Ingeniería Investigación y Tecnología, ISSN 2594-0732, VI. 1. 13-17, 2005
(artículo arbitrado)
DOI:http://dx.doi.org/10.22201/fi.25940732e.2005.06n1.002

Análisis de eficiencia de la distribución Bi-Gumbel

C.A. Escalante-Sandoval
División de Estudios de Posgrado
Facultad de Ingeniería, UNAM
E-mail: caes@servidor.unam.mx

(recibido: mayo de 2003; aceptado: diciembre de 2003)

Resumen

La mayoría de los estudios sobre eventos hidrológicos extremos se han llevado a cabo utilizando distribuciones univariadas. La gran variabilidad de los eventos estimados para ciertos períodos de retorno ha promovido la exploración de modelos de estimación conjunta, tal como la distribución bivariada con marginales de valores extremos tipo I, llamada Bi-Gumbel. Se emplea la técnica de muestreo distribucional con el propósito de determinar si los eventos estimados mediante el ajuste de la distribución bivariada son mejores que aquellos obtenidos en forma univariada. Se concluye que los eventos obtenidos en forma bivariada son menos sesgados que su contraparte univariada.

Descriptores: Distribución multivariada de valores extremos, distribución Gumbel, estimados de máxima verosimilitud, técnica de muestreo distribucional.

Abstract

Most hydrological extreme studies in the past have been analyzed through use of univariate distributions. The large variability of the T-year flood estimates has prompted exploration of joint estimation models, such as the bivariate distribution with extreme value type I marginals, named Bi-Gumbel distribution. To investigate whether the estimates of the quantiles based on bivariate distribution are better than those on univariate procedures a distribution sampling technique was used. A significant improvement occurs when the parameters are estimated using the bivariate distribution instead of univariate form and such again is more significant in relation to the shorter samples.

Keywords:*Multivariate extreme value distribution, Gumbel distribution, maximum likelihood estimates, distribution sampling technique.*

Introducción

El objetivo del análisis de frecuencias es la estimación, a través de distribuciones de probabilidad de la magnitud del gasto máximo anual de cierto período de retorno. Con frecuencia, la información que se requiere para realizar esta estimación no se encuentra disponible. En otras ocasiones los datos existen, pero no con la longitud suficiente para proveer estimadores confiables de los parámetros y el error del evento asociado al período de retorno es grande e ineficiente para propósitos de diseño.

La gran variabilidad de estos estimadores ha promovido la exploración de modelos de estimación conjunta, donde los datos de sitios vecinos en la región se combinan con el registro de longitud inadecuada para incrementar la información y proveer un estimador regional del evento de diseño.

DOI:http://dx.doi.org/10.22201/fi.25940732e.2005.06n1.002

Hay varias técnicas disponibles de estimación regional hidrológica (Cunnane, 1988), algunas de ellas requieren de la normalización de los datos, ya que están basadas en la distribución normal. Sin embargo, se han obtenido mejoras significativas al emplear procedimientos multivariados a vari ables no-normales (Raynal, 1985).

En este trabajo se presenta el modelo logístico bivariado con marginales de valores extremos tipo I, llamado Bi-Gumbel (Raynal, 1985).

Características de la distribución bivariada

La forma general del modelo logístico para las distribuciones bivariadas de valores extremos es (Gumbel, 1960):

$$F(x,y,\theta)=\exp\left\{-[(-\ln F(x))^{m}+(-\ln F(y))^{m}]^{1/m}\right\} \quad (1)$$

Donde:

- m parámetro de asociación bivariada ($m>1$).
- $F(x)$ distribución mar ginal de x tipo Gumbel.
- $F(y)$ distribución mar ginal de y tipo Gumbel.
- ? Conjunto de parámetros a estimarse(*?1, a1, ?2, a2, m)* para la distribución Bi-Gumbel.
- *x,y* gastos máximos anuales en dos estaciones vecinas.

La ecuación (1) debe satisfacer:

$$F(x)F(y) < F(x,y) < min[F(x), F(y)] \quad (2)$$

La distribución mar ginal Gumbel tiene la forma:

$$F(s)=\exp^{\exp^{-\left(\frac{s-v}{a}\right)}} \quad (3)$$

El procedimiento de estimación de parámetros de la distribución bivariada se desarrolló para permitir el caso de muestras con diferentes longitudes de registro (Figura 1).

Si $(X_1,Y_1),\ldots,(X_n,Y_n)$ es una muestra aleatoria de una densidad bivariada, la correspondiente función de verosimilitud es (Mood *et al*., 1974):

$$L(x,y,\underline{\theta})=\prod_{i=1}^{n} f(x_i,y_i,\underline{\theta}) \quad (4)$$

Con el fin de considerar todas las posibles combinaciones de los datos (Figura 1) se propone la siguiente función de verosimilitud:

$$L(x,y,\underline{\theta})=\left[\prod_{i=1}^{n_1} f(p_i,\underline{\theta}_1)\right]^{l_1}\left[\prod_{i=1}^{n_2} f(x_i,y_i\,\underline{\theta}_2)\right]^{l_2}$$

$$\left[\prod_{i=1}^{n_3} f(r_i,\underline{\theta}_3)\right]^{l_3} \quad (5)$$

Donde:

- n_1 longitud de registro antes del registro común.
- n_3 longitud de registro después del registro común.
- n_2 longitud de registro en el período común.
- p vari able con longitud n_1.
- x,y vari ables con longitud n_2.
- r vari able con longitud n_3.

Estación 1

$$x_1,\ldots,x_{n_1},x_{n_1}+1,\ldots,x_{n_1}+n_2$$

Estación 2

$$y_{n_1}+1,\ldots,y_{n_1}+n_2+1,\ldots,y_{n_1}+n_2+n_3$$

Figura 1. Máximo arreglo muestral

DOI:http://dx.doi.org/10.22201/fi.25940732e.2005.06n1.002

I_i indicador tal que $Ii=1$ si $n_i>0$ o $I_i=0$ si $n_i=0$

Dada la propiedad de que el máximo de una función y su logaritmo ocurren en el mismo punto y debido al hecho de que las expresiones obtenidas por el logaritmo de la ecuación (5) son más fáciles de manipular que su forma natural, se propone la siguiente función logarítmica de verosimilitud:

$$\ln L(x,y,\underline{\theta}) = I_1\left[\sum_{i=1}^{n_1}\ln f(p_i,\underline{\theta}_1)\right] + I_2\left[\sum_{i=1}^{n_2}\ln f(x_i,y_i,\underline{\theta}_2)\right]$$

$$+ I_3\left[\sum_{i=1}^{n_3}\ln f(r_i,\underline{\theta}_3)\right] \qquad (6)$$

Debido a que la solución del sistema de ecuaciones resultantes al derivar parcialmente la ecuación (6) con respecto a $?_1$, a_1, $?_2$, a_2 y m resulta muy complejo, se propone para el cálculo de los parámetros el algoritmo de optimación multivariado restringido de Rosenbrok (Kuester y Mize, 1973), el cual implica la directa maximización de dicha ecuación.

Confiabilidad de los eventos estimados para diferentes períodos de retorno

Una nueva aproximación para el análisis de frecuencias debe mostrar que los estimadores de los eventos asociados a cierto período de retorno son más confiables que aquellos que se obtienen con los métodos ya existentes. Esta confiabilidad se puede cuantificar a través de medir el sesgo, varianza y la raíz del error medio cuadrático. En este trabajo se realizó un estudio experimental basado en la generación de datos con el fin de comparar el sesgo de los eventos estimados mediante la distribución univariada Gumbel, con aquellos obtenidos al ajustar los datos a la distribución Bi-Gumbel. Con este propósito se generaron 99,000 números con distribución poblacional Gumbel y parámetros $?_1=14$ y $a_1=1.4$ (estación base), y fueron agrupados en muestras de tamaño 9, 19 y 49. Por lo que el número de muestras para cada tamaño es igual a 11000, 5210 y 2020, respectivamente. Tal número de muestras asegura una desviación máxima absoluta entre la distribución real y la empírica de menos de 0.016 para el tamaño más grande y de 0.036 para el más pequeño, con una probabilidad del 99% (Gnedenko, 1967).

Para el caso de la distribución Bi-Gumbel, los eventos se estimaron combinando cada muestra generada para la estación base con otra del mismo tamaño o mayor (Figura 2), así, los casos explorados tienen longitudes 9–9, 9–19, 9–49, 19–19, 19–49 y 49–49. Los números Gumbel generados para la llamada estación vecina tienen parámetros poblacionales $?_2=12$ y $a_2=1.2$.

Sea θ el evento a estimarse, $\hat{\theta}_i$, $i=1,...,n$ los eventos obtenidos de cada muestra, y n el número de muestras, las cuales varían de 11,000 a 2020, de acuerdo con lo explicado anteriormente. Entonces, el sesgo del estimador se obtiene como:

$$sesgo = m(\hat{\theta}) - \theta \qquad (7)$$

Donde $m(\hat{\theta})$ es la media de la serie $\hat{\theta}i$, $i=1,...,n$

Estación Base 1

$$x_1,...,x_{n_2}$$

Estación Vecina 2

$$y_1,...,y_{n_2},y_{n_2+1},...,y_{n_2+n_3}$$

Figura 2. Arreglo muestral propuesto para obtener los eventos de diferente período de retorno con la distribución Bi-Gumbel.

DOI:http://dx.doi.org/10.22201/fi.25940732e.2005.06n1.002

$$m(\hat{\theta}) = (1/n)\sum_{i=1}^{n}\hat{\theta}_i \qquad (8)$$

La comparación se llevó a acabo para eventos estimados con probabilidades de no excedencia de 0.50, 0.80, 0.90, 0.95, 0.999 y 0.9999, las cuales abarcan probabilidades de no excedencia por 2 a 10 000 años.

En la tabla 1 se presentan los sesgos de los eventos de diferente período de retorno, obtenidos con las expresiones (7) y (8). Se puede observar que los sesgos obtenidos con el procedimiento bivariado son más pequeños que los de origen univariado. De hecho, conforme la longitud asociada en la combinación bivariada se incrementa, el sesgo de la estación base disminuye a través del rango $0.5 \leq F \leq 0.9999$.

Esto significa que hay una ganancia en información cuando se estiman los parámetros de una serie de corta longitud con otra de igual o mayor tamaño.

También se observa que para que una muestra de 9 datos tenga la misma precisión de una de 19

Tabla 1. Sesgo de eventos de diferente período de retorno, obtenidos para la estación base considerando la estimación univariada y bivariada.

	Distribución			
Probabilidad	Gumbel 9	9 – 9	Bi-Gumbel 9 – 19	9–49
0.9999	–1.1056	–1.0708	–1.0243	–0.4824
0.9990	–0.8132	–0.8374	–0.8001	–0.3766
0.9900	–0.5204	–0.6035	–0.5754	–0.2706
0.9500	–0.3134	–0.4383	–0.4167	–0.1957
0.9000	–0.2221	–0.3653	–0.3466	–0.1626
0.8000	–0.1268	–0.2892	–0.2735	–0.1281
0.5000	0.0171	–0.1744	–0.1632	–0.0761
	19		19–19	19–49
0.9999	–0.4865		–0.3863	0.1127
0.9990	–0.3577		–0.2964	0.0921
0.9900	–0.2287		–0.2064	0.0714
0.9500	–0.1375		–0.1428	0.0568
0.9000	–0.0973		–0.1147	0.0503
0.8000	–0.0553		–0.0854	0.0436
0.5000	0.0081		–0.0411	0.0334
	49			49–49
0.9999	–0.2096			0.0277
0.9990	–0.1542			0.0246
0.9900	–0.0986			0.0214
0.9500	–0.0594			0.0192
0.9000	–0.0420			0.0182
0.8000	–0.0240			0.0172
0.5000	0.0033			0.0157

DOI:http://dx.doi.org/10.22201/fi.25940732e.2005.06n1.002

se requiere asociarla a otra que al menos cuente con 49 años de registro.

Conclusiones

El objetivo del estudio fue el de investigar el grado de mejora en la estimación de eventos de diseño cuando se emplea la distribución bivariada de valores extremos con marginales Gumbel.

El análisis de resultados sugieren que el efecto de la muestra adicional dentro del proceso de estimación de parámetros y eventos de diseño es más importante conforme su tamaño se incrementa, lo que implica una sustancial ganancia en información.

Se puede concluir que para los casos en que se requiera obtener eventos de diseño en sitios con escasa información, y se disponga de un sitio vecino, dentro de la misma región homogénea, es conveniente utilizar una distribución de probabilidad bivariada para llevar a cabo el análisis de frecuencia.

El modelo logístico bivariado permite no solo utilizar como marginales a la distribución Gumbel, sino también a distribuciones como la General de Valores Extremos (GVE), la Gumbel de dos poblaciones (Gumix) y la de Valores Extremos de dos Componentes (TCEV), lo que lo hace muy versátil dentro del análisis de frecuencias de eventos extremos hidrológicos.

Referencias

Cunnane C. (1988). Methods and merits of regional flood frequency análisis. *Journal of Hydrology*. 100, pp. 269-290.

Gnedenko B.V. (1967). *The Theory of Probability*. Chelsea.

Gumbel E.J. (1960). Multivariate extremal distributions. *Bulletin International Statist. Inst*. 39 (2), pp. 471-475.

Kuester J.L. y Mize J.H. (1973). *Optimization Techniques with FORTRAN*. McGraw-Hill.

Mood A., Graybill F. y Boes D. (1974). Intro - duction to the Theory of Statics. McGraw-Hill.

Raynal J.A. (1985). Bivariate Extreme Value Distributions Applied to Flood Frequency Análisis Ph.D Dissertation. Civil Engineering Department, Colorado State University.

Semblanza del autor

Carlos Agustin Escalante-Sandoval. Egresado como ingeniero civil en 1985 de la Universidad Autónoma de Puebla, obtuvo el grado de maestro en ingeniería en aprovechamientos hidráulicos en 1988 y el doctorado en ingeniería hidráulica en 1991, ambos en la Facultad de Ingeniería de la UNAM. Sus trabajos de hidrología, los cuales destacan los campos de fenómenos extremos (lluvias, inundaciones y sequías) le han valido el reconocimiento en el ámbito nacional e internacional. Cuenta con diversas publicaciones y ha participado en 12 proyectos de investigación, destacando al análisis hidrológico de la Costa de Chiapas con motivo de la inundaciones de 1998 (CNA), el MIA del proyecto hidroeléctrico La Parota (CFE) y el análisis nacional del fenómeno de la sequía. Ha recibido distinciones como la medalla Gabino Barreda por sus estudios de doctorado, el premio Distinción Universidad Nacional para Jóvenes Académicos en Docencia en Ciencias Exactas 1999, otorgada por la UNAM y el premio Nacional Enzo Levi a la "Investigación y Docencia en Hidráulica 2002", por la Asociación Mexicana de Hidráulica. Actualmente imparte cátedra y es jefe del Departamento de Ingeniería Hidráulica de la Facultad de Ingeniería, UNAM.