Estimación de Parámetros de una Distribución Weibull

Autor: Alberto González Almuiña

1. Distribución de Weibull

Para la caracterización estadística de la velocidad del viento se emplea comúnmente la distribución de Weibull, ya que ofrece unos buenos resultados en un amplio abanico de casos. Se puede obtener la probabilidad de que v_i tenga un valor menor o igual que v ($F(v) = P(v_i \leq v)$) mediante la función de distribución acumulada(ADF):

$$F(v) = 1 - e^{(-(\frac{v}{C})^K)} \tag{1}$$

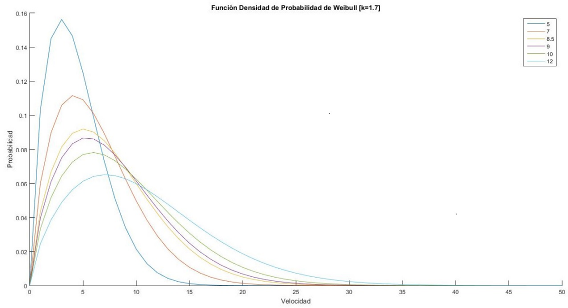
También es importante introducir la función de densidad de probabilidad (PDF), que da la probabilidad relativa de que una variable aleatoria continua (en este caso la velocidad del viento), tome un valor cercano a v. Se obtiene derivando la función de distribución acumulada respecto de la variable aleatoria, tomando el siguiente valor para la distribución de Weibull:

$$f(v) = \begin{cases} (\frac{K}{C}) \cdot (\frac{v}{C})^{K-1} \cdot e^{-(\frac{v}{C})^K} & v \geq 0 \\ 0 & v < 0 \end{cases} \tag{2}$$

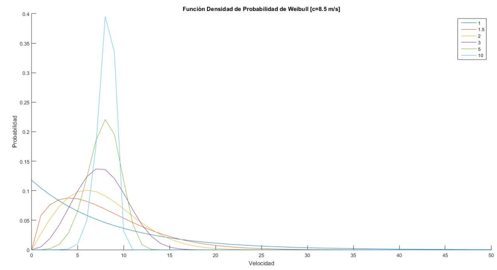
dónde:

- $K > 0 \rightarrow$ Parámetro de forma
- $C > 0 \rightarrow$ Parámetro de escala

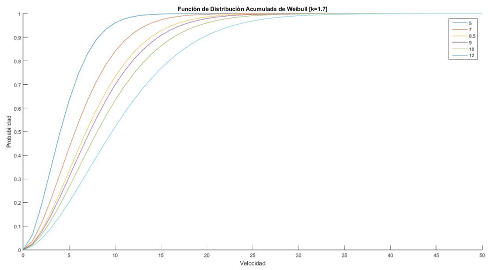
Si el régimen de viento se ajusta adecuadamente a la distribución de Weibull, estos dos parámetros serán fundamentales para poder caracterizar de una forma sencilla el comportamiento del viento en el emplazamiento. Para valores elevados de K, la distribución se asemeja a una campana de Gauss, con altas probabilidades en torno a los valores centrales y con pocos valores extremos. Al disminuir el parámetro K, la forma de la distribución se va abriendo, distribuyéndose las probabilidades centrales hacia la cola de la distribución. Por otro lado, existe una relación entre el parámetro de escala y la velocidad media, lo que implica que para valores de C pequeños, la probabilidad aumenta en velocidades pequeñas mientras se reduce para velocidades mayores. En cuánto a la distribución acumulada, al aumentar C, se observa como cada vez tarda más en alcanzarse la probabilidad total acumulada, lo que significa que las velocidades más grandes tienen mayor probabilidad que antes. Sin embargo, en el caso de variar K, se puede ver como prácticamente la probabilidad total acumulada se alcanza en valores muy cercanos, observándose claramente como cuando la probabilidad aumenta para velocidades mayores se reduce para pequeñas y viceversa. A continuación se muestran unas gráficas realizadas mediante el software Matlab dónde se ve la influencia de estos parámetros:



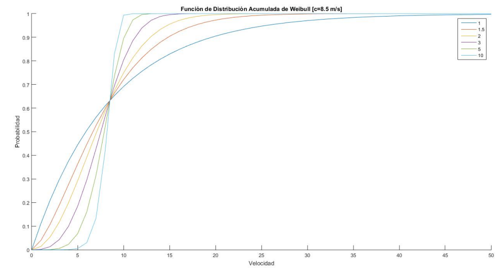
(a) PDF con K constante variando C



(b) PDF con C constante variando K



(c) ADF con K constante variando C



(d) ADF con C constante variando K

Figura 1: Gráficas ilustrativas sobre las funciones

Cómo se ha visto, es muy importante determinar los parámetros de la distribución para poder caracterizar correctamente el viento.

1.1. Método de mínimos cuadrados o gráfico

Este método se basa en la idea de linealizar la función acumulada para escribirla de la forma $y = a + b \cdot x$. Partiendo de la ecuación 1 y tomando logaritmos obtenemos:

$$ln(1 - F(v)) = ln(e^{-(\frac{v}{C})^K}) \Longrightarrow \ln(1 - F(v)) = -\left(\frac{v}{C}\right)^K$$

$$\Downarrow$$

$$\ln(-ln(1 - F(v))) = K \cdot ln(\frac{v}{C})$$

$$\Downarrow$$

$$\ln(-ln(1 - F(v))) = K \cdot ln(v) - K \cdot ln(C) \tag{3}$$

Haciendo el siguiente cambio de variables, se llega a la ecuación de una recta:

$$y = ln(-ln(1 - F(v))) \qquad\qquad x = ln(v)$$

Se puede representar en un eje de coordenadas los valores (x,y) y ajustar la gráfica mediante una recta cuya pendiente será $b = K$ y cuya intersección con el eje de ordenadas vendrá dada por $a = -K \cdot ln(C)$.

1.2. Método de los momentos

La función generadora de momentos del logaritmo de la distribución de Weibull es:

$$M_v(t) := E(e^{t \cdot v}), \quad t \in \mathbb{R} \longleftrightarrow \exists E$$

$$E\left[e^{t \cdot log(v)}\right] = C^t \cdot \Gamma\left(1 + \frac{t}{K}\right)$$

donde

$$\Gamma(z) = \int_0^\infty x^{z-1} \cdot e^{-x} \cdot dt \qquad \forall z > 0$$

A partir de la función generatriz de momentos se puede obtener el momento n-ésimo de v [2]:

$$E(v^n) = M_v^{(n)}(0) = \frac{d^n M_v}{dt^n}(0)$$

$$m_n = C^n \cdot \Gamma\left(1 + \frac{n}{K}\right)$$

El primer y el segundo momento coinciden con la media y la varianza de la velocidad (a partir de esta última se puede obtener la desviación estándar):

$$\overline{v} = E(v) = m_1 = C \cdot \Gamma\left(1 + \frac{1}{K}\right)$$

$$\sigma^2 = C^2 \cdot \left[\Gamma\left(1 + \frac{2}{K}\right) - \Gamma^2\left(1 + \frac{1}{K}\right)\right]$$

Dividiendo el uno por el otro se llega a la siguiente ecuación:

$$\frac{\sigma^2}{\overline{v}^2} = \frac{\Gamma\left(1 + \frac{2}{K}\right) - \Gamma^2\left(1 + \frac{1}{K}\right)}{\Gamma^2\left(1 + \frac{1}{K}\right)} \tag{4}$$

La ecuación 4 se puede resolver de manera recursiva mediante el método numérico de Newton-Raphson. Poniendo la ecuación de la forma $f(k) = 0$ se puede aplicar de forma iterativa:

$$g(K_n) = K_{n+1} = K_n - \frac{f(K_n)}{f'(K_n)}$$

Una vez calculado el parámetro K, C se obtiene simplemente sustituyendo en el primer momento (m_1):

$$C = \frac{\overline{v}}{\Gamma\left(1+\frac{1}{K}\right)}$$

1.3. Método de máxima verosimilitud

La función de verosimilitud es una función de los parámetros de un modelo estadístico que permite realizar inferencias de su valor a partir de un conjunto de observaciones. Se define como sigue:

$$L(\theta|x_1,...,x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

donde:

- x_i= observaciones
- θ= proporción

La idea es buscar el valor de θ que maximice (que haga más probable) las observaciones:

$$\frac{dL}{d\theta_i} = 0 \quad \text{ó} \quad \frac{dln(L)}{d\theta_i} = 0, \quad i = 1,...,n$$

Para la distribución de Weibull se tiene la siguiente función de verosimilitud [5]:

$$L(PDF|K,C) = \prod_{i=1}^{n}\left(\frac{K}{C}\right)\cdot\left(\frac{v_i}{C}\right)^{K-1}\cdot e^{-(\frac{v_i}{C})^K}$$

Tomando logaritmos y derivando respecto a los parámetros de la distribución se obtienen las siguientes ecuaciones:

$$\frac{\partial ln(L)}{\partial K} = \left(\frac{C}{K}\right) + \sum_{i=1}^{n} ln(v_i) - \frac{1}{C}\cdot\sum_{i=1}^{n}(v_i)^K\cdot ln(v_i) = 0$$

$$\frac{\partial ln(L)}{\partial C} = -\frac{n}{C} + \frac{1}{C^2}\cdot\sum_{i=1}^{n}(v_i)^K = 0$$

Haciendo unas operaciones se pueden despejar los parámetros de la distribución, K y C:

$$K = \left(\frac{\sum_{i=1}^{n} (v_i)^K \cdot ln(v_i)}{\sum_{i=1}^{n} (v_i)^K} - \frac{\sum_{i=1}^{n} ln(x_i)}{n} \right)^{-1} \tag{5}$$

$$C = \left(\frac{\sum_{i=1}^{n} (v_i)^K}{n} \right)^{\frac{1}{K}} \tag{6}$$

Estas ecuaciones no se pueden resolver de manera analítica, por lo que se aplica el método iterativo de Newton-Raphson, como se ha explicado en apartados anteriores.

1.4. Modificación del método de máxima verosimilitud

Gillette et al [3] modificaron el método anterior puesto que en la expresión del parámetro K este aparece a ambos lados de la igualdad y necesita de múltiples iteraciones para poder estimar un valor. Este método utiliza el estadístico β:

$$\beta = \frac{1}{(n-1)} \left[\frac{1}{n} \cdot \left(\sum_{i=1}^{n} (v_i)^m \right) \cdot \left(\sum_{i=1}^{n} (v_i)^{-m} \right) - 1 \right]; \quad m > 0$$

A continuación se define la esperanza de este estadístico y se realizan algunas operaciones:

$$E(\beta|m, K) = \frac{\pi \cdot m}{K} \cdot csc \left(\frac{\pi \cdot m}{K} \right)$$

$$\Downarrow$$

$$\beta = \frac{\pi \cdot m}{[K]} \cdot csc \left(\frac{\pi \cdot m}{[K]} \right), \quad [K] \rightarrow valor\ estimado\ del\ parámetro$$

$$\Downarrow$$

$$\frac{sen(\theta)}{\theta} = \frac{1}{\beta}, \quad siendo\ \theta = \frac{\pi \cdot m}{[K]}$$

Los valores de las desviaciones estándares de la estimación [K] se hacen pequeños cuando m$\rightarrow$0. Aproximando mediante una serie de Taylor:

$$\frac{sen(\theta)}{\theta} \simeq 1 - \frac{\theta^2}{6} \Longrightarrow \theta = \left[6 \cdot \left(1 - \frac{1}{\beta} \right)^{\frac{1}{2}} \right]$$

Se puede despejar el valor estimado de K a partir del cambio de variable dónde se ha definido θ:

$$[K] = \pi \cdot m \cdot \left[6 \cdot \left(1 - \frac{1}{\beta} \right)^{\frac{1}{2}} \right]$$

Aproximando m→0 y aplicando dos veces la regla de L'Hôpital se puede obtener un valor estimado del parámetro de forma que no necesita de métodos iterativos para ser calculado:

$$[K] = \frac{\pi}{\sqrt{6}} \left[\frac{n(n-1)}{n \cdot \left(\sum_{i=1}^{n} ln^2(v_i) \right) - \left(\sum_{i=1}^{n} ln(v_i) \right)^2} \right]^{\frac{1}{2}} \tag{7}$$

Una vez conocido el parámetro de forma, se puede obtener un valor para el parámetro de escala utilizando la ecuación 6.

1.5. Método de WasP

El algoritmo de WasP no emplea el histograma de frecuencias de las medidas, si no que utiliza la densidad de potencia media (WPD) asumiendo [1]:

1. La densidad de distribución de potencia es igual que la de frecuencias de las medidas.
2. La proporción de valores por encima de la velocidad media medida es la misma en ambas distribuciones.

Asumiendo ρ →constante, se tiene:

$$WPD(teórica) = \frac{1}{2} \cdot \rho \cdot C^3 \cdot \Gamma \left(1 + \frac{3}{K} \right)$$

$$WPD(medidas) = \frac{1}{2 \cdot n} \cdot \rho \cdot \sum_{i=1}^{n} v_i^3$$

Debido al requerimiento número 1, se tienen que igualar las dos ecuaciones anteriores y despejar el valor del parámetro de escala:

$$C = \sqrt[3]{\frac{\sum_{i=1}^{n} v_i^3}{n \cdot \Gamma \left(1 + \frac{3}{K} \right)}} \tag{8}$$

Para cumplir el requerimiento número 2 se define una variable X que representa la proporción de las velocidades medidas que superan la velocidad media observada:

$$X = 1 - F(\overline{v}) \Longleftrightarrow X = exp\left[-\left(\frac{\overline{v}}{C}\right)^K\right]$$

Tomando logaritmos y sustituyendo el parámetro de escala, se obtiene una ecuación a partir del a cuál se calcula el parámetro de forma:

$$-ln(X) = \left[\frac{\overline{v}}{\left(\frac{\sum_{i=1}^{n} v_i^3}{n \cdot \Gamma\left(1+\frac{3}{K}\right)}\right)^{\frac{1}{3}}}\right]^K \tag{9}$$

WasP resuelve la ecuación 9 de forma iterativa mediante el método de Brent, sin embargo, nosotros hemos resuelto esta ecuación con el método de Newton-Raphson. Una vez calculado el parámetro K, se sustituye simplemente en la ecuación 8 para obtener C.

1.6. Análisis estadístico

Para poder comprobar la validez de los métodos anteriores y poder compararlos entre sí, se necesita la ayuda de estimadores estadísticos que nos muestren, por un lado, la bondad de nuestros modelos y por otro el error entre los distintos métodos. Para estimar la bondad de los ajustes se han utilizado los siguientes estadísticos:

$$\chi^2 = \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N - n} \tag{10}$$

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{11}$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N}(y_i - x_i)^2\right]^{\frac{1}{2}} \tag{12}$$

donde:

- N →nº de observaciones
- n →nº de constantes (2 en este caso)
- y_i →frecuencia medida
- $\overline{y}$ →frecuencia media medida
- x_i →frecuencia teórica

Una vez calculados los parámetros para cada modelo, se ha ajustado la función de densidad de probabilidad (utilizando los valores calculados) a los datos medidos. Los estimadores RMSE y χ^2 indican un mejor ajuste cuánto menores son (siendo 0 un ajuste perfecto), justo lo contrario que sucede con R^2, que es mejor cuánto mayor sea (siendo 1 el ajuste perfecto).
Para comparar los resultados obtenidos se han utilizado los estadísticos ABias (Absolute Bias) y RBias (Relative Bias), que se definen del modo que sigue:

$$ABias = \frac{\sum_{i=1}^{N} |x_i - yi|}{N} \tag{13}$$

$$RBias = \frac{\sum_{i=1}^{N} \frac{x_i - y_i}{y_i}}{N} \tag{14}$$

dónde:

- N→ Nº de observaciones
- x_i → Valor i-ésimo de un parámetro
- y_i → Valor i-ésimo del otro parámetro a comparar

Tanto para la estimación de los parámetros de Weibull como para el cálculo de los estadísticos anteriormente descritos, se ha programado una función denominada WeibullParam mediante el software R [4]. Además, la función también nos proporciona un gráfico dónde se puede ver la distribución original de las medidas (en forma de histograma) junto con la distribución teórica calculada a partir de los parámetros obtenidos (representada sobre el histograma en forma de línea).

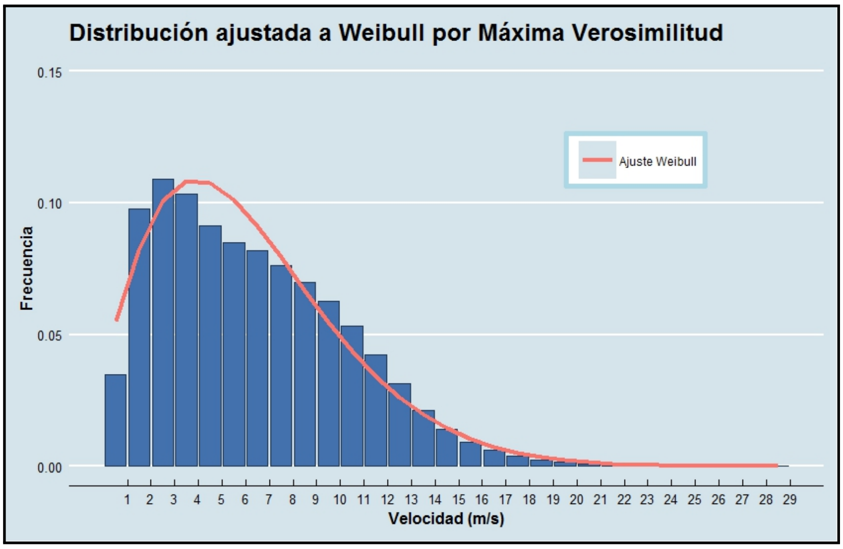


Figura 2: Gráfica de salida de la función WeibullParam

Referencias

[1] Ahmed, S.A. "Comparative study of four methods for estimating Weibull parameters for Halabja, Iraq", *International Journal of Physical Sciencies*, Vol.8 (5), 2013, pp. 186-192.

[2] Akdag, Seyit A. y Dinler, Ali. "A new method to estimate Weibull parameters for wind energy applications", *Energy Conversion and Management*, Vol. 50, 2009, pp. 1761-1766.

[3] Gillete, D.A. y Christofferson, R.D. "A simple estimator of the shape factor of the two-parameter Weibull distribution", *Journal of Climate and Applied Meteorology*, Vol 26 (5), pp. 323, 1987.

[4] R Core Team. "R: A language and environment for statistical computing", *R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/, 2017.

[5] Serrano Rico, J.C. "Comparación de métodos para determinar los parámetros de Weibull para la generación de energía eólica", *Scientia et Technica Año XVIII*, Vol. 18 (2), 2013.