Analytic Expressions for Life Expectancy in Gamma-Gompertz Mortality Settings

Trifon Ivanov Missov

Max Planck Institute for Demographic Research

Abstract

In a population with Gamma-distributed individual frailty and Gompertz-distributed mortality there is a closed-form analytic expression for calculating the life-expectancy integral. Several simplifications of the resulting formula serve as plausible approximations.

Analytic Expression for Gamma-Gompertz Life Expectancy

Let $X \in [0, +\infty)$ be a lifespan with a probability density function $p(x)$, survival function $S(x)$, and hazard $\mu(x)$. Assume hidden individual-specific mortality peculiarities are captured by a heterogeneity parameter $Z \in [0, +\infty)$, known as frailty, with a probability density function $\pi(z)$. In line with [4], we index individuals' $p(x)$, $S(x)$, and $\mu(x)$ by Z to construct a proportional hazards model:

$$p(x, Z) = Z\, S^{Z-1}(x)p(x) \qquad S(x, Z) = S^Z(x) \qquad \mu(x, Z) = Z\,\mu(x) \quad (1)$$

As a result, the population's survival functionwill be

$$\bar{S}(x) = \int\limits_{0}^{+\infty} S^Z(x)\pi(z)dz \quad (2)$$

Suppose the individual lifetime distribution is Gompertz(a, b), i.e.

1

$$p(x) = ae^{bx}e^{\frac{a}{b}\left(1-e^{bx}\right)} \qquad S(x) = e^{\frac{a}{b}\left(1-e^{bx}\right)} \qquad \mu(x) = ae^{bx} \qquad (3)$$

and the frailty distribution is Gamma(k, λ), i.e.

$$\pi(z) = z^{k-1}\lambda^k \frac{e^{\lambda x}}{\Gamma(k)} \qquad (4)$$

Then, the population's survival function will be

$$\bar{S}(x) = \left(1 + \frac{a}{b\lambda}\left(e^{bx} - 1\right)\right)^{-k} \qquad (5)$$

Life expectancy at birth e_0 can be expressed by several equivalent formulae, among which we will consider the one that depends just on the population's survival function $\bar{S}(x)$, i.e.

$$e_0 = \int_0^{+\infty} \bar{S}(x)dx = \int_0^{+\infty} \left(1 + \frac{a}{b\lambda}\left(e^{bx} - 1\right)\right)^{-k} \qquad (6)$$

Then, life expectancy at birth e_0 is given explicitly by the following formula:

$$e_0 = \frac{1}{bk} \, {}_2F_1\left(k, 1; k+1; 1 - \frac{a}{b\lambda}\right) \qquad (7)$$

where ${}_2F_1(\alpha, \beta; \gamma; z)$ is the Gaussian hypergeometric function, i.e.

$${}_2F_1(\alpha, \beta; \gamma; z) = \sum_{j=0}^{+\infty} \frac{\alpha(\alpha+1)\ldots(\alpha-j+1)\,\beta(\beta+1)\ldots(\beta-j+1)}{\gamma(\gamma+1)\ldots(\gamma-j+1)\,j!} z^j \qquad (8)$$

which is defined for $\gamma > \beta > 0$ (see, for example, [1]).

Proof. We will prove an even more general result. Namely, we will show that the indefinite integral corresponding to the right-hand side of (6) equals:

$$-\frac{1}{bk}\left(\frac{b\lambda}{a}e^{-bx}\right)^k {}_2F_1\left(k, k; k+1; \left(1 - \frac{a}{b\lambda}\right)e^{-bx}\right) \qquad (9)$$

Thus, (7) will follow directly from (9) as

$$\lim_{x \to +\infty}\left\{-\frac{1}{bk}\left(\frac{b\lambda}{a}e^{-bx}\right)^k {}_2F_1\left(k, k; k+1; \left(1 - \frac{a}{b\lambda}\right)e^{-bx}\right)\right\} = 0 \qquad (10)$$

2

and, in addition, $_2F_1(\alpha, \beta; \gamma; z) = (1 - z)^{-\alpha}\,_2F_1\left(\alpha, \gamma - \beta; \gamma; \frac{z}{z-1}\right)$ (see [3], p.247). Let us represent e_0 in the following way:

$$e_0 = \int \left(1 + \frac{a}{b\lambda}\left(e^{bx} - 1\right)\right)^{-k} = \int \left(\frac{b\lambda}{a}e^{-bx}\right)^k \left(1 - \left(1 - \frac{b\lambda}{a}\right)e^{-bx}\right)^{-k} \tag{11}$$

A $y = e^{-bx}$ substitution will result in

$$e_0 = -\frac{1}{b}\left(\frac{b\lambda}{a}\right)^k \int y^{k-1}\left(1 - \left(1 - \frac{b\lambda}{a}\right)y\right)^{-k} \tag{12}$$

Taking into account

$$\left(1 - \left(1 - \frac{b\lambda}{a}\right)y\right)^{-k} = \,_2F_1\left(k, C; C; \left(1 - \frac{b\lambda}{a}\right)y\right) \qquad \forall C \equiv const \tag{13}$$

(see [3], p.258) and

$$_2F_1(\alpha, \beta; \gamma; z) = \,_2F_1(\beta, \alpha; \gamma; z), \tag{14}$$

which follows directly from (8), we have

$$\left(1 - \left(1 - \frac{b\lambda}{a}\right)y\right)^{-k} = \,_2F_1\left(k, k; k; \left(1 - \frac{b\lambda}{a}\right)y\right) \tag{15}$$

Finally, using

$$z^{\gamma-1}\,_2F_1(\alpha, \beta; \gamma; z)dz = z^{\gamma}\,_2F_1(\alpha, \beta; \gamma + 1; z)dz \tag{16}$$

(see [1]), we reduce (12) to (9). Q.E.D. $\qquad\square$

The analytical closed-form solution (7) of the life-expectancy integral has helpful properties and implications, as well as distinct drawbacks. The findings are classified into the following remarks.

Remark 1 (On dropping the maximum-age assumption):
In Demography it is often assumed there exists a highest, yet unknown, age ω to which one can live. In this case the life-expectancy integral simplifies as its upper boundary is no longer infinite. Following this argument, Keyfitz

and Caswell ([2], 2005) derived a number of ω-dependant relationships, which illustrate the sensitivity of e_0 to mortality changes. Expression (7), on the other hand, allows for extending these (and other e_0-related) findings to the case when there is no upper boundary for life expectancy. The resulting formulae will no longer contain the unknown ω.

Remark 2 (On dropping the assumption of Gamma frailty with *unit* expectation):

Gamma(k, λ) frailty models usually, for simplicity of interpretation, assume unit expectation, i.e. $k = \lambda$. Formula (7) provides an expression for life expectancy at birth for any Gamma-distribution parameters k and λ, not necessarily equal to each other.

Remark 3 (Simplification in the special case of integer k):

The case when k is an integer ($k \in \mathbb{Z}$) deserves special attention for at least two reasons. First of all, the derivation of e_0 in this case is easier as the right-hand side of (6) can be integrated k times by parts. Thus, the expression for life expectancy at birth takes the following form:

$$e_0 = \frac{1}{b}\left[\left(1 - \frac{a}{b\lambda}\right)^{-k}\ln\frac{b\lambda}{a} - \sum_{j=1}^{k-1}\frac{1}{j}\left(1 - \frac{a}{b\lambda}\right)^{j-k}\right] \qquad (17)$$

Secondly, if the $k \in \mathbb{Z}$ assumption is supported by a certain mortality dataset, we might use the simpler, in terms of calculation, expression (17) as an approximation of (7).

Remark 4 (On some properties of the hypergeometric function):

Expression (7) deserves three additional technical remarks regarding the hypergeometric function $_2F_1(\alpha, \beta; \gamma; z)$ appearing on the right-hand side. First of all, it is finite as the sufficient condition $\gamma > \beta > 0$ ([3], pp.239-240) is fulfilled. Indeed, in our case we have $\gamma = k + 1 > 1 = \beta > 0$. Secondly, if k is integer, expression (17) follows directly from (7) by reorganizing the hypergeometric series appropriately. Namely, as the β-parameter of $_2F_1$ equals 1, (8) simplifies to

$$_2F_1\left(k, 1; k+1; 1 - \frac{a}{b\lambda}\right) = 1 + k\left(1 - \frac{a}{b\lambda}\right)^{-k}\sum_{j=1}^{\infty}\frac{1}{k+j}\left(1 - \frac{a}{b\lambda}\right)^{k+j} \qquad (18)$$

Furthermore, as $k \in \mathbb{Z}$, we can represent e_0, as it follows:

$$e_0 = \frac{1}{bk}\left(1 + k\left(1 - \frac{a}{b\lambda}\right)^{-k}\sum_{j=1}^{\infty}\frac{1}{j}\left(1 - \frac{a}{b\lambda}\right)^{j} - k\sum_{j=1}^{k}\frac{1}{j}\left(1 - \frac{a}{b\lambda}\right)^{j-k}\right) \quad (19)$$

The series in the parentheses constitute the Taylor-series representation of the function $\ln\frac{b\lambda}{a}$, and the term in the negative sum corresponding to $j = k$ equals 1. As a result, (19) is equivalent to (17). Finally, for human populations the z-argument of the hypergeometric function approaches 1. In this case the Gauss's theorem postulates that

$$\lim_{z \to 1-} {}_2F_1(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} \quad (20)$$

given $\gamma - \alpha - \beta > 0$. However, the latter necessary condition is not fulfilled for ${}_2F_1\left(k, 1; k + 1; 1 - \frac{a}{b\lambda}\right)$ as $\gamma - \alpha - \beta = k + 1 - k - 1 = 0$. Thus, formula (20) cannot serve as approximation of (7). As a result, although formula (7) represents the solution of the life-expectancy integral in a closed analytical form, it contains a term, ${}_2F_1\left(k, 1; k + 1; 1 - \frac{a}{b\lambda}\right)$, that can be evaluated only numerically. In general, the hypergeometric function is a complex object to work with in practical applications. That is why it would be reasonable to study (17), or some other approximation of (7), which would be easier for computation, on the one hand, and assessment of the Gamma-Gompertz parameters' impact on e_0, on the other. The latter question will be discussed in detail in the following section.

Remark 5 (On extending the Gompertz mortality model to Gompertz-Makeham):

When the mortality pattern is Gompertz-Makeham with a $\mu(x) = ae^{bx} + c$ hazard, the corresponding life-expectancy integral

$$e_0 = \int_0^{+\infty}\left(1 - \frac{a}{b\lambda} + \frac{c}{\lambda}x + \frac{a}{b\lambda}e^{bx}\right)^{-k} dx \quad (21)$$

cannot be solved analytically, even for integer values of k. As a result, life expectancy in mortality models with a significant non-zero Makeham term c can be approximated only numerically.

Plausible Approximations of Gamma-Gompertz Life Expectancy

Roland Rau studied the Gamma-Gompertz model for Swedish female mortality data on a both period and cohort basis for different decades. He focused on ages 80+ and assumed a Gamma frailty distribution with unit expectation, i.e. $k = \lambda$. The results of this study were first presented by James W. Vaupel during his Johann Süßmilch Lecture "Supercentenarians and the Theory of Heterogeneity" on October 21, 2008 at the Max Planck Institute for Demographic Research. Table 1 presents the maximum likelihood estimates for the Gompertz parameters a and b, as well as the Gamma parameter k for the indicated data types, decades of interest, and ages.

Case	Type	Decades	Ages	a	b	$k = k$ (est.)
1	Cohort	1780-1790	80-105	$1.44 \cdot 10^{-6}$	0.147	4.71
2	Cohort	1880-1890	80-111	$3.22 \cdot 10^{-6}$	0.129	5.46
3	Period	1895-1905	80-104	$3.00 \cdot 10^{-7}$	0.163	4.30
4	Period	1995-2005	80-111	$4.70 \cdot 10^{-7}$	0.143	4.90

Table 1: Gamma-Gompertz maximum-likelihood parameter estimates for Swedish female mortality data (Data source: Human Mortality Database, www.mortality.org; estimation performed by Roland Rau).

Table 2 presents the life expectancies at birth corresponding to the Gamma-Gompertz parameter estimates in these four cases. In addition, assuming $\lambda = k$(est.), it compares e_0 values resulting from formulae (7) and (17). It turns out that using (17) instead of (7) is justified as the relative estimation error is in the 0.21%–0.46% range.

The results would not differ significantly even if we take in (17) $\lambda = k$(est.) and $k = k$(rounded) (see Table 2). The relative errors in this approximation are bigger than the ones for $\lambda = k = k$(rounded) for cohort data, and vice versa for period data. Further approximation accuracy can be attained by taking the respective weighted average of e_0 values, calculated for $\lfloor k \rfloor$ and $\lceil k \rceil$. For instance, taking this average by using (17) for period data reduces the relative errors to less than 0.1%. Note that for human populations the ratio $1 - a/b\lambda$ is very close to 1 as $a/b\lambda \propto 10^{-6}$. As a result, if we further assume $1 - a/b\lambda \approx 1$, we can simplify (17) even more. Namely,

Case	e_0 by(7)	k (rd.)	e_0 by(17)	$\frac{e_0\text{rd.}}{e_0\text{real}}$	k(rd.) $\lambda = k$(est.) $e_0(\lambda)$by(17)	$\frac{e_0(\lambda)}{e_0\text{real}}$
1	75.22	5	75.24	1.000279	74.83	0.994876
2	78.37	5	78.48	1.001498	79.17	1.010202
3	78.63	4	78.27	0.995452	78.72	1.001095
4	84.69	5	84.98	1.003344	84.84	1.001676

Table 2: Life expectancy e_0 values, based on formulae (7) and (17), in the four cases presented in Table 1. When applying formula (17), we round the estimated k to its closest integer value.

$$e_0 \approx \frac{1}{b}\left[\ln\frac{b\lambda}{a} - \sum_{j=1}^{k-1}\frac{1}{j}\right] \qquad (22)$$

Life-expectancy values, calculated by (17) and (22) will differ just after the fifth decimal (as $a/b\lambda \propto 10^{-6}$). Note that formula (22) not only provides plausible approximation for (7), but also allows for easier interpretation of the Gamma-Gompertz parameters' impact on e_0. For instance, it turns out that b is more important than a. Indeed, life-expectancy is inversely proportional to b, whereas the effect of changes in a on e_0, no matter how small a's values are, is weakened on the logarithmic scale.

Finally, it is not surprising that the life-expectancy estimates in Table 2 exceed the corresponding values in the Human Mortality Database as Gamma-Gompertz parameters (a, b, and $k = \lambda$) were estimated for ages 80+.

Acknowledgements

References

[1] W. N. Bailey. *Generalised Hypergeometric Series*. University Press, Cambridge, England, 1935.

[2] N. Keyfitz and H. Caswell. *Applied Mathematical Demography*. Statistics for Biology and Health. Springer, 3rd ed. edition, 2005.

[3] N.N. Lebedev. *Special Functions and Their Applications*. Prentice-Hall, Englewood Cliffs, N.J., 1965. Translated from the revised Russian edition (Moscow, 1963) by Richard A. Silverman.

[4] J.W. Vaupel, K.G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454, 1979.