

Análisis de la frecuencia de las avenidas utilizando información histórica

FELIX FRANCES (*)
JOSE D. SALAS (**)
DUANE C. BOES (**)

RESUMEN. En el análisis de la frecuencia de las avenidas para una distribución LogGumbel, se ha incorporado información histórica del tipo binomial censurada, censurada y de máxima avenida. Se han estudiado las ganancias estadísticas asintótica y muestral (SG y SG^*), obteniéndose que SG no depende de los parámetros de la distribución. Otras importantes propiedades son comentadas en este trabajo.

ABSTRACT. *Binomial censored, censored, and maximum flood historical information is incorporated to the flood frequency analysis for the Log-Gumbel distribution, using the MLE technique. The asymptotic and sample statistical gain (SG and SG^*) are studied, and it was found that SG is not a function of the distribution parameters. Other important properties are commented in this work.*

INTRODUCCION

Uno de los objetivos del análisis de la frecuencia de las avenidas es la estimación de los cuantiles correspondientes a diversos periodos de retorno. Un método clásico de estimación es el de ajustar una distribución de probabilidad a los caudales máximos anuales registrados en una estación de aforos (denominados registro sistemático en este trabajo). Sin embargo, este método está limitado por diversos factores, como: i) la incertidumbre del modelo seleccionado, ii) errores en los datos (especialmente de las más grandes avenidas), y iii) poblaciones con coeficientes de variación y sesgo altos (que dan lugar a fuertes variaciones de las propiedades estadísticas muestrales).

En la estimación de cuantiles de avenidas, se puede incrementar la fiabilidad utilizando el método de estimación más eficiente, es decir, con menor error cuadrático medio (MSE), y/o extendiendo la longitud de la información disponible. Esto último puede ser realizado mediante un análisis regional y/o la incorporación al modelo de información adicional acerca de las avenidas. En este estudio vamos a explorar los beneficios de añadir la información histórica disponible en la estimación de los cuantiles. En este artículo se van a reflejar los resultados preliminares de un estudio por terminar (Frances et al., 1991).

INFORMACION HISTORICA

Dependiendo de la fuente de información se puede distinguir entre información histórica e información sobre paleoavenidas. La información histórica está constituida por todas aquellas medidas previas al registro sistemático. Estas medidas, habitualmente caudales máximos, pueden ser localizadas en los periódicos locales, marcas en edificios, etc. Ejemplos de utilización de información histórica pueden ser encontrados en Benson (1950), Leese (1973), Gerard y Karpuk (1979) y Condie (1986). Por otra parte, las paleoavenidas son previas a la información histórica y pueden ser obtenidas por estudios geológicos de los sedimentos de los ríos, datación con carbono 14 de la materia orgánica depositada, etc. Más detalles describiendo estos métodos y su aplicación en el análisis de la frecuencia de las avenidas pueden ser encontrados en Stedinger y Baker (1987). Sin embargo, desde el punto de vista estadístico no existe ninguna diferencia entre avenidas históricas y paleoavenidas, por lo que en este estudio el término «información histórica» engloba a las dos.

Como fue remarcado por Leese (1973), Condie y Lee (1982) y Stedinger y Cohn (1987), los registros históricos son el resultado de dos tipos de censura estadística: i) Censura tipo 1 en la que el límite censor es conocido y las magnitudes de las avenidas históricas por encima del límite pueden ser conocidas o desconocidas, y ii) Censura tipo 2 en la que el número de valores censurados es conocido *a priori*.

De esta clasificación y de nuestra experiencia en estudios sobre avenidas, tres tipos de información histórica puede ser encontrada y utilizada: 1) Binomial censa-

(*) Universidad Politécnica de Valencia.

(**) Hydrology and Water Resources Program, Colorado State University, Fort Collins, Colorado.

rada, cuando las magnitudes de las avenidas no son conocidas con precisión y sólo utilizamos el hecho de que durante M años el límite de percepción histórico Q_H fue superado k veces; 2) Censurada, cuando los caudales de las avenidas históricas son conocidos si superan el límite Q_H , y 3) De máxima avenida, cuando la única información disponible es el valor de la máxima avenida z durante el período histórico.

APROXIMACION AL PROBLEMA

El objetivo de este trabajo es la mejora de los estimadores de los cuantiles de avenidas utilizando estos tres tipos de información histórica frente al uso exclusivo de la información sistemática. En un primer paso obtendremos la ganancia estadística asintótica (SG) de los estimadores de los cuantiles utilizando el Límite Inferior de Cramer-Rao (CRLB) y, en un segundo paso, calcularemos la ganancia estadística asintótica (SG*) por simulación tipo Montecarlo. La mayoría de los trabajos previos sólo consideran simulaciones de Montecarlo para obtener las ganancias estadísticas. El cálculo de los resultados asintóticos es muy importante, ya que es posible descubrir cuáles son los principales factores que afectan a dicha ganancia estadística.

Las hipótesis realizadas en este estudio son:

- Los datos disponibles son locales,
- la distribución de la población es LogGumbel, cuya función de densidad de probabilidad acumulada es:

$$F_X(x) = e^{-(\theta/x)^\gamma} \quad (1)$$

donde:

X = variable aleatoria positiva (= avenida anual),
 θ = parámetro de escala,
 γ = parámetro de forma, y

- El método de estimación es el de Máxima Verosimilitud (MLE).

ESTIMADORES POR MAXIMA VEROSIMILITUD DE LOS CUANTILES

Assumiendo que las avenidas son independientes e idénticamente distribuidas, la función de verosimilitud utilizando el registro sistemático es:

$$L(\underline{\theta}; \underline{x}) = \prod_{i=1}^N f_X(x_i; \underline{\theta}) \quad (2)$$

donde:

$\underline{\theta}$ = conjunto de parámetros de la distribución de la población,
 $\underline{x}$, x_i = registro sistemático,
 $f_X(\cdot)$ = función de densidad de probabilidad de X ,
 N = longitud del registro sistemático.

Si se utiliza la información binomial censurada, la función de verosimilitud queda:

$$L(\underline{\theta}; \underline{x}) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i; \underline{\theta}) \quad (3)$$

donde:

p = probabilidad de excedencia del límite de percepción Q_H .

La función de verosimilitud utilizando información censurada es:

$$L(\underline{\theta}; \underline{x}) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i; \underline{\theta}) \prod_{j=1}^k f_Y(y_j; \underline{\theta}) \quad (4)$$

donde:

y_j = registro histórico,
 $f_Y(\cdot)$ = función de densidad de probabilidad de Y .
 Se puede demostrar fácilmente que $f_Y(\cdot) = f_X(\cdot)/p$.

Si la única información histórica durante el período histórico es la máxima avenida z , la función de verosimilitud resulta ser:

$$L(\underline{\theta}; \underline{x}) = \prod_{i=1}^N f_X(x_i; \underline{\theta}) f_Z(z; \underline{\theta}) \quad (5)$$

donde:

$f_Z(\cdot)$ = función de densidad de probabilidad de Z . Se puede demostrar que $f_Z(\cdot) = M f_X(\cdot) F_X^{M-1}(\cdot)$.

Diferentes MLE de los parámetros resultan maximizando estas funciones. Por ejemplo, para el caso de utilizar exclusivamente información sistemática, el sistema de ecuaciones resultante es:

$$\theta^\gamma = \frac{N}{\sum_{i=1}^N x_i^{-\gamma}} \quad (6)$$

$$\gamma = \frac{N}{\sum_{i=1}^N \ln(\theta/x_i) [(\theta/x_i)^\gamma - 1]} \quad (7)$$

Este sistema puede ser resuelto iterativamente. Los estimadores resultantes los denominaremos *MLE0*. Expresiones similares pueden obtenerse para los estimadores *MLE1*, *MLE2* y *MLE3*.

GANANCIA ESTADISTICA ASINTOTICA

El *CRLB* es un límite inferior de la varianza de todos los estimadores insesgados, es decir, es lo mejor que se puede hacer en términos de la varianza. También, el *CRLB* es la varianza asintótica de los *MLEs*. La ganancia estadística se define como:

$$SG_i = 1 - \frac{CRLB_i}{CRLB_0} \quad \text{for } i = 1, 2, 3 \quad (8)$$

Los *CRLBs* son obtenidos en Francés *et al.* (1991) y las ganancias para los diferentes niveles de información son:

$$SG1 = 1 - R_1 \left[1 + \frac{rqD^2 \ln^2(D/K)/p}{a - 2b \ln K + \ln^2 K} \right] \quad (9)$$

$$SG2 = 1 - R_2 \cdot \quad (10)$$

$$\cdot \left[1 + r \frac{a' + p + qD \ln^2 D - 2(b' + qD \ln D) \ln K + p \ln^2 K}{a - 2b \ln K + \ln^2 K} \right]$$

$$SG3 = 1 - R_3 \left[N + 1 + \frac{\ln M (\ln M - 2b + 2 \ln K)}{a - 2b \ln K + \ln^2 K} \right] \quad (11)$$

donde:

$$\frac{1}{R_1} = 1 + \frac{rqD^2 (\ln^2 D - 2b \ln D + a)/p}{a - b^2} \quad (12)$$

con: $q = 1 - p$, $a = 1 + \Gamma^*(2) = 1.8237$, $b = \Gamma^*(2) = 0.4228$, $D = -\ln(1 - 1/H)$, $K = -\ln(1 - 1/T)$, $r = M/N$, y $\Gamma(\cdot)$ la función gamma.

$$\frac{1}{R_2} = 1 + \frac{r}{a - b^2} [(a' + p + qD \ln^2 D) (rp + 1) - 2(b' + qD \ln D) (rb' + b) + r(b'^2 - (qD \ln D)^2 + ap)] \quad (13)$$

con: $a' = \Gamma^*(2, D)$, $b' = \Gamma^*(2, D)$, y $\Gamma(\cdot, \cdot)$ la función gamma incompleta. Y, por último,

$$\frac{1}{R_3} = \frac{(N + 1)^2}{N} + \frac{\ln^2 M}{a - b^2} \quad (14)$$

Algunas propiedades de las *SGs* son:

1. *SG1* y *SG2* son solamente función de tres factores: i) el ratio r entre las longitudes de los períodos histórico y sistemático, ii) el período de retorno H del límite de percepción Q_H , y iii) el período de retorno T del cuantil X_T . Las figuras 1 y 2 representan las *SGs* frente a T y H .
2. *SG3* es función exclusiva de: i) la longitud del registro sistemático N , ii) la longitud del período histórico M , y iii) el período de retorno del cuantil X_T .
3. Existe un punto de máxima ganancia estadística utilizando información histórica. Este punto pudo ser obtenido analíticamente y se corresponde con:

para *SG1*:

$$T = H \quad (15)$$

para *SG2* (normalmente T próximo y mayor que H):

$$\ln K = \frac{-A + (A^2 + 4BC)^{1/2}}{2B} \quad (16)$$

donde:

$$\begin{aligned} A &= a' + (1 + a)p + qD \ln^2 D, \\ B &= pb - b' - qD \ln D, \\ C &= b(a' + p + qD \ln^2 D) - a(b' + qD \ln D) \end{aligned}$$

para *SG3* (T mayor que H):

$$\ln K = b - \frac{\ln M + [\ln^2 M + 4(a - b^2)]^{1/2}}{2} \quad (17)$$

4. Existe un punto donde no existe ganancia estadística utilizando información histórica para *MLE1* y *MLE3*, y muy pequeña para *MLE2*. Este punto se corresponde con el mínimo de *SG*, y se corresponde con muy bajos períodos de retorno (menos de tres años). Por tanto, no es de interés en Hidrología.

5. *SG3* es normalmente menor que *SG1* y *SG2*. La excepción es cuando la información histórica es pobre, es decir, H grande y/o r pequeño.

MSE MUESTRAL DEL MLE DEL CUANTIL

Simulaciones por Montecarlo fueron realizadas para tratar de contestar las siguientes preguntas: a) ¿La información histórica produce algún tipo de ganancia estadística en el sesgo de los estimadores, varianza y *MSE* muestrales? b) ¿Es el *CRLB* una buena aproximación del *MSE* muestral? c) ¿Las ganancias estadísticas asintóticas *SG* son una buena aproximación de las ganancias estadísticas muestrales *SG** (obtenidas utilizando los *MSEs* en lugar de los *CRLBs* en la ecuación 8)?

Los resultados de la simulación pueden ser observados en las figuras 1 y 2. Estos indican que:

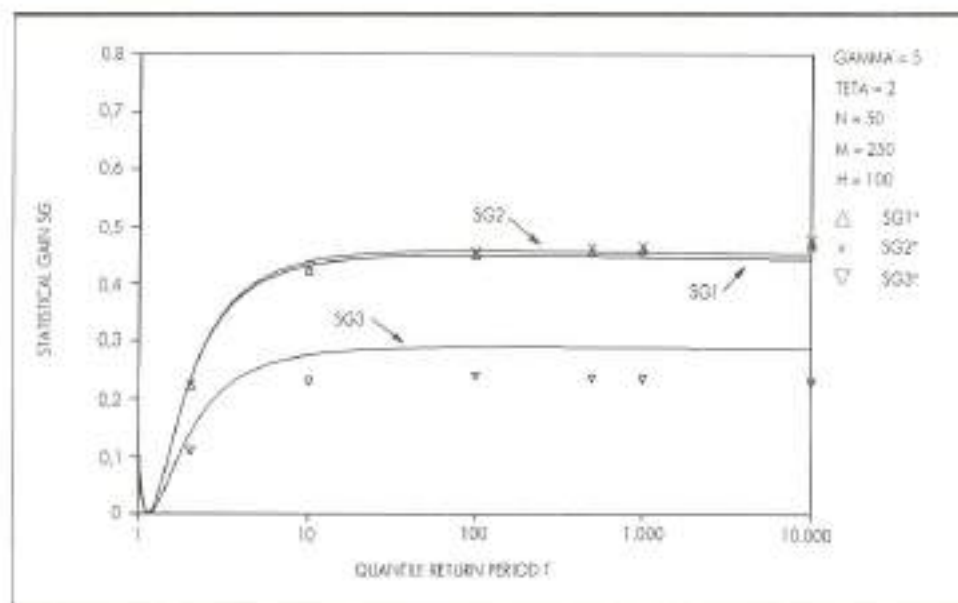


FIGURA 1. Ganancias estadísticas asintótica y media muestral versus el período de retorno del cuantil, para un límite de percepción histórica de 100 años.

1. Los *CRLB*s efectivamente son una buena aproximación de los *MSE*s. Para las semillas iniciales utilizadas (y dependiendo de ellas) la diferencia entre *MSE* y *CRLB* es menor del 5 % para *MLE1* y *MLE2*, y menor del 10 % para *MLE3* (siempre *MSE* mayor que *CRLB*).
2. En general, *SG1** y *SG2** son ligeramente mayores que *SG1* y *SG2*, y *SG3** menor que *SG3*. La razón es que *MSE1* y *MSE2* se aproximan a sus valores asintóticos más rápido que *MSE0*, mientras que *MSE3* lo hace más lentamente.
3. La ganancia estadística muestral se muestra bastante independiente del parámetro γ (con pequeñas va-

riaciones sólo para coeficientes de sesgo muy altos) y claramente independiente del parámetro θ . Esta independencia se reproduce para *SG1** y *SG2** frente a la longitud del registro sistemático N (para un r constante).

CONCLUSIONES

Un modelo estadístico ha sido desarrollado para poder incorporar diferentes tipos de información histórica (binomial censurada, censurada y de máxima avenida) al registro sistemático, basado en los *MLE* para una distribución LogGumbel.

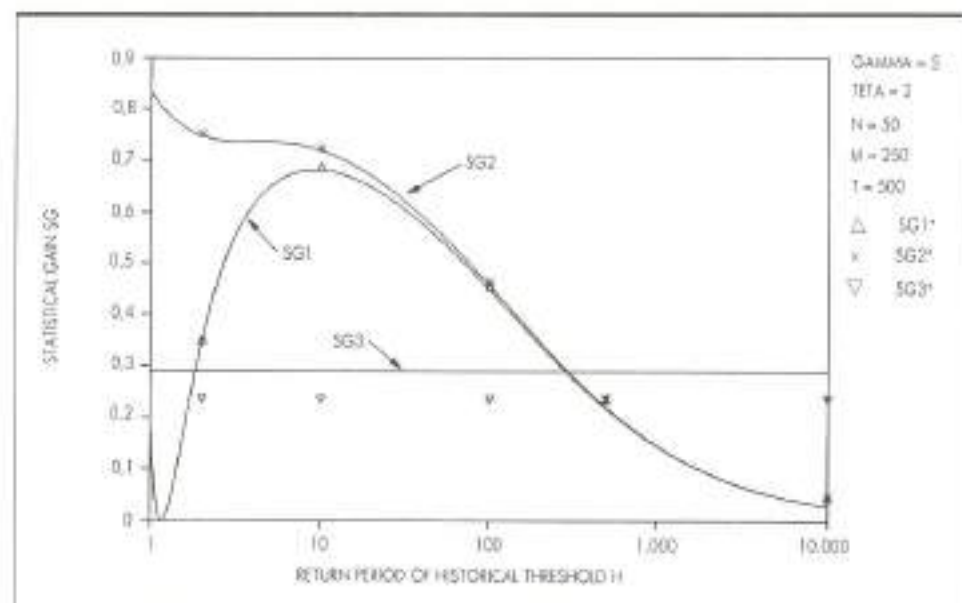


FIGURA 2. Ganancias estadísticas asintótica y media muestral versus el límite de percepción histórica, para un r de 5.

El modelo fue analizado comparando los resultados estadísticos utilizando información histórica y sin utilizar ésta. Primero considerando las varianzas asintóticas de los estimadores inescuadrados, y después mediante los *MSE*s muestrales. Todos los cálculos se realizaron para un amplio rango de posibles datos.

Los principales resultados para el rango habitual de los factores (N de 20 a 200 años, r de 1 a 5, H menor de 1.000 años y cuantiles de 10 a 10.000 años) son:

1. Siempre se obtienen mejores estimadores utilizando la información histórica disponible. Se puede considerar desde un punto de vista práctico que las ganancias estadísticas muestrales son independientes de los parámetros, es decir, independientes de las características estadísticas de la población.
2. Los *MSE*s asintóticos y *SG*s son una buena aproximación de sus correspondientes valores muestrales. La única excepción es *SG3*, para el que el error es del 15 % comparado con *SG3**.
3. El más importante factor para *SG1** y *SG2** es el límite de percepción histórico. Debido al comportamiento asintótico respecto de r , con grandes incrementos del período histórico (r mayor de 500) se producen muy pequeños incrementos para *SG1* y *SG2**. En el caso de *SG3** evidentemente el factor más importante es r .
4. Existen pequeñas diferencias entre *SG1** y *SG2**, como fue obtenido por Stedinger y Cohn (1986), excepto cuando existe gran abundancia de información (H es muy pequeño).
5. *MLE3* produce mejores estimadores que *MLE1* y *MLE2* sólo con información histórica muy pobre (H elevado y/o r pequeño).

AGRADECIMIENTOS

Esta investigación se ha llevado a cabo con la subvención en parte de la «NSF Grant on Regional Flood Hazard Analysis» con los profesores J. D. Salas y D. C. Boes como principales investigadores. Se agradecen también los comentarios y sugerencias del profesor Juan Marco (Universidad Politécnica de Valencia), muy importantes durante la realización de este trabajo.

REFERENCIAS

- BENSON, M. A. (1950). Use of Historical Data in Flood Frequency Analysis, *Eos. Trans. AGU*, 31(3), 419-424.
- CONDIE, R. (1986). Flood Samples from a Three-Parameter Lognormal Population with Historic Information; the Asymptotic Standard Error of Estimate of the T-year Flood, *Jour. of Hydrol.*, 85, 139-150.
- CONDIE, R., y LEE, K. H. (1982). Flood Frequency Analysis with Historic Information, *Jour. of Hydrol.*, 58, 47-61.
- FRANCES, F.; BOES, D. C., y SALAS, J. D. (1991). Flood Frequency Analysis by Using Historical Information for a Log-Gumbel Model, to be submitted to *Water Resources Research*.
- GERARD, R., y KARPUR, E. W. (1979). Probability Analysis of Historical Flood Data, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, 105(HY9), 1153-1165.
- LEESE, M. N. (1973). Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Maximum Flood Series, *Water Res. Res.*, 9, 1534-1542.
- STEDINGER, J. R., y COHN, T. A. (1986). Flood Frequency Analysis with Historical and Paleoflood Information, *Water Resour. Res.*, 22(5), 785-793.
- STEDINGER, J. R., y BAKER, V. R. (1987). Surface Water Hydrology: Historical and Paleoflood Information, *Reviews of Geophysics*, 25(2), 119-124.