IX CONGRESS CONGRESO
ISKO-SPAIN ISKO-ESPAÑA
Valencia 11th, 12th, 13th March 2009 11, 12 y 13 de Marzo de 2009
New Perspectives for the organisation and dissemination of knowledge
Nuevas perspectivas para la difusión y organización del conocimiento

MEDIDA DE LA OBSOLESCENCIA DEL CONOCIMIENTO. APLICACIÓN DE LA FÓRMULA DE INTERPOLACIÓN DE LAGRANGE

Antonio Pulgarín Guerrero
Universidad de Extremadura. Facultad de Biblioteconomía y Documentación. Departamento de Información y Comunicación. 06071 Badajoz (Spain). pulgarin@unex.es

María Isabel Escalona Fernández
Universidad de Extremadura. Unidad de Documentación. Biblioteca Central. Campus Universitario de Cáceres. 10071 Cáceres (Spain). escalona@unex.es

RESUMEN

La obsolescencia del conocimiento científico no es constante, sino que viene dada en función del tiempo. Por lo general, hay un incremento inicial del número de referencias o citas, seguido de una caída exponencial. Se han intentado ajustar distribuciones estadísticas, a la obsolescencia de la literatura científica, como la exponencial negativa, binomial negativa, Weibull, lognormal, etc. Creemos que ajustar distribuciones como las citadas no soluciona el problema, pues sería como tratar de encajar un traje previamente confeccionado a una persona. En el presente trabajo, el objetivo es otro, sería como confeccionar un traje a medida, para cada distribución de referencias o citas. Es un modelo diferente. Se trata de aplicar la fórmula de interpolación de Lagrange que, utilizando una serie de puntos de la distribución de referencias o citas, calcula un polinomio para cada distribución, a través del cual se puede determinar cualquier valor del segmento considerado por el polinomio. Mientras que, con los modelos anteriores, solo se pueden obtener valores de la tasa de obsolescencia para valores discretos de tiempo, con Lagrange podemos hacerlo de forma continua. Los resultados se ajustan más a la realidad, independientemente de la forma que tenga la gráfica o el polinomio correspondiente. Se han empleado tres revistas como material de estudio: *Anales de Documentación*, vol. 10 (2007); *Electronic Journal of Theoretical Physics*, volumen 4, n^{os}. 14, 15 y 16 (2007) y *Journal of Agricultural Meteorology* (versión online, japonesa), volumen 63, n^{os} 1 al 4 (2007). Las tres presentan una gráfica parecida, conforme con la hipótesis planteada.

ABSTRACT

The obsolescence of the scientific knowledge is not constant, but it is due to the time. Generally, there is an initial increase in the number of references or citations, followed

by an exponential fall. It has been tried to fit statistical distributions with the obsolescence of scientific literature, such as the negative exponential, negative binomial, Weibull, lognormal, etc. We think that to fit distributions as the mentioned above do not solve the problem, as it would be as to try to fit a suit previously made to a person. The objective is different in this work; it would be as to make a custom-made suit for each distribution of references or citations. It is a different model. It consists in applying the formula of interpolation of Lagrange that, using a series of points of the distribution of references or citations, calculates a polynomial for each distribution, with which it can be determinated any value of the segment considered by the polynomial. Whereas, with the previous models, it can only be obtained values of the obsolescence rate for discreet values of time, we can do it in a continuous form with Lagrange. The results adjust more to the reality, independently of the form that the graph or the corresponding polynomial has. Three journals have been used as material of study: *Anales de Documentación*, vol. 10 (2007); *Electronic Journal of Theoretical Physics*, vol. 4, numbs. 14, 15 and 16 (2007) and *Journal of Agricultural Meteorology* (version online, Japanese), vol. 63, numbs 1 to 4 (2007). The three displays a similar graph, according to the raised hypothesis.

PALABRAS CLAVE

Obsolescencia; factor de envejecimiento; literatura científica; función exponencial; interpolación de Lagrange.

INTRODUCCIÓN

Revisión de la literatura

Gosnell (1941), en uno de sus primeros artículos, trabajando con tres listas de referencias, de otros tantos libros, halló una semejanza entre las curvas de la obsolescencia y otras conocidas como "*curvas de decaimiento*". También, propuso el término "*tasa de obsolescencia*" y una fórmula exponencial para expresar la obsolescencia en el uso de libros.

Tres años más tarde, en otro artículo, Gosnell (1944) expresaba la obsolescencia de forma matemática e introducía el término "*omega*" [illegible], como la "*tasa de obsolescencia*", presentando la "*curva de decaimiento*" como la expresión de la obsolescencia:
$y = y_0 \cdot b^x$,[1]

donde y es el número de títulos, y_0 el número de títulos en el momento inicial y x sería el período de tiempo estudiado. Cuando [illegible], omega es la tasa anual de disminución en la curva, o tasa de obsolescencia. En este mismo trabajo, Gosnell, declara que existen "*amplias diferencias en las pendientes o tasas de decaimiento*", entre los diferentes campos científicos, y que tales diferencias "*tienden a variar con la velocidad de investigación y desarrollo del campo*". Crea la ecuación exponencial para la obsolescencia de la literatura científica y, desde este punto de partida, declara la existencia de una analogía entre la obsolescencia de la literatura científica y la desintegración de las sustancias radiactivas. El número de títulos que quedan del conjunto (y_r), después del tiempo [illegible], viene dado por $y_r = y_0 \cdot (-\ \ldots)$,[2]

Igualmente, utiliza el término "*half-life*" como índice de obsolescencia.

En 1958, en la "*Internacional Conference on Scientific Information*", celebrada en Washington, D.C., surgió, nuevamente, el problema de la obsolescencia de la literatura científica, refiriéndose a la corta duración de la "*vida media*" de ciertos campos científicos. King (1959), reflexionando sobre este congreso, dice que "*el promedio de la vida media de los documentos en algunas materias era tan corta como tres años…*"

Sin embargo, tanto la noción de "*half-life*" como la adopción de la física nuclear y su adaptación a la ciencia de la información han sido atribuidas, en la literatura citante, al bibliotecario Burton y al físico Kebler. En realidad, el trabajo de Burton y Kebler (1960) es una revisión crítica hacia la legitimidad de la expresión y de su uso en ciencia de la información "*el concepto de half-life es más familiar a los físicos e ingenieros nucleares, quienes lo emplean para describir el decaimiento de las sustancias radiactivas. Recientemente, no obstante, la expresión ha sido usada por documentalistas, algunos bibliotecarios y otros… Desafortunadamente, a diferencia* *del uso de la expresión por los físicos, que está limitada por una definición precisa, el uso por los documentalistas ha sido impreciso, sin verificar según la evidencia y, generalmente, sujeto a crítica*".

Burton y Kebler observan las diferencias y dificultades entre la desintegración nuclear y la obsolescencia de la literatura; declarando que la obsolescencia no puede ser medida: "*Desafortunadamente, esta obsolescencia no puede ser medida de forma precisa. A diferencia de una sustancia radiactiva que se convierte en otra sustancia diferente tras desintegrarse, la literatura simplemente llega a ser no usada, pero no inservible. Es obsolescente, pero no desintegrada*".

Burton y Kebler piensan que podrían resolver tanto estos problemas teóricos como las dificultades prácticas, redefiniendo el trabajo como: "*el tiempo durante el cuál la mitad de toda la literatura actual fue publicada…*". De esta forma se podría obtener una medida útil de una cantidad, de forma parecida a *half-life*.

Tras su primera contradicción con el principio de analogía con la física nuclear, Burton y Kebler mantuvieron su definición modificada y nunca fueron capaces de usar el término *half-life* de forma análoga al nuclear. Ambos autores determinaron las vidas medias en nueve campos de literaturas científicas y técnicas, y, de repente, en unas palabras finales, llegan a reconocer que la idea de una analogía, entre la *half-life* en la literatura y la de la física nuclear, tiene alguna validez. Pero al final del trabajo, Burton y Kebler afirman que "*no es la obsolescencia sino la longevidad de la literatura la que permite una función de tipo exponencial*".

Un año más tarde, Burton y Green (1961) publican un nuevo artículo en el que analizan el carácter y el estado de los informes técnicos publicados en cuatro revistas de física. En este trabajo, Burton intenta sustituir el término *half-life* por uno nuevo denominado *"median age*", que se refiere a la mediana del envejecimiento, al valor central de un conjunto de observaciones o al cuantil 50 % de una distribución.

Para más información sobre el origen del término *half-life*, su adopción de la física nuclear y adaptación a la ciencia de la información ver Száva-Kováts (2002).

En 1974 aparece la ya clásica y extensa revisión de Line y Sandison (1974), sobre obsolescencia, que recoge toda la literatura existente hasta ese momento sobre el tema.

Line y Sandison definen la obsolescencia "*como la caída a lo largo del tiempo de la validez o utilidad de la información*". La obsolescencia de la literatura científica se puede originar por diversas causas, entre las que podemos citar el crecimiento y el progreso de la ciencia, fundamentalmente. El conocimiento es registrado en documentos, por lo general, y estudiar los cambios en la validez o utilidad de los documentos ha sido lo usual, aunque la naturaleza de la relación entre el uso del documento y la validez de la información permanezca oscura, para estos autores. Por tanto, el estudio de la obsolescencia llega a ser, en términos prácticos, una interpretación de los cambios en el uso de documentos a lo largo del tiempo. Un artículo que nunca fue utilizado no puede ser obsoleto.

Así pues, la obsolescencia implica una relación entre uso y tiempo, pero los efectos del tiempo son revelados de diferentes formas, lo que hace que se confunda con facilidad. En primer lugar está la influencia del tiempo pasado, como lo expresan las edades de los ítems de la literatura cuando son estudiados. En segundo lugar, el paso del tiempo presente incrementa la edad de cada ítem. En tercer lugar, de gran interés práctico, está el efecto del tiempo futuro, que tratamos de pronosticar por extrapolación de los efectos del tiempo pasado y del tiempo presente.
El mismo año de la revisión anterior, Sandison (1974), tras analizar la densidad del uso de revistas de física en el MIT, considera que el fenómeno del envejecimiento no existe, provocando con ello una extraordinaria controversia en la comunidad científica.

Hasta 1981 no aparece otra revisión sobre la obsolescencia, tras la de Line y Sandison de 1974. En esta nueva revisión, Gapen y Milner (1981), se refieren a que "*la obsolescencia ha producido muchas fórmulas matemáticas pero que, desafortunadamente, ninguna simple ni universalmente aplicable*". Y añaden: "*después de 1944, año en que Gosnell publicara su principal trabajo, hubo más de dos décadas en que esta materia tuvo menos interés por el mundo bibliotecario. Que en los años 70 y más en los 80, debido a los ajustes presupuestarios, ha resurgido el interés por la obsolescencia. Pues el costo de las revistas ha hecho necesario cancelar suscripciones y esto, de alguna forma, ha beneficiado la vuelta a los estudios de la obsolescencia*".

Una de las últimas revisiones se debe a Line (1993), donde examina la relevancia del concepto de obsolescencia en un posible mundo donde las bibliotecas son virtuales. Line dice que no se conoce qué factores determinan que ítems, de una colección, sean leídos o citados, ni tampoco la relativa importancia que tiene el envejecimiento en todo ello.

En España, aunque no abunden los estudios, también se ha estudiado el fenómeno del envejecimiento u obsolescencia de la literatura científica, desde distintos puntos de vista. Entre ellos podemos destacar el de Terrada y col. (1979), donde analizan la información científica de las publicaciones médicas españolas por países e idiomas. Se trata de uno de los primeros trabajos sobre obsolescencia en España.

Casi dos décadas después, Ruiz Baños y Jiménez Contreras (1996) publican un trabajo donde aplican el modelo de Brookes a un conjunto de revistas especializadas en el ámbito de la documentación. Los resultados revelan la existencia de dos tipos de envejecimiento distintos: uno moderado, para las revistas de origen anglosajón, y otro más rápido para el resto de revistas.

Álvarez y col. (2002) introducen el término "*actualidad*" (topicality), como alternativo del factor de envejecimiento, considerándolo como una variable latente para ser medida con el modelo de Rasch. Se trata de un estudio diacrónico de 45 revistas del Subject Category Listing "*Physics*" de ISI, desde 1985 a 1994.

Recientemente, la ASLIB (2007) publicaba un artículo que relaciona la obsolescencia con el conocimiento. Discute las diferentes tasas y grados de obsolescencia, entre los diferentes tipos de conocimientos. Considera la idea de que el conocimiento, manejado por los profesionales de la información, puede llegar a ser obsolescente o que algunos tipos de conocimiento pueden llegar a ser más obsolescente que otros. Y, añade, que mientras algunos conocimientos llegan a ser totalmente obsoletos, debido a que las ideas que representan no se mantienen válidas a lo largo del tiempo, otros, en cambio, conservan cierta validez y utilidad a lo largo del tiempo.

Revisión de los modelos matemáticos sobre obsolescencia

En 1970, Brookes establece una ley matemática que describe la pérdida temporal de la utilidad de un conjunto de documentos, sentando las bases para los estudios de obsolescencia. En este estudio, Brookes, expresa su famosa ley exponencial negativa sobre la obsolescencia (Apéndice 1 de su trabajo), así como una serie de conceptos importantes como: utilidad de un volumen de una revista, factor de utilidad y factor de envejecimiento anual. Introduce el método gráfico de medición de la obsolescencia, así como un análisis de errores y recomendaciones en la toma de muestras.

Para Brookes, la distribución hacia la que tiende la medida de la obsolescencia es la exponencial negativa.
$f(t) = \frac{1}{\alpha} e^{-t/\alpha}$, $(0 \leq t < \infty)$, donde α es una constante que determina la tasa de decaimiento y t es la variable continua, tiempo.

La suma de la distribución anterior, desde cualquier valor arbitrario de t hasta infinito es

$$\int_t^{\infty} \frac{1}{\alpha} e^{-t/\alpha} dt = \left[-e^{-t/\alpha} \right]_t^{\infty} = e^{-t/\alpha}.$$

Si hacemos $e^{-1/\alpha} = a$, entonces $e^{-t/\alpha} = a^t$. Esta a^t expresa la fracción de la distribución de la cola que comienza en t y se extiende hasta infinito. Si le vamos dando valores a t igual a 0, 1, 2, 3,…, obtenemos la serie geométrica $1, a, a^2, a^3, \ldots$

Al representar $T(t)$, número de referencias con una antigüedad de t o más años, se tiene que $T(t) = R \cdot a^t$, donde $R = T(0)$. Esto es, el número total de referencias de la muestra. Tomando logaritmos, se tiene $\log T(t) = \log R + t \cdot \log a$.

Si representamos los valores de $T(t)$, en una escala logarítmica, frente a t, en escala lineal, la serie de puntos tenderá hacia una línea recta, si la obsolescencia sigue la exponencial negativa. Esta línea recta tendrá por pendiente $\log a$, que es negativa por ser $a < 1$. El valor de a puede ser determinado gráficamente, y cualquier valor de a^t puede ser leído desde el gráfico.

Cuando se conoce a, se puede estimar el valor de la vida media α desde la relación $1/\alpha = \ln(1/a)$.

El valor de la mediana o *half-life,* valor de t cuando $a^t = 1/2$ (50 %), también puede ser leído directamente desde la referida gráfica.

En cuanto al concepto de utilidad, Brookes afirma que si la obsolescencia de una revista se ajusta a una exponencial negativa, entonces si se hacen C referencias a uno de los volúmenes de esa revista durante su primer año de vida, en el segundo año se espera que haya Ca referencias, en el tercero el número de referencia será de Ca^2, en el año t tendrá Ca^{t-1} referencias y así sucesivamente.

Si se le da, a un nuevo volumen de la revista (t = 0) una esperanza de vida infinita, el número total de referencias que se esperan sean hechas a ese volumen desde la fecha de publicación hacia delante, la utilidad total $U(0)$, viene dado por

$$U(0) = C + Ca + Ca^2 + \dots + Ca^{t-1} + \dots\infty = C(1 + a + a^2 + \dots)$$

Como $a < 1$, esta suma infinita converge hacia un límite finito y la ecuación puede ser rescrita como

$$U(0) = C + Ca(1 + a + a^2 + \dots) = C + a \cdot U(0)$$

De aquí,

$$U(0) = C/(1-a) = u \cdot C, \qquad [3]$$

donde $u = 1/(1-a)$ es el factor de utilidad de una revista.

Entonces, el factor de utilidad depende solo del factor de envejecimiento anual.

Cuando un volumen de una revista se corresponde con el año t, el número de referencias que se pueden esperar es de Ca^t. En sucesivos años el número de referencias irá cayendo a Ca^{t+1}, Ca^{t+2} y así sucesivamente.

La utilidad que le queda a la revista en el año t $U(t)$ es

$$U(t) = Ca^t + Ca^{t+1} + Ca^{t+2} + \dots\infty = a^t C(1 + a + a^2 + \dots\infty) = a^t \cdot U(0) \qquad [4]$$

Si t fuera el número de años a los que se descarta la revista (se quita), la ecuación anterior se puede refundir en la siguiente

$$a^t = \frac{U(t)}{U(0)} = \frac{utilidad\ \ de\ \ descarte}{utilidad\ \ inicial}$$, que es de aplicabilidad general. [5]

En 1973, Brookes, afirmaba que la teoría del proceso de representación de la utilidad frente a la edad, la cual genera una línea recta, deriva del hecho de que la linealidad indica que las citas o referencias contabilizadas año a año pueden ser expresadas como una secuencia geométrica

$R, Ra, Ra^2, Ra^3 \dots Ra^{t-1} \dots,$

donde R es el número de citas o referencias durante el primer año, algunas de las cuales no emergen de inmediato en publicación.
Pero como $a < 1$, la suma de esa serie converge hacia el límite finito $R/(1-a)$.
De esa forma, la secuencia dibujada, que es la suma decreciente de citas o referencias de años previos, se puede expresar como

$$U(0) = R(1 + a + a^2 + \ldots + a^{t-1} + \ldots) = R/(1-a).$$

$$U(1) = R(a + a^2 + a^3 + \ldots + a^{t-1} + \ldots) = Ra(1 + a + a^2 + \ldots + a^{t-1} + \ldots) = Ra/(1-a) = U(0)a^2$$

Y así sucesivamente.

Este resultado implica que la secuencia dibujada, $U(0), U(0)a, U(0)a^2, \ldots, U(0)a^{t-1}$ es, también, una serie geométrica de la misma razón a que la secuencia original.

De aquí, tras t años, la utilidad residual, $U(0)a^{t-1}$, es la fracción a^t de la utilidad inicial $U(0)$.
El valor de la fracción puede ser leído directamente de la gráfica, bien como fracción o como porcentaje.

Autores posteriores, como Griffith y col. (1979), han utilizado métodos alternativos basados en el de Brookes y han rebatido el trabajo de Sandison (1974), confirmando la existencia del fenómeno de la obsolescencia. En este trabajo los autores realizan un análisis de citas, determinando el factor de envejecimiento anual para numerosas revistas del Sciencie Citation Index. Afirman que, a nivel particular, quienes trabajan constantemente con estos modelos pueden que cuestionen el papel de la acumulación en el modelo, cuando ellos tienen que reconocer que el modelo genera una serie: $U(0)a^0, U(0)a^1, U(0)a^2, U(0)a^3$

En realidad, para Giffitth y col., las citas o referencias en los sucesivos años tienen una relación *1/a,* de la forma siguiente:

$$\frac{1}{a} = \frac{U(0)a^n - U(0)a^{n+1}}{U(0)a^{n+1} - U(0)a^{n+2}} = \frac{a^n - a^{n+1}(\;)}{a^n - a^{n+1}(\;)}, \qquad [6]$$

Años después, Egghe y Rao (1992), volviendo al fenómeno de la obsolescencia, a través de estudios más meticulosos y complejos, afirman que el envejecimiento no es una constante, sino que está en función del tiempo, lo que hace de este factor una medida útil. Muestran que la función a (factor de envejecimiento) tiene un mínimo, que es obtenido en el tiempo t, después del tiempo en que se alcanza el máximo del número de citas. Estudian los máximos y mínimos de una serie de funciones aplicadas a una serie de distribuciones, constituidas por las referencias de tres libros distintos; ajustándose en los tres casos a una distribución de tipo log-normal. El ajuste lo realizan con el test de Kolmogorov-Smirnov, test de bondad de ajuste no paramétrico (α = 0,01 y α =0,05).

Egghe y Rao añaden: a) que solo el factor *a* debe ser determinado, ya que los demás factores para medir obsolescencia como: "*hal-life*", "*factor de utilidad*", etc., son funciones de *a*; b) que la mayoría de los estudios están basados en que la distribución del envejecimiento de las revistas citadas siguen una distribución exponencial; c) que si *t* representa el envejecimiento discreto de las revistas citadas y $C(t)$ al número relativo de revistas, entonces,

$$C(t) = b \cdot e^{-ct}, \; t \geq 0, b > 0 \qquad [7]$$

En base a lo expresado antes, se define el factor de envejecimiento, *a*, como:

$$a(t) = \frac{C(t+1)}{C(t)} \qquad [8]$$

De hecho, en [4] $a(t)$ es independiente de *t*.

$$a(t) = \frac{b \cdot e^{-c(t+1)}}{b \cdot e^{-ct}} = e^{-c}, \text{ que es el factor de envejecimiento anual, } a = e^{-c}. \qquad [9]$$

Asumiendo [7] y [9] se tiene $C(t) = b \cdot a^t$. [10]

Pero, para Egghe y Rao, la realidad no está de acuerdo con que exista una caída exponencial desde el principio, sino que existe un incremento inicial de citas, seguido de una corta caída exponencial.

Como consecuencia de esto no hay forma de encontrar un factor de envejecimiento, independiente del tiempo; solo en el caso de que se trate de una distribución de tipo exponencial. Para los demás casos *a* sería una función del tiempo.

Una forma elegante de solucionar este problema la ofreció Brookes (1973). Su método parte de [10], indicando con *m* el número total de citas o referencias a publicaciones con una antigüedad igual o mayor a *t* años.

De aquí,
$m(t) = ba^t + ba^{t+1} + \ldots = a^t(b + ba + ba^2 + \ldots) = a^t T$, indicándose con *T* el número total de citas o de referencias y, por lo tanto,

$$a = \left(\frac{m(t)}{T}\right)^{1/t} \qquad [11]$$

En un trabajo posterior, Egghe (1993) se centra en la influencia del crecimiento en la obsolescencia, obteniendo diferentes resultados para los estudios sincrónicos que para los diacrónicos. Para el caso de estudios sincrónicos, un incremento del crecimiento implica un crecimiento de la obsolescencia, mientras que para el caso diacrónico el mecanismo que encuentra es el opuesto. Estas pruebas se basan en modelos de crecimiento y de obsolescencia exponenciales. En este trabajo Egghe hace referencia a la nota editorial de Vickery, aparecida en el trabajo de Line (1970), en la que decía que "*la obsolescencia es una función del crecimiento…*".

Como se ha podido observar, durante esta breve descripción, han sido varios los intentos por medir la tasa de obsolescencia de la literatura científica. Hoy, ciertos conceptos, están más claros que lo estaban hace unas décadas. Así, sabemos que la tasa de obsolescencia es dependiente del tiempo, esto es, que varía de unos años a otros, por ejemplo. Por lo tanto, no podemos considerar una única tasa para la totalidad de un período de tiempo considerado, deberemos calcularla para cada unidad de tiempo.

El método propuesto por Brookes (1973) puede ser una forma de calcular la obsolescencia del conocimiento. La distribución log-normal de Egghe y Rao (1992) pudiera ser otra solución. Ambas tendrían en consideración el tiempo.

Introducidos en esta dinámica, y con objeto de aportar nuestra contribución a la consecución de una medida más propicia del fenómeno de la obsolescencia, a igual que lo hicieran otros autores, anteriormente, proponemos un método no utilizado hasta ahora, con objeto de intentar dar solución al problema. Este método es un clásico dentro del cálculo numérico, muy conocido para interpolar datos. Se trata de la fórmula de interpolación de Lagrange (1736-1811).

FÓRMULA DE INTERPOLACIÓN DE LAGRANGE

Presentación del modelo

Supongamos que el fenómeno a estudiar sea el envejecimiento de la literatura científica, fenómeno en el que se ha demostrado (Egghe y Rao, 1992) la existencia de una dependencia funcional entre las magnitudes ***t*** (=x) y ***a*** (=y), dependencia que caracteriza el aspecto cuantitativo de dicho fenómeno.

La función $y = f(t)$ es desconocida, sin embargo, mediante una serie de experiencias, se pueden determinar los valores de esta función, $y_0, y_1, y_{2,...}$, para ciertos valores de ***t:*** $t_0, t_1, t_2,...,t_n$, pertenecientes al segmento [***a***, ***b***].

El problema consiste en hallar la función más simple, para facilitar los cálculos (sea un polinomio, por ejemplo), que sea la expresión exacta o aproximada de la función desconocida $y = f(t)$, en el segmento [***a***, ***b***].

De una forma más abstracta, el problema se puede formular del modo siguiente: los valores de una función desconocida $y = f(t)$, se dan en *n+1* puntos diferentes: $t_0, t_1, t_2,...,t_n$, del segmento [***a***, ***b***].

$y_0 = f(t_0);\ \ y_1 = f(t_1);\ \ y_2 = f(t_2);\ ,\dots,\ y_n = f(t_n).$

Es preciso hallar un polinomio $P(t)$, de grado inferior o igual a ***n***, que exprese, aproximadamente, la función $Q(t)$.

Para hacer esto elijamos un polinomio cuyos valores, en los puntos $t_0, t_1, t_2, ..., t_n$, coincidan con los correspondientes valores de $y_0, y_1, y_2 ..., y_n$ de la función $Q(t)$. En este caso, el problema planteado, llamado "*problema de interpolación de la función*", se puede formular del modo siguiente: hallar, para una función dada $Q(t)$, un polinomio $P(t)$, de grado $\leq n$, que tome, en los puntos dados $t_0, t_1, t_2, ..., t_n$, los valores $y_0 = Q(t_0)$; $y_1 = Q(t_1)$; $y_2 = Q(t_2)$; ,…., $y_n = Q(t_n)$.

Tenemos para esto un polinomio de n-ésimo grado y de la forma:

$$P(t) = C(t-t_1)(t-t_2).....(t-t_n) + C_1(t-t_0)(t-t_2).....(t-t_n) + C_2(t-t_0)(t-t_1)(t-t_3)....$$
$$(t-t_n) + ...+ C_n(t-t_0)(t-t_1).....(t-t_{n-1}).$$

Ahora determinemos los coeficientes $C_0, C_1, C_2, ..., C_n$, de tal manera que se cumplan las condiciones: $P(t_0) = y_0$, $P(t_1) = y_1$, $P(t_2) = y_2$,…, $P(t_n) = y_n$.

A continuación hacemos $t = t_0$, resultando lo siguiente:

$y_0 = C_0(t_0 - t_1)(t_0 - t_2).....(t_0 - t_n)$, de donde:

$$C_0 = \frac{y_0}{(t_0 - t_1)(t_0 - t_2)....(t_0 - t_n)}$$

Si ahora hacemos $t = t_1$, resultará:

$y_1 = C_0(t_1 - t_0)(t_1 - t_2).....(t_1 - t_n)$, de donde:

$$C_1 = \frac{y_1}{(t_1 - t_0)(t_1 - t_2)....(t_1 - t_n)}$$

De igual forma se procede con los siguientes:

$y_n = C_n(t_n - t_0)(t_n - t_1).....(t_n - t_{n-1})$, de donde:

$$C_n = \frac{y_n}{(t_n - t_0)(t_n - t_1)(t_n - t_2)....(t_n - t_{n-1})}.$$

Volviendo atrás y poniendo los valores de los coeficientes en la primera ecuación del polinomio, obtenemos:

$$P(t) = \frac{(t-t_1)(t-t_2)....(t-t_n)}{(t_0-t_1)(t_0-t_2)....(t_0-t_n)} \cdot y_0 + \frac{(t-t_0)(t-t_2)....(t-t_n)}{(t_1-t_0)(t_1-t_2)....(t_1-t_n)} \cdot y_1 +$$
$$+ \frac{(t-t_0)(t-t_1).(t-t_3)...(t-t_n)}{(t_2-t_0)(t_2-t_1).(t_2-t_3)...(t_2-t_n)} \cdot y_2 + ... + \frac{(t-t_0)(t-t_1)....(t-t_{n-1})}{(t_n-t_0)(t_n-t_1)....(t_n-t_{n-1})} \cdot y_n$$

Esta fórmula, tan larga, se llama "*fórmula de interpolación de Lagrange*".

Error en la interpolación de Lagrange

Admitamos, sin demostrarlo, que si $\varphi(t)$ tiene una derivada de (***n+1***)-ésimo orden en el segmento [***a***, ***b***], el error cometido, al reemplazar la función $\varphi(t)$ por el polinomio $P(t)$, es decir, la magnitud $R(t) = \varphi(t) - P(t)$ satisface a la siguiente igualdad:

$$|R(t)| < |(t-a_0)(t-a_1)\ldots(t-a_{n-1})(t-a_n)|\frac{1}{(n+1)!}\,máx\left|\varphi^{(n+1)}(t)\right|$$

T*eorema*: "Si los valores de dos polinomios $P_1(t)$ y $P_2(t)$, de grado ***n***, coinciden para (***n+1***) valores diferentes $a_0,\ a_1,\ a_2,\ldots,\ a_n$, del argumento ***t***, los polinomios enunciados son iguales".

Observación: Del teorema anterior se deduce que el polinomio $P(t)$ es el único que satisface a las condiciones del problema planteado.

EJEMPLOS DE APLICACIÓN

Comenzaremos dando detalle de los datos que hemos seleccionado para la aplicación. Estos datos se corresponden con las referencias bibliográficas obtenidas del análisis de las siguientes revistas:

1. *Anales de Documentación*, volumen 10 (2007). Revista con contenido en ciencia de la información/documentación y periodicidad anual (597 referencias).

2. *Electronic Journal of Theoretical Physics*, volumen 4, n^{os}. 14, 15 y 16 (2007). Revista electrónica internacional para la publicación rápida de artículos de investigación original, sobre todos los aspectos de física teórica (828 referencias).

3. *Journal of Agricultual Meteorology* (versión online), volumen 63, n^{os} 1 al 4 (2007). Revista electrónica, con publicaciones originales en inglés o japonés, sobre aspectos de agricultura y meteorología (348 referencias).

La razón por la que hemos elegido estas tres revistas ha sido para disponer de datos heterogéneos a analizar, ya que tanto las temáticas como los objetivos, periodicidad o zona de influencia, de cada una de ellas, son distintas.

Los datos se muestran en las tablas 1 a 3 (en el mismo orden que los datos bibliográficos indicados antes). Representamos con *t* la edad de las referencias (diferencia entre el año de publicación de la referencia y el de la fuente citante), *C(t)* el número de referencias con una edad *t*, *m(t)* el número total de referencias con una edad $\geq t$ y *a(t)* el factor de obsolescencia o de envejecimiento, calculado para cada *t*.

En los tres ejemplos se han representado los datos correspondientes a los tiempos $t_0, t_1, ..., t_9$, con objeto de recoger las mayores variaciones que tienen lugar, precisamente, en estos primeros años.

En los tres casos, se observa cierta irregularidad al ser representada *C(t)* frente a la edad (Figs. 1-3).

Lo mismo se puede observar respecto al factor de envejecimiento, *a(t),* calculado para estos primeros años (tablas 1-3).

Tabla 1.- Distribución de la edad de las referencias de *Anales de Documentación*

t	C(t)	m(t)	*a(t)*
0	19	597	
1	73	578	*0.9681*
2	71	505	*0.9197*
3	57	434	*0.8991*
4	67	377	*0.8914*
5	42	310	*0.8771*
6	30	268	*0.8750*
7	30	238	*0.8768*
8	25	208	*0.8765*
9	*11*	*183*	*0.8768*

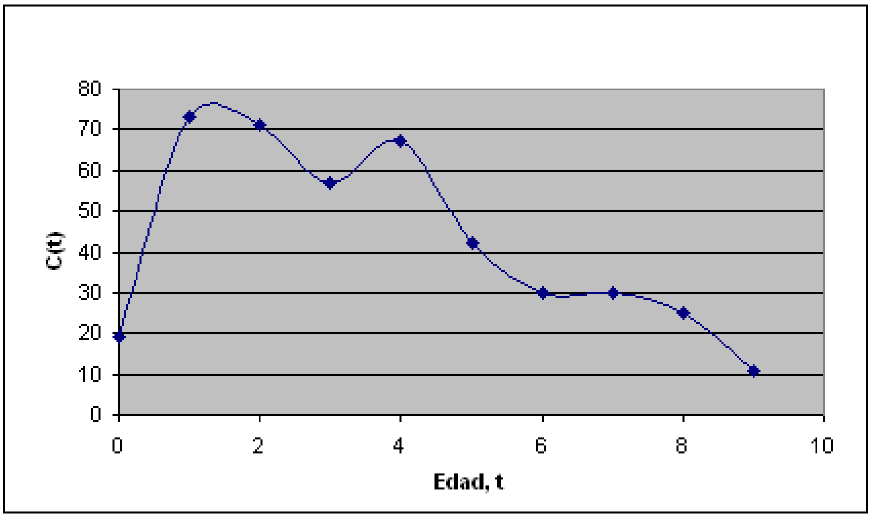

Figura 1.- Representación de *C(t)* frente a *t*, para la revista *Anales de Documentación.*

Tabla 2.- Distribución de la edad de las referencias de la revista *Electronic Journal of Theoretical Physics.*

t	C(t)	m(t)	a(t)
0	26	828	
1	72	802	0.9685
2	48	730	0.9389
3	45	682	0.9373
4	37	637	0.9365
5	45	600	0.9376
6	33	555	0.9354
7	36	522	0.9362
8	32	486	0.9355
9	30	454	0.9354

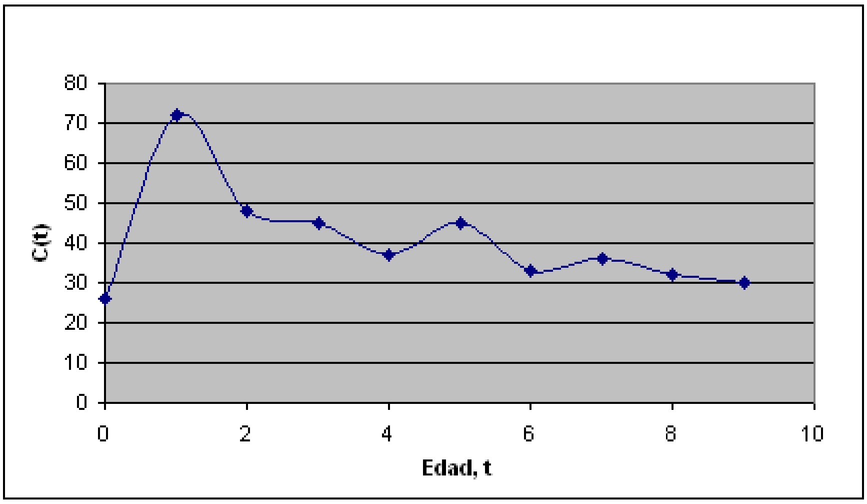

Figura 2.- Representación de *C(t)* frente a *t*, para la revista *Electronic Journal of Theoretical Physics.*

Tabla 3.- Distribución de la edad de las referencias de la revista *Journal of Agricultural Meteorology.*

t	C(t)	m(t)	a(t)
0	9	348	
1	34	339	0.9741
2	32	305	0.9361
3	37	273	0.9222
4	21	236	0.9074
5	22	215	0.9081
6	21	193	0.9064

7	14	172	0.9042
8	11	158	0.9060
9	19	147	0.9086

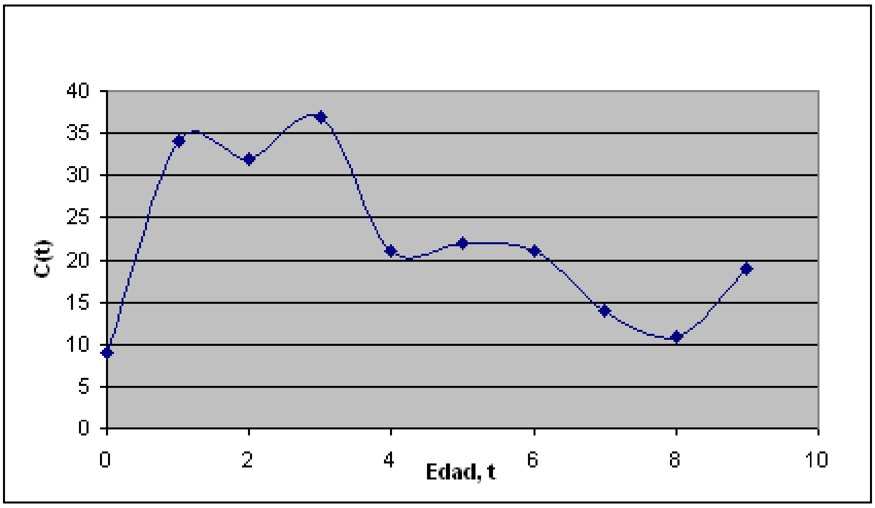

Figura 3.- Representación de *C(t)* frente a *t*, para la revista *Journal of Agricultural Meteorology.*

Estos resultados parecen estar conforme con la hipótesis planteada por Egghe y Rao (1992), esto es, un incremento en los primeros años (máximo) y un decaimiento a partir de esos primeros años, hasta alcanzar el mínimo.

Si en lugar de representar *C(t)* se representa *m(t)*, frente a *t*, el resultado es idéntico en los tres casos, presentando un buen ajuste ($R^2 = 0.99$) a una distribución exponencial. Esto es lógico, ya que al restar pequeñas cantidades a otra mucho mayor, siempre la forma de la curva va a ser de tipo exponencial.

Sin ser muy exigente, se podrían tolerar estas pequeñas diferencias y asumir que, en los tres casos, la obsolescencia sigue una distribución exponencial negativa, tal y como propusiera Brookes (1970). En ese caso, también habría que asumir que la tasa es constante e independiente del tiempo.

Pero si mostramos cierto grado de exigencia y no aceptamos esas diferencias como normales, deberemos buscar otra solución al problema.

Esta solución nos la podría ofrecer la fórmula de interpolación polinomial de Lagrange.

Para calcular los polinomios, se toman los valores de $t_0, t_1, t_2, ..., t_9$, y los correspondientes valores de $y_0, y_1, y_2 ..., y_n$, para cada caso en particular.

El cálculo de los polinomios, para los tres casos, se ha llevado a cabo utilizando Java Applet Window, como software.

1) Para *Anales de Documentación*, el polinomio resultante fue:

$$C(t) = \frac{(t-\ )(t-\ )....(t-\ )}{-\ 62880} \cdot 31 + \frac{(t-\ )(t-\ )....(t-\ )}{40320} \cdot 76 + \frac{(t-\ )(t-\ )(t-\ )...(t-\ )}{-\ 0080} \cdot 73 +$$
$$+ \frac{(t-\ )(t-\ )(t-\ )(t-\ )....(t-\ )}{4320} \cdot 52 + \frac{(t-\ )(t-\ )(t-\ )(t-\ )(t-\ )....(t-\ )}{-\ 880} \cdot 66 +$$
$$+ \frac{(t-\ )(t-\ )(t-\ )(t-\ )(t-\ )(t-\ )....(t-\ )}{2880} \cdot 46 + \frac{(t-\ )(t-\ )...(t-\ )(t-\ )....(t-\ )}{-\ 320} \cdot 29 +$$
$$\frac{(t-\ )(t-\ )....(t-\ )(t-\ )(t-\ )}{10080} \cdot 34 + \frac{(t-\ )(t-\ )....(t-\ )(t-\ )}{-\ 0320} \cdot 24 + \frac{(t-\ )(t-\ )....(t-\ )}{362880} \cdot 13$$

Y el polinomio reducido es:

$$C(t) = -\ ,0040t^9 + \ ,181t^8 - \ ,271t^7 + 2,101t^6 - 86,08t^5 +$$
$$+ \ 43,964t^4 - 273,946t^3 + 270,251t^2 - \ 38,196t + \ 1$$

Si en lugar de tomar como $y_0, y_1, y_2 ..., y_n$ los valores de *C(t)*, se toman los de *a(t)* multiplicados por 100, el polinomio reducido resultantes es:

$$a(t) = \ ,028t^7 - \ ,496t^6 + \ ,761t^5 - \ 6,912t^4 + \ 0,491t^3 - 73,386t^2 + 66,563t - 3,899$$

2) Para la revista *Electronic Journal of Theoretical Physics*, el polinomio reducido resultante, al tomar los valores de *C(t)* fue el siguiente:

$$C(t) = \ ,005t^9 - \ ,218t^8 + \ ,72t^7 - 4,738t^6 + 93,076t^5 -$$
$$- \ 50,888t^4 + 294,37t^3 - 389,255t^2 + \ 29,927t - \ 6$$

Y tomando los valores de *a(t) × 100*, se tiene el polinomio:

$$a(t) = -\ ,017t^7 + \ ,291t^6 - \ ,704t^5 + 4,982t^4 - 0,511t^3 + 00,96t^2 - 09,431t + 43,279$$

3) Y, por último, para la revista *Journal of Agricultural Metereology*, los polinomios reducidos resultantes, tanto para los valores de *C(t)* como para los de *a(t) × 100*, respectivamente, fueron:

$$C(t) = \ ,002t^9 - \ ,12t^8 + \ ,146t^7 - \ 0,883t^6 + 20,689t^5 -$$
$$- \ 20,225t^4 + 51,017t^3 - \ 08,67t^2 + \ 01,035t - \ $$
$$a(t) = -\ ,028t^7 + \ ,507t^6 - \ ,868t^5 + \ 7,644t^4 - \ 4,371t^3 + 87,436t^2 - 97,761t + 78,849$$

CONSIDERACIONES FINALES

A la vista de los resultados del análisis de las distribuciones de referencias para las tres revistas, no podemos sino estar de acuerdo con la hipótesis de Egghe y Rao (1992): "*no hay una caída exponencial desde el principio, por el contrario, hay un incremento inicial, seguido de una corta caída exponencial*". Esto suele ocurrir casi siempre.

Esta hipótesis nos hace reflexionar en el sentido de que no hay forma de calcular un factor de envejecimiento, de forma independiente del tiempo. Únicamente, en el caso exponencial, se puede considerar independiente. Para el resto de los casos, sería una función dependiente del tiempo.

Por tanto, mostrar el problema e intentar darle solución fue el objetivo de nuestro estudio.

Ya anunciaron, Gapen y Milner (1981), que a pesar de la existencia de muchas fórmulas matemáticas, ninguna de ellas ha resultado ser simple ni universalmente aplicable.

Egghe y Rao proponen la distribución log-normal, como solución, añadiendo así una más a otras experimentadas con anterioridad.

Creemos que aplicar un modelo, como puede ser una distribución estadística es encajar los datos a una forma establecida, es como buscar un traje para la serie de datos, pero el traje ya está confeccionado, lo que significa que cuando la forma de estos datos no tiene una forma típica, el traje no va a quedar bien ajustado.

Sin embargo el modelo que aquí se propone, la fórmula de Lagrange, sería como confeccionar un traje a medida. La idea subyacente es distinta a cualquier otra solución.

El concepto al que nos enfrentamos es distinto, es un concepto de cualidad, ya que es pasar de un concepto discreto a otro continuo.

Hay que advertir que el polinomio de interpolación de Lagrange se ajusta a la totalidad de puntos que le son especificados. Por eso, en aquellas situaciones con gran cantidad de datos, se obtiene un polinomio de grado muy alto, lo que lo hace impracticable.

Otro problema que suele presentar el polinomio de Lagrange es lo que se conoce como *overfitting,* esto es, a medida que crece el grado del polinomio, se percibe una creciente variación entre los puntos de control consecutivos, lo que hace que la aproximación entre dos puntos continuos sea muy distinta a la esperada. De modo que es preferible ajustar los datos lo mejor posible a un polinomio de menor grado, aunque no pase por los puntos especificados.

REFERENCIAS

ÁLVAREZ, P.; ESCALONA, M. I.; PULGARÍN, A. (2000). "What is wrong with obsolescence?" *Journal of the American Society for Information Science*. 51(9): 812-815.

ASLIB, the Association for Information Management. (2007). "The concept of obsolescence in relation to knowledge". *Managing Information*. 14(1): 50-53.

BROOKES, B. C. (1970). "Obsolescence of special library periodicals: sampling errors and utility contours". *Journal of the American Society for Information Science.* 21(5): 320-329.

BROOKES, B. C. (1973). "Numerical methods of bibliography analysis". *Library Trends*. 22(1): 18-43.

BURTON, R. E.; KEBLER, R. W. (1960). "The 'half-life' of some scientific and technical literatures". *American Documentation*. 11: 18-22.

BURTON, R. E.; GREEN, B. A. (1961). "Technical reports in physics literature". *Physiscs Today*. 14(10): 35-37.

EGGHE, L. (1993). "On the influence of growth on obsolescence". *Scientometrics*. 27(2): 195-214.

EGGHE, L.; RAVICHANDRA RAO, I. K. (1992). "Citation age data and the obsolescence function: fits and explanation". *Information Processing & Managements*. 28(2): 201-217.

GAPEN, D. K.; MILNER, S. P. (1981). "Obsolescence". *Library Trends*. 30(1): 107-124.

GOSNELL, C. F. (1941). "Values and dangers of standard book and periodical lists for college libraries". *College and Research Libraries*. 2: 216-220.

GOSNELL, C. F. (1944). "Obsolescence of books in college libraries". *College and Research Libraries*. 5: 115-125.

GRIFFITH, B. C.; SERVI, P. N.; ANKER, A. L.; DROTT, M. C. (1979). "The aging of scientific literature: a citation analysis". *Journal of Documentation*. 35(3): 179-196.

KING, A. (1959). "Some reflections on the International Conference on Scientific Information, Washington, November 1958". *Revue de la Documentation*. 26: 1-5.

LINE, M. B. (1970). “The ‘half-life’ of periodical literature: apparent and real obsolescence”. *Journal of Documentation*. 26(1): 46-54.

LINE, M. B. (1993). “Changes in the use of literature with time – obsolescence revisited”. *Library Trends*. 41(4): 665-683.

LINE, M. B.; SANDISON, A. (1974). “Obsolescence and changes in the use of the literature with time”. *Journal of Documentation*. 30(3): 283-350.

RUÍZ BAÑOS, R.; JIMÉNEZ CONTRERAS, E. (1996). “Envejecimiento de la literatura científica en documentación. Influencia del origen nacional de las revistas. Estudio de una muestra”. *Revista Española de Documentación Científica*. 19(1): 39-49.

SANDISON, A. (1974). “Densities of use, and absence of obsolescence, in physic journal at MIT”. *Journal of the American Society for Information Science*. 25: 172-182.

SZAVA-KOVATS, Endre (2002). “Unfounded attribution of the ‘Half-life’ index-number of literature obsolescence to Burton and Kebler. A literature science study”. *Journal of the American Society for Information Science and Technology*. 53(13): 1098-1105.

TERRADA, M. L.; CUEVA, A. D. L.; AÑON, R. (1979). “La obsolescencia de la información científica en las publicaciones médicas españolas”. *Revista Española de Documentación Científica*. 2(1): 9-32.