MÓDULO 1

LEYES DE DISTRIBUCIÓN DE PROCESOS HIDROLÓGICOS

Autores:
Dr. Ing. Roberto Pizarro T.
Ing. Juan Pablo Flores V.
Ing. Claudia Sangüesa P.
Ing. Enzo Martínez A.

1. INTRODUCCIÓN

El presente documento fue extraído y reeditado del Instructivo N° 5 del libro "Elementos de Hidrología I" (Pizarro *et al*, 1986), el cual pretende entregar una metodología que permita predecir con cierta probabilidad los valores que puede tomar una variable hidrológica, en función de la información de que se disponga, planteándose lo anterior, en valores máximos probables, aplicando la ley de distribución de Gumbel, y asociado esto, a algunas pruebas de bondad de ajuste.

Se plantea la utilización de la ley de distribución de Gumbel, dado que ella ha demostrado poseer una adecuada capacidad de ajuste, a valores máximos de caudales, precipitación en distintos períodos de tiempo, aportaciones anuales, etc. Además, se entrega una prueba de bondad de ajuste de Kolmogorov-Smirnov, un cálculo del coeficiente de determinación, y ello asociado a un ejemplo práctico.

2. AJUSTE A LA FUNCION DE GUMBEL

2.1 Definición de la función de Gumbel

Una variable aleatoria E sigue una distribución de probabilidad de Gumbel, si:

$$F(X) = P[\varepsilon \le X] = e^{-e^{-d(x-u)}} \quad \text{con } -\infty \le x \le \infty \qquad (1)$$

y donde x presenta el valor a asumir por la variable aleatoria, con d y u parámetros y e base de los logaritmos neperianos.

Despejando x de (1), queda;

$$x = u - \frac{\ln(-\ln F(X))}{d} \qquad (2)$$

Para la determinación de los parámetros d y u, se utilizan las siguientes expresiones que los definen:

$$u = \bar{x} - 0.450047S \quad (3)$$

$$\frac{1}{d} = 0.779696S \quad (4)$$

donde:
$\bar{x}$ = media aritmética de la serie de datos considerados.
S = desviación típica de la muestra de datos considerados.

Los valores 0,450047 y 0,779696, son válidos para un número de cincuenta datos. Sin embargo, Heras (s/a), los señala como admisibles para cualquier tamaño de población, en virtud de la escasa relevancia que poseen. Luego, es posible determinar la función de Gumbel, con la información entregada precedentemente.

De la ecuación (2), se desprende que es dable encontrar, para una probabilidad determinada, un valor para la variable aleatoria. Así, si se le aplica una probabilidad, de al menos 0,9 y se obtiene un valor K, implica que en el noventa por ciento de los casos cabe esperar un valor de x ≤ K.

2.2 Determinación de la Probabilidad

Para conseguir definir la probabilidad implícita es preciso consignar dos conceptos previos, que son el período de retorno y la probabilidad de excedencia.

Período de Retorno: se define como el tiempo que transcurre entre dos sucesos iguales. Sea ese tiempo, T.

Probabilidad de Excedencia: es la probabilidad asociada al período de retorno.

Así,

$$P_{excedencia} = P(x) = \frac{1}{T}$$

En otras palabras, la probabilidad de que la variable aleatoria tome un valor igual o inferior a cierto número X, está dado por la función de distribución de probabilidad F (X).

$$F(x) = \int_{-\infty}^{x} f(x)dx = P(x \leq X) = 1 - \frac{1}{T}$$

luego, la probabilidad de que x sea mayor que X está dada por la función complementaria.

$$P(x > X) = 1 - F(X) = \frac{1}{T}$$

3. BONDAD DEL AJUSTE

Se entiende por bondad de ajuste, la asimilación de datos observados de una variable, a una función matemática previamente establecida y reconocida. A través de ésta es posible interpolar y extrapolar información; en otras palabras, predecir el comportamiento de la variable en estudio (Pizarro *et al*, 1986).

Para la estimación de la bondad de ajuste, existen variadas pruebas, las cuales poseen distinto grado de efectividad.

En el presente documento se entrega el test de Kolmogorov-Smirnov y el coeficiente de determinación (Cid ***et al***, 1990; Shao, 1990).

3.1 Test de Kolmogorov-Smirnov:

Para la aplicación del test señalado, es necesario determinar la frecuencia observada acumulada.

Para la frecuencia observada en el caso especial de Gumbel, se ordena la información de menor a mayor y se aplica:

$$F_n = \frac{n}{N+1}$$

donde:
Fn (x): frecuencia observada acumulada.
n: N° total de orden
N: N° total de datos.

En el caso de la frecuencia teórica acumulada, ésta se determina a través de la función de Gumbel.

$$F(x) = e^{-e^{-d(x-u)}}$$

Una vez determinadas ambas frecuencias, se obtiene el supremo de las diferencias entre ambas, en la i-ésima posición de orden, que se denomina D.

$$D = Sup\left|F_n(x)_i - F(x)_i\right|$$

Luego, asumiendo un valor de significancia, se recurre a la tabla de valores críticos de D en la prueba de bondad de ajuste de Kolmogorov-Smirnov, y considerando el tamaño de la muestra, se establece lo siguiente:

Si D < D tabla, se acepta que(.el ajuste es adecuado, con el nivel de con- fiabilidad asumido.

3.2 Coeficiente de Determinación.

Se encuentra definido por la siguiente expresión

$$R^2 = 1 - \frac{\sum (F_n(x))_i - F(x)_i)^2}{\sum (F_n(x)_i - \overline{F_n(x)}_i)^2}$$

donde:
R^2: Coeficiente de determinación $0 \leq R^2 \leq 1$
$\overline{F_n(x)}_i$: Media de las frecuencias observadas acumuladas.

El coeficiente de determinación señala qué proporción de la variación total de las frecuencias observadas, es explicada por las frecuencias teóricas acumuladas.

4. EJEMPLO PRACTICO

Ajuste a Gumbel:

Se desea conocer la ley de distribución de las precipitaciones máximas en 24 horas, de la estación Monte Patria provincia de Limarí. Para ello, se dispone de los siguientes datos.

CUADRO N' 1. Precipitaciones máximas en 24 horas.

AÑO	**70**	**71**	**72**	**75**	**77**	**78**
Pp máxima en 24 hrs.	18.0	35.5	47.5	65.0	21.0	30.0
AÑO	**79**	**80**	**81**	**82**	**83**	**84**
Pp máxima en 24 hrs.	3.5	56.0	40.0	42.5	78.0	82.0

De lo expuesto, se deduce que se cuenta con una información de doce años, y además que los montos denotan una extrema variabilidad.

En relación al primer aspecto, es un denominador común en muchas estaciones del país, la carencia de series hidrológicas consistentes, por lo cual es difícil soslayarlo. En cuanto a la variabilidad, es preciso destacar que las zonas áridas se caracterizan por presentar este elemento como característica de la distribución y monto de las precipitaciones.

No obstante lo anterior, y como se tiende a estimar valores máximos, se puede obviar este último aspecto considerando las dos o tres precipitaciones máximas anuales, para con esta nueva serie de datos elegir un número mayor de años a considerar.

Luego, el enfrentamiento de este problema es resorte del criterio que el ingeniero utilice para tomar la decisión, y la cual sólo podrá ser calificada a la luz de los antecedentes que cada situación denote. Así, para el caso en cuestión, se trabajará con la información de precipitación máxima anual en 24 horas, toda vez que se trata de un ejercicio metodológico.

CUADRO N° 2. Datos para Gumbel y Bondad de Ajuste.

Precipitación máxima en 24 hrs.		**Frecuencia Relativa Acumulada**	**Frecuencia Teórica Acumulada**
1	3.5	0,077	0,090
2	18.0	0,154	0,114
3	21.0	0,231	0,158
4	30.0	0,308	0,320
5	35.4	0,385	0,427
6	40.0	0,462	0,513
7	42.5	0,538	0,557
8	47.5	0,615	0,639
9	56.0	0,692	0,753
10	65.0	0,769	0,839
11	78.0	0,846	0,916
12	82.0	0,923	0,932

Con los datos de la columna 1, se determina que:

$\bar{x}$ = 43.25
S = 23.97

Luego, los parámetros u y d quedan:

$u = \bar{x} - 0.450047S$ =43.25 – 0.450047 x 23.97 = 32.46

$d = \dfrac{1}{0.779696x23.97}$ = 0.0535

Por consiguiente, la función de Gumbel se define como:

$$F(X) = e^{-e^{-0.0535\cdot(x-32.46)}}$$

De lo expuesto, se deduce que se cuenta con una información de doce años, y además que los montos denotan una extrema variabilidad.

Por otra parte, aplicando la expresión n/N+1, se obtiene la frecuencia observada acumulada, la cual se expresa en la columna (2) del cuadro N° 2. Asimismo, reemplazando en la ecuación (1) los valores de x, se obtienen las frecuencias teóricas acumuladas, las cuales constituyen la columna (3) del cuadro N° 2.

Aplicación de Kolmogorov-Smirnov.

Con la información del cuadro N° 2, se busca el $Sup\left|F_n(x)_i - F(x)_i\right| = D$. En este caso, corresponde a D = 0.073 en el tercer valor del cuadro mencionado. Con un 95% de confiabilidad y n = 12, se obtiene un valor de tabla Dt = 0.375.

Luego D < Dt, por consiguiente se acepta con 95% de seguridad que el ajuste es bueno.

Aplicación del Coeficiente de determinación (R^2):

Utilizando la ecuación descrita en 3.2., y las columnas 2 y 3 del cuadro N° 2, queda:

$$R^2 = 0{,}988$$

Luego se considera que el modelo elegido, explica en un 98,8% las variaciones de las frecuencias observadas, lo cual es muy bueno.

Utilidad Práctica del Ajuste a Gumbel:

Una vez que se ha validado el ajuste a la función de Gumbel, resta definir la utilidad que esto puede determinar.

En este marco, si de la ecuación,

$F(X) = e^{-e^{-0.0535\cdot(x-32.46)}}$ se despeja x, queda:

$$x = 32.46 - \frac{\ln(-\ln F(x))}{0.0537}$$

Por consiguiente, para determinar los montos de precipitación en 24 horas, asociados a un período de retorno y a una probabilidad, se aplica la ecuación anterior, y se obtienen los X_i.

CUADRO N° 3. Precipitaciones máximas en 24 horas asociadas a un período de retorno.

Período de retorno T (años)	$1-\frac{1}{T} = F(x)$	X_i (pp máximas en 24 hrs.)
10	0.900	74.52
20	0.950	87.98
30	0.967	95.91
40	0.975	101.17
50	0.980	105.39
100	0.990	118.44

Luego, se puede deducir del cuadro anterior, que existe un 1% de probabilidad, de que sean superados los 118.44 mm. en 24 horas de precipitación, y lo cual corresponde a un evento centenario; en otras palabras, existe un 99% de probabilidades de que el año 1985, la precipitación en 24 horas sea menor o igual a 118.44 mm. Similar análisis, puede realizarse para todos los períodos de retorno involucrados.

No obstante lo anterior, se recomienda que los períodos de retorno considerados, no incluyan un número mayor de información que el doble o el triple como máximo, de la longitud de la serie de datos en estudio. En este caso, como la información base corresponde a 12 años, se recomienda no exceder de 24 años o un máximo de 36 años, dado que la serie estadística no presenta una longitud adecuada. Por ello, se recomienda el valor de T = 30, corno intermedio de lo señalado precedentemente; el considerar mayor número de años no posee sentido desde el punto de vista estadístico.

Conforme se recopile una mayor información, las predicciones a realizar poseerán mayor consistencia, y por ende, una mayor probabilidad de acierto. Por ello, año a año deben ser revisadas y completadas.

Por otra parte, se plantea que este tipo de estudios posee validez en un gran cúmulo de variables hidrológicas, y además, puede ser aplicado en cálculo de caudales máximos.

Finalmente, es importante señalar que sobre la metodología empleada puede surgir algún cuestionamiento, hecho totalmente válido, si se considera que no existen vías de solución de problemas totalmente definidas, por ello, lo expresado anteriormente más que una metodología, constituye una proposición.

5. REFERENCIAS BIBLIOGRÁFICAS.

- CID, L.; MORA, C.; VALENZUELA, M. 1990. Estadística matemática. Probabilidades e Inferencia Estadística. Universidad de Concepción. Facultad de Ciencias. Departamento de Matemáticas. Chile. 319 p.

- HERAS, R.s/a. Recursos Hidráulicos. Sintésis, metodología y normas. Cooperativa e Publicaciones del Colegio e Ingenieros e Caminos, Canales y Puertos. España, 361 p.

- PIZARRO, R.; NOVOA, P. 1986. Instructivo n° 5. Determinación de valores probabilísticos para variables hidrológicas. Elementos técnicos de Hidrología. Corporación Nacional Forestal (CONAF). Chile. 78 p.

- SHAO, S. 1970. Estadística para economistas y administradores de empresas. Editorial Herrero Hermanos, SUCS. S.A. México. 786 p.