras en el caso de distribuciones asimétricas examinando la robustez en la estimación del coeficiente de asimetría.

Las distribuciones de las poblaciones que generaron las muestras fueron la PT3 y la distribución logarítmica Normal. Sus resultados mostraron que este nuevo estimador es más robusto que las correcciones arriba descritas, especialmente cuando el tamaño de las muestras es pequeño y el  $\delta$  es grande.

Finalmente, es importante enfatizar el hecho que un adecuado conocimiento de la distribución que genera las muestras tiene una fuerte implicación en la estimación del coeficiente de asimetría, Wallis et al., [1974], Yevjevich y Obeysekera [1984].

## BIBLIOGRAFIA

1. BERAN, M.A., HOSKING, J.R.M. and ARNELL, N.: Comment on "Two Component Extreme Value Distribution for Flood Frequency Analysis" by Rossi et al. (1984), *Water Resour. Res.*, 22(2), 263-266, 1986.
2. BOBEE, B. AND ROBITAILLE, R.: Correction of Bias in the Estimation of the Coefficient of Skewness, *Water Resour. Res.*, 11(6), 851-854, 1975.
3. CUNNANE, C.: Condition of Separation of Skewness of Random Samples from the General Extreme Value Distribution, Unpub. Rep., Dept. Eng. Hydrol., Univ. Coll. Galway, 7 pag., 1984.
4. CUNNANE, C., Stochastic Hydrology Section (A) and (B) for M.Sc Course, Univ. Coll. Galway, Ireland, 1985.
5. CUNNANE, C.: Review of Statistical Models for Flood Frequency Estimation, Keynote Address Presented at the International Symposium on Flood Frequency and Risk Analysis, U.S.A., 1986.
6. GREIS, N.P.: Flood Frequency Analysis: A Review of 1979-1982, *Reviews of Geophysics and Space Physics*, U.S., Vol 21(3), 699-706, April 1983.
7. HAAN, T.C.: *Statistical Methods in Hydrology*, Iowa State University Press, Ames, 1977.
8. HAZEN, A.: *Flood Flows*, John Wiley, N.Y., 1930.
9. HOUGHTON, J.C.: Birth of a Parent: The Wakeby Distribution for Modelling Flood Flows, *Water Resour. Res.*, 14(4), 1105-1109, 1978.
10. Interagency Committee on Water Resources, Subcommittee on Hydrology: *Methods of Flood Frequency Analysis*, Bull. 13, 42 pag., Washington, D.C., April 1966.
11. KING, W.: A Low Frequency for Irish Rivers, M. Sc. Thesis, Univ. Coll. Galway, Ireland, May. 1985.
12. KIRBY, W.: Algebraic Boundness of Sample Statistics, *Water Resour. Res.*, 10(2), 220 - 222, 1974.
13. MATALAS, N.C., SLACK, J.R. and WALLIS, J.R.: Regional Skew in Search of a Parent, *Water Resour. Res.*, 11(6), 815-826, 1975.
14. ROSSI, F., FIORENTINO, M., and VERSACE, P.: Two Component Extreme Value Distribution for Flood Frequency Analysis, *Water Resour. Res.*, 20(7), 847-856, 1984.
15. WALLIS, J.R., MATALAS, N.C. and SLACK, J.R.: Just a Moment!, *Water Resour. Res.*, 10(2), 211-219, 1974.
16. Water Resources Council: *A Uniform Technique for Determining Flood Flow Frequencies*, Bull. 15, 15 pag., Washington, D.C., Dec. 1967.
17. WU, B. and GOODRIDGE, J.D.: Selection of Frequency Distributions for Hydrologic Frequency Analysis, Dept. of Water Resources, State of California, 85 pag., Sacramento, June 1976.
18. YEVJEVICH, V. and OBEYSEKERA, J.T.B.: Estimation of Skewness of Hydrological Variables, *Water Resour. Res.*, 20(7), 1984.

## 5. CONCLUSIONES

Usualmente las muestras hidrológicas de caudales máximos son asumidas como provenientes de una población cuya función de distribución es conocida, es analíticamente expresable y, contiene un número finito de parámetros. El método más común de estimación de parámetros ha sido el de momentos, debido primordialmente a la facilidad en la computación. Generalmente estas estimaciones resultan sesgadas y con una variabilidad grande. Paralelamente se tiene que los límites en la estimación de estos no permitirá obtener los momentos de la población que supestandamente genera las muestras hidrológicas.

La condición de separación ha llevado a justificar el hecho de introducir nuevas distribuciones en el campo de hidrología, tales como la Wakeby y la TCEV; ya que de acuerdo a lo mostrado en la simulación de Monte-Carlo y sustentado en dicha condición, parece ser que ninguna distribución hasta el momento usada en hidrología sirve para modelar caudales máximos.

Todos estos problemas inherentes en la escogencia de la distribución más apropiada para modelar caudales máximos, han llevado a determinar la distribución no sólo desde el punto de vista de observar su habilidad para describir la muestra, sino también su habilidad para predecir (Obtener modelos robustos), tales como los métodos regionales de nuevo en auge en el campo de análisis de frecuencia.