

GRADUACIÓN ACTUARIAL NO PARAMÉTRICA. UNA REVISIÓN CRÍTICA

*Irene Albarrán Lozano y José M^a Sánchez López
Actuaria y profesora Titular de la Universidad de Extremadura
Profesor de la Universidad Rey Juan Carlos*

RESUMEN

Este artículo trata de la graduación como técnica para rectificar estimaciones anuales de mortalidad en base a ideas previas sobre el comportamiento de toda la secuencia de los datos. Se ofrece un enfoque no paramétrico para conseguir un desarrollo más detallado del proceso.

Se destacan los métodos de medias móviles, fórmula de Whittaker y el enfoque bayesiano. Estos métodos se utilizan menos que los paramétricos aunque, teóricamente, tienen mayor justificación.

La recopilación que se presenta supone una novedad al recoger los últimos aportes que ofrecen nuevas soluciones a problemas antiguos.

ABSTRACT

In this article we are concerning with the graduation as a form to revise annual mortality estimations in an attempt to improve them according to the whole set of data. Basically, the graduation methods developed are Moving Weighted Average methods, Whittaker formula and Bayesian graduation.

In particular, a complete and exhaustive revision of the main non parametric graduation methods is presented emphasising on some of them in order to present the subject with more detail.

PALABRAS CLAVE: Métodos actuariales de graduación, graduación no paramétrica, mortalidad y predicción, funciones de Kernel, fórmula de Whittaker, graduación bayesiana, pruebas de ajuste.

KEYWORDS: Actuarial graduation methods, non parametric graduation, mortality and forecasting, Kernel functions, Whittaker formula, bayesian graduation, test of fit (goodness of fit).

1.- INTRODUCCIÓN.

La graduación ha sido y es uno de los temas más tratados en investigaciones y artículos de naturaleza actuarial. Actualmente, gracias a la tecnología informática que reduce tiempo y dinero, estos métodos, apenas utilizados hace años debido a sus laboriosos cálculos, se efectúan cada vez más.

El tema de la graduación, aunque fundamentalmente se aplica al campo actuarial, está íntimamente ligado a la Ciencia Estadística.

Existen diversas definiciones que complementan el concepto de graduación. Andrews y Nesbitt (1965) lo consideran un “esfuerzo por representar un fenómeno físico a través de una revisión sistemática de algunas observaciones de ese fenómeno”. Dicha representación pasa de una serie irregular de valores observados de una variable continua, a una serie de valores observados.

El objetivo de la graduación consiste en revisar valores observados con el fin de producir una mejor representación. Es decir, supone un proceso de rectificación o cambio de una secuencia de estimaciones iniciales que se justifica por las relaciones existentes en el conjunto y la incorporación de creencias previas a las observaciones.

A veces se ha identificado a la graduación con un simple mecanismo de suavización (alisamiento) de una secuencia de datos. London (1984) reconoce la validez del objetivo de alisamiento, pero extiende la idea hacia el concepto más general que refleja todos los elementos de opinión previa en un proceso de graduación.

Este proceso consta, al menos, de dos etapas: la obtención del conjunto de datos observados y la posterior graduación de los mismos para obtener una mejor representación de la ley implícita que subyace de acuerdo con un criterio.

Una tarea importante que realiza el actuario es construir modelos de supervivencia a partir de los cuales se calculan primas, reservas, etc... Estos modelos generalmente aparecen tabulados en lo que se conoce comúnmente como **tabla de mortalidad**. Así, con la graduación no sólo se consigue una "curva suave" sino, más bien, las muertes más probables. Es decir, estimar las verdaderas tasas que realmente prevalecen en la población.

Sin embargo, hay que señalar que la graduación se aplica a estas tablas y a otros muchos campos dentro del ámbito actuarial.

Estas técnicas que proporcionan una mejora en la estimación de tasas anuales de mortalidad, tienen una naturaleza bayesiana aunque no siempre se basen en un proceso bayesiano formal. Esto requiere, además de una opinión previa, la valoración de esta creencia en forma de una distribución de probabilidad.

2.- NATURALEZA ESTADÍSTICA Y ANÁLISIS PREVIO.

Existe una conexión importante entre la graduación y la Ciencia Actuarial, aunque, debido a los procedimientos y el fundamento que lo motiva, está íntimamente unida a la Estadística.

Kimeldorf y Jones (1967) consideran que "no es meramente un suavizado, sino que un proceso de estimación de las verdaderas tasas de mortalidad que realmente prevalecen en la población". Así se puede afirmar, tal y como considera Whittaker (1923), que "pertenece esencialmente a la Teoría Matemática de la Probabilidad". De esta forma, se sitúa a la graduación en el contexto de la estimación estadística, considerando en su planteamiento variables aleatorias, medias, varianzas, etc..

Los resultados graduados diferirán mucho de las estimaciones

iniciales ya que éstas, si están bien construidas, son indicadores razonables del verdadero valor. Las rectificaciones en base al conocimiento previo sobre el fenómeno serán menores cuanto mayor sea la muestra (y menor sea la varianza de la estimación inicial).

El planteamiento de la graduación se puede establecer con los siguiente elementos:

x : índice que define la secuencia (según la edad en el caso de seguro de vida),

t_x : secuencia de valores verdaderos desconocidos que se deben estimar,

n_x : tamaño de la muestra,

w_x : ponderaciones,

U_x : variable aleatoria estimador de t_x ,

E_x : variable aleatoria error de la estimación,

u_x : realizaciones de U_x ,

e_x : realizaciones de E_x ,

v_x : valores graduados.

La relación entre el fenómeno verdadero y las estimaciones iniciales se pueden establecer de la forma:

$$U_x = t_x + E_x \Rightarrow u_x = t_x + e_x .$$

Si se emplea un operador de graduación:

$$G(u_x) = v_x = G(t_x + e_x) = t_x + e'_x \Rightarrow |e'_x| < |e_x| \quad \forall x .$$

Esto es, se supone que disminuyen los errores cometidos.

Como medida numérica de las rectificaciones efectuadas en un proceso de graduación, con el fin de comparar el ajuste en varias graduaciones y para los mismos datos, se pueden utilizar las expresiones siguientes:

$$\begin{aligned}F_1 &= \sum_x w_x(v_x - u_x) \\F_2 &= \sum_x w_x(v_x - u_x)^2 \\F_3 &= \sum_x x \cdot w_x(v_x - u_x)\end{aligned}$$

El valor que se toma para la ponderación estará relacionado con la exposición al riesgo como aproximación al tamaño muestral.

El análisis básico previo al ajuste puede realizarse a partir de la construcción de un gráfico que presente los datos disponibles, o bien, las tasas de decrementos (diferencias de algún orden) de los datos disponibles, mediante puntos en unos ejes cartesianos (en abscisas las edades y en ordenadas los valores alcanzados).

A continuación, se traza una curva suave que pase tan cerca como sea posible a dichos puntos. Puede servir de referencia utilizar junto a los puntos un indicador (en forma de barra) que exprese la desviación típica correspondiente:

$$\text{Var}(U_x) = \frac{t_x(1-t_x)}{n_x} \approx \frac{u_x(1-u_x)}{n_x}$$

Las ventajas e inconvenientes de un análisis gráfico son evidentes¹ apreciándose su utilidad ante situaciones en las que la cantidad de datos disponibles sea escasa (para edades avanzadas por ejemplo) permitiendo incorporar el juicio subjetivo en base a la experiencia previa (especialmente útil en los extremos de la tabla).

Los inconvenientes aparecen por la falta de soporte formal del método: no se establecen hipótesis precisas sobre las ideas previas a

¹ Ver Benjamin y Pollard (1980).

incorporar en los datos (pueden aparecer sesgos debido a prejuicios no oportunos) lo que provoca que dos investigadores experimentados obtengan resultados diferentes con los mismos datos.

La utilización simultánea de una tabla estándar derivada de experiencias de mortalidad con características similares y consistente con la opinión previa sobre el fenómeno puede informar sobre la escala idónea del gráfico y la tendencia general de los decrementos. El proceso consistiría en utilizar una variable de la forma:

$$R_x = \frac{U_x}{s_x} \quad \text{con} \quad \text{Var}(R_x) = \frac{t_x(1-t_x)}{(s_x)^2 \cdot n_x} \approx \frac{u_x(1-u_x)}{(s_x)^2 \cdot n_x}$$

siendo

s_x los valores de mortalidad anual en la tabla estándar,

u_x las estimaciones iniciales de mortalidad anual y

r_x los valores que se gradúan con un operador determinado,

$G(r_x)$, evitando los problemas de escala y que permiten obtener los valores finales perseguidos:

$$v_x = s_x \cdot G(r_x)$$

También cabe utilizar la tabla estándar como elemento de referencia o método de graduación. Se compararían las estimaciones iniciales resultado del análisis de los datos de mortalidad de la aseguradora con las probabilidades de muerte que aparezcan en una tabla para el conjunto de la población. Una forma simple de incorporar la información de la tabla estándar consiste en calcular los valores

$$h_x = \log\left(\frac{1-s_x}{1-u_x}\right)$$

que manifiestan una secuencia más suave que las estimaciones iniciales de probabilidades de fallecimiento debido a la relación con la tabla estándar y al logaritmo. Sobre esta nueva secuencia se aplicará el método de graduación elegido para a continuación calcular, deshaciendo la transformación, las tasas de mortalidad graduadas

$$v_x = u_x^* = 1 - (1 - s_x)e^{-hx}$$

Dado que la tabla utilizada de referencia será suficientemente suave (y los resultados del proceso serán suaves) hay que centrarse en las pruebas que garanticen la adherencia de los datos a los resultados obtenidos.

Las tablas generales, con datos del conjunto de la población y no sólo de individuos asegurados, son las usadas actualmente por las compañías de seguros. Supone una actitud prudente combinar los resultados derivados de la información de una cartera de pólizas (especialmente si la cartera no es muy amplia) con los habituales resultados (más utilizados por los actuarios) reflejados en las tablas de mortalidad generales.

3.- CLASIFICACIÓN DE LOS MÉTODOS DE GRADUACIÓN.

Con la graduación se pretende producir un modelo de supervivencia que sea la mejor estimación del verdadero “patrón” que sigue la mortalidad de una determinada población. Para ello, existen distintos métodos.

Una primera clasificación de dichos métodos consiste en resaltar los fundamentos teóricos utilizados. En Hickman (1.987) se distingue entre fundamentos basados en programación matemática, cuando se especifica una función objetivo que debe alcanzar algún valor óptimo, y fundamentos basados en análisis estadístico, cuando se especifica un modelo estadístico que se relaciona con objetivos en términos estadísticos.

Más habitual resulta la segunda clasificación que recurre a la forma en que se presenta el resultado distinguiéndose entre graduación no paramétrica y paramétrica. Con la graduación no paramétrica se obtienen las estimaciones revisadas en forma tabular. Con la graduación paramétrica se obtienen las estimaciones revisadas junto con una función o funciones que relacionan toda la secuencia (cada función expresa en la forma y en el método de ajuste la opinión previa).

La graduación paramétrica, también llamada de fórmula matemática, es el método más utilizado². Ejemplos son Prieto Pérez y Fernández Plasencia (1.994) y Felipe Checa y Guillén Estany (1.999).

En este artículo se desea ofrecer una panorámica de las alternativas no paramétricas analizándose: el método de las medias móviles junto con las correcciones para valores extremos, la generalización de las funciones de Kernel, los desarrollos basados en la fórmula de Whittaker y la aplicación de procesos bayesianos formales.

4.- MÉTODO DE LAS MEDIAS PONDERADAS MÓVILES Y EL PROBLEMA DE LOS VALORES EXTREMOS.

El método de las medias móviles fue uno de los primeros métodos desarrollados que se inició de forma intuitiva. Sólo buscando un objetivo práctico y de fácil aplicación para la suavización de la secuencia de estimaciones, se llega a una expresión más formal en estudios posteriores, tal como se encuentra en Borgan (1.979). Para dar solución al tratamiento de los extremos de la secuencia aparecen ciertas estructuras matriciales en Hoem y Linnemann (1.988).

Los métodos de las medias móviles tiene la característica de ser operativos incluso si las restricciones no se cumplen, siendo esta propiedad de robusted uno de los motivos más fuertes para aplicarlos.

² Para un análisis detallado de graduación paramétrica y su desarrollo empírico actual destaca el artículo "Modelos de tablas de mortalidad en España y situación actual" de Betzuen Zalbidegoitia, Felipe Checa y Guillén Estany (1.997).

Se comienza con el planteamiento y la solución para el problema de la determinación de coeficientes en los valores centrales de la secuencia.

El valor graduado se halla mediante una media ponderada de un cierto número de valores no graduados consecutivos. Siendo v_x los valores ajustados, u_x las estimaciones iniciales, t_x el verdadero valor subyacente, e_x el error cometido en la estimación inicial y a_r las ponderaciones simétricas en torno al año "x" considerado, se representa:

$$v_x = \sum_{r=-n}^n a_r \cdot u_{x+r} = \sum_{r=-n}^n a_r \cdot t_{x+r} + \sum_{r=-n}^n a_r \cdot e_{x+r} = \sum_{r=-n}^n a_r \cdot t_{x+r} + e_x'$$

con $u_x = t_x + e_x$

$$e_x' = \sum_{r=-n}^n a_r \cdot e_{x+r}$$

tomando en consideración un rango de $2n+1$ estimaciones iniciales para formar cada valor graduado.

Se intenta minimizar el error estadístico representado por e_x' , tras reproducir t_x . La suavización en los saltos entre las estimaciones consecutivas debe ser el inicio y la consecuencia del proceso.

En primer lugar, con una hipótesis funcional (guiada por la opinión previa de suavidad) de la secuencia de los t_x en el rango $[x-n, x+n]$ según un polinomio de grado tres se obtiene la deseada reproducción en los verdaderos valores t_x (propiedad reproductiva) con sólo dos restricciones:

$$t_x = \sum_{r=-n}^n a_r \cdot t_{x+r} \quad \text{si} \quad \begin{cases} \sum_{r=-n}^n a_r = 1 \\ \sum_{r=-n}^n r^2 \cdot a_r = 0 \end{cases}$$

Además, considerando el carácter simétrico de los coeficientes, $a_r = a_{-r}$ para los valores de r en el rango, se tienen:

$$\left\{ \begin{array}{l} \sum_{r=-n}^n r \cdot a_r = 0 \\ \sum_{r=-n}^n r^3 \cdot a_r = 0 \\ \sum_{r=1}^n r^2 \cdot a_r = 1 \\ a_0 + 2 \sum_{r=1}^n a_r = 1 \end{array} \right.$$

En segundo lugar, para minimizar e_X' se considera que es una concreción de la variable aleatoria E_X' . El valor probable será:

$$\begin{aligned} E[E_X'] &= E[t_x - V_x] = t_x - E[V_x] = t_x - E\left[\sum_{r=-n}^n a_r \cdot U_{x+r}\right] = \\ &= t_x - \sum_{r=-n}^n a_r \cdot E(U_{x+r}) = t_x - \sum_{r=-n}^n a_r \cdot t_{x+r} = t_x - t_x = 0 \end{aligned}$$

Para esto se admite U_X como proporción binomial con $E[U_X] = t_X$ y que t_X cumple la propiedad reproductiva. Como la variable aleatoria E_X' es insesgada, se centra el problema en obtener la menor varianza posible:

$$\begin{aligned} \text{Var}[E_X'] &= \text{Var}[t_x - V_x] = \text{Var}[V_x] = \text{Var}\left[\sum_{r=-n}^n a_r \cdot U_{x+r}\right] = \\ &= \text{Var}\left[\sum_{r=-n}^n a_r \cdot E_{x+r}\right] = \sigma^2 \cdot \sum_{r=-n}^n (a_r)^2 \\ \text{con } \text{Var}(E_{x+r}) &= \text{Var}(U_{x+r}) = \sigma^2 \quad \forall r \end{aligned}$$

Como se observa se incorpora la hipótesis de homocedasticidad en la variante de estimación.

La expresión final que se conoce como ratio de varianzas porque puede obtenerse como cociente entre las varianzas de la variante graduada y sin graduar:

$$R_0^2 = \sum_{r=-n}^n (a_r)^2 = \frac{\text{Var}[V_x]}{\text{Var}[U_x]}$$

Debe ser menor que uno para que el proceso de graduación mejore la estimación estadística.

En la práctica, para aumentar la suavidad en el ajuste, se define una expresión a minimizar de la forma:

$$R_z^2 = \frac{\text{Var}(\Delta^z V_x)}{\text{Var}(\Delta^z U_x)}$$

considerándola una generalización de la expresión inicial y recomendándose $z=2,3,4$.

El numerador y denominador de la expresión anterior se obtiene (con supuestos de incorrelación y equivarianza) y se llega a:

$$R_z^2 = \frac{1}{\binom{2z}{z}} \sum_{r=-n-z}^n (\Delta^z a_r)^2$$

El siguiente problema que se presenta al utilizar este método de graduación consiste en su aplicación en las colas o extremos de la secuencia. Unas soluciones intentan ajustar un polinomio a los puntos extremos de la secuencia de estimaciones iniciales y usar los puntos del polinomio como valores graduados. Otras intentan hallar puntos más allá de los extremos estimados (por ajuste, extrapolación, etc.) para utilizar las medias móviles en toda la secuencia deseada. Además

de críticas parciales, el mayor problema que presentan se deriva de la falta en estos métodos de un proceso único global basado en principios simples y de general aceptación. Son procesos con soluciones específicas para los extremos que permiten una transición suave de los puntos obtenidos por las medias móviles a los puntos hallados con el método específico empleado en los extremos.

Por el contrario, Hoem y Linnemann (1.988) minimizan una función de riesgo que es una generalización de

$$R_z^2 = \frac{1}{\binom{2z}{z}} \sum_{v=a-z}^b (\Delta^z r_v)^2$$

siendo “z” el nivel de las diferencias, “r” coeficiente real, “a” extremo inferior y “b” extremo superior para secuencia a considerar en la ponderación. Se utiliza un proceso unitario para el total de la secuencia que no ajusta previamente los valores centrales sino que se determinan conjuntamente con las colas mediante el criterio elegido. No obstante, cuando sea necesario puede usarse a partir de unos valores preseleccionados para las medias móviles centrales.

El modelo de partida utiliza la expresión matricial que relaciona “m-a-b” valores graduados en forma de matriz fila, la matriz de las ponderaciones [(m-a-b),m] o matriz de graduación, y la matriz columna de los “m” valores iniciales a graduar:

$$\lambda_c^* = G \cdot \hat{\lambda}$$

A continuación se completa el modelo añadiendo las colas o extremos para los valores graduados y aumentando el tamaño de la matriz de graduación:

$$\lambda^* = \begin{pmatrix} \Phi_{(a,a+b+1)}, O_{(a,m-a-b-1)} \\ G_{(m-a-b,m)} \\ O_{(b,m-a-b-1)}, \Psi_{(b,a+b+1)} \end{pmatrix} \cdot \hat{\lambda} = \Gamma \cdot \hat{\lambda}$$

siendo los elementos de O siempre ceros.

Se deben encontrar los coeficientes de la matriz de graduación para resolver el problema. Estos coeficientes deben dotar a los valores graduados de propiedades estadísticas deseables. Para encontrar un criterio se emplea la función de pérdida de Borgan (1.979) extendida al caso actual:

$$L(\lambda^*, \lambda) = N \sum_{z=0}^Z \frac{\alpha_z}{m-a-b-z} \sum_{t=1}^{m-z} (\Delta^z \lambda_t^* - \Delta^z \lambda_t)^2$$

El valor “ N ” es un número entero que representa el tamaño de la población que da origen a los datos, “ Z ” es menor que “ $m-a-b$ ”, “ α_z ” son ponderaciones no negativas elegidas por el usuario cuya suma será uno.

Además, se acostumbra a exigir que un método de graduación lineal reproduzca los valores graduados considerando algún subespacio “ V ” lineal (por ejemplo el espacio de los polinomios cúbicos). Por ello se debe cumplir:

$$\Gamma V = \Gamma$$

siendo “ V ” el vector de valores graduados.

El criterio busca minimizar la función de pérdida sujeto a la condición de reproducción. Esta idea inicial puede desarrollarse utilizando distintos supuestos en la matriz de covarianzas entre elementos de la secuencia entre los que destaca la sustitución de la matriz de covarianzas por una matriz “ S ” según propiedades deseables para el proceso de cálculo y para su interpretación (matriz cuadrada y definida positiva, homocedasticidad, etc.).

Como se puede intuir, para poder aplicar el criterio y obtener las soluciones deseadas, se deben tomar decisiones respecto a varios puntos: los valores de “ a ” y de “ b ” para determinar las colas generadas en las medias móviles, la elección de la matriz “ S ”, los coeficientes “ α_z ”, y el subespacio “ V ”.

La evidencia empírica sugiere que “a” sea igual a ”b”, que se de que “ $m > 2(a+b+1)$ ”, que la matriz “S” puede ser varianza por matriz identidad, que todos los coeficientes sean cero menos uno, que se utilice el subespacio referido a los polinomios cúbicos y que basta con referirse a las diferencias primeras ($\min-R_1$).

Tratado el tema de los extremos de la secuencia mediante la extensión en la matriz de graduación, se continua con otra generalización del método de las medias móviles. La extensión empieza esta vez en el diseño de las expresiones que incluyen las ponderaciones.

5.- MÉTODO DE LAS MEDIAS PONDERADAS MÓVILES Y LAS FUNCIONES DE KERNEL.

En general, los estudios sobre medias móviles ponderadas aplicadas a la graduación en mortalidad se basan en las deseables propiedades teóricas que se alcanzan mediante determinadas ponderaciones. Estas fórmulas pueden obtenerse a partir de las llamadas funciones de Kernel³. Estas últimas suponen una mejora y proporcionan una aproximación más flexible al problema de la graduación. Además, en la graduación de Kernel la elección del ancho de banda (o parámetro de suavidad) presenta mayor importancia que la propia elección de la función.

El método de estimación de Kernel se utiliza para una distribución de densidad que se pretende estimar utilizando ciertas observaciones disponibles:

$$\hat{f}(x) = (nb)^{-1} \sum_{i=1}^n K\left(\frac{x - x_i}{b}\right) = (nb)^{-1} \sum_{i=1}^n K_b(x - x_i)$$
$$\text{con } \int_{-\infty}^{\infty} K(x) dx = 1$$

³ Ver Ramlau-Hansen (1.983) y Gavin, Haberman y Verral (1.993).

Los valores $\{x_i : i = 1, 2, \dots, n\}$ representan las observaciones y “b” el parámetro de suavidad o ancho de banda (mayores valores producen mayor suavidad).

La aplicación en graduación actuarial se tiene mediante la descomposición de la probabilidad de muerte según el Teorema de Bayes:

$$q_x = P(E = d / X = x) = \frac{P(X = x / E = d)}{P(X = x)} P(E = d)$$

A continuación, la probabilidad de que la variante edad tome un valor $P(X=x)$ y la misma probabilidad condicionada a la aparición del suceso muerte $P(X=x/E=d)$ pueden estimarse mediante funciones de Kernel. La probabilidad de la aparición del suceso muerte $P(E=d)$ se obtiene con una estimación más simple.

De esta forma, si se estima:

$$P(E = d) = \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n e_i}$$

se obtiene el estimador de Copas-Haberman

$$\hat{q}_x^{CH} = \frac{\sum_{i=1}^n d_i K_b(x - x_i)}{\sum_{i=1}^n e_i K_b(x - x_i)}$$

El estimador de Nadaraya-Watson supone observaciones espaciadas uniformemente con lo que a cada edad concreta se tiene la relación d_i/e_i lo que lleva a

$$\hat{q}_x^{NW} = \frac{\sum_{i=1}^n \left(\frac{d_i}{e_i} \right) K_b(x - x_i)}{\sum_{i=1}^n K_b(x - x_i)}$$

Ambos estimadores eliminan las fluctuaciones aleatorias mediante la suavización de los datos, pero pueden conducir a sesgos en las estimaciones resultantes.

Los datos de mortalidad tienen dos características que influyen en el sesgo del estimador: la mortalidad presenta tendencia exponencial con la edad y el número de datos o exposiciones a riesgo recogidas de empresas aseguradoras varía con la edad.

El sesgo del estimador de Copas-Haberman será:

$$\frac{\sum_{i=1}^n (x_i - x) K(x - x_i)}{\sum_{i=1}^n K(x - x_i)} \cdot q_{x'+(1/2)} - \frac{\sum_{i=1}^n (x_i - x)^2 K(x - x_i)}{\sum_{i=1}^n K(x - x_i)} \cdot q_{x''+R}$$

El sesgo del estimador de Nadaraya-Watson será:

$$\frac{\sum_{i=1}^n (x_i - x) e_i K_b(x - x_i)}{\sum_{i=1}^n e_i K_b(x - x_i)} \cdot q_{x'+(1/2)} - \frac{\sum_{i=1}^n (x_i - x)^2 e_i K_b(x - x_i)}{\sum_{i=1}^n e_i K_b(x - x_i)} \cdot q_{x''+R}$$

En ambos casos las derivadas de la función de la probabilidad de muerte se entienden respecto de “x” y el término R se considera pequeño al contener derivadas de orden mayor debidas a un desarrollo en serie de Taylor.

En cuanto a la elección de la función de Kernel, se comentó su importancia relativa respecto a la más importante cuestión que es la fijación del ancho de banda. Se puede recurrir al Kernel normal:

$$K^N(x) = \frac{\exp\left\{-\frac{(x/2b)^2}{2}\right\}}{\sqrt{2\pi b^2}} \quad \text{con} \quad -\infty < x < +\infty$$

También presenta interés para la graduación un Kernel que busca minimizar la varianza de las primeras diferencias de los valores graduados (filosofía habitual en la graduación por medias móviles ponderadas):

$$K(x) = \begin{cases} 15(b^2 - x^2)(3b^2 - 7x^2)/32b^5 & \text{con } |x| \leq b \\ 0 & \text{con } |x| > b \end{cases}$$

La elección del parámetro “b” o de ancho de banda busca ajustar el modelo a la vez que comprobar la “suavidad” que proporciona a las estimaciones iniciales. Para ello, se emplea el método de validación transversal, esto es, se minimiza la expresión:

$$CV(b) = n^{-1} \sum_{j=1}^n [\dot{q}_j - \hat{q}_j^{(-j)}]^2$$

con

$$\dot{q}_i = \frac{d_i}{e_i}$$

y si se considera el estimador de Copas-Haberman:

$$\hat{q}_j^{(-j)} = \frac{\sum_{\substack{i=1 \\ i \neq j}}^n d_i K_b(x_j - x_i)}{\sum_{\substack{i=1 \\ i \neq j}}^n e_i K_b(x_j - x_i)}$$

mientras que si se considera el estimador de Nadaraya-Watson:

$$\hat{q}_j^{(-j)} = \frac{\sum_{\substack{i=1 \\ i \neq j}}^n \hat{q}_i K_b(x_j - x_i)}{\sum_{\substack{i=1 \\ i \neq j}}^n K_b(x_j - x_i)}$$

Esta minimización obtiene un valor de $CV(b)$ aproximadamente equivalente al error cuadrado integrado medio que recoge el sesgo cuadrado y la varianza del estimador:

$$MISE(\hat{q}) = E\left[\int (\hat{q}_x - q_x)^2 dx\right] = \int (E\{\hat{q}_x\} - q_x)^2 dx + \int V\{\hat{q}_x\} dx$$

Se consigue así un equilibrio entre el sesgo y la varianza.

Según Gavin, Haberman y Verral (1.994) el estimador de Nadaraya-Watson es mejor que el estimador de Copas-Haberman para datos de mortalidad ya que presenta menor sesgo sistemático como resultado de las distintas exposiciones a cada edad. Además, sugiere que no aparecen mejoras si se utiliza un Kernel que busca minimizar la varianza de las primeras diferencias de los valores graduados frente a un Kernel normal justificándose por la mayor importancia del ancho de banda frente a la función elegida.

6.- LA FÓRMULA DE WHITTAKER.

Método iniciado por Whittaker (1.923) y Henderson (1.924). Posteriores trabajos, como los de Taylor (1.992) y Verrall (1.993), obtienen el desarrollo bayesiano, el parámetro de suavidad y generalizan e interpretan los resultados.

Se observa que subyace una racionalidad bayesiana formal.

Tal como expresan sus autores iniciales, la opinión previa sobre la secuencia de los verdaderos valores de t_x , realizaciones de la secuencia de variables aleatorias T_x , se concreta expresando la densidad de probabilidad de una secuencia dada de t_x , sería:

$$f_T(t_x) = c_1 \cdot e^{-\lambda S} \text{con} S = \sum_{x=1}^{n-z} (\Delta^z t_x)^2$$

Siendo λ una constante que puede determinar la medida de la confianza que se tiene en la opinión previa y c_1 otra constante que toma el valor que hace de la función una verdadera función de densidad.

En principio, una secuencia de valores de t_x es más probable si "S" toma valores menores (más suave). Ahora bien, Whittaker considera que la secuencia más probable previa a la información suministrada por el dato observado, está indefinida.

Por otra parte, la variable aleatoria E_x , que informa de la desviación entre la variable aleatoria estimación inicial de t_x y el verdadero valor subyacente desconocido, se considera con distribución normal de media cero y desviación típica finita. Por ello, la función de densidad para un valor determinado "y" es:

$$f_{E/T}(e_y / t_y) = c_y \cdot e^{-\frac{1}{2} \left(\frac{e_y}{\sigma_y} \right)^2}$$
$$f_{U/T}(u_y / t_y) = c_y \cdot e^{-\frac{1}{2} w_y (t_y - u_y)^2}$$

siendo " w_y " la recíproca de la varianza de " U_y ".

Para toda la secuencia, asumiendo la independencia para todos los valores de "x" (todas las edades):

$$f_{U/T}(u_x / t_x) = c_2 \cdot e^{-\frac{1}{2} \sum_{x=1}^n w_x (t_x - u_x)^2}$$

interpretado como la probabilidad de obtener la secuencia entera de estimaciones iniciales supuesto que la secuencia de los valores verdaderos es la secuencia "t_x".

A continuación se puede obtener la función de densidad de "t_x" condicionada a las estimaciones iniciales:

$$f_{T/U}(t_x / u_x) = \frac{f_{U/T}(u_x / t_x) \cdot f_T(t_x)}{f_U(u_x)} = \frac{c_3 \cdot e^{-\lambda S - \frac{1}{2} \sum_{x=1}^n w_x (t_x - u_x)^2}}{f_U(u_x)}$$

con $c_3 = c_1 \cdot c_2$.

Se considera, entonces, la secuencia más probable de verdaderos valores condicionado por la información que suministran las estimaciones iniciales. Para ello se maximiza la expresión anterior minimizando:

$$\lambda S + \frac{1}{2} \sum_{x=1}^n w_x (t_x - u_x)^2 = \lambda S + \frac{1}{2} F$$

Finalmente se opta por minimizar una expresión de la forma:

$$2\lambda S + F = hS + F$$

Se llega a una fórmula que recoge los niveles de ajuste y de suavidad en la graduación. Se obtienen los valores rectificadas de la estimación (valores de v_x) minimizando, entonces, la expresión:

$$M = F + hS = \sum_{x=1}^n w_x (v_x - u_x)^2 + h \sum_{x=1}^{n-z} (\Delta^z v_x)^2$$

El primer sumando valora el nivel de ajuste entre las estimaciones iniciales y las estimaciones corregidas. Pondera las desviaciones cuadráticas asignando pesos distintos a cada desviación. El segundo sumatorio es una medida de la suavidad de la secuencia de las estimaciones rectificadas (considerada como cualidad necesaria en la secuencia de estimaciones).

Una forma sencilla de establecer las ponderaciones es identificarlas con el tamaño de la muestra disponible en cada estimación u_x (a mayor tamaño más peso) o establecer una relación de la forma $w_x=(n_x):(n)$, siendo "n" la media aritmética de todos los tamaños de muestras utilizados para la estimación de toda la serie de valores considerada. Con datos de mortalidad, cuando se identifica ponderación y tamaño de la muestra, resulta que

$$\sum_{x=1}^n w_x(v_x - u_x) = \sum_{x=1}^n x \cdot w_x(v_x - u_x) = 0$$

Esto es, el número total de muertes, las edades totales de muerte y la edad media de muerte coinciden para los datos observados y los datos graduados (considerando " $n_x v_x$ " como las muertes esperadas).

Otra manera de proceder para fijar los valores de w_x es seguir las consideraciones ya manifestadas sobre el uso del recíproco de la varianza estimada de u_x , para lo que se emplean los valores de v_x en lugar de los desconocidos verdaderos valores t_x (una simplificación aun mayor nos podría llevar a utilizar las estimaciones iniciales u_x):

$$w_x = \frac{n_x}{v_x(1 - v_x)} \approx \frac{n_x}{u_x(1 - u_x)}$$

El parámetro "h" (número real positivo) es un elemento de control de la importancia relativa que se atribuye a los dos sumandos de la expresión, es decir, sirve para dar mayor o menor relevancia a la suavidad sobre el ajuste o viceversa. Si se hiciera $h=0$ no se corregirían las estimaciones iniciales. Si "h" tomara un valor

suficientemente grande se prescindiría del proceso de ajuste.

6.1.- INTERPRETACIÓN BAYESIANA DE LA FÓRMULA DE WHITTAKER.

Para una interpretación moderna y una derivación bayesiana formal de la fórmula de Whittaker pueden consultarse Taylor (1.992) y Verrall (1.993). En ambos se considera la fórmula de Whittaker como una función de pérdida con semejanzas a una verosimilitud penalizada. Esto es, se llega a un compromiso entre ajuste y suavidad mediante una función de pérdida.

Expresada en forma matricial sería:

$$L(U, \hat{V}, \lambda) = F(U, \hat{V}) + \lambda S(\hat{V}) = (U - \hat{V})' W (U - \hat{V}) + \lambda (K \hat{V})' (K \hat{V})$$

La matriz “U” es la matriz de la secuencia de valores sin graduar, la matriz “ \hat{V} ” es de la secuencia de valores graduados, la matriz “W” es de las ponderaciones en el ajuste, la matriz “K” es de coeficientes para obtener las diferencias utilizadas en el mecanismo de suavidad seleccionado y el valor “ λ ” asigna peso relativo al ajuste frente a la suavización. Los valores para “ \hat{V} ” se obtienen minimizando la función de pérdida:

$$\frac{\partial L(U, \hat{V}, \lambda)}{\partial \hat{V}} = 0 \Rightarrow \hat{V} = (I + \lambda W^{-1} K' K)^{-1} U$$

Para llegar a la interpretación bayesiana de la función de pérdida (considerando primero el ajuste) se recuerda que

$$\begin{aligned} u(x) &= v(x) + e(x) \\ E[e(x)] &= 0 \end{aligned}$$

y se considera la relación entre ponderación en el ajuste, varianza de las variantes iniciales y tamaño muestral (siendo “N(x)” el tamaño y “z(.)” una función):

$$\frac{1}{w(x)} = \text{Var}[u(x)] = \frac{z(x)}{N(x)}$$

Entonces, se tiene para el rango $(a, a+m)$ que

$$E[\hat{V}] = (I + \lambda W^{-1} K' K)^{-1} V = V + O\left(\min\{[N(x)]^{-1} : x = a + 1, \dots, a + m\}\right)$$

esto es, \hat{V} es asintóticamente insesgado (con muestras grandes).
También, asintóticamente

$$F(\hat{V}, U) = \sum_{x=a+1}^{a+m} \frac{[u(x) - E\{u(x)\}]^2}{\text{Var}\{u(x)\}}$$

lo que representa el logaritmo de la verosimilitud de “ $u(x)$ ” cuando se eliminan constantes y se suponen las variantes normales con i.i.d. El criterio de ajuste se puede considerar como una verosimilitud.

En segundo lugar se muestra que el criterio de suavidad también se puede considerar como una verosimilitud. La expresión que mide la suavidad es:

$$S(\hat{V}) = (K\hat{V})(K\hat{V}) = \sum_{x=a+1}^{a+m-n} [\Delta^n \hat{v}(x)]^2$$

A continuación se utiliza $\xi(x)$ para representar $\Delta^n v(x)$ y se considera la distribución previa $\xi(x) \approx N(0, \tau^2)$. Si las $\xi(x)$ se suponen estocásticamente independientes, entonces el logaritmo de la verosimilitud (omitiendo el término constante) es $\xi' \xi / \tau^2$.

Siendo $\hat{\xi}(x) = \Delta^n \hat{v}(x)$, entonces $\hat{\xi}(x)$ es asintóticamente igual a $\xi(x)$ dado que \hat{V} es asintóticamente insesgada. Por esto, el logaritmo de la verosimilitud anterior es asintóticamente igual a $\hat{\xi}' \hat{\xi} / \tau^2 = S(\hat{V}) / \tau^2$.

La asignación de una distribución previa a $\Delta^n v(x)$ en vez de asignarla a $v(x)$ necesita sólo del conocimiento previo de la forma de $v(x)$ en cuanto a función de x (no necesita la determinación exacta de la función).

Para resumir la notación se establece que $G(X)$ será el logaritmo de la verosimilitud de una variable aleatoria X .

Considerando las verosimilitudes, y su expresión asintótica, obtenidas a partir del criterio de ajuste y a partir del criterio de suavidad se tiene que:

$$G(U/V) = -F(\hat{V}, U)$$

$$G(\xi) = -\tau^{-2}S(\hat{V}),$$

expresiones que incluyen el signo verdadero y prescinden de las constantes.

Utilizando ambas expresiones:

$$\begin{aligned} G(U, V) &= G(U/V) + G(V) = G(U/V) + G(\xi) = \\ &= -F(\hat{V}, U) + \tau^{-2}S(\hat{V}) = -L(\hat{V}, U, \tau^{-2}). \end{aligned}$$

Lo que permite llegar a

$$G(V/U) = G(U, V) - G(U) = -L(\hat{V}, U, \tau^{-2}) - G(U).$$

Todas las verosimilitudes involucradas son normales y entonces también $G(V/U)$. Por ello, el estimador de máxima verosimilitud para $E[V/U]$ se encuentra maximizando $G(V/U)$ o, de forma equivalente, minimizando $L(\hat{V}, U, \tau^{-2})$. Esto es lo mismo que se realiza en una graduación de Whittaker-Henderson con constante relativa λ igual a τ^{-2} .

Como conclusión resumen consideramos que suponiendo unas tasas

de mortalidad $u(x)$ estocásticamente independientes para las distintas edades y con distribuciones asintóticamente normales en grandes muestras, y si se supone también que $V=E[U]$ sigue una distribución previa según la cual las $\xi(x)=\Delta^n v(x)$ se consideran independientes e igualmente distribuidas como $\xi(x) \approx N(0, \tau^2)$; entonces, asintóticamente, el estimador de máxima verosimilitud de la esperanza a posteriori bayesiana $E[V/U]$ se da por la graduación de Whittaker-Henderson sobre U , siendo la ponderación $w(x) = 1/\text{Var}[u(x)]$ y la constante que relaciona ajuste y suavización $\lambda = 1/\tau^2$.

Con lo tratado hasta aquí se puede interpretar el parámetro $1/\lambda$ como la varianza a priori de una medida de suavidad que se aplica, habitualmente un orden de diferencias finitas. Según las circunstancias, puede ser posible expresar esta varianza previa con razonable precisión (sería el caso de una tabla de mortalidad nacional que aproveche la información de la secuencia de tablas anteriores sobre las medidas de suavidad).

7.- APLICACIÓN DE PROCESOS BAYESIANOS.

La estadística bayesiana permite formalizar la opinión previa que se tenga mediante la utilización de distribuciones de probabilidad "a priori" y "a posteriori" de la muestra de datos obtenida mediante observación. Se trata de aprovechar, de forma explícita, la opinión previa formada a partir de los numerosos y extensos estudios de mortalidad anteriores a la muestra con la que se trabaje.

La información suministrada por una muestra elegida puede cambiar la idea del comportamiento probabilístico que se tuviera sobre el parámetro (debido a investigaciones anteriores y/o a otras muestras utilizadas en el pasado) y, en este sentido, es posible aceptar que exista una distribución "a posteriori" del parámetro, donde se recoge la modificación de la distribución de probabilidad de dicho parámetro cuando se dispone de la información muestral.

Por ello, el proceso bayesiano aplicado a la estimación en el análisis de mortalidad requiere asignar distribuciones de probabilidad “a priori” al parámetro desconocido " t_x " (según la opinión previa a la obtención de la muestra) que se considera variable aleatoria. La selección de la función de densidad previa de " t_x " es una manera de caracterizar o explicitar el grado de conocimiento sobre el parámetro " t_x ".

Se podría plantear también la idea de un parámetro " t_x " fijo aunque desconocido (no aleatorio en el sentido conceptual estricto) al que se tratase matemáticamente como una variable aleatoria. Se considerará " T_x " como variable aleatoria del parámetro " t_x ", " T " como el vector aleatorio para todos los valores de " x ", y " t " como el vector de realizaciones o concreciones de " T ". Dado el carácter multidimensional del problema se obtendrá una función multivariante de densidad previa para los parámetros que expresan el comportamiento aleatorio de la variante tasa de mortalidad para los distintos " x " (normalmente edades enteras y para un periodo de un año). Esta función debería contener en su estructura, teóricamente, todos los conocimientos y opiniones previas sobre " T ". En la práctica se consigue una aproximación a esa densidad ideal utilizando un miembro de una familia adecuada de distribuciones de probabilidad que presente propiedades matemáticas convenientes.

Tras formular la distribución de probabilidad previa se precisa seleccionar una expresión para la distribución condicionada de los datos observados, supuesta la secuencia " t_x ". Se trata del modelo experimental o modelo estocástico que siguen los datos teniendo en cuenta la presentación de " t ".

Los datos de este modelo muestral serán las estimaciones iniciales obtenidas con la información almacenada por las compañías de seguros (recogidos en sus pólizas). En estos datos se puede considerar su distribución (sin apoyarse en un modelo probabilístico según cierto " t ").

Con todo ello, se llega a la obtención de la distribución "a posteriori"

de "T" condicionada a las estimaciones iniciales:

$$f_{T/U}(t/u) = \frac{f_{U/T}(u/t) f_T(t)}{f_U(u)}$$

Para el análisis y tratamiento posterior de esta distribución "a posteriori" es conveniente emplear distribuciones conjugadas. El vector de valores graduados se obtiene a partir de la información contenida en esta distribución. Si interesa el valor esperado se hallará la media, si interesa el valor más probable se hallará la moda, etc.

Se considera que los trabajos de Kimeldorf y Jones (1.967) suponen el primer método formal de graduación bayesiana al aplicar las líneas generales expuestas, concretar la forma de las distintas funciones de densidad que intervienen, analizar el significado de los parámetros y ofrecer directrices para su estimación.

Se propone una distribución previa para "T" de la forma:

$$f_T(t) = \left[(2\pi)^n |A| \right]^{-1/2} e^{-(1/2)(t-m)'A^{-1}(t-m)}$$

siendo "n" el número variables contempladas en la normal multivariante, "m" el vector de las medias y "A" la matriz de covarianzas (simétrica, definida positiva y no singular).

Para fijar "m" se puede recurrir a la información que suministre una tabla estándar con población similar a la que interese en el estudio. Se busca la mejor estimación posible sin recurrir al dato experimental.

Para la matriz "A" se introduce un importante factor subjetivo (se intenta alcanzar una guía racional al tratar el problema). En primer lugar, se atiende a la naturaleza de la graduación (relación entre tasas desconocidas), al papel de "A" como matriz de covarianzas de "T" y se recurre al concepto más intuitivo de coeficiente de correlación. En segundo lugar, se opta por admitir que la correlación disminuye con la distancia (más real) o por la posibilidad de considerar no

correlacionadas significativamente las variables (cálculos más simples).

Siguiendo la hipótesis más realista podrían quedar como condiciones deseables:

$$a_{ij} = \text{Cov}(T_i T_j) = c_{ij} \sqrt{a_{ii} a_{jj}}$$
$$j > k \Rightarrow c_{ij} > c_{ik} \Rightarrow \frac{a_{ij}}{\sqrt{a_{jj}}} > \frac{a_{ik}}{\sqrt{a_{kk}}}$$

que son cumplidas por una expresiones del tipo

$$a_{ij} = p^2 r^{|i-j|} \quad \text{con } p > 0 \quad \text{y} \quad 0 \leq r < 1$$

“ p^2 ” es la varianza de “ T_i ” (que se supone constante) y “ r ” el coeficiente de correlación cuando $|i - j| = 1$.

Se comprueba que resulta una matriz “A” positiva y simétrica. Incluso, si se pueden fijar todas las correlaciones bastaría con utilizar una expresión de la forma

$$a_{ij} = p^2 \cdot r_i \cdot r_{i+1} \cdot r_{i+2} \cdot \dots \cdot r_{j-1}$$

y si, además, las varianzas no se mantienen constantes para todas las variantes “ T_i ” se utilizaría (cuando se puedan estimar)

$$a_{ij} = p_i \cdot p_j \cdot r_i \cdot r_{i+1} \cdot \dots \cdot r_{j-1}$$

Siguiendo la hipótesis más simple se consideraría no significativo el grado de correlación entre “ T_i ” y “ T_j ” para todo “ i ” distinto de “ j ”. Se llega a una matriz diagonal de varianzas.

Se propone, también, una distribución normal multivariante para la distribución de las estimaciones iniciales condicionada a los verdaderos valores subyacentes (puede interpretarse como una aproximación desde la proporción binomial de “ U_x ”) de la forma:

$$f_{U/T}(u/t) = \left[(2\pi)^n |B| \right]^{-1/2} e^{-(1/2)(u-t)'B^{-1}(u-t)}$$

siendo "t" el vector de las medias y "B" la matriz de covarianzas. La fijación de los elementos de la matriz "B", si se supone la hipótesis de independencia, lleva a una matriz diagonal con elementos "b_{ii}" que son las varianzas de la variable aleatoria "U_i". Si se admite la proporción binomial como su comportamiento:

$$\text{Var}(u_i) = \frac{t_i(1-t_i)}{n_i} \quad \text{estimándose} \rightarrow \quad b_{ii} = \frac{m_i(1-m_i)}{n_i}$$

donde "m_i" es la mejor estimación para "t_i" y "n_i" es el tamaño muestral.

Dado que "f_U(u)" es constante respecto de "t", se supone f_U(u)=k₁ como forma conveniente para la función.

Utilizando todas las funciones analizadas, se halla la distribución "a posteriori" para "T":

$$f_{T/U}(t/u) = \frac{(2\pi)^{-n} \left[|A||B| \right]^{-1/2}}{k_1} e^{-(1/2)[(t-m)'A^{-1}(t-m)+(u-t)'B^{-1}(u-t)]}$$

que puede ser expresada de esta manera:

$$f_{T/U}(t/u) = k_2 \cdot e^{-(1/2)(t-v)'C^{-1}(t-v)}$$

con

$$\begin{aligned} v &= (A^{-1} + B^{-1})^{-1} (A^{-1}m + B^{-1}u)^{-1} \\ C &= (A^{-1} + B^{-1})^{-1} \\ k_2 &= \frac{\left[(2\pi)^n |A| \right]^{-1/2} \left[(2\pi)^n |B| \right]^{-1/2}}{k_1} e^{-(1/2)(m'A^{-1}m+u'B^{-1}u-v'C^{-1}v)} \end{aligned}$$

(siendo "k₂" una constante que no depende de "t").

La forma de la función corresponde a un modelo multinormal de media el vector " v " y de matriz de covarianzas " C ". Por ello, media, moda y mediana coinciden y pueden representar, a la vez, el vector de los valores graduados. Además, el vector " v " pondera los valores de " m ", que recogen parte de la opinión previa de " T ", y los valores de " u ", que contienen los datos observados, según las matrices de covarianzas.

La elección de los parámetros es guiada por la interpretación que se hace de ellos en este método. En la práctica, en situaciones poco claras, se gradúan los datos usando varias combinaciones de parámetros y eligiendo según la coherencia de los resultados.

8.- GRADUACIÓN BAYESIANA ISOTÓNICA.

Otro sistema de graduación basado en la metodología bayesiana es el aplicado por Broffit (1.987, 1.988). Se corrige una primera función de verosimilitud obtenida a partir de los datos con la distribución previa de los parámetros (obligados a un cierto orden creciente) para obtener la distribución corregida que conduce a un estimador bayesiano. La influencia del modelo (que representa la información previa) evita que los resultados se apoyen exclusivamente en los datos, que ante una muestra poco representativa lleva a una estimación no adecuada.

El proceso que se explicará puede interpretarse como el uso de una estimación inicial y a continuación la rectificación mediante la graduación. Otra interpretación, más global, supone entender el proceso como un problema de estimación estadística: los datos se interpretan según un modelo que integra la información previa que se tiene de los parámetros y la estimación obtenida ya no necesita rectificación.

La función de verosimilitud para los datos se tiene analizando una muestra de " N " individuos, con edades entre " x " y " $x+k$ ", que pueden salir de la observación por muerte, retirada o finalización del periodo de observación. Se supone independencia entre tiempos de muerte y retirada, fuerza de mortalidad anual constante para obtener la función de verosimilitud como proporcional a

$$L(\theta_1, \dots, \theta_k) \approx \prod_{j=1}^k [\theta_j^{d_j} e^{-e_j \theta_j}]$$

y llegar a un estimador de máxima verosimilitud (sin restricciones por información previa) de la forma

$$\theta_j^M = \frac{d_j}{e_j}$$

siendo “ d_j ” el número de muertes observadas entre “ $x+j-1$ ” y “ $x+j$ ”, y “ e_j ” la exposición o cantidad de tiempo total (medida en años) que son observados los distintos individuos.

La noción de suavidad se incorpora suponiendo una relación creciente, conocida como hipótesis isotónica, de la forma:

$$q_x < q_{x+1} < \dots < q_{x+k-1}$$

Para cada periodo anual se supone una fuerza de mortalidad constante y, por ello, $q_{x+j-1} = 1 - e^{-\theta_j}$, con lo que la hipótesis isotónica es equivalente a $\theta_1 < \theta_2 < \dots < \theta_k$.

En busca de garantizar esta situación, se desarrollará un estimador bayesiano que garantice $P[\Theta_1 < \Theta_2 < \dots < \Theta_k] = 1$ y, entonces, $P[\Theta_1 < \Theta_2 < \dots < \Theta_k / X] = 1$ (siendo X la muestra de datos que se tenga).

El estimador se tomará como:

$$\begin{aligned} \theta_j^{IB} &= E[\Theta_j / X] \Rightarrow \theta_1^{IB} < \dots < \theta_k^{IB} \\ \Rightarrow q_x^{IB} &< \dots < q_{x+k-1}^{IB} \quad \text{con} \quad q_{x+j-1}^{IB} = 1 - e^{-\theta_j^{IB}} \end{aligned}$$

Para la distribución previa se utilizan distribuciones gamma independientes “ Y_j ”, dos posibilidades son la “previa ordenada” o la “previa aditiva”:

$$\begin{aligned} (\Theta_1, \dots, \Theta_k) &\equiv (Y_1, \dots, Y_k / Y_1 < \dots < Y_k) \quad \text{o} \\ (\Theta_1, \dots, \Theta_k) &\equiv (Y_1, Y_1 + Y_2, \dots, Y_1 + \dots + Y_k) \end{aligned}$$

La segunda opción será la elegida debido a la mayor facilidad para la especificación de parámetros.

En la denominada “previa aditiva” se pueden considerar las ecuaciones:

$$\begin{aligned} \phi_1 &= \theta_1 & \theta_1 &= \phi_1 \\ \phi_2 &= \theta_2 - \theta_1 & \theta_2 &= \phi_1 + \phi_2 \\ \dots & & \dots & \\ \phi_k &= \theta_k - \theta_{k-1} & \theta_k &= \phi_1 + \dots + \phi_k \end{aligned}$$

Y llegarse a que son equivalentes las dos situaciones siguientes:

$$\begin{aligned} (\Theta_1, \dots, \Theta_k) &\equiv (Y_1, Y_1 + Y_2, \dots, Y_1 + \dots + Y_k) \\ (\Phi_1, \dots, \Phi_k) &\equiv (Y_1, \dots, Y_k) \end{aligned}$$

Ahora se toman las distribuciones gamma con sus parámetros $Y_j \equiv G(a_j, r_j)$ para llegar a una expresión proporcional a la verosimilitud de la distribución previa:

$$prior_A(\phi_1, \dots, \phi_k) \approx \prod_{j=1}^k \phi_j^{a_j-1} e^{-r_j \phi_j} \quad \phi_j > 0, \quad j = 1, \dots, k$$

Finalmente la expresión proporcional a la verosimilitud de la distribución posterior queda:

$$\begin{aligned} post_A(\phi_1, \dots, \phi_k) &\approx L(\phi_1, \dots, \phi_k) \cdot prior_A(\phi_1, \dots, \phi_k) \approx \\ &\approx \prod_{j=1}^k (\phi_1, \dots, \phi_j)^{d_j} \phi_j^{a_j-1} e^{-b_j \phi_j} \quad \text{con } \phi_j > 0, \quad j = 1, \dots, k \end{aligned}$$

siendo

$$b_j = r_j + \frac{e_j}{\phi_j} (\phi_1 + \dots + \phi_j).$$

Esto nos lleva a una gamma:

$$\text{post}_A(\phi_1, \dots, \phi_k) \approx \prod_{j=1}^k (\phi_1, \dots, \phi_j)^{d_j} g(\phi_j, a_j, b_j)$$

Para expresar la “posterior aditiva” de forma más precisa se definen “k” gammas independientes $Z_j \equiv G(a_j, b_j)$ para $j=1, \dots, k$; $b = (b_1, \dots, b_k)$; $a = (a_1, \dots, a_k)$; $d = (d_1, \dots, d_k)$; y, $B(d; a; b) = E[Z_1^{d_1} (Z_1 + Z_2)^{d_2} \dots (Z_1 + \dots + Z_k)^{d_k}]$. Con estos elementos se puede llegar a:

$$\text{post}_A(\phi_1, \dots, \phi_k) = \frac{\prod_{j=1}^k (\phi_1, \dots, \phi_j)^{d_j} g(\phi_j, a_j, b_j)}{B(d; a; b)}$$

Para hallar el estimador de Bayes isotónico según la aditiva previa se procede recurriendo a la relación entre parámetros:

$$\theta_j^{\text{IBA}} = \sum_{i=1}^j \phi_i^{\text{IBA}}$$

Por ello, se encuentra la expresión:

$$\begin{aligned} \phi_i^{\text{IBA}} &= E(\Phi_i / X) = \int_0^\infty \dots \int_0^\infty \phi_i \text{post}_A(\phi_1, \dots, \phi_k) d\phi_1 \dots d\phi_k = \\ &= \frac{a_i}{b_i} \frac{B(d; a^{(i)}; b)}{B(d; a; b)} \end{aligned}$$

considerando que

$$\begin{aligned} a^{(i)} &= (a_1, \dots, a_{i-1}, a_i + 1, a_{i+1}, \dots, a_k) \\ \phi g(\phi; a, b) &= \frac{a}{b} g(\phi; a + 1, b) \end{aligned}$$

Dado que el parámetro que interesa se obtiene fácilmente a partir del anterior, el problema está resuelto:

$$\theta_j^{IBA} = \sum_{i=1}^j \phi_i^{IBA} = \sum_{i=1}^j \frac{a_i}{b_i} \frac{B(d; a^{(i)}; b)}{B(d; a; b)}$$

Sólo queda por discutir el valor asignable a los parámetros previos. Los parámetros (a_1, \dots, a_k) y (r_1, \dots, r_k) se pueden determinar una vez que la media y la varianza de $(\Theta_1, \dots, \Theta_k)$ se hayan determinado. Si los valores dados a la media y a la varianza de " Θ_j " son " θ_j^0 " y " v_j^0 " se tiene:

$$\theta_j^0 - \theta_{j-1}^0 = E(\Phi_j) = \frac{a_j}{r_j}, \quad \text{con } j = 1, \dots, k$$

$$v_j^0 - v_{j-1}^0 = \text{Var}(\Phi_j) = \frac{a_j}{r_j^2}, \quad \text{con } j = 1, \dots, k$$

$$\text{donde } \theta_0^0 = v_0^0 = 0$$

Resolviendo las ecuaciones

$$r_j = \frac{\theta_j^0 - \theta_{j-1}^0}{v_j^0 - v_{j-1}^0} \quad \text{y} \quad a_j = \frac{(\theta_j^0 - \theta_{j-1}^0)^2}{v_j^0 - v_{j-1}^0}$$

El cálculo de $B(d; a; b)$ requiere el uso de un algoritmo o una aproximación.

9.- PRUEBAS PARA EVALUAR EL PROCESO DE GRADUACIÓN.

Una interesante aplicación de las pruebas para aceptar o rechazar una determinada graduación puede encontrarse en Navarro (1.992). Otros trabajos que aportan más tests con el señalado fin se encuentran en

Benjamin y Pollard (1.980) y en Forfar, McCutcheon y Wilkie (1.988).

El análisis se dedica a examinar las desviaciones entre valores estimados y valores graduados a lo largo de toda la secuencia. Se considera que las desviaciones tienen que estar distribuidas de forma aleatoria (no hay errores sistemáticos en toda la secuencia de valores graduados) y su distribución concreta será coherente con las hipótesis iniciales utilizadas (según el modelo binomial, Poisson o normal).

Atendiendo a estas directrices generales se elaboran unas pruebas para las desviaciones absolutas y relativas entre estimaciones iniciales y corregidas, considerándolas individualmente y dentro de la secuencia total. Los tests o pruebas que se mencionan son: test de intervalos de confianza, test de desviaciones relativas, test de desviaciones acumuladas, test de signos, test de cambio de signo, test de la chi-cuadrado, test de Kolmogorov-Smirnov.

El *test de intervalos de confianza* consiste en comprobar si los valores obtenidos tras el proceso de graduación se encuentran dentro de un intervalo de confianza generado a partir de la distribución del estadístico utilizado para estimar la probabilidad de muerte. Se debe realizar la prueba para toda la secuencia ya que las conclusiones dependen del número de estimaciones corregidas o graduadas que queden fuera de sus respectivos intervalos. Si es muy elevado el número se tienen indicios de que los valores graduados no recogen de forma adecuada la mortalidad del colectivo estudiado. Si es muy reducido el número se tienen indicios de que el problema proviene de los datos en los tramos de edad que se produzca esta situación. El número de estimaciones corregidas que queden fuera de sus respectivos intervalos se considera numeroso en función del nivel de significación tomado. La decisión, en todo caso, tiene un margen amplio de subjetividad.

El *test de las desviaciones relativas* consiste en comprobar si las desviaciones, para cada edad, entre el número real de muertes y el valor esperado según la estimación corregida, dividido entre la desviación típica según la estimación corregida (desviaciones relativas o tipificadas) se comportan o son compatibles con la hipótesis de

normalidad.

El *test de las desviaciones acumuladas*, basado en la independencia de las muertes a las respectivas edades, consiste en comprobar si la variable obtenida con las desviaciones sumadas entre el número real de muertes y el valor esperado atendiendo a la estimación corregida, para todas las edades, se comporta con ley normal de media cero y de varianzas la suma de las varianzas a cada edad (exposición por probabilidad de muerte y por la complementaria).

El *test de signos* consiste en analizar el signo de las desviaciones relativas admitiendo como hipótesis inicial que la probabilidad de que exista una desviación positiva es la misma que la probabilidad de que la desviación sea negativa. Por ello, siendo independientes los signos para las distintas edades, la distribución en la secuencia del número de signos positivos o negativos sigue una ley binomial de parámetros “N” (número de edades consideradas en la secuencia) y “0’5” (probabilidad de signo positivo o negativo para una edad cualquiera). Con esta información, se pueden construir los intervalos de confianza para el número de desviaciones positivas o negativas.

El *test de cambios de signo*, supuesta la normalidad del número de fallecimientos a cada edad, se basa en la independencia entre las desviaciones de edades sucesivas con lo que se tiene una distribución aleatoria de los signos de las desviaciones (número no muy alto ni bajo de cambios de signo). Es posible construir intervalos de confianza para el número de cambios de signo.

El *test de la chi-cuadrado* comprueba que la suma de todas las desviaciones relativas o tipificadas elevadas al cuadrado se distribuyen como una chi-cuadrado de “N-n” grados de libertad, siendo “N” el número de edades considerado y “n” el números de parámetros estimados en el proceso de ajuste.

El test de Kolmogorov-Smirnov se basa en la convergencia casi segura de la función de distribución empírica de la muestra a la función de distribución poblacional. Se plantea el estadístico con una distribución asintótica libre de la función de distribución poblacional, dada la distribución uniforme en $[0,1]$ para la variable obtenida con

los valores de la mencionada función de distribución poblacional.

Interesa definir la variable número de fallecidos a cada edad, identificar la distribución de la población con las muertes esperadas según el modelo de graduación e identificar la distribución empírica con el número de muertes realmente ocurridas (información muestral). El mayor problema que se presenta al aplicar este test en el proceso de graduación radica en que la distribución poblacional (según los valores graduados) se derivan (en parte) de las muertes reales (los mismos datos reales que generan la distribución empírica). Para admitir la validez del test debería existir independencia entre la obtención de fallecimientos esperados a cada edad y los datos de fallecimientos reales. En la mayoría de los casos esto no se produce. A pesar de ello, sigue aportando una prueba negativa: si se obtienen discrepancias significativas entre las dos funciones el proceso de graduación no será satisfactorio.

10.- CONCLUSIONES.

Con el objetivo de finalizar esta revisión crítica de los diferentes métodos no paramétricos de graduación actuarial incluyendo los temas más recientes conviene, a modo de resumen, señalar varias cuestiones. La graduación es un tema actuarial aunque la exposición y desarrollo de estos métodos confirma, obviamente, que su naturaleza es compartida por dos Ciencias: la Actuarial y la Estadística.

Es importante, además, no sólo realizar las estimaciones estadísticas pertinentes, sino también, establecer si éstas pueden o no considerarse correctas gracias a la aplicación de distintos tests comentados en el último epígrafe de este artículo.

Se trata de un campo de estudio aún en desarrollo, muy interesante y en el que se debería continuar investigando para poder ofrecer cada vez mejores estimaciones que ayuden a elaborar tablas de mortalidad más ajustadas al fenómeno actuarial que describen y, por tanto, con una precisión y fiabilidad mayores.

11.- BIBLIOGRAFÍA.

- [1]- ANDREWS, G. y NESBITT, C.J. (1965). **Periodograms of graduation operators.** Transactions of the Society of Actuaries. (Núm.17, págs.166-177).
- [2]- BENJAMIN, B. y POLLARD, J.H. (1.980). **The analysis of Mortality.** Ed. Heinemann. Londres.
- [3]- BETZUEN ZALBIDEGOITIA, A.; FELIPE CHECA, A. y GUILLÉN ESTANY, M. (1.997). **Modelos de tablas de mortalidad en España y situación actual.** Anales del Instituto de Actuarios Españoles. (Tercera época, Núm.3, págs.79-104).
- [4]- BORGAN, O. (1.979). **On the theory of moving average graduation.** Scandinavian Actuarial Journal. (Núm.2-3, págs.83-105).
- [5]- BROCKETT, P. y ZHANG, J. (1.986). **Information theoretical mortality table graduation.** Scandinavian Actuarial Journal. (Núm.3-4, págs.131-140).
- [6]- BROFFIT, J.D. (1.987). **Isotonic bayesian graduation with an additive prior.** Actuarial science. (Vol.6, págs.19-39).
- [7]- BROFFIT, J.D. (1.988). **Increasing and increasing convex bayesian graduation (with discussion).** Transactions of the Society of Actuaries (Vol.XL, págs.115-148).
- [8]- BROFFIT, J.D. (1.996). **On smoothness terms in multidimensional Whittaker graduation.** Insurance: Mathematics and Economics. (Vol.18, nº1, págs.13-27). North-Holland.
- [9]- CHAN, F.Y.; CHAN, L.K. y MEAD, E.R. (1.982). **Properties and modifications of Whittaker-Henderson graduation.** Scandinavian Actuarial Journal. (Núm.1, págs.57-61).
- [10]- FELIPE CHECA, A. y GUILLÉN ESTANY, M. (1.999). **Evolución y predicción de tablas de mortalidad dinámicas para la población española.** Cuadernos de la Fundación. Editorial Mapfre Estudios.
- [11]- FORFAR, D.O.; MCCUTCHEON, J.J. y WILKIE, A.D. (1.988). **On the graduation by mathematical formula.** Journal of the Institute of Actuaries. (Núm.115, págs.1-135).

- [12]- GAVIN, J.; HABERMAN, S. y VERRAL, R. (1.993). **Moving weighted average graduation using kernel estimation.** Insurance, mathematics and economics. (Núm.12, págs.113-126). North-Holland.
- [13]- GAVIN, J.B.; HABERMAN, S. y VERRAL, R.J. (1.994). **On the choice of bandwidth for Kernel graduation.** Journal of the Institute of Actuaries. Vol.121, nº 478. pp: 119-134.
- [14]- GERBER, H. (1.990). **Life insurance mathematics.** Swiss Association of Actuaries. Berlin.
- [15]- HENDERSON, R. (1.924). **A new method of graduation.** Transactions of the Actuarial Society of America. Vol. 25, págs. 29-40.
- Hickman, J.C. (1.987)
- [16]- HOEM, J.M. y LINNEMANN, P. (1.988). **The tails in moving average graduations.** Scandinavian Actuarial Journal. (Núm.4, págs.193-229).
- [17]- JORDAN, C. W. (1.991). **Society of actuaries's textbook on life contingencies.** Society of Actuaries. Chicago.
- [18]- KENNETH, B. (1.987). **Life insurance.** Prentice-Hall. Nueva Jersey.
- [19]-KIMELDORF, G.S. y JONES, D.A. (1.967). **Bayesian graduation.** Transactions of the Society of Actuaries (Vol.XIX, págs.66-112).
- [20]- KLEINBAUM, D.G. (1.995). **Survival Analys.** Springer. New York.
- [21]- LE, C.T. (1.997). **Applied survival analysis.** Wiley. New York.
- [22]- LEE, E.T. (1.992). **Statistical methods for survival data analysis.** Wiley. New York.
- [23]- LONDON, R. (1.984). **Graduation: the revision of estimates,** Itasca, Illinois: Society of Actuaries.
- [24]- LÓPEZ CACHERO, M. y LÓPEZ DE LA MANZANARA BARBERO, J. (1.996). **Estadística para actuarios.** Editorial Mapfre. Madrid.
- [25]- MALLER, R. y XIAN, Z. (1.996). **Survival Analysis with Long Term Survivors.** J.Wiley. Nueva York.

- [26]- NAVARRO, E. (1.992). **Tablas de mortalidad de la población española 1.982. Metodología y fuentes.** Editorial Mapfre. Madrid.
- [27]- NEILL, A. (1.989). **Life contingencies.** Ed. Heinemann. Londres.
- [28]- PARMAR, M.K.B. (1.995). **Survival analysis.** J. Wiley. Nueva York.
- [29]- PRIETO PÉREZ, E. y FERNÁNDEZ PLASENCIA, M.J. (1.994). **Tablas de mortalidad de la población española de 1.950 a 1.990. Tabla proyectada del año 2.000. Tablas con y sin margen de seguridad.** UNESPA. Madrid.
- [30]- RAMLAU-HANSEN, H. (1.983). **The choice of a Kernel function in the graduation of counting process intensities.** Scandinavian Actuarial Journal. (Núm.3, págs.165-182).
- [31]- TAYLOR, G. (1.992). **A bayesian interpretation of Whittaker-Henderson graduation.** Insurance, Mathematics and Economics. (Núm.11, págs.7-16). North-Holland.
- [32]- VERRALL, R.J. (1.993). **A state space formulation of Whittaker-Henderson graduation, with extensions.** Insurance, Mathematics and Economics. (Núm.13, págs.7-14). North-Holland.
- [33]- WHITTAKER, E.T. (1.923). **On a new method of graduation.** Proceedings of the Edinburg Mathematical Society. Vol. 41, págs. 63-75.