Application of Goodness-of-fit Tests for the Gompertz Distribution to Identify Structural Deviations from the Exponential Pattern of Human Death Rates

Trifon I. Missov
Max Planck Institute for Demographic Research

Laszlo Nemeth
Max Planck Institute for Demographic Research

Extended abstract submitted to the 2020 EPC Conference, Padova

November 1, 2019

1 Introduction

Plotting log-mortality against age reveals a large segment of (seemingly) linear increase that implies a Gompertz risk of dying. Determining the actual range of Gompertz mortality, though, requires a formal procedure. Using Gompertz goodness-of-fit tests, we can check whether or not mortality between ages x_1 and x_2 increases exponentially. Lenart and Missov (2016) adapt general goodness-of-fit tests like the Anderson-Darling test (Anderson and Darling, 1952), the correlation coefficient test (Filliben, 1975), and the likelihood ratio test against the truncated generalized extreme value distribution (Elandt-Johnson, 1976; Wilks, 1938) to the case of the Gompertz distribution assuming that individual lifetimes are fully observed. Aggregate mortality data in the age interval $[x_1, x_2]$, though, are subjected to a different observation scheme: 1) individual lifetimes are interval censored (age at death is given in completed years and to reconstruct the "actual" age at death we need to assume a distribution of deaths within the interval, e.g. uniform distribution), and 2) individuals that die outside of $[x_1, x_2]$ should be addressed accordingly (if we do not consider them in the sample, then we have left and right truncation). We focus on the Anderson-Darling goodness-of-fit for the Gompertz distribution and modify the Anderson-Darling statistic (Chernobai et al., 2015: formula 20.17, p.583) so that it accounts for both left and right truncation.

Country	Year	Onset	Gompertz Ending
England & Wales	1900	63	83
England & Wales	2000	76	96
France	1900	63	87
France	2000	75	101
Italy	1900	62	82
Italy	2000	73	98
Japan	1950	65	92
Japan	2000	76	107
Sweden	1900	67	84
Sweden	2000	74	99

Table 1: The age range (columns 3 and 4) of pure Gompertz death-rate increase (determined by applying a correlation coefficient goodness-of fit test) for England & Wales, France, Italy, Japan, and Sweden in 1900 (1950 for Japan) and 2000.

2 Preliminary Findings

Figures 1 and 2 show the results of applying Anderson-Darling goodness-of-fit test for the Gompertz distribution in all possible intervals $[x_1, x_2]$, $x_1 \in [20, 90]$, $x_2 \in [20, 90]$, $x_1 < x_2$ for two populations: France 1972 and Italy 1987. Blue circles denote intervals $[x_1, x_2]$ with Gompertz mortality, while orange circles mark intervals with non-Gompertz risk of dying. For all four populations the vertical right border of blues areas denotes the endpoint of each interval with Gompertz mortality, e.g. for France 1972 death rates increase exponentially in the intervals $[25, 35]$, $[40, 52]$, and $[59, 72]$. The "accident hump" seems to be located between ages 35 and 40, while some excess mortality, perhaps cancer-related, causes deviation from the Gompertz pattern between ages 52 and 59. The estimated slope of the log-Gompertz line in the aforementioned three intervals is almost identical (0.12, 0.122, and 0.121), which implies that fitting a Gompertz line with a slope of 0.121 will separate Gompertzian ("senescent") from non-Gompertzian ("external") mortality.

Populations characterized by increasing longevity postpone deaths to later ages. Table 1 reflects this by showing increasing (with time) onset ages of the purely Gompertz part of mortality (column 3) as well as increasing ages of mortality deceleration (column 4). For countries with fluctuating life-expectancy patterns, e.g. Eastern European countries in the 1990s, we expect year-to-year fluctuation in extrinsic mortality and stagnation of the age of mortality deceleration.

3 Further Steps

This research aims to further study the time trend by country for the three measures of interest (the onset of aging, the age-ranges of non-Gompertz mortality at young adult and later adult ages, as well as the age of mortality deceleration). A by-product will be the separation of age-specific non-Gompertz mortality that can provide insight into the way to adequately model non-aging related mortality (we know that using a Makeham term is an oversimplified approach to quantify this). A third goal of this study is to estimate the slope of the Gompertz segments for each population, i.e. the rate of population aging, and draw inference about the rate of individual aging in a frailty model setting.

4 Method

Fitting Gompertz distribution with the assumption that death counts are Poisson distributed. Maximizing the likelihood:

$$\ln L = \sum_x [D(x) \ln \mu(x) - E(x)\mu(x)] \,, \tag{1}$$

Results in parameters a, b for a given (year, sex, country,) start-last age combination for all starting ages in [0,109] with last ages [(startage+1),110]. This is like a triangular matrix. Samples are always age transformed to starting age 0.

4.1 Anderson-Darling test

Based on the Anderson-Darling test for left-truncated distributions detailed in Chernobai et al. (2015). Eq. 20.17 on page 583 the AD statistic with left truncation at H is given by

$$AD^{2*} = -n + 2n \log(1 - z_H) - \frac{1}{n} \sum_{j=1}^{n} (1 + 2\,(n - j)) \log\,(1 - z_j) + \frac{1}{n} \sum_{j=1}^{n} (1 - 2j) \log\,(z_j - z_H)$$

Testing

$$H_0 : F_n(x) \in \hat{F^*}(x)$$

$$H_A : F_n(x) \notin \hat{F^*}(x)$$

where:

$$\hat{F^*}(x) := \frac{\hat{F}_\theta(x) - \hat{F}_\theta(H)}{1 - \hat{F}_\theta(H)}$$

$$z_j := F^*\left(x_{(j)}\right) \quad z_H := F^*\left(H\right)$$

Therefore z_j is the ordered cumulative distribution function transform of the sample values. With small modification, adding z_G for right truncation:

$$AD = -n + 2n\log(z_G - z_H) - \frac{1}{n}\sum_{j=1}^{n}\left(1 + 2\left(n - j\right)\right)\log\left(z_G - z_j\right) + \frac{1}{n}\sum_{j=1}^{n}\left(1 - 2j\right)\log\left(z_j - z_H\right)$$

In order to calculate AD, we need lifespans. Lifespans are reconstructed between the starting age and last age 1) first truncating to integer death counts 2) and lifespans uniformly distributed in a given year.

References

Anderson, T.W. and D.A. Darling. 1952. "Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes." *The Annals of Mathematical Statistics* 23:193–212.

Chernobai, A., S.T. Rachev, and F.J. Fabozzi. 2015. *Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples*, pp. 575–596. New York, NY: Springer New York.

Elandt-Johnson, R.C. 1976. "Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates." *Scandinavian Actuarial Journal* 1976:37–51.

Filliben, J.J. 1975. "The probability plot correlation coefficient test for normality." *Technometrics* 17:111–117.

Lenart, A. and T.I. Missov. 2016. "Goodness-of-fit tests for the Gompertz distribution." *Communications in Statistics - Theory and Methods* 45:2920–2937.

Wilks, S.S. 1938. "The large-sample distribution of the likelihood ratio for testing composite hypotheses." *The Annals of Mathematical Statistics* 9:60–62.

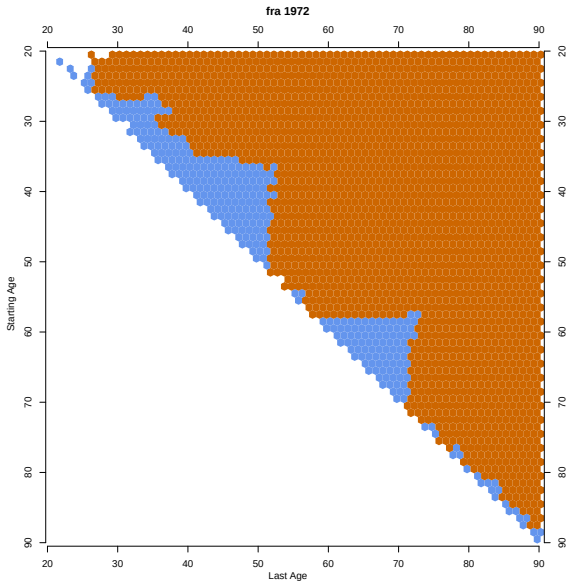

Figure 1: Application of the Anderson-Darling goodness-of-fit test for the Gompertz distribution in all intervals $[x_1, x_2]$, where $x_1 \in [20, 90]$ (vertical axis), $x_2 \in [20, 90]$ (horizontal axis), and $x_1 < x_2$ for France, 1972. Blue circles designate age ranges with Gompertz mortality, while orange circles point at age ranges with non-Gompertz risk of dying.

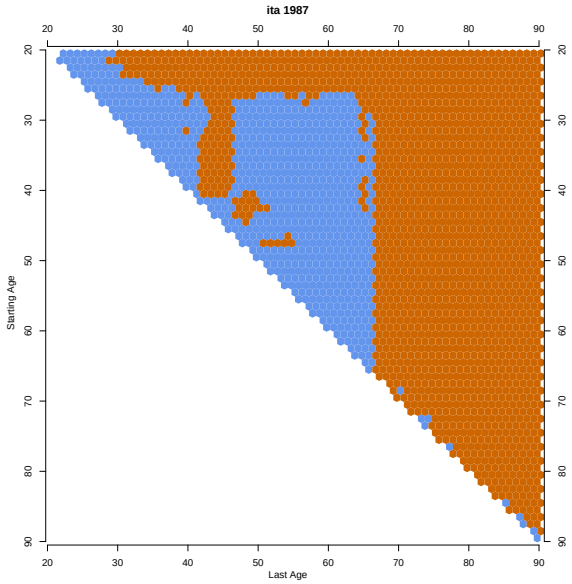

Figure 2: Application of the Anderson-Darling goodness-of-fit test for the Gompertz distribution in all intervals $[x_1, x_2]$, where $x_1 \in [20, 90]$ (vertical axis), $x_2 \in [20, 90]$ (horizontal axis), and $x_1 < x_2$ for Italy, 1987. Blue circles designate age ranges with Gompertz mortality, while orange circles point at age ranges with non-Gompertz risk of dying.