$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[}(T_x)$$

DURATION MODELS

UNIL | Université de Lausanne

Lecture Notes 2015-2016

Part 5

Smoothing and fitting methods

Frédéric PLANCHET / Olivier DURAND

April 2016

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{I}_{]t;\infty[}(T_x)$$

Table of content

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{]t;\infty[}(T_x)$$

1. Introduction

When building an experience distribution based on exit or survival data, the first step consists in estimating gross rates[1]; this is an essential step, whether the adopted approach is parametric or non-parametric.

The resulting obtained values show some irregularity, which can be thought as not reflecting the underlying studied phenomenon, and is rather the outcome of the imperfection of conditions of experiment; sampling fluctuations thus induce a "parasitic" variability in estimated values.

One then wishes to apply "fitting" or "smoothing" to gross rates, in order to represent in a more faithful way the (unknown) distribution which one wants to estimate.

In a formal way, and while being placed in the standard case of incidence rates estimation (mortality, exit of sick leave, long term care, etc.), the initial estimation procedure led to a value $\hat{q}_x$ for estimating q_x, and thus to an error $e_x = \hat{q}_x - q_x$; the purpose of the estimate update which is to be undertaken is to decrease this error, while building a curve of rates as a function of x more "smooth" than the curve of gross rates.

It can also be seen that the gross rates estimation process generally considers ages (or seniorities) independently from one another, and thus does not take into account the interactions which obviously exist: for example, in the case of mortality and starting from a certain age, it seems natural that the theoretical series of q_x is increasing. The series of $\hat{q}_x$ is in general not increasing.

1.1. Regularity and precision

The choice of a gross data revision procedure sees two types of constraints which will have to be taken into account in a joint way:

- ✓ precision (or fidelity / "fit"): it is natural to expect revised rates which are close to initial rates;

- ✓ regularity ("smoothness"): the series of adjusted rates will be seeked as regular as possible.

As for any statistical approach, the relevance of the fitting procedure will have to be validated by tests.

1.2. Various approaches

The revision process of the initial estimate can be carried out in two ways:

- ✓ one can set an *a priori* form for the underlying distribution, for example assuming that the hazard function is a Makeham function; this is a fitting approach defined by a parameter θ; examples of such procedures are spline smoothing and "smooth junction" interpolations;

---

[1] Or the empirical estimate of a characteristic function of the survival distribution: cumulated hazard function, survival function, *etc*.

✓ one can live without a parametric representation, and simply define processes to be applied to initial gross data in order to "smooth" them; examples of such non-parametric methods are moving averages[2], Whittaker-Henderson type smoothing as well as their extension to a more general Bayesian framework.

From a semantic point of view, one will tend to refer to "fitting" in the parametric framework and "smoothing" in the non-parametric framework – this terminology not actually being fixed.

The two approaches are developed below.

1.3. Definitions and notations

Smoothing methods are often based on the series of term to term differences of the original series. In practice one finds three discrete differentiation operators:

✓ forward difference: $\Delta u(x) = u(x+1) - u(x)$

✓ backward difference: $\nabla u(x) = u(x) - u(x-1)$

✓ central difference: $\delta u(x) = u\left(x+\frac{1}{2}\right) - u\left(x-\frac{1}{2}\right)$

These operators can be applied in a recursive way: $\Delta^2 u = \Delta(\Delta u)$; thus $\Delta^2 u(x) = u(x+2) - 2u(x+1) + u(x)$; more generally $\Delta^n u(x)$, is written with the binomial coefficients:

$$\Delta^n u(x) = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} u(x+j)$$

One obtains similar expressions with ∇ and δ.

2. Parametric fitting

The common approach described in lecture notes about "parametric models" consists in carrying out a fitting to a generally continuous parametric distribution; it will therefore not be developed here again.

It can be noted that in this case the maximum likelihood method directly leads from observations to fitted values *via* obtaining the parameters of the underlying distribution; the two-steps reasoning – obtaining gross rates, then fitting – can therefore be simplified.

In the case of sick leave, methods initially developed for construction of prospective mortality tables, like the Lee-Carter method, can be adapted to the fitting of instantaneous exit rates μ_{xt} where *x* is the seniority of sick leave and *t* the age at the beginning of sick leave[3].

---

[2] Which are in a way the ancestors of smoothing procedures.
[3] See the lecture notes on "mortality tables" for the presentation of these methods

2.1. Spline smoothing

2.1.1. Presentation

Fitting to a continuous distribution implicitly assumes that the curve of incidence rates can be represented for all considered ages or seniorities by a single parametric function. In practice, for example because of ruptures in the evolution of gross rates, this condition is rather restrictive.

The idea of spline smoothing is to divide the support of the function to be fitted in sub-intervals, then fit a simple function on each sub-interval, and be cautious of proper connections at junction points. Dividing in a smart way should indeed allow the use on each sub-interval of a function materially more simple than the function corresponding to a global fit.

Polynomials are simple functions and can for this reason be used to carry out spline smoothing; in practice, one only considers from now on polynomials of degree 3 which will allow us to build cubic splines. The connection of these arcs will be done by imposing continuity as well as equality of slopes and curvatures, at the junction points.

2.1.2. Cubic spline with two arcs

Initially, it is supposed that dividing the support of variation of ages (or seniorities) in two parts is sufficient, and one therefore writes:

$$q_x = \begin{cases} p_0(x) & x_0 \le x \le x_1 \\ p_1(x) & x_1 \le x \le x_2 \end{cases}$$

$p_i(x)$ being a polynomial of degree 3, with the following constraints at the junction point:

$$p_0(x_1) = p_1(x_1) \quad \frac{d}{dx} p_0(x_1) = \frac{d}{dx} p_1(x_1) \quad \frac{d^2}{dx^2} p_0(x_1) = \frac{d^2}{dx^2} p_1(x_1)$$

Which results in posing $p_0(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$ and $p_1(x) = p_0(x) + c_5(x - x_1)^3$. The problem thus comprises 5 unknown factors (8 polynomials coefficients minus 3 regularity constraints). One solves it by using a criterion of weighed least squares, on the basis of weights (w_x), which results in looking for the parameters which minimise

$M = \sum_{x=x_0}^{x_2} w_x (q_x - \hat{q}_x)^2$.

It can be noted that in this approach it is not necessary to have all of the gross values $\hat{q}_x$ as the spline could be interpolating for missing values. If one then notes $\bar{x}_1$ the greatest value of x lower or equal to x_1 for which one has a value of $\hat{q}_x$, then one breaks down the sum involved in criterion M into two sums and one writes the normal equations by cancelling the derivative with respect to the parameters: $\frac{\partial M}{\partial c_i} = 0$.

After some calculations, these equations can be put in the form:

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[}(T_x)$$

$$X'wXc = X'w\hat{q}$$

the matrix X of size $(m,5)$ for m $\hat{q}_x$ values available on $[x_0, x_2]$ being defined by:

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & 0 \\ .. & .. & .. & .. & .. \\ 1 & \overline{x}_1 & \overline{x}_1^2 & \overline{x}_1^3 & 0 \\ 1 & \overline{x}_1^1 & \left(\overline{x}_1^1\right)^2 & \left(\overline{x}_1^1\right)^3 & \left(\overline{x}_1^1 - x_1\right)^3 \\ .. & .. & .. & .. & .. \\ 1 & x_2 & x_2^2 & x_2^3 & \left(x_2 - x_1\right)^3 \end{bmatrix}$$

with $\overline{x}_1^1$ the index value posterior to $\overline{x}_1$ for which $\hat{q}_x$ is known.

2.1.3. Cubic splines: the general case

The above expressions can easily be generalised to the n nodes case $x_1, ..., x_n$, with a matrix of size $(m, n+4)$; the coefficients are obtained through the ordinary least squares estimator:

$$c = \left(X'wX\right)^{-1} X'w\hat{q}.$$

2.1.4. Choice of parameters

The choice of parameters is reduced to the choice of nodes' locations. Since the fitting function is a polynomial of degree 3, it can be noticed that if only 4 values of $\hat{q}_x$ are available in the interval $[x_i, x_{i+1}]$, the fitted values will be equal to the gross ones: it is thus advisable to choose sufficiently broad intervals; in practice, changes in gross curve curvature will provide indications on locations of the nodes.

2.1.5. Generalisation

The spline smoothing method can be presented in a more general non-parametric regression framework, which allows the introduction of a fidelity vs regularity arbitrage through a criterion similar to that of Whittaker-Henderson (see section 3.3 below).

This presentation will not be developed here, but the interested reader on the matter will be able to consult Besse and Cardot [2001].

2.2. Smooth junction interpolations

If, instead of calculating a gross estimate for each age, individuals are gathered by age groups, then one obtains gross estimates by "steps" (a step corresponds to an age group). This situation can for example be seen in the case of the construction of sick leave retention distributions, with age at entry being regrouped.

One then wishes to get back to estimates for each age. The "smooth junction interpolations" technique meets this objective. It will be noted that the objective is not here to smooth or fit

erratic values, but rather to supplement missing values. The cubic splines fitting technique also allows to interpolate missing values, and it will have to be used preferably to these interpolations.

2.2.1. Everett interpolation formula

One now looks into symmetrical interpolation procedures, in the sense that the direction of the interpolation does not modify the result. It is then possible to write the interpolation formula in the following general form:

$$q_{x+s} = F(s)\hat{q}_{x+1} + F(1-s)\hat{q}_x$$

for $0 \le s \le 1$ and $F(s)$ a differentiation operator of the form:

$$F(s) = p_0(s) + p_1(s)\delta^2 + p_2(s)\delta^4 + ..$$

With δ the central difference operator and p_i some polynomials. One can for example imagine that one has quinquennial rough rates, and one will use this approach with $s = 0,2; 0,4; ..$ in order to obtain annual rates through interpolation. If if in the expression of $F(s)$ above the last term is in δ^{2m}, then q_{x+s} is determined based on the $2(m+1)$ values $\hat{q}_{x-m}, .., \hat{q}_{x+m+1}$. This is referred to as "interpolation formula at $2(m+1)$ points". When $m = 0$ and $p_0(s) = s$ one obtains the traditional linear interpolation formula. Lastly, one can note that q_{x+s} is a polynomial function in *s*. This is therefore a spline-type interpolation, as in 2.1 above.

Regularity conditions on the obtained interpolation function are imposed:

- ✓ continuity implies that $p_i(0) = 0$;

- ✓ equality of the derivative at the junction points, which implies the (formal) equality $2F'(1) = F'(0)(2+\delta^2)$;

- ✓ equality of the second derivative at the junction points, which implies that $p_i''(0) = 0$.

The above conditions do not impose that the interpolation formula provides original values to integer points, *i.e.* one can have $q_x \ne \hat{q}_x$. In order to actually keep original values at integer points, it is necessary to also have $p_0(1) = 1$ and $p_i(1) = 0$, $i \ge 1$.

In the same way one can determine conditions on polynomials p_i so that the formula is exact for the interpolation of polynomials of fixed degree z.[4]

2.2.2. Four points formulae

In practice, the 4 points formulae, hence with $m = 2$, are the most commonly used. One thus seeks to express q_{x+s} as a function of to $\hat{q}_{x-1}, \hat{q}_x, \hat{q}_{x+1}, \hat{q}_{x+2}$; one has $F(s) = p_0(s) + p_1(s)\delta^2$

---

[4] Cf. London [1995].

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{I}_{]t;\infty[}(T_x)$$

and it is thus necessary to determine the polynomials p_0 and p_1. The constraints of continuity, derivability and invariance for the polynomials of degree one easily imply that: $p_0(s) = s$, $p_1(0) = 0$, $p_1'(0) = 0$ and $p_1'(1) = \frac{1}{2}$; one controls regularity via $l = p_1(1)$. Hence there are 4 constraints and the polynomial of minimal degree which satisfies them is of degree 3, with the following expression:

$$p_1(s) = \left(3l - \frac{1}{2}\right)s^2 + \left(\frac{1}{2} - 2l\right)s^3$$

The typical case $l = 0$ led to the Karup-King formula, often used in practice:

$$p_1(s) = \frac{1}{2}s^2(s-1).$$

3. Non-parametric smoothing methods

3.1. Moving averages

Moving averages have the advantage of simplicity of implementation; however, they present some disadvantages, essentially related to the sensitivity of the arithmetic mean to extreme values, which will often eventually result in moving averages not being used. At the very least, moving averages will not be the only means of revision of gross rates.

Provided one restricts oneself to symmetrical moving averages, the basic formula is as follows:

$$q_x = \sum_{i=-r}^{+r} a_i \hat{q}_{x+i}$$

With $a_{-i} = a_i$. The major limitation of moving averages is that their use at the edges is problematic.

When aiming to decrease estimation error within the framework of the measurement of theoretical rates q_x, constraints can be imposed that consist in requiring that should the series of q_x present the regularity of a polynomial, for example of degree 3, application of the moving average would not modify values of q_x; in other words, one writes $q_x = \sum_{i=-r}^{+r} a_i q_{x+i}$ which leads to:

$$\sum_{i=-r}^{+r} a_i = 1 \quad \sum_{i=-r}^{+r} i^2 a_i = 0$$

In some cases (sick leave with age at entry and leave seniority as dimensions, or prospective mortality with age and year as dimensions), two-dimensional tables push to look for the smoothing method best adapted to each of both dimensions.

The idea is that, should smoothing in one direction and then the other be chosen, one is likely to deteriorate the first smoothing. One therefore seeks to smooth the series in the two directions at once. The moving average method is well adapted to that purpose. It is for

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[}(T_x)$$

example used to smooth reserves calculated based on regulatory gross tables. In this case the BCAC uses:

$$q(i,j) = \frac{1}{2}\hat{q}(i,j) + \frac{1}{16}\sum_{k,l}\hat{q}(k,l),$$

the summation being extended to the 8 points bordering (i,j).

This approach can be generalised to any moving average; however, this process is not well adapted to the treatment of the edges of the table, and results in practice in letting some irregularities remain.

3.2. Kernel methods

When estimating a survival function through a non-parametric method, the obtained function is non differentiable, and it is therefore not possible to recompute the hazard function simply. When one wishes to regularise the empirical cumulative distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i \le x\}}$, a traditional method consists in choosing a function K regular, positive and of unit integral, called kernel, and in posing:

$$\hat{f}_{h,n}(x) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{X_i - x}{h}\right).$$

When $h \to 0$ and $nh \to +\infty$, $\hat{f}_{h,n}(x)$ converges towards $f(x)$ and one has:

$$\sqrt{nh}\left(\hat{f}_{h,n}(x) - f(x)\right) \to N\left(0, f(x)\int K(u)^2 \, du\right).$$

Among the kernels often used one can quote the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$

and the Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)\mathbf{1}_{\{|u|\le 1\}}$. The practical main difficulty consists in determining the optimal "bandwidth" h.

In the context of a censored survival model, the direct application of these formulae introduces a bias because of (right) censoring. Marron and Padgett [1987] proposed to use the following estimator:

$$\hat{f}_{h,n}(t) = \frac{1}{h}\sum_{i=1}^{n}\frac{d_i}{r_i}K\left(\frac{T_i - t}{h}\right)$$

in which the uniform jump $\frac{1}{n}$ of the non-censored case is replaced by the jump of the Kaplan-Meier estimator at the time of occurrence of a non-censored exit.

For a complete presentation of these methods, the reader can refer to Wand and Jones [1995].

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbb{1}_{]t;\infty[} (T_x)$$

3.3. Whittaker-Henderson smoothing

The principle of the Whittaker-Henderson method is to combine a fidelity criterion with a regularity criterion and to look for fitted values that minimise the sum of both criteria.

3.3.1. The one-dimensional case

One sets weights (w_i) and one poses for the fidelity criterion:

$$F = \sum_{i=1}^{p} w_i (q_i - \hat{q}_i)^2$$

and for the of regularity criterion:

$$S = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$$

z being a model parameter. The criterion to be minimised is a linear combination of fidelity and regularity, the weight of each of the two terms being controlled by a second parameter h:

$$M = F + h \times S$$

The solution to this optimisation problem satisfies the conditions $\dfrac{\partial M}{\partial q_i} = 0$, $1 \leq i \leq p$; the resolution of this system of equations can be carried out by means of a few matrix processing steps. For that purpose, one poses $q = (q_i)_{1 \leq i \leq p}$, $\hat{q} = (\hat{q}_i)_{1 \leq i \leq p}$ and $w = diag(w_i)_{1 \leq i \leq p}$; with these notations one can write $F = (q - \hat{q})' w (q - \hat{q})$; as far as the regularity criterion is concerned, if one notes $\Delta^z q = (\Delta^z q_i)_{1 \leq i \leq p-z}$, then $S = (\Delta^z q)' \Delta^z q$. In order to further detail this writing, one introduces the matrix K_z of size $(p-z, p)$, the terms of which are the binomial coefficients of order z of alternating sign and starting positively for z even[5]:

$$\Delta^z q(i) = \sum_{j=0}^{z} \binom{z}{j} (-1)^{z-j} q(j+i).$$

For example, for $z = 2$ and $p = 5$ one gets:

$$K_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

If $p = 3$ and $z = 1$, one obtains $K_2 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$. It is easily verified that $\Delta^z q = K_z q$, which finally makes it possible to write criterion M in the following form:

$$M = (q - \hat{q})' w (q - \hat{q}) + h q' K_z' K_z q$$

---

[5] Cf. section 1.3

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[} (T_x)$$

By developing the above expression one finds that:

$$M = q'w\,q - 2q'w\,\hat{q} + \hat{q}'w\,\hat{q} + hq'K_z'K_z q$$

which leads to:

$$\frac{\partial M}{\partial q} = 2w\,q - 2w\,\hat{q} + 2hK_z'K_z q\ .$$

The resolution of $\dfrac{\partial M}{\partial q} = 0$ leads to the following expression for fitted rates:

$$q^* = \left(w + hK_z'K_z \right)^{-1} w\,\hat{q}$$

However, the inversion of matrix $C = w + hK_z'K_z$ requires some care, since $hK_z'K_z$ is not invertible, and adding the term w makes C invertible, but the inversion of C can then be delicate. In practice, the Cholesky decomposition of the positive symmetrical matrix C can be used to invert it.

3.3.2. Extension to the two-dimensional case

The extension of the Whittaker-Henderson method to the dimension 2 (or more) does not pose major problems. One thus has estimates $\hat{q} = \left(\hat{q}_{ij} \right)_{1 \le i \le p, 1 \le j \le q}$; the fidelity criterion can immediately be generalised:

$$F = \sum_{i=1}^{p} \sum_{j=1}^{q} w_{ij} \left(q_{ij} - \hat{q}_{ij} \right)^2 .$$

The extension to dimension 2 of the regularity criterion is a little more delicate; one distinguishes initially the vertical regularity *via* the operator Δ_v^z (which acts on q_{ij} with j fixed, seen as a series with i indices) which allows the calculation of a vertical regularity index:

$$S_v = \sum_{j=1}^{q} \sum_{i=1}^{p-z} \left(\Delta_v^z q_{ij} \right)^2 .$$

In the same manner, one calculates the horizontal regularity index S_h, then one poses:

$$M = F + \alpha \times S_v + \beta \times S_h$$

which must be minimised. The resolution of the optimisation problem is carried out by rearranging the elements in order to get back to the unidimensional case. For that purpose, one defines the size vector $p \times q$, u such that: $u_{q(i-1)+j} = \hat{q}_{ij}$; that amounts to taking for the first q elements of vector u the first row of matrix $\hat{q}$, then the elements of the second row, and so on. In the same way, one manufactures a weight matrix by copying on the diagonal rows of matrix $\left(w_{ij} \right)$. One thus poses $w^*_{q(i-1)+j, q(i-1)+j} = w_{ij}$. One proceeds in the same manner to define matrices K_z^v and K_y^h (*cf.* Knorr [1984]). Smoothed values are then obtained by:

$$q^* = \left(w^* + \alpha K_z^{v}{}' K_z^{v} + \beta K_y^{h}{}' K_y^{h}\right)^{-1} w^* u.$$

An example

Here is a simple and hands on case, which illustrates this method. Gross rates form a matrix $p \times q$ with $p = 4$ and $q = 3$. One chooses $z = 2$ (resp. $y = 1$) as degree of vertical regularity (resp. horizontal), one will have K_v^z, of dimensions $(q(p-z), m) = (6,12)$ and K_h^y, of dimensions $(p(q-z), m) = (8,12)$. One-dimensional construction was already carried out previously. The breakdown of the vertical matrix yields:

$$K_h^2 = \begin{pmatrix} 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 \end{pmatrix}$$
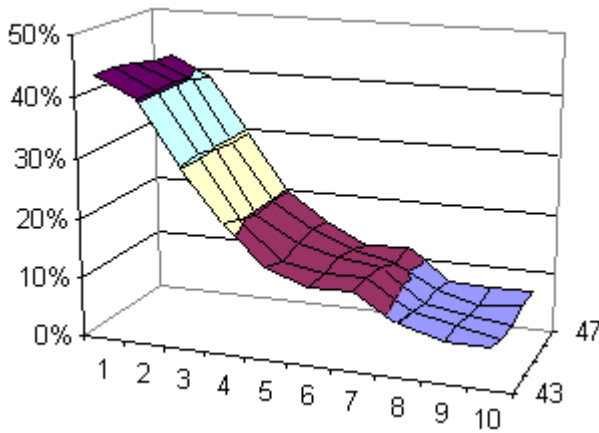
The breakdown of the horizontal matrix leads to:

$$K_h^1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

The Whitaker-Henderson method in dimension 2 can in particular be applied to sick leave data, which are usually estimated as a function of entry age and leave seniority. This method allows a joint smoothing in both directions, more effective than separate smoothing upon each variable. It is illustrated below:
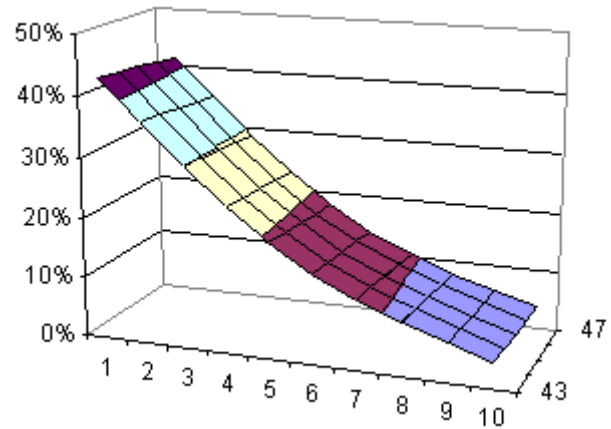
$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[}(T_x)$$

Gross rates

Smoothed rates

3.3.3. Whittaker-Henderson method as Bayesian[6] smoothing

The regularity measure $S(q) = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$ can be used to define an *a priori* distribution for the vector $q = (q_x)$; indeed, one is naturally inclined to assume that small values of S are associated with "more probable" values of q, which leads to writing for *a priori* density:

$$f_q(q) = c\,\mathbf{exp}(-\lambda S)$$

with c *a* constant of standardization[7] and λ a parameter. If moreover one makes the assumption that the estimation error $e_x = \hat{q}_x - q_x$ is distributed according to a centered normal distribution of variance σ_x^2, and that the various ages (or seniorities) are independent, one finds that:

$$f_{\hat{q}|q}(\hat{q}|q) = c\,\mathbf{exp}\left(-\frac{1}{2}\sum_{i=1}^{p}\frac{(\hat{q}_i - q_i)^2}{\sigma_i^2}\right)$$

One then happens to be under the conditions for application of the Bayes theorem to write the density of $q = (q_x)$ conditionally to the observations $\hat{q} = (\hat{q}_x)$:

$$f_{q|\hat{q}}(q|\hat{q}) = c\,\mathbf{exp}\left(-\lambda S - \frac{1}{2}\sum_{i=1}^{p}\frac{(q_i - \hat{q}_i)^2}{\sigma_i^2}\right)$$

Maximising this expression in q is equivalent to minimising $M = \lambda S + \frac{1}{2}\sum_{i=1}^{p}\frac{(q_i - \hat{q}_i)^2}{\sigma_i^2}$; one recognises the Whittaker-Henderson criterion; the Whittaker-Henderson smoothing thus has a

---

[6] See in particular Taylor [1992].

[7] The notation c indicates a constant term which can change depending on the formula considered.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t;\infty[} (T_x)$$

https://www.ressources-actuarielles.net

probabilistic interpretation within the framework of Bayesian smoothing, described in a more general way hereafter.

3.4. Bayesian smoothing

It was seen that the Whittaker-Henderson method could be justified by a probabilistic reasoning in making an assumption on the *a priori* distribution of true values and on the form of the conditional distribution of errors. This approach can be generalised, leading to the definition of Bayesian smoothing – the most famous of which is the Kimeldorf-Jones method.

3.4.1. General presentation

One considers vector $q = (q_x)$ of true values which one assumes result from an *a priori* distribution $f_q(.)$. Within the framework of the experiment, one carries out observations resulting from the conditional distribution knowing q, $f_{\hat{q}|q}(.)$. In the case of mortality rates estimation, this distribution is a product of binomial distributions of parameters[8] (n_x, q_x); in a general manner, this distribution is determined by the context of the experiment[9].

Based on these two distributions, Bayes formula is used to determine the *a posteriori* distribution of *q*:

$$f_{q|\hat{q}}(q|\hat{q}) = \frac{f_{\hat{q}|q}(\hat{q}|q) f_q(q)}{f_{\hat{q}}(\hat{q})}$$

Lastly, starting from this *a posteriori* distribution one can define a revised version of the estimator $\hat{q}$, for example through the maximum likelihood method by keeping the mode of the *a posteriori* density as adjusted value.

The difficulty of this approach is to define the characteristics of the *a priori* distribution. The Kimeldorf-Jones method, presented below, provides a relatively general framework for the Bayesian approach implementation, well adapted to the context of incidence distributions (or retention), in which some useful indications for the definition of the characteristics of the *a priori* distribution are provided.

3.4.2. Kimeldorf-Jones[10] method

The context is the estimation of incidence rates. The assumption is made that $f_q(.)$ is a multidimensional normal law of parameters (m, Σ); hence one has:

$$f_q(q) = \left[(2\pi)^p |\Sigma| \right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(q-m)'\Sigma^{-1}(q-m) \right)$$

---

[8] Which will be approximated by a multidimensional normal distribution when the number of observations is large.

[9] i.e. by the procedure of gross rates estimation.

[10] See Kimeldorf and Jones [1967].

with $q = (q_1, ..., q_p)$ the p incidence rates to be estimated. At this stage of the approach, one assumes (m, Σ) given and known. In this context, and as it was seen in 3.4.1 above, the distribution of the experiment is a product of binomial distributions, which can each be approximated by a normal law, and one can thus write:

$$f_{\hat{q}|q}(\hat{q}|q) = \left[(2\pi)^p |V| \right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\hat{q}-q)'V^{-1}(\hat{q}-q) \right)$$

since the initial gross estimators are without bias; the matrix V is diagonal, with elements $n_i q_i (1 - q_i)$. The *a posteriori* density is thus of the form:

$$f_{q|\hat{q}}(q|\hat{q}) = C \exp\left(-\frac{1}{2}\left[(q-m)'\Sigma^{-1}(q-m) + (\hat{q}-q)'V^{-1}(\hat{q}-q) \right] \right),$$

C being a constant; after some processing one can put this expression in the form:

$$f_{q|\hat{q}}(q|\hat{q}) = C' \exp\left(-\frac{1}{2}(q-z)'W^{-1}(q-z) \right)$$

with $z = \left(\Sigma^{-1} + V^{-1} \right)^{-1} \left(V^{-1}\hat{q} + \Sigma^{-1}m \right)$ and $W^{-1} = \Sigma^{-1} + V^{-1}$. This shows that this distribution is also Gaussian. The natural expression of fitted rates[11] can be deduced:

$$q^* = \left(\Sigma^{-1} + V^{-1} \right)^{-1} \left(V^{-1}\hat{q} + \Sigma^{-1}m \right).$$

This expression means that fitted rates are a weighted average of gross rates and *a priori* rates. This expression can be written in the following way, useful for calculations:

$$q^* = \hat{q} + \left(I_p + \Sigma V^{-1} \right)^{-1} (m - \hat{q})$$

<u>Note</u>: the elements of the diagonal matrix V are $n_i q_i (1 - q_i)$ which are unknown; they can be replaced in the above formula by the estimator $n_i \hat{q}_i (1 - \hat{q}_i)$, that is to say by the *a priori* estimates $n_i m_i (1 - m_i)$.

Beyond a smoothed version of gross rates, this approach provides a simple framework of mortality stochastic model, by providing a (normal) distribution for the model's exit rates.

However, further clarification of smoothing parameters (m, Σ) remains to be done; the values of m are the idea that one can have about incidence rates without observations; one can for example use a reference like the BCAC tables for sick leave, an INSEE mortality table for mortality, etc. The situation is more delicate for the variance-covariance matrix Σ.

The idea is to describe a dependency amongst rates, according to differences in age or decreasing seniority; this leads to proposing the following form for the coefficients of the matrix Σ:

$$\Sigma_{ij} = \sigma_i \sigma_j \rho^{|i-j|}$$

---

[11] The expectancy of a normal distribution is also its mode.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \, \boldsymbol{I}_{]t;\infty[} (T_x)$$

with $\sigma_i^2 = \mathbf{var}(q_i)$ and ρ the correlation coefficient of 2 consecutive terms. Determining numerical values for these $p+1$ elements remains to be done.

4. Fit validation

Once the fit is finished, it is appropriate to verify its validity through some traditional statistical tests, which will in particular allow to make sure that the revised rates are not too far away from the gross rates. However, in certain situations (construction of an experience mortality table, for example), the application of rules of prudence can lead to values of adjusted rates deliberately different from gross rates (for example higher in the case of a contract in the event of death), and the statistical tests will then have to be adapted to this situation. Concretely, it will be necessary to give up bilateral tests and rather use unilateral tests better adapted to the situation.

The objective of this section is not to describe in an exhaustive way the tests to be done, but rather to provide some hints and illustrations in order to carry out this validation process.

4.1. Chi-squared test

Once fitting (or smoothing) has been carried out, the Chi-squared test allows the verification of the global quality of revised rates, by making sure that they are not "too far" from estimated rates. The statistics is calculated:

$$Z = \sum_{i=1}^{p} n_i \frac{(\hat{q}_i - q_i)^2}{q_i (1 - q_i)}$$

In the case of a parametric fit through maximum of likelihood with r parameters, then the (asymptotic) distribution of Z is a $\chi^2(p-r-1)$ in the case of a non-parametric smoothing the number of degrees of freedom is less easy to determine.

4.2. *Standardised mortality ratio* (SMR)

SMR is defined as the ratio of observed death to the number of deaths predicted in a reference population, with the objective to decide whether the mortality of the observed group is identical to that of the reference group; one has as follows:

$$SMR = \frac{D}{E} = \frac{\sum_{i=1}^{p} D_i}{\sum_{i=1}^{p} n_i q_i} \; .$$

In this expression, E is a constant and D a binomial random variable which one can approximate through a Poisson distribution, which leads to the following confidence interval for the *SMR* variable, in the case where the reference mortality does not depend on the studied group:

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{]t;\infty[}(T_x)$$

$$\frac{D}{E}\left(1 - \frac{1}{9 \times D} - \frac{u_{1-\frac{\alpha}{2}}}{3\sqrt{D}}\right)^3 \leq SMR \leq \frac{D+1}{E}\left(1 - \frac{1}{9 \times (D+1)} + \frac{u_{1-\frac{\alpha}{2}}}{3\sqrt{D+1}}\right)^3$$

In the case of a reference mortality depending on the data (thus following a fitting process), similarly to Chi-squared, the above formula must be adapted according to the context.

4.3. Changes of sign test

The sign of the difference $d_x = q_x - \hat{q}_x$ has, under the conditions for application of the normal approximation, a ½ probability of being positive. If ages are independent, the probability that d_x and d_{x+1} do not have the same sign is thus also equal to ½. If the considered age segment contains p values, then it can be deduced that the number of sign changes in the series d_x follows a binomial distribution of parameters $\left(p-1, \frac{1}{2}\right)$. In particular, the average number of sign changes is $\frac{p-1}{2}$. A test can easily be deduced, either with finite distance[12] or through Gaussian approximation, based on the following statistics:

$$S = \frac{2n - (p-1)}{\sqrt{p-1}}$$

where n is the observed number of sign changes.

5. References

Bess P., Cardot H. [2001] "*statistical Modelling of functional data*", University of Toulouse, lecture notes

Kimeldorf G.S, Jones D.A. [1967] "*Bayesian graduation*", TSA, XIX

Knorr F.E. [1984] "Multidimensional Whittaker-Henderson Graduation", TSA, vol. 36.

London D. [1995] "*The revision of estimates*", Actex

Saporta G., [1990] *Probabilities, data analysis and statistics*", Technip

Taylor G. [1992] "*with Bayesian interpretation of Whittaker-Henderson graduation*", IME, 11,7-16

Wand M.P., Jones M.C. [1995] Kernel Smoothing. London: Chapman and Hall.

---

[12] See the algorithm of construction of confidence intervals for a binomial distribution.