Mortality modeling for Romanian population

Elena-Corina Cipu and Carmen Dragnea

Abstract. We study different models of mortality for Romanian population using an abridged life table obtained from data of year 2002. We estimate the parameters of the Perks distribution for individual force of mortality. Remarks upon results of different models used were made.

M.S.C. 2000: 62N01, 62N02.
Key words: life tables, force rate of mortality, survival function.

1 Starting of the problem

The study of survival function is essential both for knowing the evolution and the structure of the population depending on its main indicators, and also for demographic research.

The demographers and biologists are especially interested in the structure study and evolution of a population. For this many methods were developed to present statistics about populations. One of the methods is the survival table, used to describe and understand the dynamic of a species. The gathered information are important in the fauna study, in agriculture and human health.

Depending on the length of the elementary age intervals, the survival tables are of 2 types: ***complete:*** the tables are created for 1 year periods; ***abridged:*** on age classes, usually from 5 to 5 years (0-4, 5-9, 10-14...) - used when age, year or dates from the table are incomplete.

To be able to put them in practice, the mortality tables are created based on the mortality rate independence at age X regarding the total number of people from that population.

2 Biometrical functions. The survival function

The intensity of the mortality is measured by certain numeric coefficients which are called biometric functions. For the construction of the survival tables we use the following set of parametric functions: l_x - the number of survivors from the arbitrary number l_0 corresponding to age class x, ${}_nd_x$ - the number of deaths in the interval

Proceedings of The 4-th International Colloquium "Mathematics in Engineering and Numerical Physics" October 6-8 , 2006, Bucharest, Romania, pp. 38-45.
© Balkan Society of Geometers, Geometry Balkan Press 2007.

$[x, x+n)$, ${}_nq_x$ - the probability of dying in the interval $[x, x+n)$, ${}_nL_x$ - the number of person - years that survived , with age in the interval $[x, x+n)$, T_x - the number of person - years that survived after age x, e_x - life expectancy, ${}_n\mu_x$ - the force of mortality, ${}_np_x$ - the probability of survival in the interval $[x, x+n)$.

The functions ${}_nq_x$, l_x and ${}_nd_x$ are measures of frequency (or intensity), and T_x, ${}_nL_x$, e_x are measures of the lengths expressed in person – years. The survival function l_x which describes the number of survivors at age x satisfies the relation $dl_x/dx = -l_x\mu_x$, where μ_x is the force of mortality and is a non negative function. This differential equation has the solution: $l_x = l_0 \exp(-\int_0^x \mu_s ds)$, l_0 being the value of the function at the moment considered as reference.

The mortality tables contain values of l_x for age classes, and for the intermediary a method of interpolation is used. For this reason l_0 is chosen such that the interpolation is as accurate as possible. From practical reasons the mortality tables use a maximum age ω= 86, 100, 105 or 110 for x greater than ω, $l_x$is 0. The probability that a person at age x to live another **t** years is denoted with ${}_tp_x$ and the probability that a person dies in this period is denoted with ${}_tq_x = 1 -_t p_x$. We get:

$${}_tp_x = l_{x+t}/l_x = \exp(-\int_0^t \mu_{x+s} ds), \quad {}_tq_x = (l_x - l_{x+t})/l_x = \int_0^t {}_sp_x\mu_{x+s}ds.$$

Through convention when the survival period ranges from x to $x+1$, the left index dropped out eliminated such that $p_x =_1 p_x$ and $q_x =_1 q_x$.

3 Computation methods. Fill in the abridged tables

A survival table is a rectangular matrix formed from biometric functions (co- lumns) and age classes (rows). We shall use the following notations: N – the total number of individuals of a population, X – age class for interval $[x, x+n)$, n – the length of an elementary age class, D – the number of deceased with the age in the interval $[x, x+n)$, ω **-** the upper age limit.

For the construction of the table we use each class with the lengths of 5 years except the class 0-1, 1-4 and 100+ class (the interval of the last class is open at the upper bound). Taking into account these considerations we build the survival table in the following way:

The scheme of the survival table *Table 1:*

X	D	lx	dx	Qx	px	Lx	Tx	ex
0–1								
1–4								
5–9								
10–14								
.....								

In D column we introduced the census observed values from the year 2002, showing the number of deceased persons with age within age class X.

In order to fill in the survival table (Table 1) we use the following algorithm for row calculus: ***1:*** l_0 is initialized with the value 100.000; ***2:*** The values of d_x are computed by multiplying the column D with $100000/N$, namely ${}_nd_x = 1000000 D_x/N$. For verification, we sum the "D" column and obtain 99999; ***3:*** The number of people alive at the age x, l_x is obtained by rounding the difference $l_x - {}_nd_x$: $l_{x+n} = ROUND(l_x - {}_nd_x)$; ***4:*** For the rest of the biometric functions, we calculate the corresponding values as follows: ${}_nq_x = {}_nd_x/l_x$; ${}_np_x = 1 -_n q_x$; ${}_nL_x = n \cdot (l_x + l_{x+n})/2$; $T_i = \sum_{j=i}^{x} {}_nL_j$; $e_x = T_x/l_x$.

We constructed the survival table for Romania based on the row calculus algorithm corresponding to the Romanian population (Table 2).

Life table for Romanian population in the year 2002. *Table 2:*

class X	D	lx	dx	qx	px	Lx	Tx	ex
0–1	3648	10^5	1352.785	0.0135	0.9865	495940	7178478	71.78
1–4	731	98647	271.0761	0.0027	0.9973	492153	6682538	67.74
5–9	436	98376	161.6815	0.0016	0.9984	490745	6190385	62.93
10–14	787	98214	291.8425	0.003	0.9970	489580	5699640	58.03
15–19	820	97922	304.0799	0.0031	0.9969	487653	5210060	53.21
20–24	1291	97618	478.7404	0.0049	0.9951	485410	4722408	48.38
25–29	1598	97139	592.5849	0.0061	0.9939	481723	4236998	43.62
30–34	2687	96546	996.4178	0.0103	0.9897	477518	3755275	38.90
35–39	2936	95550	1088.754	0.0114	0.9886	469533	3277758	34.30
40–44	5926	94461	2197.533	0.0233	0.9767	457123	2808225	29.73
45–49	10450	92263	3875.164	0.042	0.9580	439025	2351103	25.48
50–54	13595	88388	5041.422	0.057	0.9430	416303	1912078	21.63
55–59	14060	83347	5213.857	0.0626	0.9374	383070	1495775	17.95
60–64	22254	78133	8252.431	0.1056	0.8944	340920	1112705	14.24
65–69	31404	69881	11645.52	0.1666	0.8334	283148	771785	11.04
70–74	40064	58235	14856.9	0.2551	0.7449	211108	488637.5	8.39
75–79	46302	43378	17170.13	0.3958	0.6042	144198	277530	6.40
80–84	32110	26208	11907.32	0.4543	0.5457	79108	133332.5	5.09
85–89	23909	14301	8866.153	0.62	0.3800	38005	54225	3.79
90–94	12227	5435	4534.127	0.8342	0.1658	13778	16220	2.98
95–99	2224	901	824.7239	0.9153	0.0847	2253	2442.5	2.71

The data used were given by the National Institute of Statistics, using "The population census" made at 1 July 2002 and "Demographic statistics bulletin" completed by the registry and the health departments from each district. At the 2002 census, the stable population has been determined according to CEE/UNO recommendation for population and living places censuses.

4 The construction of the life table for Romanian population at 1^{st}July 2002

For construction of the survival table we will be using the algorithm presented in the third paragraph. We initialize the variables according to the given data from the census, such that: N = 269666 – the total number of deceased people from the population, X – the age class for the interval $[x, x+n)$, n =5 – the length of an elementary age interval, D – the number of deceased, with age included in the interval $[x, x+n)$, ω =100 – upper age limit. The initial data were distributed by age classes like this: 0-4, 5-9, 10-15 ... 95-100, 100+. Due to the high infantile mortality rate, we divided the first age class in the intervals 0-1 and 1-4. This division was possible because of the separate statistics supplied by NSI about infantile mortality.

In the construction of the survival tables for the dying probability we have the following constraints:${}_{\infty}q_{\omega} = 1$, if the last age interval is open, $q_x \leq 1$ if the last age interval is closed. The last column represents the modified life expectancy.

Using Sprague interpolation formula we can obtain the annual survival probabilities on the side with survival function.

5 Annual data determination for the survival function

Once the survival probabilityp_x determined, we can find the survival function l_x for $l_{x+t} = l_x \cdot {}_tp_x$, $t \in [1, 4]$ (if we use the left string values of survival probabilities and ${}_{t_1+t_2}p_x = \quad {}_{t_1+t_2}p_{x+t_1} \cdot {}_{t_1}p_x$, $\forall t_1, t_2 > 0$, $x > 0$), respectively $l_x = l_{x+t}/{}_tp_x$, $t \in [1, 4]$ (using the right string values for ${}_tp_x$), on each five years interval and determining the mean of those values. For all the obtained data we consider a cubic interpolation and the result is presented in Figure 1.

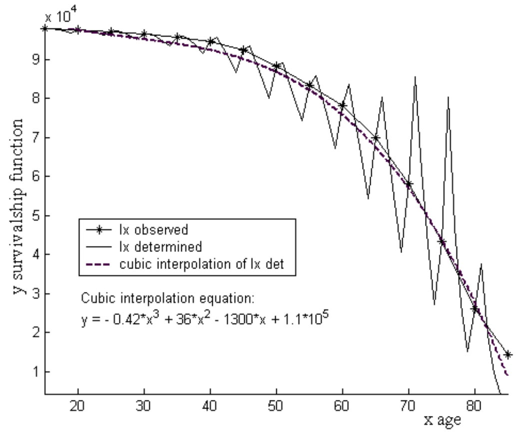

Figure 1: The survival function for the Romanian population-2002; comparison between observed data at reduced level, annual determined data and interpolated data.

Our goal is to find the survival function as $l_x = K \cdot s^x \cdot g^{c^x}$, $K = l_{x_0} \cdot e^{ax_0} \cdot e^{b \cdot c^{x}{}^{0}/\ln c}$, $s = e^{-a}$, $g = e^{-b/\ln c}$ if we consider for the **force of mortality** a function type **Makeham - Gompertz**: $\mu_x = a + b \cdot c^x$.

Using the relation $\ln\, p_x = \ln\, s + c^x(c-1)\ln\, g$ for estimation of parameters a, b, c by King and Hardy method (with notation $A_x = \sum_{y=x}^{x+n-1} \ln\, p_y$, the parameters are $c = \sqrt[n]{\dfrac{A_{x+2n} - A_{x+n}}{A_{x+n} - A_x}}$, $g = \exp(\dfrac{A_{x+n} - A_x}{c^x(c^n-1)^2})$, $a = -\ln(p_x g^{c^x(1-c)})$, $b = -\ln g \cdot \ln c)$. The result for $n \in \overline{10, 23}$ and for the start age $x = 15$, with an averaging of the obtained values, is described in Figure 2.

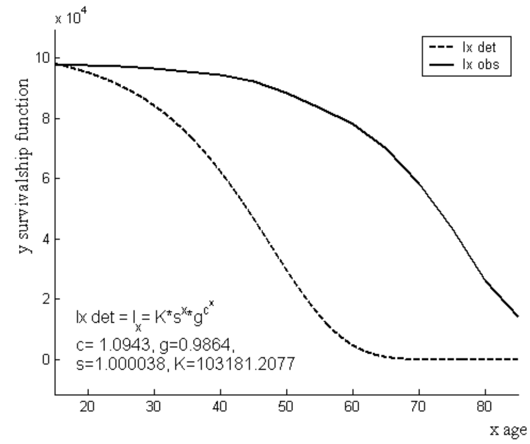

Figure 2: The survival function for the Romanian population-2002; observed data at reduced level and determined data.

6 Estimation of parameters for Perks distribution

From Figure 2 we observe that the Makeham-Gompertz function is not a very good expression for the force rate of mortality for Romania population. We try an estimation of parameters for Perks distribution: $\mu_x = \dfrac{a + b \cdot c^x}{1 + d \cdot c^x}$, $c > 0$ (see Horiuchi, Coale [5]).

In this case the survival function has the form: $l_x = K \cdot s^x \cdot g^{\ln(1+d\cdot c^x)}$, with $s = e^{-a}$, $K = l_{x_0} \cdot e^{ax_0} \cdot \exp\left(-\dfrac{b - a\cdot d}{d \cdot \ln c}\ln(1 + d \cdot c^{x_0})\right)$, $g = e^{(b-a\cdot d)/(d\cdot \ln c)}$.

We obtain the relation $\ln\, p_x = \ln\, s + \ln\left(\dfrac{1 + d\cdot c^{x+1}}{1 + d\cdot c^x}\right) \cdot \ln\, g$ for the estimation of parameters a, b, c, d. By King and Hardy method (where are considered four series of observations at the same length n, for persons who have consecutively ages, with $S_x = \sum_{y=x}^{x+n-1} \ln\, p_y$, one obtains two nonlinear equations for parameters c, d, respectively:

$$(6.1)\quad \begin{aligned} F(c,d) &= (1+d\cdot c^{x+4n})(1+d\cdot c^{x+2n})^{1+2B_1} - \\ &\quad -(1+d\cdot c^{x+3n})^{2+B_1}(1+d\cdot c^{x+n})^{B_1} = 0 \\ G(c,d) &= (1+d\cdot c^{x+3n})(1+d\cdot c^{x+n})^{1+2B_2} - \\ &\quad -(1+d\cdot c^{x+2n})^{2+B_2}(1+d\cdot c^{x})^{B_2} = 0. \end{aligned}$$

where $B_1 = \dfrac{S_{x+3n} - S_{x+2n}}{S_{x+2n} - S_{x+n}}$, $B_2 = \dfrac{S_{x+2n} - S_{x+n}}{S_{x+n} - S_x}$.

The parameters a, b, are determined from:

$$\ln g = \frac{S_{x+n} - S_x}{\ln(1+d\cdot c^{x+2n})(1+d\cdot c^{x}) - 2\ln(1+d\cdot c^{x+n})}$$
$$\ln s = \frac{S_x}{n} - \frac{\ln g}{n}\ln\left(\frac{1+d\cdot c^{x+n}}{1+d\cdot c^{x}}\right),\ a = -\ln s,\ b = d\ln g\cdot \ln x + ad,$$

starting with age $x = 15$ and n=17 (using data from$x = 15$ to $x = 84$) .

We solve (6.1) graphically in order to find the intersection of the surfaces $z = F(c,d)$, and $z = G(c,d)$ with the plane $z = 0$.

Using the notations $SP1 = S_x$, $SP2 = S_{x+n}$, $SP3 = S_{x+2n}$, $SP4 = S_{x+3n}$ we have the numerical values:

B1 =5.0263; B2 =3.3102; and SP1 =-0.10507; SP2 =-0.45797; SP3 = -1.6261; SP4 = -7.4977.

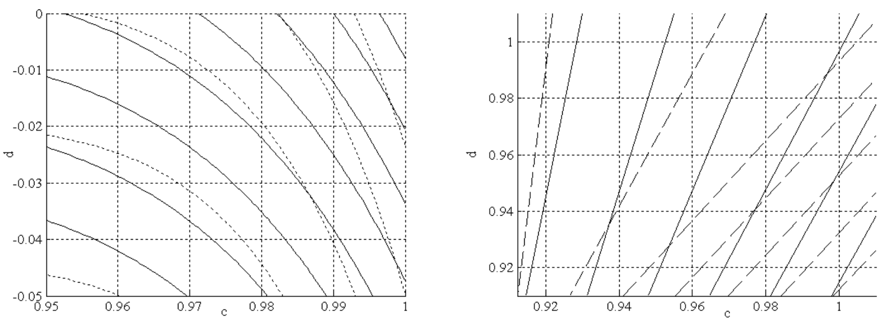

Figure 3: Graphical plot of the parameters c,d.

From Figure 3 we find different points:

c=0.9265;d=0.9258; c=1.087;d=1.07; c=1.097;d=0.93;c=1.0995;d=1.0967; c=1.4993;d=1.447; c=1.643;d=1.554; c=2.987;d=2.98; c=2.984;d=2.75; c=-0.99;d=-0.37; c=0.91;d=-0.007; **c=0.987;d=-0.032.**

Using c=0.987;d=-0.032, we express the survival function as in Figure 4.

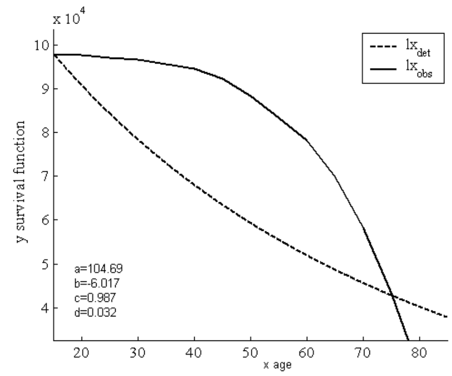

Figure 4: The survival function for the Romanian population. Observed data at reduced level and determined data.

7 Conclusions

We present our algorithm in order to fill in the abridged survival table, according to the available data for Romania. This table contains: the survival function, the number of deaths, dying probability, the number of persons – years that survived, the number of person – years that survived after age x, the life expectancy, the force of mortality and the probability of survival. Using Sprague interpolation for observed function p_x we obtain the annual data for survival probabilities.

Modeling the force rate of mortality for Romania population one can see from Figure 2 and Figure 4 that a **Makeham** − **Gompertz** function type or Perks distribution are not sufficiently suitable for the Romanian population data.

We can also use the obtained annual data for survival probabilities in order to provide the force of mortality annual data by formula $\mu_x = (-p_x + 1/p_{x-1})/2$. The fourth polynomial function interpolating this data is presented in Figure 5.

The survival function using this polynomial function for the force of mortality can be seen in Figure 6.

Moreover, it should be noted that the force of mortality is not an exponential distribution and the survival function is not a Weibull distribution.

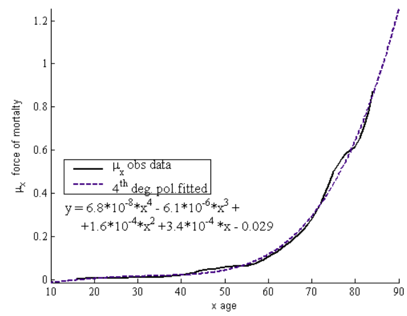

Figure 5: The force of mortality for the Romanian population.

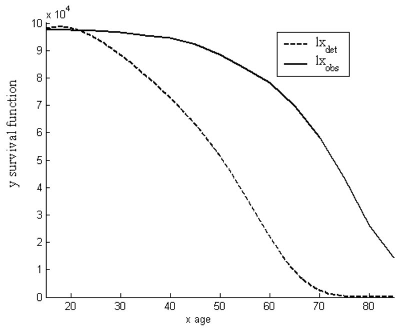

Figure 6: The survival function.

References

[1] T. Haavelmo, *A note on the variate difference method*, Econometrica 9, 1 (1941), 74-79.

[2] H.C. Herrero, em An experience constructing a complete and an abridged life table using a mathematical formula for a small population, Departamento de Economía Aplicada, Universidad de Valladolid, 2002.

[3] C. Horiuchi, *Age pattern of mortality for older women: an analysis using the age-specific rate of mortality change with age*, Mathematical Population Studies 2, 4 (1990), 245-267.

[4] I. Purcaru, I. Mircea, Gh. Lazar, *Asigurari de persoane si de bunuri. Aplicatii. Cazuri. Solutii*, Ed. Economica, Bucuresti, 1998.

[5] N.D. Shyamalkumar, *Analysis of mortality data using smoothing spline poisson regression*, Dept. of Stat., Act.Sci. The University of Iowa, 2002.

[6] H. Tudor, O. Popescu, *Matematici financiare si actuariale*, Ed. Albastra, Cluj-Napoca, 2004.

Authors' address:

Elena-Corina Cipu and Carmen Dragnea
University Politehnica of Bucharest, Mathematics III Department,
Splaiul Independentei 313, RO-060042, Bucharest, Romania
e-mail: corinac@math.pub.ro, d_c_dragnea@yahoo.com