DETERMINACIÓN DE LA PROBABILIDAD DE EXCEDENCIA EMPIRICA DE LOS CAUDALES ANUALES

Este análisis se realiza con la serie de caudales anuales. Dada una serie de caudales anuales con n años observados, esta se ordena de mayor a menor y se evalúa la probabilidad de excedencia para cada dato de caudal Q_i de la serie ordenada como:

$$F(Q_i) = p(Q \geq Q_i) = \frac{i - \alpha}{n + 1 - 2\alpha}$$

Ecuación 1

Donde i es la posición de ploteo del caudal Q_i en la serie ordenada de mayor a menor y n es el número de datos en la serie de caudales. A su vez i recorre valores desde 1 hasta n.

Se han adelantado interpretaciones prácticas de la Ecuación 1 por diferentes autores. Algunas de ellas están relacionadas con el tipo de distribución propuesta cómo modelo teórico, otras con la esperanza matemática de la probabilidad de excedencia estimada (considerando está característica cómo una variable aleatoria también). Entre las interpretaciones más conocidas se tienen las siguientes:

Tabla 1. Probabilidad de excedencia empirica

Nombre	α	Expresión
Weibull ; Kritskiy - Menkel	0	$F(Q_i) = \frac{i}{n+1}$ **Ecuación 2**
Gringorton	0,44	$F(Q_i) = \frac{i - 0.44}{n + 0.12}$ **Ecuación 3**
Hazen	0,5	$F(Q_i) = \frac{i - 0,5}{n}$

Nombre	α	Expresión
		Ecuación 4
Chegodaiev	0.3	$F(Q_i) = \frac{i - 0.3}{n + 0.4}$ Ecuación 5
Blókhinov	0.4	$F(Q_i) = \frac{i - 0.4}{n + 0.2}$ Ecuación 6
Distribución Gumbel	0.44	$F(Q_i) = \frac{i - 0.44}{n + 0.12}$ Ecuación 7
Distribución normal	$3/8$	$F(Q_i) = \frac{i - 3/8}{n + 1/4}$ Ecuación 8

Las ecuaciones presentadas en la Tabla 1 producen resultados que tienen diferencias en los extremos de la función de distribución para series de datos con longitudes menores a $n = 40$. En la práctica hidrológica actual la formula más utilizada es la Ecuación 6.

AJUSTE DE FUNCIONES DE DISTRIBUCIÓN TEÓRICAS A LA PROBABILIDAD DE EXCEDENCIA EMPIRICA DE LOS CAUDALES ANUALES

Con el fin de caracterizar a la muestra de caudales anuales que definen la oferta hídrica del territorio de interés es necesario ajustar una función de distribución teórica a la probabilidad de excedencia empírica hallada en el punto anterior. Existen varias funciones de distribución que ya han sido aplicadas en hidrología y dentro de las cuales se debe escoger la que mejor representa a la probabilidad de excedencia definida empíricamente. Entre los modelos teóricos de distribución probabilística se encuentran:

- Distribución normal;

Una variable aleatoria Q está distribuida normalmente, con media $\bar{Q}$ y varianza σ_Q^2 si tiene una densidad de distribución probabilística del tipo:

$$p(Q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left[(Q-\bar{Q})/\sigma_Q\right]^2}; \; (-\infty < Q < \infty);$$

Ecuación 9

Dado que el teorema central del límite establece que la suma de n variables aleatorias independientes se distribuye normalmente es factible que en algunos casos el caudal promedio anual (que representa la suma de los caudales medios diarios divididos por el número de días del año) resulte distribuido en forma normal.

- Distribución Log-normal;

Una variable aleatoria Q está distribuida de forma Log – normal si $Y = Log(Q)$ sigue una distribución normal con media μ_Y y desviación estándar σ_Y; si se ajusta a la expresión:

$$p(Q) = \frac{1}{Q\sigma_Y\sqrt{2\pi}} e^{-(1/2)\left[(\ln Q-\mu_Y)/\sigma_Y\right]^2} \quad \forall Q > 0$$
$$p(Q) = 0; \quad \forall Q \leq 0.$$

Ecuación 10

El valor esperado para esta distribución es: $E(X) = \mu_X = e^{\mu_Y + (1/2)\sigma_Y^2}$ y su varianza se obtiene cómo : $V(X) = \sigma_X^2 = e^{2\mu_Y + \sigma_Y^2}(e^{\sigma_Y^2} - 1)$. Esta distribución tiene como ventaja sobre la distribución normal que sus valores están restringidos a los valores positivos y que su forma no es simétrica, dos características que le permiten representar mejor a las series de caudales anuales.

- Distribución Gamma;

Una variable aleatoria Q tiene distribución Gamma si su función de distribución tiene la siguiente forma:

$$p(Q) = \frac{\lambda}{\Gamma(r)}(\lambda Q)^{r-1} e^{-\lambda x}; \forall Q > 0$$
$$p(Q) = 0; \quad \forall Q \leq 0.$$

Ecuación 11

Aquí: Γ(n) es la función Gamma que se define cómo: $\Gamma(n) = \int_0^\infty Q^{n-1} e^{-Q} dQ$.

Los parámetros de esta distribución son: λ, que se conoce cómo el parámetro de forma y r, que es reconocido como el parámetro de escala. A Través de estos parámetros el valor esperado y la varianza se representan cómo $\bar{Q} = r/\lambda$ y $\sigma_Q = \frac{\sqrt{r}}{\lambda^2}$. (Montgomery and Runger, 2003).

- Distribución Weibull.

Una variable aleatoria Q sigue una distribución Weibull si su función de densidad probabilística es:

$$p(Q) = \frac{\beta}{\delta}\left(\frac{Q-\gamma}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{Q-\gamma}{\delta}\right)^{\beta}\right]; \forall Q > \gamma$$

$$p(Q) = 0; \quad \forall Q \leq 0.$$

Ecuación 12

Los parámetros de esta distribución son el parámetro de localización γ, el parámetro de escala δ y el parámetro de forma β. El valor esperado y la varianza de una variable aleatoria Weibull son:

$$E(X) = \gamma + \delta\Gamma(1+1/\beta) \text{ y } V(X) = \delta^2\left\{\Gamma(1+2/\beta) - \left[\Gamma(1+1/\beta)\right]^2\right\}.$$

- Familia de Curvas de Pearson;

Existe un amplio conjunto de curvas de densidad probabilística que contiene a todas las curvas de densidad probabilística descritas anteriormente. Este conjunto de curvas se describe a través de la siguiente ecuación diferencial:

$$\frac{dp(Q)}{dQ} = \frac{(Q-a)p(Q)}{b_0 + b_1 Q + b_2 Q^2}$$

Ecuación 13

Entre la familia de curvas de Pearson sobresale la ecuación de Pearson Tipo III, para la cual el parámetro $b_2 = 0$. Esta subfamilia de la familia de curvas de Pearson incluye distribuciones importantes como las distribuciones Binomial y Gamma.

Para validar el ajuste de la distribución teórica a los datos empíricos se deben tomar en cuenta los siguientes criterios:

- Análisis visual;
- Error medio absoluto de ajuste;
- Validación de hipótesis de concordancia.

Aunque el análisis visual es subjetivo no se debe descartar. Usualmente si los criterios del error medio absoluto y de validación de hipótesis de concordancia están bien realizados la concordancia visual de las funciones de distribución empírica y teórica serán evidentes. Como error medio de ajuste máximo se recomienda el 15%. El error medio absoluto de ajuste se calcula cómo:

$$\varepsilon[\%] = \frac{|F^{*}(Q) - F(Q)|}{F^{*}(Q)} 100$$

Ecuación 14

Donde $F^{*}(Q)$ es la función de distribución empírica ó probabilidad de excedencia empírica.

El ajuste de una función de distribución teórica a los datos de probabilidad de excedencia empíricos se puede realizar a través del método de los momentos, o del de máxima verosimilitud.

Los métodos habituales para el ajuste de una función de distribución teórica a los datos de probabilidad de excedencia empíricos son los de momentos (MOM), máxima verosimilitud (ML) y momentos ponderados probabilísticamente (PWM), siendo menos utilizado el de mínimos cuadrados a partir de una de las formulas de gráfico anteriormente indicada. El método de MOM (en el espacio real o en el logarítmico LMOM) obtiene el valor de los parámetros, planteando el sistema de ecuaciones que resulta de igualar la expresión teórica de los momentos de la población en función de los parámetros y las estimaciones de dichos momentos obtenidos a partir de la muestra. La resolución del sistema con tantas ecuaciones como parámetros a estimar, permite obtener la distribución que reproduce los momentos de la muestra. Este método es fácil de aplicar pero no utiliza de forma exhaustiva toda la información contenida en la muestra. No obstante, en ciertos casos puede ser una buena elección por su menor sensibilidad ante elecciones incorrectas del modelo de distribución.

El método de ML obtiene el valor de los parámetros que maximizan la probabilidad de que se presente la muestra observada. Para ello se deriva el funcional de máxima verosimilitud (o su logaritmo) respecto a los distintos parámetros y se resuelve el sistema que resulta de igualar a cero dichas derivadas, calculándose el valor de los parámetros. Este método es considerado habitualmente como el más eficiente, es decir aquel que produce una varianza menor a los parámetros estimados, pero es bastante sensible a una incorrecta elección del modelo de distribución.

El método de PWM fue desarrollado por Greenwood et al. (1979) para distribuciones en las cuales la función de distribución $F(Q)$ sea expresable de forma explícita en forma inversa, es decir Q= Q(F), por lo que no es aplicable a la ley de distribución Log-Pearson tipo III. Este método calcula unas funciones lineales (los momentos ponderados probabilísticamente) de los datos e iguala dichas cantidades con las expresiones teóricas en función de los parámetros para la ley considerada, de forma análoga al método de los momentos. Este método confiere mayor peso a los mayores valores de la serie, resultando valores más conservadores.

Para los ejemplos de la presente guía se aplica el método de los momentos como primera aproximación que luego es ajustada a través de un método de optimización de gradiente conjugado que busca los parámetros de la distribución teórica que minimizan alguna función objetivo, por ejemplo el error cuadrático o el error medio absoluto presentado en la Ecuación 14.

Una vez seleccionada y calculada una función de distribución teórica se evalúa la hipótesis de su correspondencia con la función de distribución empírica. Para verificar esta hipótesis de concordancia se deben aplicar las pruebas de Pearson (χ^2), Kolmogorov, Smirnov (ω^2). Cómo requisito mínimo se exige que la hipótesis nula no sea rechazada por al menos dos de las pruebas propuestas. En esta evaluación de la bondad de ajuste como hipótesis nula se establece H_0: la función de distribución teórica corresponde a la empírica. Como hipótesis alternativa se establece. H': la función de distribución teórica no corresponde a la empírica.

Prueba de Kolmogorov λ

La prueba de Kolmogorov verifica la bondad de ajuste entre la función de distribución teórica -$F(Q)$- y la empírica -$F^*(Q)$- a través del estadístico λ. Para llegar a este estadístico se determina la diferencia máxima "D" entre las ordenadas de la función teórica y empírica cómo:

$$D = max|F^*(Q) - F(Q)|$$

Ecuación 15

A su vez el estadístico λ se determina cómo:

$$\lambda = \mathrm{D}\sqrt{\mathrm{n}};$$

Ecuación 16

Donde n representa la longitud de la serie de caudales anuales.

A continuación, de la distribución de Kolmogorov se establece el valor crítico λ_q. Si $\lambda \leq \lambda_q$ la hipótesis nula no se rechaza y la distribución teórica propuesta concuerda con la distribución empírica. La distribución del criterio de Kolmogorov se puede aproximar cómo:

$$F(\lambda) = 1 - \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2\lambda^2}$$

Ecuación 17

En forma de resumen los valores críticos λ_q se pueden obtener de la siguiente Tabla:

Tabla 2. Valores críticos λ_q para el criterio de Kolmogorov

Nivel de Significación α	0.40	0.30	0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Valor Crítico λ_q	0.89	0.97	1.07	1.22	1.36	1.48	1.63	1.73	1.95	2.03

Prueba de Pearson (χ^2)

La prueba de Pearson compara las frecuencias observadas y teóricas a través del estadístico:

$$\chi^2 = \mathrm{n} \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{(p^*(Q) - p(Q))^2}{p(Q)}$$

Ecuación 18

Donde n es el número de datos de la muestra y k es el número de clases en el histograma de frecuencias. χ^2 sigue una distribución Ji-cuadrado con $\nu = k - r - 1$ grados de libertad. $r$es el número de parámetros necesarios para la función de distribución teórica. Valores pequeños de χ^2 muestran una buena concordancia entre las probabilidades empíricas y la función de distribución teórica para la que se evalúa el ajuste. Por el contrario valores grandes de χ^2 denotan discrepancia entre la función teórica y la probabilidad empírica. El valor crítico χ_q^2 se obtiene de las tablas de la distribución Ji cuadrado con base en los grados de libertad $\nu = k - r - 1$ y el nivel de significación α. Las tablas de la distribución Ji-cuadrado se pueden consultar en (Bendat & Piersol, 1986) o en (Tomas Morer, González Sabaté, Fernández Ruano, & Cuadros Margarit, 2004). Cuando $\chi^2 \geq \chi_q^2$ se debe rechazar la hipótesis nula (H_0: la función de distribución teórica corresponde a la empírica) y de lo contrario esta debe ser aceptada

Prueba de Cramer-Von Mises-Smirnov ($n\omega^2$)

Esta prueba aplica el estadístico:

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} [F^*(Q) - F(Q)]^2$$

Ecuación 19

Para $n \geq 40$ la distribución de $n\omega^2$ no depende del tipo de distribución teórica que se está ajustando y tiende a una distribución límite cómo la presentada en el siguiente cuadro:

Tabla 3. Valores críticos para el estadístico $n\omega^2$

Nivel de significación $q = p(n\omega^2 > z_q)100$	50	40	30	20	10	5	3	2	1	0.1
Valor crítico z_q	0.1184	0.1467	0.1843	0.2412	0.3473	0.4614	0.5489	0.6198	0.7435	1.1679

Si $n\omega^2 > n\omega_q^2$ la hipótesis nula sobre la concordancia de la función de distribución teórica con los datos empíricos se rechaza con un nivel de significación q.

Varios trabajos recomiendan a la función Gamma como una distribución teórica que describe en forma adecuada los caudales anuales (Haan T. C., 2002; Rozhdenstvenskiy & Chevotariov, 1974; Druzhinin & Sikan, 2001). Adicionalmente, en Colombia, una análisis sobre cerca de 420 estaciones hidrológicas señala a la distribución Gamma como la que con mayor frecuencia ofrece el mejor ajuste para series de caudales anuales medios, máximos y mínimos (Domínguez, Hassidoff, León, Ivanova, & Rivera, 2009), en función de estos resultados, sin eliminar la posibilidad de analizar otras distribuciones, se recomienda utilizar la distribución Gamma cómo la distribución base para el análisis estadístico de la oferta hídrica superficial en Colombia.