ANÁLISIS DE FRECUENCIAS

EXPRESIONES PARA EL CÁLCULO DE LOS EVENTOS PARA EL PERÍODO DE RETORNO *T* Y DE LOS RESPECTIVOS ERRORES ESTÁNDAR DE ESTIMACIÓN REQUERIDOS PARA LA DETERMINACIÓN DE LOS INTERVALOS DE CONFIANZA DE LOS ESTIMADOS DE LOS VALORES ESPERADOS

JULIAN DAVID ROJO HERNÁNDEZ

1. EXPRESIÓN GENERAL PARA EL CÁLCULO DEL ESTIMADO DEL VALOR ESPERADO DE UN EVENTO PARA UN PERÍODO DE RETORNO T (expresión de Ven T. Chow para análisis de frecuencias de eventos extremos):

Sea la serie de tiempo X:

$$X = \{x_1, x_2, x_3, \ldots, x_N\}$$

la muestra observada de los N valores medidos de un fenómeno o evento cualquiera (lluvias, caudales, temperaturas, sismos, etc.), tiene las siguientes características estadísticas (muestrales):

$\overline{X}$: estimado (muestral) del valor esperado del fenómeno analizado X, equivalente al valor medio o promedio de los N valores observados de la serie X; se le llama también “media” de X. Corresponde al denominado primer momento muestral del arreglo X con relación al origen de los reales (con respecto a cero).

S_X : estimado (muestral) de la desviación típica (desviación estándar) muestral (raíz cuadrada de la varianza), obtenido a partir de la utilización de los N valores observados de la serie X, empleando el estimador insesgado de este parámetro. Corresponde a la raíz cuadrada del segundo momento muestral del arreglo X con respecto de la media.

g_X : estimado del coeficiente de asimetría muestral (raíz cuadrada de la varianza), obtenido a partir de la utilización de los N valores observados de la serie X, empleando el estimador insesgado de este parámetro. Corresponde a la relación entre el tercer momento muestral del arreglo X con relación a la media, y el cubo del estimado sesgado de S_X.

Debido a la incertidumbre en los estimados de parámetros estadísticos muestrales de orden superior al tercer momento muestral, no se tienen en cuenta parámetros como el de kurtosis y similares.

Si los valores de la serie X son eventos aleatorios independientes entre sí, pertenecientes a una misma función de distribución (densidad) de probabilidades *f(x;* $\alpha, \beta, \gamma, ...$*)*, tal que $\alpha, \beta, \gamma, ...$ son los parámetros de la función de distribución de probabilidades, se puede obtener los valores estimados de la variable X para cualquier probabilidad de ocurrencia dada ***p*** a partir de la siguiente expresión general (propuesta por Ven T, Chow):

$$X_p = \overline{X} + K_p.S_X$$

Donde K_p es un número adimensional, aleatorio, obtenido de la función de distribución de probabilidades cualquiera f(x; $\alpha, \beta, \gamma, ...$) a la que pertenecen las ocurrencias aleatorias del fenómeno X.

Como es sabido, en el manejo ingenieril tradicional de las probabilidades de ocurrencia de fenómenos, el concepto de definición de la probabilidad de acurrencia ha sido asociado al concepto de período de retorno o recurrencia del evento en consideración (T), el cual, desde el punto de vista de su formalización matemática ha sido definido de la siguiente forma:

$$T = 1/\boldsymbol{p}$$

Así, cuando se habla de eventos extremos máximos, la probabilidad de interés es la probabilidad de excedencia de un evento (G(x) = 1 - F(x)), mientras cuando se trata de eventos extrmos mínimos dicha probabilidad de interés es la de no excedencia ($F(x) = P[X \leq x]$). Por esta razón, aunque para algunos pueda ser más ilustrativo utilizar el concepto del "período de retorno", su empleo en el manejo numérico de probabilidades debe ser consecuente para evitar errores de cálculo graves.

De esta forma, en la práctica, cuando se habla de períodos de retorno para el caso de eventos máximos, la expresión de T es el inverso de la *probabilidad de excedencia* del evento de interés, o sea:

$$T = 1/G(x) = 1/[1 - F(x)]$$

mientras que si se trata de períodos de retorno para el caso de eventos mínimos, la expresión de T es el inverso de la *probabilidad de no - excedencia* del evento de interés, o sea:

$$T = 1/F(x)$$

Con estas advertencias, la forma más popular de la expresión de Ven T. Chow para el análisis de frecuencias (de máximos o de mínimos) es:

$$X_T = \overline{X} + K_T.S_X \qquad (1)$$

El valor de X_T obtenido con la expresión (1) corresponde al estimado del *valor esperado del evento X para el período de retorno T*, el cual se encuentra exactamente dentro de la función matemática que (se supone) describe el comportamiento probabilístico de la variable X que se está considerando.

2. INTERVALO DE CONFIANZA PARA EL ESTIMADO DE X_T

Como se ha dicho en clase, X_T de la expresión (1) es solo el *valor esperado del evento X para el período de retorno T*, ya debe recordarse que realmente X_T es una variable aleatoria a causa, entre otras, de la incertidumbre originada en la estimación de los parámetros de la distribución de probabilidades de X (y en su selección, por supuesto), de manera que pueden existir (son probables) valores de X diferentes al calculado con la expresión (1), el cual se supone que es simplemente el valor esperado de X para ese T; es más, ni siquiera puede afirmarse que el resultado de la expresión (1) sea el valor más probable, ya que la distribución de probabilidades de los valores de X para dicho T (distribución marginal de X para T) no necesariamente es simétrica (en cuyo caso el valor esperado sí es a la vez el más probable). Por lo tanto, el valor X_T tiene incertidumbre, la cual se calcula a partir de la estimación de la varianza de X para dicho período de retorno T, cuya raíz cuadrada es conocida como “Error Estándar”, $S_E(T)$.

En términos generales, $S_E(T)$ es función de S_X, T, N y la $f(x;\ \alpha, \beta, \gamma, ...)$ a la que se supone pertenece X.

De esta forma, conocidos X_T y el correspondiente valor de $S_E(T)$ (para el mismo T), es posible calcular el intervalo de confianza del estimado del evento X para el período de retorno T y un nivel de significancia α, de acuerdo con la siguiente expresión:

$$(X_T + \omega_{\alpha/2}\,.\, S_E(T)) < X_T < (X_T + \omega_{1-\alpha/2}\,.\, S_E(T)) \quad (2)$$

donde $\omega_{\alpha/2}$ ó $\omega_{1-\alpha/2}$ son respectivamente los valores (adimensionales) representativos de las probabilidades de no excedencia $\alpha/2$ y de excedencia 1 - $\alpha/2$ de la función de distribución de probabilidades marginal de los eventos de la variable X para el período de retorno T, la cual tiene como valor medio X_T y como desviación típica $S_E(T)$.; naturalmente, $\omega_{\alpha/2}$ y $\omega_{1-\alpha/2}$ pueden ser positivos o negativos, según el nivel de probabilidad $\alpha/2$ y 1 - $\alpha/2$ que sea definido.

Lo anterior quiere decir que existe una probabilidad α de que X_T se encuentre por fuera del intervalo de confianza definido en la expresión (2).

En este caso ($X_T + \omega_{\alpha/2} . S_E(T)$) corresponde al denominado "límite inferior de confianza", y ($X_T + \omega_{1-\alpha/2} . S_E(T)$) al "límite superior de confianza", con una "confiabilidad" de (1- α)" (nivel de significancia α).

Estrictamente hablando, para calcular los valores ω se debe conocer la distribución de probabilidades de los valores de X para el período de retorno T en consideración (distribución marginal), la cual no obstante es desconocida, por lo cual se utiliza habitualmente la distribución "t" (Student), la cual tiene la ventaja de considerar el número de datos de la muestra X.

No obstante, algunos investigadores conceptúan que para propósitos prácticos aplicados en hidrología, para el cálculo de los límites de confianza del un evento X en el período de retorno T se puede utilizar la distribución de probabilidades normal [(1)], simplificando de esta manera la utilización de las diferentes herramientas de cálculo, aprovechando la existencia de similares condiciones de incertidumbre y precisión de los resultados.

Teniendo en cuenta esta consideración sobre la distribución probabilística de los valores de X para el período de retorno T, los intervalos de confianza para los estimados de un evento X en el período de retorno T utilizando la distribución de probabilidades normal, queda de la siguiente forma:

Límite inferior del intervalo = $X_T + z_{\alpha/2} . S_E(T)$

Límite superior del intervalo = $X_T + z_{1-\alpha/2} . S_E(T)$

Donde $z_{\alpha/2}$ y $z_{1-\alpha/2}$ corresponden a los valores de la variable Normal estandarizada o tipificada para probabilidades de no excedencia y de excedencia de $\alpha/2$ y 1- $\alpha/2$, respectivamente.

Si se adopta un nivel de significancia del 5% (α=0,05) normalmente utilizado en estudios hidrológicos, los límites de confianza quedarán de la siguiente forma:

Límite inferior del intervalo = $X_T - 1{,}96 . S_E(T)$

Límite superior del intervalo = $X_T + 1{,}96 . S_E(T)$

3. EXPRESIONES PARA EL CÁLCULO DE K_T Y $S_E(T)$ PARA DIFERENTES DISTRIBUCIONES DE PROBABILIDADES UTILIZADAS EN HIDROLOGÍA OBTENIDAS CON BASE EN EL MÉTODO DE LOS MOMENTOS

Se utilizan las expresiones derivadas en diferentes textos de estadística aplicada a la hidrología en el análisis de eventos extremos, obtenidas a partir de la aplicación de técnicas de estimación por el método de los momentos.

---

(1) KITE, G. W. Confidence Limits for Design Events. Water Resources Research, Vol 11, No. 1, pp. 48-53. 1975.

Distribución de probabilidades	**Expresiones para K_T y $S_E(T)$**
Normal	$K_T = z_T$ (*)
	$S_E(T) = S_X . \{[1 + (z_T)^2/2] /N\}^{1/2}$
Lognormal de dos parámetros (espacio real)	$K_T = \dfrac{\exp\{z_T . [\ln(1 + C_{vX}^2)]^{1/2} - 0,5 . \ln(1 + C_{vX}^2)\} - 1}{C_{vX}}$
	$S_E(T) = S_X . \{ [1 + (C_{vX}^3 + 3 C_{vX})K_T + (1/4) (C_{vX}^8 + 6 C_{vX}^6 + 15 C_{vX}^4 + 16 C_{vX}^2 + 2)K_T^2] /N\}^{1/2}$
Eventos Extremos Tipo I, o Gumbel	$K_T = - 0,45 - 0,7797 \ln \{ - \ln[F(x)]\}$ (**)
	$S_E(T) = S_X . \{ [1 + 1,1396 K_T + 1,1 K_T^2] /N\}^{1/2}$
Pearson Tipo III	$K_T = z_T + (z_T^2 - 1) (g_X/6) + (1/3) (z_T^3 - 6z_T)(g_X/6)^2 - (z_T^2 - 1) (g_X/6)^3 + (z_T)(g_X/6)^4 - (1/3)(g_X/6)^5$
	$S_E(T) = S_X .\{ \{1 + g_X K_T + (K_T^2/2) (3 g_X^2/4 + 1) + (3 K_T)(W) (g_X + g_X^3/4) + 3 (W^2) (2 + 3 g_X^2 + 5 g_X^4/8)\}/N\}^{1/2}$ (***)

(*) z_T : Variable normal tipificada asociada a una probabilidad $p = 1/T$

(**) Recuérdese que para el caso de análisis de eventos máximos $p = 1 - F(x) = 1/T$, y para el análisis de eventos mínimos $p = F(x) = 1/T$. Por lo tanto, para el caso de análisis de ***máximos***, $\ln [F(x)] = \ln(1-1/T)$, mientras que para los eventos ***mínimos*** $\ln [F(x)] = \ln(1/T)$

(***) $W = (z_T^2 - 1)/6 + 4 g_X (z_T^3 - 6z_T)/ 6^3 - 3(g_X^2)(z_T^2 - 1)/6^3 + 4(g_X^3)(z_T)/6^4 - 10 (g_X^4)/6^6$

En el caso de las distribución de probabilidades lognormal de dos parámetros, en la tabla anterior se presentan las expresiones correspondientes al cálculo de X_T y $S_E(T)$ directamente en el espacio de los valores medidos de la variable X (no transformados). No obstante debe recordarse que para esta distribución el análisis se puede realizar también, en su totalidad, en el espacio de los logaritmos, caso en el cual ya no se utilizan el valor medio y la desviación típica de los valores de X, sino los parámetros estadísticos de sus logaritmos naturales (neperianos); en este caso se utilizan directamente las expresiones mostradas para el caso de la distribución normal, solo que X_T y $S_E(T)$ corresponden al valor esperado del estimado del logaritmo natural del evento para el período de retorno T y su correspondiente error estándar (en el campo de los logaritmos, por supuesto), de manera que los intervalos de confianza se determinan en el espacio de los logaritmos, de forma tal que el resultado en el espacio de medición de la variable original X se obtiene a partir del cálculo del antilogaritmo de los límites de confianza antes obtenidos.

En el caso de la distribución de probabilidades lognormal de dos parámetros, la variable $\mathbf{C_{vX}}$ representa el coeficiente de variación de la variable X, el cual se calcula como la relación entre la desviación típica y la media de X.

El caso de la popular distribución de probabilidades Log Pearson Tipo III es similar al de la distribución Lognormal, ya que también bastaría con utilizar las expresiones antes mostradas correspondientes a la distribución Pearson Tipo III, pero aplicada a los logaritmos de la variable de interés, de manera que los antilogaritmos de los resultados así obtenidos corresponden a los resultados de la distribución Log Pearson Tipo III.