The Gompertz distribution and Maximum Likelihood Estimation of its parameters - a revision

Adam Lenart (lenart@demogr.mpg.de)

The Gompertz distribution and Maximum Likelihood Estimation of its parameters - a revision

Adam Lenart

November 28, 2011

Abstract

The Gompertz distribution is widely used to describe the distribution of adult deaths. Previous works concentrated on formulating approximate relationships to characterize it. However, using the generalized integro-exponential function Milgram (1985) exact formulas can be derived for its moment-generating function and central moments. Based on the exact central moments, higher accuracy approximations can be defined for them. In demographic or actuarial applications, maximum-likelihood estimation is often used to determine the parameters of the Gompertz distribution. By solving the maximum-likelihood estimates analytically, the dimension of the optimization problem can be reduced to one both in the case of discrete and continuous data.

Keywords: Gompertz distribution, moment-generating function, expected value, variance, skewness, kurtosis, maximum-likelihood estimation

1 Introduction

The Gompertz distribution is important in describing the pattern of adult deaths (Wetterstrand 1981; Gavrilov and Gavrilova 1991). For low levels of infant (and young adult) mortality, the Gompertz force of mortality extends to the whole life span (Vaupel 1986) of populations with no observed mortality deceleration.

The Gompertz distribution has received considerable attention from demographers and actuaries. Pollard and Valkovics (1992) were the first to study the Gompertz distribution thoroughly. However, their results are true only in the case when the initial level of mortality is very close to 0. Kunimura (1998) arrived at similar conclusions. They both defined the moment generating function of the Gompertz distribution in terms of the incomplete or complete gamma function and their results are either approximate or left in an integral form. Willemse and Koppelaar (2000) reformulated the Gompertz force of mortality and derived relationships for this new formulation. Later, Marshall and Olkin (2007) described the negative Gompertz distribution; a Gompertz distribution with a negative rate of aging parameter. Willekens (2002) provided connections between the Gompertz, the Weibull and other Type I extreme value distributions. In this paper, I will keep the most often used

formulation of the Gompertz' law of mortality (Gompertz 1825) in demography (Preston et al. 2001:9.1). The Gompertz force of mortality at age x, $x \geq 0$, is

$$\mu(x) = ae^{bx} \quad a, b > 0 \,, \tag{1}$$

where a denotes the level of the force of mortality at age 0 and b the rate of aging. The moments of the Gompertz distribution can be explicitly given by the generalized integro-exponential function (Milgram 1985), that offers exact power series representation of them.

2 The Gompertz distribution

The Gompertz distribution has a continuous probability density function with location parameter a and shape parameter b,

$$f(x) = ae^{bx - \frac{a}{b}\left(e^{bx} - 1\right)} \,, \tag{2}$$

with support on $(-\infty, \infty)$. In actuarial or demographic applications, x usually denotes age which cannot be negative, leading to bounded support on $[0, \infty)$. The truncated distribution yields a proper density function by rescaling the a parameter to correspond to $x = 0$ (Garg et al. 1970). The distribution function is

$$F(x) = 1 - e^{-\frac{a}{b}\left(e^{bx} - 1\right)} \,.$$

The moment generating function of the Gompertz distribution is given by its Laplace transform.

Proposition 1. *The Laplace transform of the Gompertz distribution is*

$$L(s) = \frac{a}{b} e^{\frac{a}{b}} E_{\frac{s}{b}}\left(\frac{a}{b}\right) \quad a, b > 0 \,, \tag{3}$$

where $E_n(z) = \int\limits_1^\infty \frac{e^{-zt}}{t^n}\, dt$ (Abramowitz and Stegun 1965:5.1.4).

Proof. The Laplace transform of the Gompetz pdf of (2) is

$$L(s) = \int\limits_0^\infty ae^{bx - \frac{a}{b}\left(e^{bx} - 1\right)} e^{-sx}\, dx \,. \tag{4}$$

By substituting $q = e^{bx}$ in (4)

$$L(s) = \frac{a}{b} e^{\frac{a}{b}} \int\limits_1^\infty e^{-\frac{a}{b} q} q^{-\frac{s}{b}}\, dq$$

2

Note that the exponential integral, $E_n(z)$, is defined as (Abramowitz and Stegun 1965:5.1.4)

$$E_n(z) = \int_1^\infty \frac{e^{-zt}}{t^n}\, dt \quad n > 0, Re(z) > 0\,,$$

therefore

$$L(s) = \frac{a}{b} e^{\frac{a}{b}} E_{\frac{s}{b}}\left(\frac{a}{b}\right)$$

□

The moments of the Gompertz distribution are expressed as the derivatives of its Laplace transform that can be expressed in a general formula by using the generalized integro-exponential function (Milgram 1985).

Proposition 2. *The nth moment of a Gompertz distributed random variable X is*

$$E\left[X^n\right] = \frac{n!}{b^n}\, e^{\frac{a}{b}} E_1^{n-1}\left(\frac{a}{b}\right)\,, \tag{5}$$

where $E_s^n(z) = \frac{1}{n!} \int_1^\infty (\ln x)^n x^{-s} e^{-zx}\, dz$ is the generalized integro-exponential function (Milgram 1985).

Proof. The moments of random variable X of a Gompertz distribution can be calculated with the aid of the Laplace transform. From (3),

$$E\left[X\right] = -\frac{d}{ds}L(s)\Big|_{s=0} = \frac{a}{b} e^{\frac{a}{b}} \int_1^\infty \frac{1}{b} e^{-\frac{a}{b}x} \ln(x)\, dx$$

$$E\left[X^2\right] = \frac{d^2}{ds^2}L(s)\Big|_{s=0} = \frac{a}{b} e^{\frac{a}{b}} \int_1^\infty \frac{1}{b^2} e^{-\frac{a}{b}x} [\ln(x)]^2\, dx$$

$$\vdots$$

$$E\left[X^n\right] = (-1)^{n+1}\frac{d^n}{ds^n}L(s)\Big|_{s=0} = \frac{a}{b} e^{\frac{a}{b}} \int_1^\infty \frac{1}{b^n} e^{-\frac{a}{b}x} [\ln(x)]^n\, dx$$

By integration by parts

$$E\left[X^n\right] = e^{\frac{a}{b}} \int_1^\infty \frac{n}{b^n} x^{-1} e^{-\frac{a}{b}x} [\ln(x)]^{n-1}\, dx$$

3

Note that the intergral representation of the generalized integro-exponential function (Milgram 1985) is

$$E_s^j(z) = \frac{1}{\Gamma(j+1)} \int\limits_1^\infty [\ln x]^j x^{-s} e^{-zx} \, dx$$

or by definition:

$$E_s^j(z) = \frac{(-1)^j}{j!} \frac{\partial^j}{\partial s^j} E_s(z) \,, \tag{6}$$

and

$$E_s^0(z) \equiv E_s(z).$$

Therefore, $E[X^n]$ can be represented by (6):

$$E[X^n] = \frac{n!}{b^n} e^{\frac{a}{b}} E_1^{n-1}\left(\frac{a}{b}\right)$$

or by the more widely used Meijer G-function (Milgram 1985):

$$E[X^n] = \frac{n!}{b^n} e^{\frac{a}{b}} G_{n,n+1}^{n+1,0}\left(\frac{a}{b} \left| \begin{array}{c} ; 1, \ldots, 1 \\ 0, \ldots, 0; \end{array} \right.\right) \,, \tag{7}$$

where the Meijer G-function is a generalized hypergeometric function. It is defined by the contour integral

$$G_{p,q}^{m,n}\left[z \left| \begin{array}{c} a_1, \ldots, a_n; a_{n+1}, \ldots, a_p \\ b_1, \ldots, b_m; b_{m+1}, \ldots, b_q \end{array} \right.\right] = \frac{1}{2\pi i} \int_C \frac{\displaystyle\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^n \Gamma(1 - a_j + s)}{\displaystyle\prod_{j=m+1}^q \Gamma(1 - b_j + s) \prod_{j=n+1}^p \Gamma(a_j - s)} z^s \, ds$$

along contour C (Erdélyi 1953). □

Series expansion of the integral form offers a simpler, but lengthier, representation of the moments of the Gompertz distribution. The power series expansion of the generalized integro-exponential function (Milgram 1985:2.10) in the Gompertz case, ie. for $s = 1$ and $z = \frac{a}{b}$ is,

$$E_1^{n-1}\left(\frac{a}{b}\right) = \left[\sum_{j=1}^\infty \frac{1}{(-j)^n} \frac{(-\frac{a}{b})^j}{j!}\right] + \frac{(-1)^n}{n!} \sum_{j=0}^n \binom{n}{j} \ln\left(\frac{a}{b}\right)^{n-j} \Psi_j \,, \tag{8}$$

where

$$\Psi_j = \lim_{t \to 0} \frac{d^j}{dt^j} \Gamma(1 - t) \,. \tag{9}$$

4

For $j \in (0, 1, 2, 3, 4)$

$$\Psi_0 = 1 \, ;$$
$$\Psi_1 = \gamma \, ;$$
$$\Psi_2 = \gamma^2 + \frac{\pi^2}{6} \, ;$$
$$\Psi_3 = \gamma^3 + \gamma\frac{\pi^2}{2} + 2\zeta(3) \, ;$$
$$\Psi_4 = \gamma^4 + \gamma^2\pi^2 + 8\gamma\zeta(3) + \frac{3}{20}\pi^4 \, ,$$

where $\gamma \approx 0.57722$ is the Euler-Mascheroni constant and $\zeta(s) = \sum\limits_{n=1}^{\infty} n^{-s}$ is the Riemann zeta function (Abramowitz and Stegun 1965:23.2.18). Note that $\frac{\pi^2}{6} = \zeta(2)$ and $\frac{\pi^4}{90} = \zeta(4)$. Also note that $\zeta(3) \approx 1.2021$ is known as Apéry's constant (Apéry 1979). Please refer to Appendix A.1 for derivation of the values of the Ψ function.

2.1   Expected value, variance, skewness and kurtosis

Corollary 1. *The expected value of random variable X of a Gompertz distribution is*

$$E[X] = \frac{1}{b}e^{\frac{a}{b}}E_1\left(\frac{a}{b}\right) \tag{10}$$
$$\approx \frac{1}{b}e^{\frac{a}{b}}\left(\frac{a}{b} - \ln\left(\frac{a}{b}\right) - \gamma\right) \, .$$

The precision of the approximate result depends on the ratio of a to b. For low values of $\frac{1}{b}e^{\frac{a}{b}}\frac{\left(\frac{a}{b}\right)^2}{4}$, the approximation is precise.

Proof.

Exact   From (5), the expected value of random variable X of a Gompertz distribution, or life expectancy, is given by

$$E[X] = \frac{1}{b}e^{\frac{a}{b}}\int\limits_{1}^{\infty} x^{-1}e^{-\frac{a}{b}x}\,dx$$
$$= \frac{1}{b}e^{\frac{a}{b}}E_1\left(\frac{a}{b}\right)$$

This result has already been shown by Missov and Lenart (2011).

Approximate The approximate result for the expected value in a Gompertz model is also well-known, when a is close to 0, by Abramowitz and Stegun (1965:5.1.11) the exponential integral, $E_1(z)$, can be approximated by $-\gamma - \ln(z)$, and (10) appears as $\frac{1}{b}e^{\frac{a}{b}}\left(-\gamma - \ln\frac{a}{b}\right)$, where $\gamma = 0.57722$ denotes the Euler-Mascheroni constant (see e.g. Pollard and Valkovics 1992; Kunimura 1998; Missov and Lenart 2011). However, the precision of the previous approximations can be improved by adding new terms to it.

From (8) the power series expansion of the expected value is

$$E[X] = \frac{1}{b}e^{\frac{a}{b}}\left(-\sum_{j=1}^{\infty}\frac{(-\frac{a}{b})^j}{j!} - \sum_{j=0}^{1}\binom{1}{j}\ln\left(\frac{a}{b}\right)^{1-j}\Psi_j\right),$$

and by (24)

$$E[X] = \frac{1}{b}e^{\frac{a}{b}}\left(-\sum_{j=1}^{\infty}\frac{(-\frac{a}{b})^j}{j!} - \ln\left(\frac{a}{b}\right) - \gamma\right).$$

As a is usually close to 0, $\frac{a}{b} \approx -\sum_{j=1}^{\infty}\frac{(-\frac{a}{b})^j}{j!}$ and

$$E[X] \approx \frac{1}{b}e^{\frac{a}{b}}\left(\frac{a}{b} - \ln\left(\frac{a}{b}\right) - \gamma\right). \tag{11}$$

Note that because the sign of $-\sum_{j=1}^{\infty}\frac{(-\frac{a}{b})^j}{j!}$ changes, (11) will be slightly overestimated for relatively large a but by adding new terms of the sum, the expected value, or life expectancy at birth, can be approximated by arbitrary precision. For the error of the approximation, please see Fig. 5 in the Appendix. $\square$

Corollary 2. *The variance of random variable X of a Gompertz distribution is*

$$Var[X] = \frac{2}{b^2}e^{\frac{a}{b}}\left\{-\frac{a}{b}\,_3F_3\left[\begin{array}{c}1,1,1\\2,2,2;\end{array};\frac{a}{b}\right] + \frac{1}{2}\left[\frac{\pi^2}{6} + \left(\gamma + \ln\left(\frac{a}{b}\right)\right)^2\right]\right\} - \left[\frac{1}{b}e^{\frac{a}{b}}E_1\left(\frac{a}{b}\right)\right]^2 \tag{12}$$

$$\approx \frac{1}{b^2}\frac{\pi^2}{6} - 2\frac{a}{b^3}.$$

The approximate result holds for $a \approx 0$. $_pF_q\left[\begin{array}{c}a_1,\ldots,a_p\\b_1,\ldots,b_q;\end{array};z\right]$ denotes the generalized hypergeometric function (e.g. Askey and Daalhuis 2010).

Proof.

Exact The variance of a random variable X is

$$Var[X] = E[X^2] - E[X]^2.$$

6

Concentrating on $E\left[X^2\right]$, if X is Gompertz distributed, from (5):

$$E\left[X^2\right] = \frac{2}{b^2} e^{\frac{a}{b}} E_1^1\left(\frac{a}{b}\right) \tag{13}$$

An exact representation of (13) can be reached by the G function by (7):

$$E\left[X^2\right] = \frac{2}{b^2} e^{\frac{a}{b}} G_{2,3}^{3,0}\left(\frac{a}{b} \left|\begin{array}{c} ;1,1 \\ 0,0,0; \end{array}\right.\right).$$

Power series expansion of E_s^j yields (Milgram 1985:2.10),

$$E_1^j(z) = \sum_{l=1}^{\infty} \frac{1}{(-l)^{j+1}} \frac{(-z)^l}{l!} + \frac{(-1)^{j+1}}{(j+1)!} \sum_{l=0}^{j+1} \binom{j+1}{l} \ln(z)^{1+j-l} \Psi_l, \tag{14}$$

where

$$\Psi_l = \lim_{t \to 0} \frac{d^l}{dt^l} \Gamma(1-t).$$

Note that

$$\sum_{l=1}^{\infty} \frac{1}{(-l)^{j+1}} \frac{(-z)^l}{l!} = (-1)^j z \,_{j+2}F_{j+2}\left[\begin{array}{c} 1_1, \ldots, 1_{j+2} \\ 2_1, \ldots, 2_{j+2}; \end{array} ; z\right],$$

where $_pF_q\left[\begin{array}{c} a_1, \ldots, a_p \\ b_1, \ldots, b_q; \end{array} ; z\right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_p)_k} \frac{z^k}{k!}$ denotes the generalized hypergeometric function (Askey and Daalhuis 2010).

Therefore

$$E\left[X^2\right] = \frac{2}{b^2}\left\{-\frac{a}{b} \,_3F_3\left[\begin{array}{c} 1,1,1 \\ 2,2,2; \end{array} ; \frac{a}{b}\right] + \frac{1}{2}\left[\frac{\pi^2}{6} + \left(\gamma + \ln\left(\frac{a}{b}\right)\right)^2\right]\right\} \tag{15}$$

and

$$Var[X] = \frac{2}{b^2} e^{\frac{a}{b}}\left\{-\frac{a}{b} \,_3F_3\left[\begin{array}{c} 1,1,1 \\ 2,2,2; \end{array} ; \frac{a}{b}\right] + \frac{1}{2}\left[\frac{\pi^2}{6} + \left(\gamma + \ln\left(\frac{a}{b}\right)\right)^2\right]\right\} - \left[\frac{1}{b} e^{\frac{a}{b}} E_1\left(\frac{a}{b}\right)\right]^2$$

Approximate By juxtaposing (11) with (15), it is easy to see possible approximations. Pollard and Valkovics (1992) derived that the variance of the Gompertz distribution is $\frac{1}{b^2}\frac{\pi^2}{6}$. This result only holds when a is very close to 0. However, from (12), it is clear that the variance depends on the parameter a, i.e. for fixed b, as a increases, the variance decreases.

When $a \approx 0$, by (11)

$$E\left[X^2\right] \approx \frac{1}{b^2} e^{\frac{a}{b}} \frac{\pi^2}{6} - e^{-\frac{a}{b}} E\left[X\right]^2$$

and for $e^{\frac{a}{b}} = 1$

$$Var(X) \approx \frac{1}{b^2} \frac{\pi^2}{6} . \tag{16}$$

Note that (16) crudely overestimates the variance for relatively larger a. The value of $_3F_3 \begin{bmatrix} 1,1,1 \\ 2,2,2; \end{bmatrix} \frac{a}{b}$ for $a \approx 0$ is 1 and for any higher likely values for human force of mortality (ie. $a \approx 0.1$ about age 90), it is equal to about 0.9, or approximately 1. However, $-2\frac{a}{b^3}$ can be significantly large and is the only term that reduces the overestimated variance of (16) in (12). Therefore,

$$Var(X) \approx \frac{1}{b^2} \frac{\pi^2}{6} - 2\frac{a}{b^3}$$

always gives a better approximation to the variance than (16). For the error of the approximation, please see Fig. 5 in the Appendix. □

Corollary 3. *The skewness, γ_1 of the probability distribution of random variable X of a Gompertz distribution is*

$$\gamma_1 = \frac{\frac{6}{b^3}e^{\frac{a}{b}}G_{3,4}^{4,0}\left(\frac{a}{b} \middle| \begin{array}{c} ;1,1,1 \\ 0,0,0,0; \end{array}\right) - 3m_1\frac{2}{b^2}e^{\frac{a}{b}}G_{2,3}^{3,0}\left(\frac{a}{b} \middle| \begin{array}{c} ;1,1 \\ 0,0,0; \end{array}\right) + 2(m_1)^3}{\left(\frac{2}{b^2}e^{\frac{a}{b}}\left\{-\frac{a}{b}\,_3F_3\begin{bmatrix} 1,1,1 \\ 2,2,2; \end{bmatrix}\frac{a}{b}\right\} + \frac{1}{2}\left[\frac{\pi^2}{6} + \left(\gamma + \ln\left(\frac{a}{b}\right)\right)^2\right]\right\} - \left[\frac{1}{b}e^{\frac{a}{b}}E_1\left(\frac{a}{b}\right)\right]^2\right)^{\frac{3}{2}}}$$

$$\approx \begin{cases} 4.15a^{0.3} - 5b^{0.49} - 1.48a + 4.31b - 4.96ab, & \text{if } a > 0 \\ \frac{-12\sqrt{6}}{\pi^3}\zeta(3), & \text{if } a \approx 0 \end{cases},$$

where m_1 denotes the expected value of the distribution.

Proof.

Exact Skewness is given as the third standardized moment,

$$\gamma_1 = \frac{E[X^3] - 3E[X]E[X^2] + 2E[X]^3}{Var[X]^{\frac{3}{2}}} .$$

From (7) and (12) the exact result is immediately given. An equivalent power series representation can be easily acquired from (8) and (9).

Approximate The approximate result is given by the limit of the power series expansion of the third central moment, m_3, when a goes to 0

$$m_3|_{a=0} = \lim_{a \to 0} m_3 = -\frac{2}{b^3}\zeta(3)$$

8

and from (16)

$$\gamma_1|_{a=0} = \lim_{a \to 0} \frac{m_3}{Var\,[X]^{\frac{3}{2}}} = \frac{-12\sqrt{6}}{\pi^3}\zeta(3) \qquad (17)$$

and for the other cases in the $0 < a, b \leq 0.2$ surface, the approximation is the result of fitting the

$$\gamma_1|_{a>0} \approx c_1 a^{c_2} + c_3 b^{c_4} + c_5 a + c_6 b + c_7 ab$$

expression by using (unweighted) non-linear least squares. For the error of the approximation, please see Fig. 5 in the Appendix.

Note that (17) also yields the lowest value, $\gamma_1 = -1.13955$ for the skewness of the Gompertz distribution. $\qquad \square$

Corollary 4. *The excess kurtosis, γ_2 of the probability distribution of random variable X of a Gompertz distribution is*

$$
\gamma_2 = \left\{ \frac{24}{b^4}e^{\frac{a}{b}}G_{4,5}^{5,0}\left(\frac{a}{b} \,\middle|\, \begin{matrix} ;1,1,1,1 \\ 0,0,0,0,0; \end{matrix} \right) - 4m_1\frac{6}{b^3}e^{\frac{a}{b}}G_{3,4}^{4,0}\left(\frac{a}{b} \,\middle|\, \begin{matrix} ;1,1,1 \\ 0,0,0,0; \end{matrix} \right) \right.
$$
$$
\left. +6(m_1)^2\frac{2}{b^2}e^{\frac{a}{b}}G_{2,3}^{3,0}\left(\frac{a}{b} \,\middle|\, \begin{matrix} ;1,1 \\ 0,0,0; \end{matrix} \right) + 3(m_1)^4 \right\} \Big/ \left\{ \frac{2}{b^2}e^{\frac{a}{b}}\left\{-\frac{a}{b}\,{}_3F_3\left[\begin{matrix}1,1,1\\2,2,2;\end{matrix}\,;\frac{a}{b}\right]\right.\right.
$$
$$
\left.\left. +\frac{1}{2}\left[\frac{\pi^2}{6}+\left(\gamma+\ln\left(\frac{a}{b}\right)\right)^2\right]\right\} - \left[\frac{1}{b}e^{\frac{a}{b}}E_1\left(\frac{a}{b}\right)\right]^2\right\}^2 - 3
$$
$$
\approx \begin{cases} -0.75 + 34.13a^{0.253} + 20b^{0.311} - 53.51(ab)^{0.14}, & if\ a > 0 \\ \frac{12}{5}, & if\ a \approx 0 \end{cases},
$$

Proof.

Exact Excess kurtosis is given as the difference between the fourth standardized moment and the kurtosis of the normal distribution; 3,

$$\gamma_2 = \frac{E\,[X^4] - 4E\,[X]\,E\,[X^3] + 6E\,[X]^2\,E\,[X^2] - 3E\,[X]^4}{Var\,[X]^2} - 3\,.$$

From (7) and (12) the exact result is immediately given. An equivalent power series representation can be easily acquired from (8) and (9). Note that the excess kurtosis reaches a minimal value about -0.75.

9

Approximate   The approximate result is given by the limit of the power series expansion of the fourth central moment, m_4, when a goes to 0

$$m_4\big|_{a=0} = \lim_{a \to 0} m_4 = \frac{3\pi^4}{20b^4}$$

and from (16)

$$\gamma_2\big|_{a=0} = \lim_{a \to 0} \frac{m_4}{Var[X]^2} - 3 = \frac{12}{5} \ .$$

and for the other cases in the $0 < a, b \leq 0.2$ surface, the approximation is the result of fitting the

$$\gamma_2\big|_{a>0} \approx -0.75 + c_1 a^{c_2} + c_3 b^{c_4} + c_5 (ab)^{c_6}$$

expression by using (unweighted) non-linear least squares. For the error of the approximation, please see Fig. 5 in the Appendix.   □
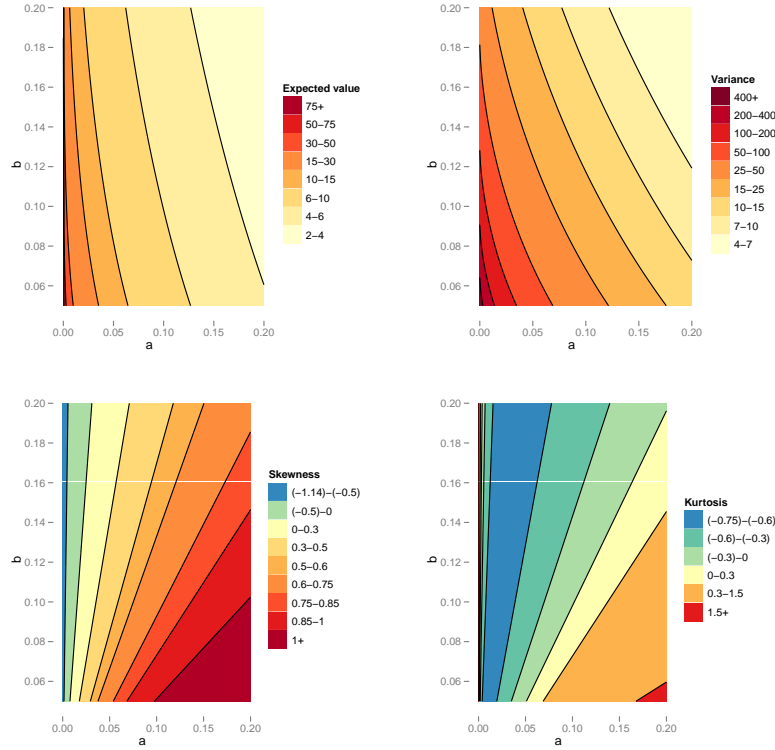

Figure 1: Expected value, variance, skewness and kurtosis of the Gompertz distribution

The Gompertz probability density function can be reformulated into a distribution with location and scale parameters. By noting that the mode, M

$$M = \frac{1}{b} \ln \left(\frac{b}{a} \right),$$

(2) can be expressed as

$$f(x) = be^{e^{-bM} - e^{b(x-M)} + b(x-M)}$$

3 Maximum Likelihood Estimation

3.1 Discrete age

Often, only discrete data is available and (1) cannot be directly estimated. However, if the number of deaths and the number of person-years exposed to the risk of dying can be observed, we can assume that the number of deaths in a given age interval follows a Poisson distribution.

Let D denote the number of deaths and λ the rate parameter of the Poisson distribution

$$f(D; \lambda) = \frac{1}{D!} \theta^D e^{-\lambda}. \tag{18}$$

The rate parameter, λ, of the Poisson distribution is consistent with $\mu(x)$ as both $\lambda, \mu(x) > 0$. The rate of occurrence of a Poisson process assumes equal observation window for each unit of observation. When applying the Poisson distribution to death-exposure data, the unit of observation is the number of deaths at each age group, D_x and the observation window is the number of people (or person-years) who are exposed to the risk of death. Clearly, the number of people at the risk of dying differs from age group to age group. Therefore, θ has to be weighted by the number of person-years exposed to death, E_x at age x. After substituting $\lambda = \lambda E_x$ in (18) and taking the log of it and eliminating the additive constants, the log-likelihood function of deaths D will be

$$l(\theta|D) \propto \sum_x D_x \ln \theta - E_x \theta. \tag{19}$$

If we substitute $\mu(x)$ for θ, we assume that we have n independent draws (for each age) from a Poisson distribution with a rate parameter that sheds light on the previously hidden (in (19)) functional relationship between rate λ and age x.

Using data for US females in 2007, aged 35-100 from the Human Mortality Database, parameters a and b can be readily estimated.

From Fig. 3, we can see that the maximal likelihood value lies somewhere in the middle of the range of plotted a's and b's ($\hat{a} = 0.0006$, $\hat{b} = 0.095$, $log\text{-}L = -4963073$).

The maximum likelihood estimator for a is (Appendix B.1)

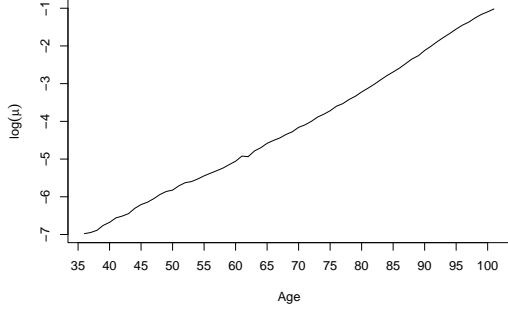$$\hat{a} = \frac{\sum_x D_x}{\sum_x E_x e^{bx}}. \tag{20}$$

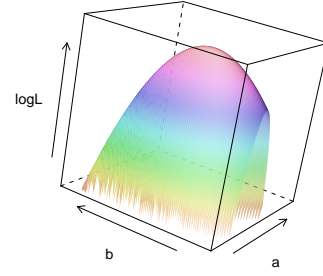Figure 2: Observed force of mortality. USA 2007, ages 35-100, females.

Figure 3: Log-likelihood values of a and b for USA 2007, ages 35-100 data

(20) can be immediately used to reduce the dimension of the optimization problem from two to one. Applying it to the US data from 2007, from Fig. 4 the optimal b, $\hat{b}$ that has the associated maximal log likelihood value can easily be seen and both $\hat{a}$, $\hat{b}$ and the maximal likelihood value is the same as in the two dimensional optimization case.
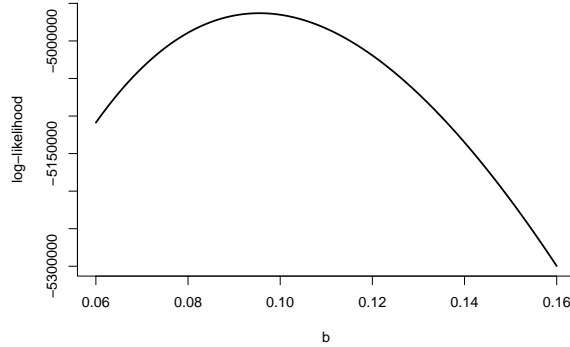

Figure 4: Log-likelihood values of b given optimal a for each b for US 2007, ages 35-100, females data

If $\hat{a}$ is the maximum likelihood estimator for a, then the maximum likelihood estimator for b has to satisfy (see Appendix B.1)

$$\frac{1}{\sum_x D_x} \sum_x D_x x = \frac{1}{\sum_x E_x e^{\hat{b}x}} \sum_x E_x e^{\hat{b}x} x \,. \tag{21}$$

As $\hat{b}$ is the only unknown in (21), it can be numerically solved and afterward, $\hat{b}$ can be substituted in the expression of $\hat{a}$ which reduces the problem of two-dimensional optimization to equation solving.

3.1.1 Example for solving b

The Newton-Raphson method is commonly used to solve (systems of) equations numerically. This method uses an iterative process to approximate the root of a function. If r denotes the root and n the number of iterations,

$$r_{n+1} = r_n - \frac{f(r_n)}{f'(r_n)},$$

r_{n+1} will give a $dr = r_{n+1} - r_n$ better approximation to the root of function $f(\cdot)$ than r_n.

In the case of discrete Gompertz mortality models, $f(b)$ is (21), and

$$f'(b) = \left(\frac{\sum\limits_x E_x e^{bx} x}{\sum\limits_x E_x e^{bx}} \right)^2 - \frac{\sum\limits_x E_x e^{bx} x^2}{\sum\limits_x E_x e^{bx}}.$$

For the US 2007, females aged 35-100 data, by taking a reasonable initial guess of $b = 0.12$ for any human population,

n	b_n	$f(b_n)$	$f'(b_n)$	b_{n+1}	db
0	0.12	-3.738	-131.264	0.0915218	-0.0284782
1	0.0915218	0.7091	-182.8819	0.0953989	3.87×10^{-3}
2	0.0953989	0.01468335	-175.3180	0.09548266	8.38×10^{-5}

Table 1: Newton-Raphson solution for b in case of US 2007, ages 35-100, females data

the solution of $\hat{b} = 0.095$ is reached after the second iteration. If higher than 8.38×10^{-5} precision is required, the iterations can be continued up to infinity. (Table 1)

3.1.2 Remarks

Please note that by letting w_D and w_E be

$$w_D = \frac{D_x}{\sum\limits_x D_x} \quad \text{and} \quad w_E = \frac{E_x e^{\hat{b}x}}{\sum\limits_x E_x e^{\hat{b}x}}$$

and by denoting the respective vectors as $\mathbf{w_D}$, $\mathbf{w_E}$ and $\mathbf{x}$, (21) can be rewritten as

$$\mathbf{w_D}^T \mathbf{x} = \mathbf{w_E}^T \mathbf{x}.$$

At the optimal b, $\hat{b}$, the two weighted averages of ages ($\mathbf{x}$); one weighted by the number of deaths ($\mathbf{w_D}$) and the other, weighted by the exposures ($\mathbf{w_E}$) are equal to each other. The weights of the exposures are themselves also weighted. Their weights correspond to the rate of aging, each consecutive age has $e^{\hat{b}}$ times higher importance than the previous one.

3.2   Continuous age

In the case when fully observed, or right-censored continuous data for the distribution of deaths is available, the Gompertz distribution can provide the likelihood function.[1] Let X_i denote the observed or right-censored lifetimes and let $\delta_i = 1$ if time of death of individual i is observed and $\delta_i = 0$ if the individual lifetime is right-censored. In this case, the joint density and the log likelihood function of the Gompertz distribution will be (Garg et al. 1970)

$$p(X_i|a, b) = \left(ae^{bX_i}\right)^{\delta_i} e^{-\frac{a}{b}\left(e^{bX_i}-1\right)}$$

and

$$l(a, b|X) = \sum_i \delta_i \left(\ln a + bX_i\right) - \frac{a}{b}\left(e^{bX_i} - 1\right)$$

respectively.

The score functions in the continuous case will be

$$\begin{pmatrix} S_a \\ S_b \end{pmatrix} = \begin{pmatrix} \sum_i \frac{\delta_i}{a} - \frac{e^{bX_i}-1}{b} \\ \sum_i \frac{a}{b^2}\left(e^{bX_i} - 1\right) + \delta_i X_i - \frac{a}{b}e^{bX_i}X_i \end{pmatrix}$$

and therefore

$$\hat{a} = \frac{Db}{\frac{1}{n}\sum_i e^{bX_i} - 1},$$

where n is the number of observations and $D = \frac{1}{n}\sum_i \delta_i$ is the proportion of deaths out of all observations.

The MLE of b is the solution to

$$\sum_i \frac{De^{bX_i}X_i}{\frac{1}{n}\sum_i e^{bX_i} - 1} = \sum_i \left(\frac{D\left(e^{bX_i} - 1\right)}{\frac{b}{n}\sum_i e^{bX_i} - b} + \delta_i X_i \right).$$

Note that Garg et al. (1970) used the discretized version of this likelihood to provide maximum likelihood estimates and when the step size of the discretization is sufficiently small, it would give the same results as the continuous form shown here.

---

[1]However, Pollard and Valkovics (1997) showed that an approximate likelihood for discrete data can be found using the truncated Gompertz distribution.

3.2.1 Remarks

If all X_i are observed times of death, by denoting the average force of mortality of the population by $\bar{\mu}$ and noting that $\bar{\mu} = \sum_i \frac{a}{n} e^{bX_i}$:

$$\bar{\mu} = \hat{a} + b \,.$$

If $b = \hat{b}$ and the data come from a Gompertz distribution, the observed average force of mortality is equal to $\hat{a} + \hat{b}$.

References

Abramowitz, M. and I. Stegun. 1965. *Handbook of Mathematical Functions*. Washington, DC: US Government Printing Office.

Apéry, R. 1979. "Irrationalité de ζ (2) et ζ (3)." *Astérisque* 61:11–13.

Askey, R. A. and A. B. O. Daalhuis. 2010. "Generalized Hypergeometric Functions and Meijer G-Function." In *NIST handbook of mathematical functions*, edited by F. Olver, D. Lozier, R. Boisvert and C. Clark, Cambridge University Press.

Erdélyi, A. 1953. *Higher transcendental functions*, vol. 1. New York-Toronto-London: McGraw-Hill.

Garg, M., B. Rao and C. Redmond. 1970. "Maximum-likelihood estimation of the parameters of the Gompertz survival function." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 19(2):152–159.

Gavrilov, L. and N. Gavrilova. 1991. *The biology of Life Span: A Quantitative Approach*. Chur: Harwood.

Gompertz, B. 1825. "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies." *Philosophical Transactions of the Royal Society of London* 115:513–583.

Gussmann, E. 1967. "Modifizierung der Gewichtsfunktionenmethode zur Berechnung der Fraunhoferlinien in Sonnen-und Sternspektren." *Zeitschrift für Astrophysik* 65:456–497.

Kunimura, D. 1998. "The Gompertz Distribution-Estimation of Parameters." *Actuarial Research Clearing House* 2:65–76.

Marshall, A. and I. Olkin. 2007. *Life Distributions*. Springer.

Milgram, M. 1985. "The generalized integro-exponential function." *Mathematics of computation* 44(170):443–458.

Missov, T. and A. Lenart. 2011. "Linking period and cohort life-expectancy linear increases in Gompertz proportional hazards models." *Demographic Research* 24(19):455–468.

Pollard, J. and E. Valkovics. 1992. "The Gompertz distribution and its applications." *Genus* 48(3-4):15–29.

———. 1997. "On the use of the truncated Gompertz distribution and other models to represent the parity progression functions of high fertility populations." *Mathematical Population Studies* 6(4):291–305.

Preston, S., P. Heuveline and M. Guillot. 2001. *Demography: measuring and modeling population processes*. Wiley-Blackwell.

Vaupel, J. 1986. "How change in age-specific mortality affects life expectancy." *Population Studies* 40(1):147–157.

Wetterstrand, W. 1981. "Parametric models for life insurance mortality data: Gompertz's law over time." *Transactions of the Society of Actuaries* 33:159–175.

Willekens, F. 2002. "Gompertz in context: the Gompertz and related distributions." In *Forecasting Mortality in Developed Countries - Insights from a Statistical, Demographic and Epidemiological Perspective, European Studies of Population*, vol. 9, edited by E. Tabeau, A. van den Berg Jeths and C. Heathcote, pp. 105–126, Springer.

Willemse, W. and H. Koppelaar. 2000. "Knowledge Elicitation of Gompertz' Law of Morality." *Scandinavian Actuarial Journal* pp. 168–180.

A   Appendix

A.1   Values of the Ψ function

Note that (Abramowitz and Stegun 1965:6.3.1,6.3.2)

$$\int\limits_0^\infty e^{-u} \ln u \, du = \psi(1) = -\gamma \,, \tag{22}$$

where $\psi(z) = \frac{d}{dz} \ln \Gamma(z)$ denotes the digamma function and let $\psi^{(n)}(z) = \frac{d^n}{dz^n} \psi(z)$ denote the polygamma function (Abramowitz and Stegun 1965:6.4.1).

As a special case of $\psi^{(n)}(z)$, when $z = 1$ (Abramowitz and Stegun 1965:6.4.2),

$$\psi^{(n)}(1) = (-1)^{n+1} n! \zeta(n+1) \,. \tag{23}$$

From (9), (22) and (23),

$$\Psi_0 = \lim_{t \to 0} -\Gamma(1-t) = 1\,;$$

$$\Psi_1 = \lim_{t \to 0} -\Gamma(1-t)\psi(1-t) = -\psi(1) = \gamma\,;$$

$$\Psi_2 = \lim_{t \to 0} -\Gamma(1-t)\left[\psi(1-t)\right]^2 + \Gamma(1-t)\psi^{(1)}(1-t) = \gamma^2 + \zeta(2)\,;$$

$$\vdots$$

$$\Psi_n = \lim_{t \to 0} \frac{d^{n-1}}{dt^{n-1}}\left[-\Gamma(1-t)\psi(1-t)\right] = \lim_{t \to 0} \sum_{k=0}^{n-1} \binom{n-1}{k}\Gamma(1-t)^{(n-1-k)}\psi^{(n-1)}(1-t)$$

(24)

Any values of Ψ_n can be derived by using relationships (22), (23), $\frac{d}{dz}\Gamma(z) = \psi(z)\Gamma(z)$ and $\Gamma(1) = 1$.[2]

A.2   Central moments as Meijer-G functions

Meijer-G functions offer the most succinct way to represent the central moments of the Gompertz distribution. The nth order central moment, m_n is given as

$$m_n = \sum_{j=0}^{n} \binom{n}{j}(-1)^{n-j}E\left[X^j\right]E\left[X\right]^{n-j}$$

and from (7)

$$m_1 = \frac{1}{b}e^{\frac{a}{b}}G_{1,2}^{2,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1 \\ 0,0; \end{matrix}\right)$$

$$m_2 = \frac{2}{b^2}e^{\frac{a}{b}}G_{2,3}^{3,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1 \\ 0,0,0; \end{matrix}\right) - (m_1)^2$$

$$m_3 = \frac{6}{b^3}e^{\frac{a}{b}}G_{3,4}^{4,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1,1 \\ 0,0,0,0; \end{matrix}\right) - 3m_1\frac{2}{b^2}e^{\frac{a}{b}}G_{2,3}^{3,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1 \\ 0,0,0; \end{matrix}\right) + 2(m_1)^3$$

$$m_4 = \frac{24}{b^4}e^{\frac{a}{b}}G_{4,5}^{5,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1,1,1 \\ 0,0,0,0,0; \end{matrix}\right) - 4m_1\frac{6}{b^3}e^{\frac{a}{b}}G_{3,4}^{4,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1,1 \\ 0,0,0,0; \end{matrix}\right)$$
$$+ 6(m_1)^2\frac{2}{b^2}e^{\frac{a}{b}}G_{2,3}^{3,0}\left(\frac{a}{b}\,\middle|\,\begin{matrix} ;1,1 \\ 0,0,0; \end{matrix}\right) + 3(m_1)^4\,.$$

---

[2]For an alternative derivation (and also for alternative power series expansion of $E_1^n(z)$) please see the Appendix of Gussmann (1967), especially equations $(A.39) - (A.41)$.

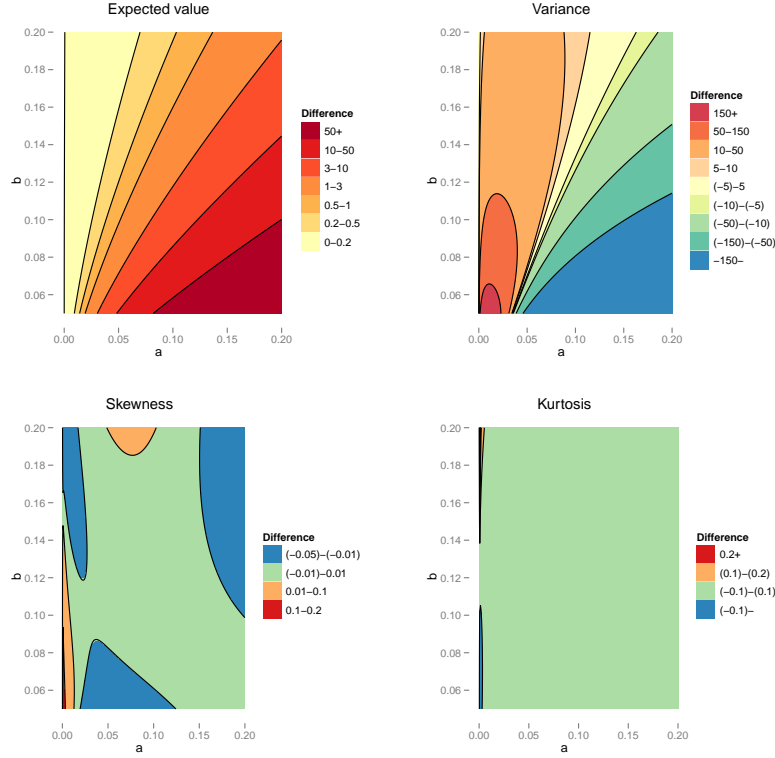## A.3 Error of the approximations

Figure 5: Approximation error for the expected value, variance, skewness and kurtosis of the Gompertz distribution

B Appendix

B.1 Maximum Likelihood Estimation

To solve the maximum likelihood estimators analytically, let S denote the scores of the likelihood function and be defined by

$$S = \frac{\partial l(\theta|D)}{\partial \theta} \, .$$

Maximizing the likelihood function implies that at the optimal parameters, $\hat{\theta}$ the system of equations defined by the score function is homogeneous,

$$S = 0 \, ,$$

and the Hessian matrix

$$H = \frac{\partial^2 l(\hat{\theta}|D)}{\partial \hat{\theta}^2}$$

18

is negative definite.

The s score functions, $i \ldots s$, in the Poisson likelihood case appear as

$$S_i = \sum_x D_x \frac{\frac{\partial f(\theta_i, x)}{\partial \theta_i}}{f(\theta_i, x)} - E_x \frac{\partial f(\theta_i, x)}{\partial \theta_i} \, , \tag{25}$$

where $f(\cdot)$ is some function of θ and x.

Substituting (1) in (25) yields[3]

$$S = \begin{pmatrix} \sum_x \frac{D_x}{a} - \sum_x E_x e^{bx} \\ \sum_x D_x x - \sum_x E_x x a e^{bx} \end{pmatrix}.$$

By setting

$$S = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

S_a can be solved for $\hat{a}$:

$$\hat{a} = \frac{\sum_x D_x}{\sum_x E_x e^{bx}}$$

Substituting $\hat{a}$ in S_b gives

$$S_b = \sum_x D_x x - \sum_x E_x \frac{\sum_x D_x}{\sum_x E_x e^{bx}} x e^{bx} \, ,$$

or

$$\frac{1}{\sum_x D_x} \sum_x D_x x = \frac{1}{\sum_x E_x e^{\hat{b}x}} \sum_x E_x e^{\hat{b}x} x \, .$$

---

[3]S will be an s × n (Jacobian) matrix