МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2000
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2000
Proceedings of Twenty Ninth Spring Conference of the Union of Bulgarian Mathematicians
Lovetch, April 3–6, 2000

GRADUATION OF MORTALITY TABLES IN THE CASE OF INSUFFICIENCY OF DATA

Kristina Paskova Pavlova, Plamen Stoyanov Mateev

The main objective of this paper consists of finding out a criterion for verification of graduated rates in the case of data insufficiency. This situation is typical of the Bulgarian life insurance companies, therefore the solution of the formulated problem is important for the practice.
The analogous problem of crude rates of mortality is also considered in the paper. A test for rejection of considerable errors is elaborated.
To resolve these problems two lemmas are formulated and demonstrated and the necessary tests are attained. A method for application of the test in the case of insufficiency of data is developed.
The method for graduation by mathematical formula and the described tests are applied to data referring to a Bulgarian insurance company. The obtained results show which formulae are more appropriate for such data and what special features the crude rates under consideration have.
The problems discussed in the paper could be resolved in the future for all types of actuarial death rates that assume graduation.

1. Unknown quantities and assumptions for their obtaining. In this paper we will mostly discuss three of the basic actuarial quantities: q_x (the probability of dying of an individual at age x within a year), μ_x (the force of mortality) and m_x (the central death rate for a given time period). For their formal definition we need the function $T(x)$ of the time until death of an individual at age x and the number of lives at the same age – l_x. Thus

$$q_x = \mathbf{P}\{T(x) \le 1\}, \mu_x = -\frac{d}{dx} \ln \mathbf{P}\{T(0) > x\}$$

(1) $$m_x = \frac{\int_x^{x+1} l_y \mu_y dy}{\int_x^{x+1} l_y dy}.^1$$

In fact, most of the actuarial functions use directly or indirectly these quantities. Therefore, for life insurance companies it is very important to estimate them as accurately as possible to prevent eventual losses. First, one should calculate the number of lives and deaths separately for each age. There are two basic methods for computation of the number of lives that survived the period of exposure: according to the *central* or to the

[1] We use notations that are part of the *International Actuarial Notation* (see [1]).

initial exposed to risk. We denote them by R_x^c and R_x^i respectively, where x indicates the age that they refer to[2]. By analogy, we denote by A_x the number of deaths recorded in the particular investigation.

Let us fix an arbitrary age x and estimate the values of q_x, μ_x and m_x. To simplify the notations we will omit the subscript x up to the end of this section. With " ˆ " we will denote the unbiased estimators of the quantities that we are interested in *(the crude rates of mortality).*

- Unbiased estimation of q and μ: $\hat{q} = \dfrac{A}{R^i}$, $\hat{\mu} = \dfrac{A}{R^c}$;
- Unbiased estimation of m_x: m_x is estimated using numerical approach based on formula (1)[3].

2. Requirements for graduation. The major goal of graduation of an experience is to estimate *the true mortality rates* at each age on the basis of the crude rates of mortality. Consequently the graduation must satisfy the following two conditions:

- *the differences between the crude rates and the true mortality rates must be considered as results from accidental errors* (so called 'adhering to data');
- *the obtained true mortality rates must be circumscribed by smoothly changed with the age curve* (so called 'smoothness')[4].

3. Method for graduation in the case of data insufficiency. The basic assumption in this paper is that the investigation has relatively few deaths. This situation is typical of the Bulgarian life insurance companies. In principle it is preferable to keep separate experience with different indications (sex, diseases etc.). Such an approach assumes that the data are split into groups that could be too small. In spite of this, the delimitation of the experience is necessary to avoid redundant risks.

Henceforth, we will discuss *experience of the two sexes that is kept separate.* It is necessary to develop such an approach that prevents wide deviations of the found out graduated rates from the true mortality rates. The problem could be formulated as follows:

Let the experience be graduated both by using initial exposures to risk and graduating q_x. Making use of relationships among separated and combined experiences, the most appropriate graduated curve for the separated crude rates must be chosen.

In order to solve this problem we will take advantage of the following

Lemma 3.1. *Let there exist n different indications for the individuals in a given investigation and ${}_tq_x^j$ is the probability of dying of an individual at age x, who possesses the indication j, within t years, and ${}_tq_x$ is the same probability for an unspecified individual*

[2]Theory, related to the practical ways for calculation of R_x^c and R_x^i, is described in [2].
[3]For detailed exposition about calculation of these estimations see [3].
[4]For more details see [2].

form the population. Let furthermore $C_x^j = \{$*The individual* (x) *possesses the indication* $j\}$. *Then a necessary and sufficient condition for*

$$ {}_tq_x = \sum_{j=1}^{n} \mathbf{P}\{C_x^j\}{}_tq_x^j \tag{2} $$

is that each individual in the investigation has exactly one indication.

The proof of this lemma is based on the formula of total probability. Because of its simplicity, we will not dwell on its exposition in detail.

Corollary 3.1.1. *Let us denote with superscripts* m *and* w *the sex of the individuals in the investigation. Then*

$$ {}_tq_x = \mathbf{P}\{C_x^m\}{}_tq_x^m + \mathbf{P}\{C_x^w\}{}_tq_x^w. \tag{3} $$

Corollary 3.1.2. *The following inequality follows from equation (3):*

$$ {}_tq_x < {}_tq_x^m + {}_tq_x^w. \tag{4} $$

By analogy with Lemma 3.1 the following relationships for the force of mortality can be proved:

Lemma 3.2. *Let the requirements of Lemma* 3.1 *hold. Then a necessary and sufficient condition for*

$$ \mu_{x+t} = -\frac{d}{ds}\ln\Big[\sum_{j=1}^{n}\Big(\mathbf{P}\{C_x^j\}\exp\Big\{-\int_0^t \mu_{x+s}^j ds\Big\}\Big)\Big] \tag{5} $$

is that each individual in the investigation has exactly one indication.

Corollary 3.2.1. *By using the notations of Corollary* 3.1.1 *the following is true for the force of mortality:*

$$ \mu_{x+t} = -\frac{d}{ds}\ln\Big[\mathbf{P}\{C_x^m\}\exp\Big\{-\int_0^t \mu_{x+s}^m ds\Big\} + \mathbf{P}\{C_x^w\}\exp\Big\{-\int_0^t \mu_{x+s}^w ds\Big\}\Big]. \tag{6} $$

4. Rejection of considerable errors test. In practice, it is difficult to reject categorically a particular graduation. Usually the applied battery of statistical tests (signs test, runs test, Kolmogorov-Smirnov test, serial correlation test, and χ^2 test) does not give an explicit answer to the question to what extent the unsatisfactory results of a graduation are due to its defects and to what extend to defects of the very tests with respect to the considered experience[5].

We will describe a test, which rejects graduations of kept separate rates that considerably deviate from satisfactory graduation of combined rates. The test could be easily implemented in actuarial software packages.

Let us consider, as it is formulated in the raised problem, the case of initial exposed to risk and graduated rates of q_x are presumed to be looked for.

[5] See [3].

By making use of inequality (4) when $t = 1$ we could reject already obtained graduated rates for which it is not true. In the cases when we deal with μ_x or m_x, the corresponding relationships are considerably more complex, therefore it is preferable to use numerical approximations instead. In practice a divergence form (4) seldom occurs.

We could deduce an analogous relationship for the crude rates of mortality:

$$\tilde{q}_x = \frac{A_x^m + A_x^w}{R_x^m + R_x^w} = \frac{A_x^m}{R_x^m + R_x^w} + \frac{A_x^w}{R_x^m + R_x^w} < \frac{A_x^m}{R_x^m} + \frac{A_x^w}{R_x^w} = \tilde{q}_x^m + \tilde{q}_x^w . \tag{7}$$

Using inequality (7) we could control the advent of considerable errors in data grouped by age, prior to begin to look for graduated rates.

5. Test for selection of the most appropriate graduations of kept separate crude rates. When an investigation contains few deaths, it is necessary to group neighboring ages instead of making use of the crude rates separately for each age. Under these circumstances, we could not seek correctly for relationships between the graduated and the crude rates. Additionally we have to dispose of equal number of age groups in the three cases to keep separate experience of the two sexes and the combined experience. Furthermore, the corresponding groups must include identical ages.

In the course of the grouping, on the one hand we should try to obtain as many groups as possible, and on the other hand to balance the number of deaths in them in the three cases, respectively.

Let us consider the case when several graduations of the crude rates are available for each combined and kept separate experience. We have to answer the question, which is the most reliable one. To great extent the answer depends on the results of the applied tests. For more completeness we will discus both cases when more or less data are available.

5.1. A large number of deaths available. In this case we assume that the excerpt of data is large enough to obtain graduated rates, which are very close to the true ones. Nevertheless, an additional verification contributes to determine the most suitable graduations of the kept separate rates. To this end we will make use of equality (3) when $t = 1$. In practice it could be true only in rare occasions. Therefore, it is advisable to fix an admissible divergence towards it.

5.2. A small number of deaths available. Under such circumstances the age groups are few, which is an additional argument in favor of even more precision. Otherwise, the divergences would be transfered also on the particular ages.

When we dispose of a small experience, very large discrepancies between both sides of (3) can be seen in practice. For that reason we propose the following algorithm that consists of three steps:

step 1 Empirical finding out of a relationship in the form of $\hat{q}_x = f(\hat{q}_x^m, \hat{q}_x^w)$. In the most simple case it has the following form: $\hat{q}_x = c(x)(\hat{q}_x^m + \hat{q}_x^w)$, where $c(x)$ is a continuous on the left step function with jump discontinuities at the points x, which is determined by $c(x) = \tilde{q}_x/(\tilde{q}_x^m + \tilde{q}_x^w)$.

step 2 Deduction of an uniform law for all x. In the above example at this step a suitable constant that concentrates already obtained $c(x)$ should be found (like for example, $c = \dfrac{1}{N+1}\sum_{x=0}^{N} c(x)$).

step 3 Choice of the most suitable functions for graduated rates of kept separate experience of the two sexes on the basis of the obtained law at step 2.

Remark 5.1. *In the equation (3) the probabilities* $\mathbf{P}\{C_x^m\}$ *and* $\mathbf{P}\{C_x^w\}$ *could be interpreted as weights. Their estimation according to the data inevitably leads to accumulation of errors. For that reason, at step 1 we propose a relationship that presumes equal weights in front of* $\hat{q}_x^m$ *and* $\hat{q}_x^w$.

Remark 5.2. *The method presented above could be considered, in a sense, as an analogue to the graduation by reference to a standard mortality table.*

6. Description of the data. The construction of actuarial mortality tables is applied rarely in Bulgaria. Instead of them, demographic mortality tables are used. To a great extent this is due to the lack of data accumulated in the insurance companies. The insured individuals represent a specific group. Therefore, it is preferable to undertake investigations especially for the mortality among them.

We will consider the graduation by mathematical formula. In this case the goal is to find out an appropriate smooth function, which gives us the needed *graduated rates of mortality*. The principal difficulty consists of finding out the very function. Accordingly to the specific case, one could apply functions already invented in similar cases that could be adjusted additionally to the particular data.

The data, that we will use, belong to a Bulgarian life insurance company. The investigation spans three-year period – from 1996 to 1998. It has relatively few deaths, 757 in all. There is no reason to suppose that a large number of duplicates exists in this experience, so that it will be assumed that each observation represents one 'life'. The subsequent results show that our assumption is not unjustified.

We will look for graduated rates of *kept separate experience*. We will look for graduation not only in the indicated two cases but also in the case of combined experience. This graduation will serve only as auxiliary means when we choose the most appropriate graduations of the kept separate experience in relation with the test described in §5.

7. Finding out graduations of the available crude rates. The experience is graduated by using initial exposures to risk and graduating q_x. To calculate them we use a formula from the $LGM_\alpha^{r,s}(x)$ family and the Bird's, Barnett's, and Heligman-Pollard's formulae[6].

More than a half of the obtained graduations pass the two most common tests: signs test and runs test. Bird's and Heligman-Pollard's functions with the two criterions for optimization are the sole that show too low values of the probabilities $p(pos)$ and $p(runs)$[7], 0.025 and 0.05 respectively. In this case, the results are due to the excessive smoothness of the graduated curves. By this reason, they should be rejected as possible graduations.

The serial correlation test shows relatively low values of ρ_1, ρ_2 and ρ_3. This fact indirectly confirms our assumption of absence of duplicates.

Most functions do not pass the Kolmogorov-Smirnov test. In those cases, again the Barnett's and Heligman-Pollard's functions are used. We could positively deem that those graduations are unlikely to be a satisfactory representation of the experience.

Very low values of the probability $p(\chi^2)$ are observed, as well. Despite this we could consider that the graduations, which pass the other tests are satisfactory.

Barnett's and Heligman-Pollard's functions in all considered cases show very unsatisfactory results. We could assume that they are inappropriate in similar circumstances. Heligman-Pollard's formula itself has too many parameters and is a sophisticated one for practical use.

On the basis of the results of all tests, we choose Bird's function with χ^2 criterion as the most appropriate one in the considered case of combined experience (see Fig. 1).

To illustrate the proposed in §5 algorithm we choose graduations in both for kept separate experience. The preferred graduations to the greatest extent satisfy the mentioned above battery of tests. In this case they are $LGM(0,5)$ and Bird's function with maximum likelihood criterion.

The rejection of considerable errors test, which is based on inequality (7), does not show errors in the calculation of the crude rates $\tilde{q}_x^m$ and $\tilde{q}_x^w$ (see Fig. 2). This justifies our assumption of absence of duplicates. After applying the corresponding test for the obtained graduated rates (inequality (4)), it is detected that it is not true for the last age group when Bird's formula is used for q_x^m. This is due to the wide confidence intervals at old ages. The same phenomenon should be expected also at young ages where there were few deaths. In our example however, it is not the case.

Let us now demonstrate how to find out the most satisfactory graduated rates of the kept separate experience:

step 1 We realize that $c(x)$ fluctuates in the interval $[0.39555902, 0.51520306]$, which is fairly small. On account of this, we choose $c(x) = c = \dfrac{1}{33}\sum_{i=1}^{33} c(x_i) = 0.45626269$.

[6]The definitions of these and other successfully applied functions could be found in [2] and [3].

[7]The principal symbols that are used in relation with the battery of statistical tests are standard. Their description could be found in [3].

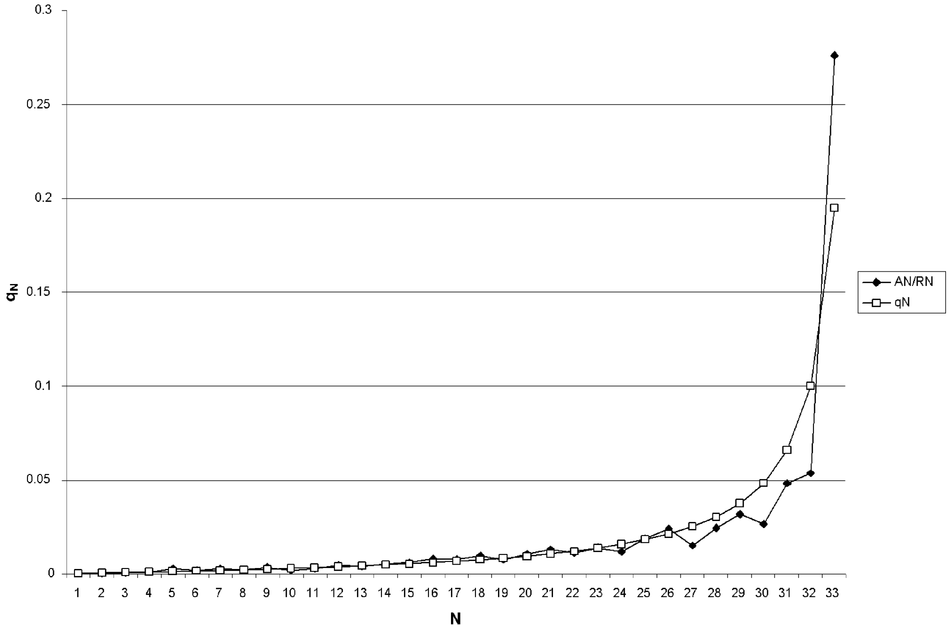

Fig. 1. Crude and graduated by Bird's formula with χ^2 criterion rates of q_x in the case of combined experience, 1996 - 1998

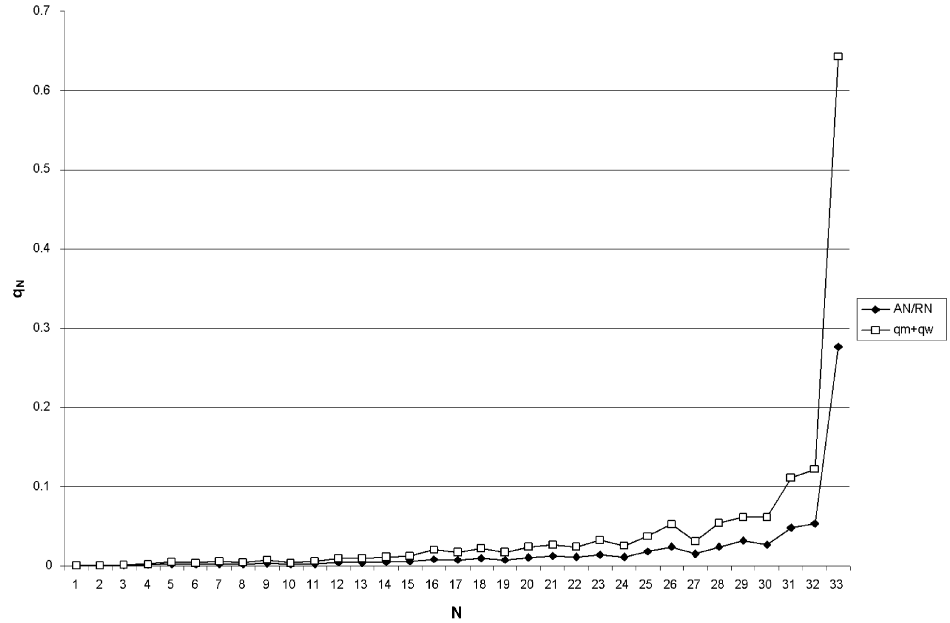

Fig. 2. $q_x < q_x^m + q_x^w$

step 2 By making use of the value of c found at the previous step, we obtain the following formula: $q_x = 0.45626269(\hat{q}_x^m + \hat{q}_x^w)$.

step 3 For the numerical data that are based on the formula above the smallest total divergence is observed when we combine $LGM(0,5)$ in the men's experience and Bird's formula in the women's experience (see Fig. 3). The divergence itself is 0.09073129, which is small enough in order not to impose a quest for more complex formula for $c(x)$.

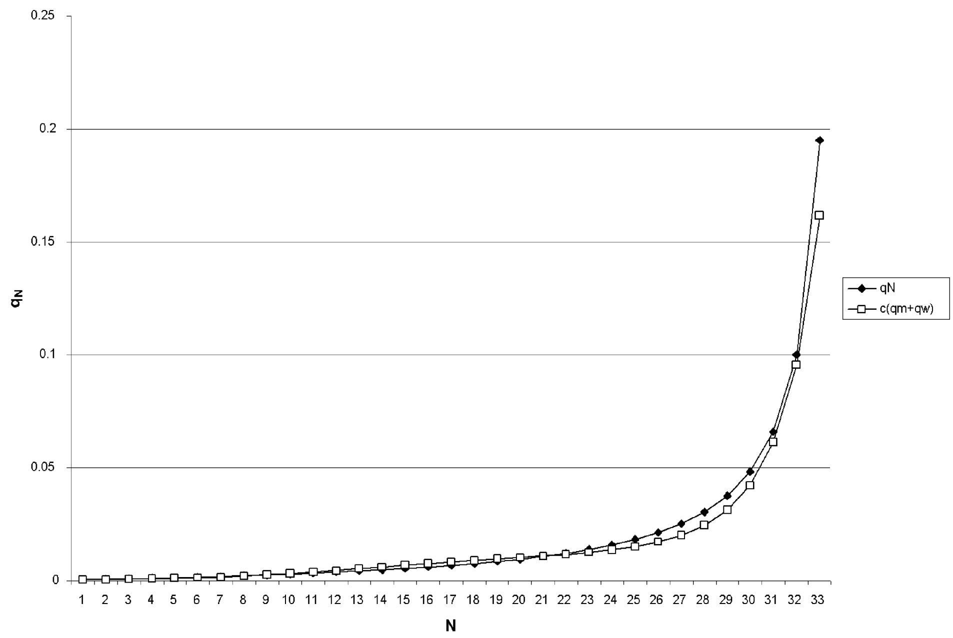

Fig. 3. Comparison among graduated rates

In conclusion we could affirm that the use of formulae with many parameters is not needed. Frequent variations are not observed among the crude rates, so that a simple formula with few parameters suffices. Barnett's and Heligman-Pollard's formulae prove to be unsuitable for this kind of data for the practice in view of the large number of parameters.

REFERENCES

[1] Bowers, Gerber, Hickman, Jones, Nesbitt. Actuarial Mathematics., Society of Actuaries, 1997.

[2] R. Chatburn, S. Haberman. Fundaments of Actuarial Mathematics, module I.

[3] D. O. Forfar, J. J. McCutcheon, A. D. Wilkie. On graduation by mathematical formula. *Journal of the Institute of actuaries*, **115**, part I, (1988).

Kristina Pavlova
Plamen Mateev
Institute of Mathematics and Informatics
Acad. G. Bonchev St., Bl. 8
1113 Sofia, Bulgaria
e-mail: pavlova@math.bas.bg
e-mail: pmat@math.bas.bg

ИЗГЛАЖДАНЕ НА ТАБЛИЦИ НА СМЪРТНОСТ ПРИ ОСКЪДНИ ДАННИ

Кристина Паскова Павлова, Пламен Стоянов Матеев

Основната цел на настоящото изследване се състои в намирането на критерий за проверка на изглаждащи степени при оскъдни данни. Този случай е характерен за българските застрахователни компании, поради което решаването на поставената задача има практическа стойност. В разработката е отделено внимание и на аналогичната задача при необработените степени на смъртност. Изработен е критерий за отхвърляне на груби грешки.
За решаването на поставените задачи са формулирани и доказани твърдения, с чиято помощ се достига до необходимите критерии. За начина на прилагане на критерия при недостатъчно данни е разработен метод.
Методът за изглаждане по формула и описаните критерии са приложени за данни от българска застрахователна компания. Получените резултати показват кои функции са най-подходящи при такъв род данни и какви особености имат разглежданите необработени степени на смъртност.
Поставените в разработката задачи могат да бъдат решавани в перспектива за всички актюерски функции, които предполагат изглаждане.