Goodness-of-fit tests for the Gompertz distribution

Adam Lenart & Trifon I. Missov

Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1
18057 Rostock, Germany

Abstract

While the Gompertz distribution is often fitted to lifespan data, testing whether the fit satisfies theoretical criteria is being neglected. Here six goodness-of-fit measures – the Anderson-Darling statistic, the Kullback-Leibler discrimination information, the correlation coefficient test, a statistic using moments, testing for the mean of the sample hazard, and a nested test against the generalized extreme value distributions, are discussed. Along with an application to laboratory rat data, critical values calculated by the empirical distribution of the test statistics are also presented.

1 Introduction

Goodness-of-fit tests determine if the empirical distribution of the data satisfies the assumptions of theoretical distributions. While the Gompertz distribution is routinely used in demography, biology, actuarial and medical science, according to our best knowledge, no studies about on goodness-of-fit tests for it have been published so far. However, the Gompertz distribution is a degenerate generalized extreme value distribution for the minima, and an abundance of goodness-of-fit tests exist in the literature for other extreme value distributions (see e.g., Hosking 1984).

In a landmark paper Shannon (1948) defined the entropy of distributions and Kullback and Leibler (1951) were the first to measure the distance between probability distributions based on their entropy. Later, Song (2002) operationalized the Kullback-Leibler distance to test the goodness-of-fit of distributions. Recently, Pérez-Rodríguez et al. (2009) applied it to the Gumbel distribution.

In another important article, Anderson and Darling (1952) developed the Anderson-Darling test that later Stephens (1977) analyzed in the context of extreme value distributions. Sinclair et al. (1990) modified the Anderson-Darling test to allow different weighting schemes that emphasize either the lower or the upper tail of the distributions.

Filliben (1975) used the Pearson correlation coefficient to check the correlation between expected statistics of a theoretical distribution and sample statistics. The correlation coefficient test was the most popular in hydrology (Vogel 1986; Kinnison 1989) to assess the fit of extreme value distributions.

The likelihood ratio test naturally arises to account for the differences between the Gompertz and other extreme value distributions. The generalized extreme value distribution is characterized by μ, location, σ, scale and ξ shape parameters. For $\xi = 0$, the generalized extreme value distribution reduces to the Gumbel, and the Gompertz distribution is a reversed and truncated Gumbel distribution with additional correlation between its parameters a and b. The different parametrization of the Gompertz distribution removes it from location-scale family of distributions.

Li and Papadopoulos (2002) proposed a goodness of fit test using moments. The test statistic is derived from an identity for the moments, and its values are compared to the z-values of the standard normal distribution.

This paper will first briefly describe each of these tests and apply them to the Gompertz distribution. Additionally, based on the observation that the maximum likelihood estimators of a and b equal the mean of the sample hazard in the Gompertz distribution (Lenart 2012), a fifth test will be defined. The final sections of the paper compare the power of the tests against alternative distributions and derive critical values of them based on Monte Carlo simulation experiments. An application of the tests to laboratory rat data is also discussed.

2 Kullback-Leibler information

Let the entropy (Shannon 1948) be defined as

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log fx \, dx \tag{1}$$

for continuous random variable X with probability density function $f(x)$, $x \in \mathbb{R}$. Entropy is a measure of information (e.g., Brissaud 2005) that shows the difficulty of predicting the outcome of a random draw from a probability distribution. Kullback and Leibler (1951) realized that probability distributions can be compared based on their entropy; the closer the values of their entropies are, the more the distributions resemble to each other.

The Kullback-Leibler distance measures the information for discrimination between two probability distributions. Let distribution F have density function $f(x)$ and $G(x, \theta)$ be a parametric family of distributions that have density $g(x; \theta)$, then the Kullback-Leibler distance

$$I(F,G;\theta) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x;\theta)}\, dx\,, \qquad I(F,G;\theta) \geq 0\,, \tag{2}$$

will measure the relative entropy of distribution F to G.

2.1 Goodness-of-fit test based on the Kullback-Leibler information

The definition of the Kullback-Leibler distance allows to develop a goodness-of-fit test for a parametric distribution (Song 2002):

$$H_0 \,:\, F(x) = G(x;\theta)\,.$$

Under the null hypothesis, $I(F,G;\theta) = 0$ and if $I(F,G;\theta) > 0$ then the alternative, $F(x) \neq G(x;\theta)$ is true. The derivation of the Kullback-Leibler discrimination information for the goodness-of-fit of the Gompertz distribution will follow the goodness-of-fit procedure implemented by Song (2002) and used by Pérez-Rodríguez et al. (2009) to test the Gumbel distribution.

Note that (1) can be substituted in (2) as

$$I(F,G;\theta) = -H(F) - \int_{-\infty}^{\infty} f(x) \log g(x,\theta)\, dx\,.$$

To estimate $H(F)$, Song (2002) advises to introduce the Vasicek (1976) entropy estimator

$$H_{mn} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{n}{2m} \left(X_{(i+m)} - X_{(i-m)} \right)\,, \qquad \frac{n}{2} - 1 \geq m \geq 1\,,$$

where $\left\{X_{(1)}, \ldots, X_{(n)}\right\}$ are n independent, ordered observations. The width of the estimation window is of size m, if $i - m < 1$, then $X_{(i-m)} = X_{(1)}$ and if $i + m > n$, then $X_{(i+m)} = X_{(n)}$.

Next, $\int_{-\infty}^{\infty} f(x) \log g(x,\theta)\, dx$ has to be estimated. Song (2002) proposes to use

$$\int_{-\infty}^{\infty} f(x) \log g(x,\theta)\, dx = \frac{1}{n} \sum_{i=1}^{n} \log g\left(X_i, \hat{\theta}\right)\,,$$

where $\hat{\theta}$ is a vector of maximum likelihood estimators.

Song (2002) advocates to choose the optimal window width, $\hat{m}$ by maximizing the sample entropy with the constraint of $I(F,G;\theta) \geq 0$:

$$\hat{m} := \min m^* :$$
$$m^* = \arg\max_m H_{mn}$$
$$\text{s.t.} \quad -H_{mn} - \frac{1}{n}\sum_{i=1}^{n} \log g\left(X_i, \hat{\theta}\right) \geq 0\,,$$

that is, choose the minimal m^* that maximizes H_{mn}.

Adding all of the pieces together, the test statistic will be

$$I_{\hat{m}n} = -H_{\hat{m}n} - \frac{1}{n}\sum_{i=1}^{n} \log g\left(X_i, \hat{\theta}\right)\,.$$

Large values of $I(F,G;\theta)$ support the alternative hypothesis. As $I_{\hat{m}n}$ is a sample estimate of $I(F,G;\theta)$, H_0 will be rejected if $I_{\hat{m}n}$ is larger than a critical value of it.

2.1.1 Asymptotic distribution

To calculate the asymptotic distribution of the Kullback-Leibler statistic, let

$$\phi(G,\theta) = \sup\{x \,:\, G(x,\theta) = 0\}$$
$$\psi(G,\theta) = \inf\{x \,:\, G(x,\theta) = 1\}\,.$$

and if the assumptions of

$$\sup_{\phi(G,\theta)<x<\psi(G,\theta)} G(x\theta)(1-G(x,\theta))\frac{|\frac{\partial g(x,\theta)}{\partial x}|}{g^2(x,\theta)} < \infty$$
$$\lim_{n\to\infty} \frac{m}{\log n} \to \infty$$
$$\lim_{n\to\infty} \frac{m(\log n)^{\frac{2}{3}}}{n^{\frac{1}{3}}} \to 0$$

hold then the standardized test statistic

$$(6\hat{m}n)^{\frac{1}{2}}\left(I_{\hat{m}n} - \log 2\hat{m} - \gamma + R_{2\hat{m}-1}\right) \to_d N(0,1)\,, \tag{3}$$

where

$$R_m = \sum_{j=1}^{m} \frac{1}{j}$$

and $\gamma \approx 0.57722$ is the Euler-Mascheroni constant converges in distribution to $N(0,1)$ (Song 2002: Theorem 1).

Note that R_m is a partial sum of the harmonic series, so $R_m = \log m + \gamma + O(\frac{1}{2m})$. If for large m, $\log(2m) \approx \log(2m-1)$ and $\frac{1}{2m} \approx 0$, then

$$6(\hat{m}n)^{\frac{1}{2}} I_{\hat{m}n} \to_d N(0,1)\,.$$

2.2 Small sample bias

$I_{\hat{m}m}$ is a biased estimator of $I(F,G;\theta)$ for small samples (Song 2002) and the bias can be corrected by changing (3) to

$$(6\hat{m}n)^{\frac{1}{2}} \left(I_{\hat{m}n} - \log 2\hat{m} - \log n + R_n - R_{2\hat{m}-1} + \frac{2\hat{m}}{n} R_{2\hat{m}-1} - \frac{1}{2n} \sum_{i=1}^{\hat{m}} R_{i+\hat{m}-2} \right) \to_d N(0,1)\,.$$

2.3 Goodness-of-fit of the Gompertz distribution

In the case of the Gompertz distribution,

$$g(x;a,b) = a e^{-\frac{a}{b}\left(e^{bx}-1\right)+bx}$$

or by noting that the mode, $M = \frac{1}{b} \log \frac{b}{a}$, hence $a = be^{-bM}$, then

$$g(x;b,M) = b e^{e^{-bM} - e^{b(x-M)} + b(x-M)}\,.$$

In general, the form that includes the mode will be preferred because Monte Carlo experiments showed that the test statistics are scale-free in this case (Fig 2) while in the parametrization with $\theta = \{a,b\}$ the distribution of the test statistics are affected either by a or b (Monte Carlo experiments for the latter case are not shown here).

The Kullback-Leibler discrimination information statistic for the Gompertz distribution is

$$I_{\hat{m}n} = e^{-\hat{b}\hat{M}} + \hat{b}\left(\bar{X} - \hat{M}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(e^{\hat{b}\left(X_i - \hat{M}\right)} - \log \frac{n}{2\hat{m}} \frac{X_{(i+\hat{m})} - X_{(i-m)}}{\hat{b}} \right)\,,$$

where $\bar{X}$ denotes the mean of $\{X_i, \ldots, X_n\}$.

If G is the Gompertz distribution, then $\phi(G,\theta) = -\infty$ and $\psi(G,\theta) = \infty$ and

$$\sup_{-\infty<x<\infty} G(x\theta)(1 - G(x,\theta)) \frac{|\frac{\partial g(x,\theta)}{\partial x}|}{g^2(x,\theta)} = 1$$

for any suitable m, this statistic converges in distribution to the standard normal. Therefore,

$$I_{\hat{m}n} \geq \log 2\hat{m} + \gamma - R_{2\hat{m}-1} + (6\hat{m}n)^{-\frac{1}{2}} Z_{1-\alpha}$$

tests H_0 at level α asymptotically. $Z_{1-\alpha}$ denotes the $100(1-\alpha)$ percentage point of the standard normal distribution. Large values of $I_{\hat{m}n}$ reject the null hypothesis. For small samples either the bias has to be corrected or the critical values of $I(F,G;\theta)$ has to be estimated by Monte Carlo simulations from $I_{\hat{m}n}$. For empirical critical values, please see Table 5.

3 Correlation coefficient test

Filliben (1975) introduced the probability plot coefficient test for normal distributions. The idea of the test is to compare the ordered observations with predicted order statistics of a theoretical distribution. Let $X_{[i]}$ denote the ith largest observed datum ,$\widetilde{X}_{[i]}$ the order statistic median, $\bar{X}$ the average observation and $\widetilde{X}$ the population median, then the probability plot correlation coefficient is given by the Pearson correlation coefficient

$$r = \frac{\sum_{i=1}^{n} \left(X_{[i]} - \bar{X}\right) \left(\widetilde{X}_{[i]} - \widetilde{X}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_{[i]} - \bar{X}\right)^2 \sum_{i=1}^{n} \left(\widetilde{X}_{[i]} - \widetilde{X}\right)^2}} .$$

Filliben (1975) estimated the order statistic medians from the quantile function and later the same approach was used for the Gumbel and other extreme-value distributions (Vogel 1986; Kinnison 1989). These approaches relied on numerical approximations to the plotting positions between the order statistics and the order statistic medians or other measures of location[1] such as the plotting position of Gringorten (1963) which is unbiased only for the largest observation.

The correlation coefficient test can be improved by comparing the ordered observations with their expected values of a distribution. Let $X_{(i)}$ denote the i^{th} smallest observation, $E\left[X_{(i)}\right]$ the expectation of it and $E\left[X\right]$ the expected value of the theoretical population.

3.1 Density and expected value of order statistics

The density of $f_{(i)}(x)$ is (see e.g., Harter 1961)

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} F^{i-1}(x) \left(1 - F(x)\right)^{n-i} f(x)$$

and

$$E\left[X_{(i)}\right] = \int_{-\infty}^{\infty} f_{(i)}(x)\, dx$$

The density of $f_{(i)}(x)$ can be simplified by

$$X_{(i)} =_d F^{-1}\left(U_{(i)}\right) , \tag{4}$$

[1]Plotting $X_{[i]}$ against $M_{[i]}$ yields and approximately linear plot.

where $U \sim U(0,1)$ and F^{-1} is the quantile function of X. Because[2]

$$f_{U_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1} (1-x)^{n-i} \,, \qquad x \in [0,1]$$

the expected value of $E\left[X_{(i)}\right]$ can be reformulated (Sen 1959) as

$$E\left[X_{(i)}\right] = \frac{n!}{(i-1)!(n-i)!} \int_0^1 F^{-1}(x) x^{i-1} (1-x)^{n-i} \, dx \,.$$

3.2 Correlation coefficient test for the Gompertz distribution

The correlation coefficient test has the null hypothesis

$$H_0 \,:\, F(x) = G(x;\theta) \,.$$

If $X \sim Gompertz(a,b)$, then

$$F^{-1}(x) = \frac{1}{b} \log\left(1 - \frac{b}{a}\log(1-x)\right) \,, \qquad a > 0, b \geq 0 \,,$$

and

$$E\left[X_{(i)}\right] = \frac{n!}{b(i-1)!(n-i)!} \int_0^1 \log\left(1 - \frac{b}{a}\log(1-x)\right) x^{i-1}(1-x)^{n-i} \, dx \,.$$

The expected value of the population is (Missov and Lenart 2011)

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

where $E_n(z) = \int_1^\infty \exp(-zt)/t^n \, dt$ denotes the exponential integral (Abramowitz and Stegun 1965:5.1.4).

The estimated correlation coefficient is then

$$\hat{r}\left(\hat{\theta}\right) = \frac{\sum_{i=1}^{n} \left(X_{(i)} - \bar{X}\right)\left(\hat{E}\left[X_{(i)};\hat{\theta}\right] - \hat{E}\left[X;\hat{\theta}\right]\right)}{\sqrt{\sum_{i=1}^{n} \left(X_{(i)} - \bar{X}\right)^2 \sum_{i=1}^{n} \left(\hat{E}\left[X_{(i)};\hat{\theta}\right] - \hat{E}\left[X;\hat{\theta}\right]\right)^2}} \,,$$

where $\hat{\theta}$ is the maximum likelihood estimate of $\theta = (a,b)$. The test statistic ranges from $[0,1]$ and the null hypothesis is rejected if $\hat{r}$ is lower than a critical value estimated by Monte Carlo simulations (Table 6). For the distribution of the correlation coefficients, please see Fig. 3.

[2]Note that the distribution function, $F_{U_{(i)}}(x)$ of the i^{th} observation of a uniform distribution would be equal to the regularized incomplete beta function, $I_x(i, n-i+1)$ (Abramowitz and Stegun 1965:26.5).

4 Anderson-Darling test

The Anderson-Darling (1952) test is based on the difference between the empirical and the theoretical distribution function $F(x)$ and $G(x)$,

$$W^2 = n \int_{-\infty}^{\infty} [F(x) - G(x)]^2 \, \psi(x) \, dG(x) \, ,$$

where $\psi(x)$ is a weight function. As Anderson and Darling (1952: p. 194) notes, for $\psi(x) := 1$ W^2 will be the same as the Cramér-von Mises test statistic

$$T = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ \frac{2i-1}{2n} - G\left[X_{(i)}\right] \right\}^2 \, ,$$

where $X_{(i)}$ is the i^{th} smallest observation (Stephens 1974). Other weight functions are also used to test the goodness-of-fit of extreme value distributions (e.g. Stephens 1977), most notably $\psi(x) := \{G(x)\,[1 - G(x)]\}^{-1}$ that gives the Anderson-Darling test statistic (Shin et al. 2011)

$$\begin{aligned} A^2 &= n \int_{-\infty}^{\infty} \frac{[F(x) - G(x)]^2}{G(x)\,[1 - G(x)]} \, dG(x) \\ &= -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left\{ \log G\left(X_{(i)}\right) + \log\left[1 - G\left(X_{(n-i+1)}\right)\right] \right\} \, . \end{aligned} \tag{5}$$

4.1 Extensions of the Anderson-Darling test

For testing the mortality of heterogeneous populations, the modified Anderson-Darling test statistic (Sinclair et al. 1990) is of interest. It attributes a different weight function for the upper and the lower tail

$$\begin{aligned} AU^2 &= n \int_{-\infty}^{\infty} \frac{[F(x) - G(x)]^2}{1 - G(x)} \, dG(x) \\ &= \frac{n}{2} - 2 \sum_{i=1}^{n} G\left(X_{(i)}\right) - \sum_{i=1}^{n} \left(2 - \frac{2i-1}{n}\right) \log\left[1 - G\left(X_{(i)}\right)\right] \end{aligned} \tag{6}$$

and

$$\begin{aligned} AL^2 &= n \int_{-\infty}^{\infty} \frac{[F(x) - G(x)]^2}{G(x)} \, dG(x) \\ &= -\frac{3n}{2} + 2 \sum_{i=1}^{n} G\left(X_{(i)}\right) - \sum_{i=1}^{n} \frac{2i-1}{n} \log G\left(X_{(i)}\right) \end{aligned} \tag{7}$$

respectively. In a model where individuals have different levels of frailty (Vaupel et al. 1979) that acts multiplicatively on their baseline level of mortality, there would be more robust individuals (lower level of frailty) that would deviate in the upper tail from the homogeneous (all individuals having the same frailty) distribution.

4.2 Anderson-Darling test for the Gompertz distribution

As previously, the null hypothesis of the Anderson-Darling test is

$$H_0 \,:\, F(x) = G(x;\theta)\,.$$

In case of the Gompertz distribution, $\theta = (a, b)$. By substituting

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

in either (5), (6) or (7), the Anderson-Darling test statistic is immediately given. Large values of the statistic reject the null hypothesis. The critical values are defined by Monte Carlo simulations (Table 7).

5 Moments test for the Gompertz distribution

An interesting, yet not very popular, goodness-of-fit test using moments was suggested by Li and Papadopoulos (2002). Suppose $X_1, \ldots, X_n$ are i.i.d. random variables characterized by a c.d.f. $F(x)$. We test a null hypothesis

$$H_0 : \text{F belongs to a parametric family } F_\theta,\ \theta \in \Theta$$

Suppose the k-th ($k \in \mathbb{N}$) moment $m_k = \int x^k dF_\theta(x)$ of F_θ exists and

$$g(m_1, \ldots, m_k) = 0 \qquad \forall \theta \in \Theta$$

for some function g. Then

$$\sqrt{n}\, g(\hat{m}_1, \ldots, \hat{m}_k) \to_d N(0, V(\theta))$$

$\hat{m}_i = \sum\limits_{i=j}^{n} X_j^i/n$ denotes the sample moment of order i ($i = 1, \ldots, k$) and

$$V(\theta) = \nabla g(m_1, \ldots, m_k)^T \; \Sigma \; \nabla g(m_1, \ldots, m_k),$$

where $\Sigma = ||\sigma_{ij}||_{i,j=1}^{k}$ has elements $\sigma_{ij} = m_{i+j} - m_i m_j$ and $\nabla g(m_1, \ldots, m_k)$ denotes the gradient of g. We can choose $g(x, y, z) = z - 3xy + x^3$ and construct the following statistic:

$$T = \frac{\sqrt{n}\,(\hat{m}_3 - 3\hat{m}_1\hat{m}_2 + 2\hat{m}_1^3)}{\sqrt{V(\hat{a}, \hat{b})}} \sim N(0, 1)$$

We use the expressions for m_1, m_2, and m_3 calculated in Lenart (2012).

6 Test for the mean of the sample hazard

Let $\mu(x)$ be a non-parametric estimator of the hazard (e.g., Müller and Wang 1994) at age x. Measure $\mu(X)$ the sample hazard where $X = (X_1, \ldots, X_n)$ are the ages at death. Lenart (2012) showed that $\hat{a} + \hat{b} = \bar{\mu}\,(X)$ in the case of the Gompertz distribution, where $\hat{a}$ and $\hat{b}$ are the maximum likelihood estimators of the Gompertz parameters and $\bar{\mu}\,(X)$ is the mean of the hazard. In a discrete setting, $\bar{\mu}\,(X)$ would be approximated by the age-specific hazards weighted by the number of events in the intervals between the discretization steps.

$$H_0 \; : \; \hat{a} + \hat{b} = \bar{\mu}\,(X)$$

To test the null hypothesis, we will use that

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \rightarrow_d N\left(0, I(\hat{\theta})^{-1}\right) ,$$

where $\theta = (a, b)$ and $I(\hat{\theta})$ is the (observed) Fisher information matrix

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} \log g(X; \theta)$$

evaluated at the maximum likelihood estimates. The sum of two correlated normally distributed random variables

$$\begin{aligned} Y &\sim N(\hat{a}, \sigma_Y^2) \\ Z &\sim N(\hat{b}, \sigma_Z^2) \\ Y + Z &\sim N(\hat{a} + \hat{b}, \sigma_Y^2 + \sigma_Z^2 + 2Cov(Y, Z)) \end{aligned}$$

is a normally distributed random variable. The inverse of the Fisher information matrix yields the variance-covariance matrix (see Appendix A for the Gompertz distribution). According to the central limit theorem, when the sample size approaches infinity

$$\bar{M} - \bar{\mu}(X) \rightarrow_d N(0, \sigma_{\bar{\mu}}^2)\,,$$

where $\bar{M}$ corresponds to the population mean hazard and $\sigma^2_{\bar{\mu}} = s^2/n$ and s^2 is the variance of the sample. For finite size samples, consult the skewness of the distribution (please see (Lenart 2012) for the Gompertz distribution) to decide whether the sample means are normally distributed at sample size n. If

$$V \sim N(\bar{\mu}(X), \sigma_V^2)$$

and the sample mean is independent from the maximum likelihood estimators, then

$$Y + Z - V \sim N(\hat{a} + \hat{b} - \bar{\mu}(X), \sigma_Y^2 + \sigma_Z^2 + 2Cov(Y, Z) + \sigma_V^2)\,. \tag{8}$$

Let $F^{-1}(x; N_{YZV})$ be the quantile function of the normal distribution with mean and variance equal to the one in (8), then according to H_0 at level α

$$F^{-1}\left(\frac{\alpha}{2}; N_{YZV}\right) \leq 0 \leq F^{-1}\left(1 - \frac{\alpha}{2}; N_{YZV}\right)\,.$$

7 Nested test against the generalized extreme value distribution

Let

$$f_{GEV}(x; m, \xi, \sigma) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-m}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)-1} e^{-\left[1+\xi\left(\frac{x-m}{\sigma}\right)\right]^{-\frac{1}{\xi}}} \qquad m, \xi \in \mathbb{R},\ \sigma > 0 \tag{9}$$

be the density function of the generalized extreme value distribution for the maxima, where m is the location, ξ is the shape and σ is the scale parameter. As Willekens (2002) noted, the Gompertz is a special case of the generalized extreme value distribution for the minima, that is, when $-x$ is substituted for x in (9). Moreover, let $f_{GEV}(x)$ be truncated at $x = 0$ (Elandt-Johnson 1976), $b := 1/\sigma$ and $m := 1/b\log(a/b)$:

$$f_{GEV}(x; a, b, \xi) = b\left[1 - \xi\left(bx - \log\frac{a}{b}\right)\right]^{-1-\frac{1}{\xi}} e^{\left(1-\xi\log\frac{a}{b}\right)^{-\frac{1}{\xi}} - \left[1-\xi\left(bx-\log\frac{a}{b}\right)\right]^{-\frac{1}{\xi}}}$$

then for $\xi \to 0$, $f_{GEV}(x; a, b, \xi)$ goes to the Gompertz distribution. To test whether the Gompertz distribution fits the data as well as the generalized extreme value distribution,

$$H_0 : \xi = 0$$

a likelihood ratio test is employed

$$-2\log\frac{L\left(g(x; \hat{a}, \hat{b})\right)}{L\left(f_{GEV}(x; \hat{a}, \hat{b}, \hat{\xi})\right)} \sim \chi^2(1)\,,$$

where $L(\cdot)$ denotes the likelihood function and $g(\cdot)$ the Gompertz distribution. The likelihood ratio is evaluated at the maximum likelihood estimates of the two log-likelihood functions and by Wilks (1938) the limiting distribution of the likelihood ratio test statistic is the χ^2 distribution with degrees of freedom equal to the number of constraints under the null hypothesis. Please see Fig. 4 for the distribution of the Anderson-Darling statistic.

Note that with $\xi = 0$, the generalized extreme value reduces to Type I extreme value, or Gumbel distributions. The Gompertz distribution is a truncated Gumbel distribution for the minima with additionally introduced correlation between the parameters. For detailed analysis of the tests of the Gumbel distribution, please see Hosking (1984).

8 Power of the tests

To compare the tests, $n = 50$ and $n = 200$ samples were simulated from alternative distributions repeated 50000 times each. These alternatives distributions were Weibull, as it is often used in survival analysis or reliability engineering, the Log-normal as another asymmetric distribution, the Normal distribution as the distribution of life times were often assumed to follow a normal distribution (Véron and Rohrbasser (2003) citing Wilhelm Lexis), the logistic distribution as it is observed as the hazard function in many biological studies (Wilson 1994), the Gamma distribution because of its flexible shape and lastly, the Gamma-Gompertz distribution (Vaupel et al. 1979) for a more complicated model in which the Gompertz distribution is nested. The simulating parameters were attained by fitting each of the alternative distributions to a random sample of a Gompertz distribution. Note that the normal and the logistic distributions are the only symmetric distributions among the alternatives.

	$\bar{\mu}$	r	KL	AD	M	LR
Weibull(10,80)	0.9664	0.0497	0.0653	0.0915	0.0974	0.0982
Log-normal(4.4,0.01)	0.8293	0.4793	0.3461	0.7822	0.3124	0.3965
Normal(80,10)	0.8641	0.5147	0.3015	0.5483	0.3791	0.4014
Logistic(80,5)	0.7225	0.4895	0.4602	0.6837	0.4101	0.4534
Gamma(71,1.1)	0.8131	0.6863	0.4391	0.7356	0.2710	0.5747
Gamma-Gompertz(0.001,0.1,0.2)	0.7049	0.0130	0.0670	0.0854	0.0869	0.0955

Table 1: Power of the goodness-of-fit statistics against alternative distributions with $n = 50$, $\alpha = 0.05$.

As Table 1 shows, the test for the mean rejected most of the alternative distributions for sample size 50. However, at such a low sample size, it is difficult to get a reliable estimate of the sample hazard and the test rejects also a similar proportion of simulated samples from Gompertz distributions. Monte Carlo simulations (50000 iterations) show that for sample sizes around 200, the probability of committing Type I error is about 20% at $\alpha = 0.05$. When the sample size reaches 400, the probability of a Type I error decreases to the expected 5% for $\alpha = 0.05$. The sample hazard was estimated using locally optimal varying kernels as described by Müller and Wang (1994). The second most powerful test was the

Anderson-Darling test for all except the Weibull and the Gamma-Gompertz distributions. Not surprisingly, the likelihood ratio test was the best to identify the differences between the Gompertz and the Weibull distribution and was also effective against the Gamma-Gompertz distribution. The modified Anderson-Darling test, with emphasis on the upper tail of the distribution could distuinguish between Gompertz and Gamma-Gompertz distributions 12% of the samples of size 50.

	$\bar{\mu}$	r	KL	AD	M	LR
Weibull(10,80)	0.2616	0.1460	0.0049	0.2843	0.3542	0.5149
Log-normal(4.4,0.01)	0.5531	1.0000	0.9471	1.0000	0.6813	0.9491
Normal(80,10)	0.5765	0.9863	0.5971	0.9950	0.5631	0.9687
Logistic(80,5)	0.4704	0.9634	0.5553	0.9983	0.4773	0.9676
Gamma(71,1.1)	0.5952	0.9999	0.9089	0.9998	0.8129	0.9632
Gamma-Gompertz(0.001,0.1,0.2)	0.2214	0.0287	0.0006	0.3009	0.3939	0.2692

Table 2: Power of the goodness-of-fit statistics against alternative distributions with $n = 200$, $\alpha = 0.05$.

The rejection rate of the tests increases for larger samples with the exception of the test for the sample mean. It seems that the most powerful tests for the Gompertz distribution are the Anderson-Darling and the correlation coefficient tests, especially if they tests against a less related distribution (log-normal, normal, logistic or gamma). If the test is against a related distribution such as Weibull or Gamma-Gompertz, the efficiency of all tests drop. Against the Weibull distribution, the likelihood ratio against the generalized extreme value distribution works the best, its efficiency is lower for the Gamma-Gompertz model as the test is not explicitly against it. It is more difficult to evaluate the power of the likelihood ratio test against non-extreme value distributions. It has a relatively high rejection rate against all of the other distribution but it is not an appropriate test against them as they are not members of the family of extreme value distributions. The moments test performs best in the Weibull and Gamma-Gompertz cases, yet has the weakest power in all other settings.

9 Application: goodness-of-fit to laboratory rat data

The goodness-of-fit tests defined above can be readily used to check if empirical data is Gompertz distributed. As an example, individual life span data of rats will be used. The analyzed data was collected by Vladimir N. Anisimov at the N.N.Petrov Research Institute of Oncology, St.Petersburg, Russia to test carcinogenicity and it is now published in the Biodemographic Database (BDB). Here we will use only the rats in the control group, $n = 51$ females and $n = 46$ males. The data is fully observed and the number of survivors was recorded every day. Please see Figure 1) for the estimated hazard and the Kaplan-Meier survival function and Table 3 for descriptive statistics of the dataset. The hazard estimation was carried out by the same varying kernel width estimation procedure as mentioned earlier.

The Gompertz fit to the data show vary wide confidence intervals which were estimated by the delta method.

Sex	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	**Max**	s	**IQR**
Female	51	192.5	477.0	649.5	603.2	729.0	891.5	177.9	252
Male	46	185.5	399.5	604.0	559.1	747.5	893.5	219.4	348

Table 3: Descriptive statistics of life spans of 51 female and 46 male rats (days)

The goodness-of-fit statistics in general do not reject the null hypothesis that both the distribution of death of both the male and the female rats is Gompertz. (Table 4) While the maximum likelihood estimate of a of the male rats is higher than $\hat{a}$ of the female rats, the estimated daily rate of aging parameter, $\hat{b}$ is lower, leading to a cross-over of mortality later in life. (Figure 1) This result is corroborated by the non-parametric estimates. However, because of the low sample size, the confidence bands are very wide. In spite of that, by looking at the goodness-of-fit statistics and their respective critical values in the appendix, it can be seen that the null is not rejected either by the Anderson-Darling ($0.384 < 0.63$ and $0.55 < 0.62$) and the correlation coefficient ($0.991 > 0.973$ and $0.983 > 0.976$) test statistics at $\alpha = 0.1$. The Kullback-Leibler statistic does not reject the null hypothesis at $\alpha = 0.1$ in the case of females ($5.6\times\ 10^{-6} < 0.25$) but would reject it in the case of males even at $alpha = 0.01$ ($0.23 > 0.201$). The likelihood ratio test also confirms that the Gompertz distribution fits the data as well as the generalized extreme value distribution (its shape parameter equals to 0) at $\alpha = 0.1$ for both females ($0.895 < 2.71$) and males ($2.165 < 2.71$). The quantiles of the normal distribution estimated by the test for the mean of the sample hazard envelops 0 only at $\alpha = 0.01$, otherwise it would reject the null. However, this test is not reliable at this low sample size because of the difficulties of estimating the sample hazard.

Sex	$\hat{a}$	$\hat{b}$	$\bar{\mu}_{\alpha=0.01}$	r	KL	AD	LR
Female	5.7×10^{-5}	0.007	-0.0014 – 0.0029	0.991	5.6×10^{-6}	0.384	0.895
Male	1.9×10^{-3}	0.005	-0.00012 – 0.0034	0.983	0.23	0.55	2.165

Table 4: Calculated Gompertz goodness-of-fit test statistics to the dataset of 51 female and 46 male rats

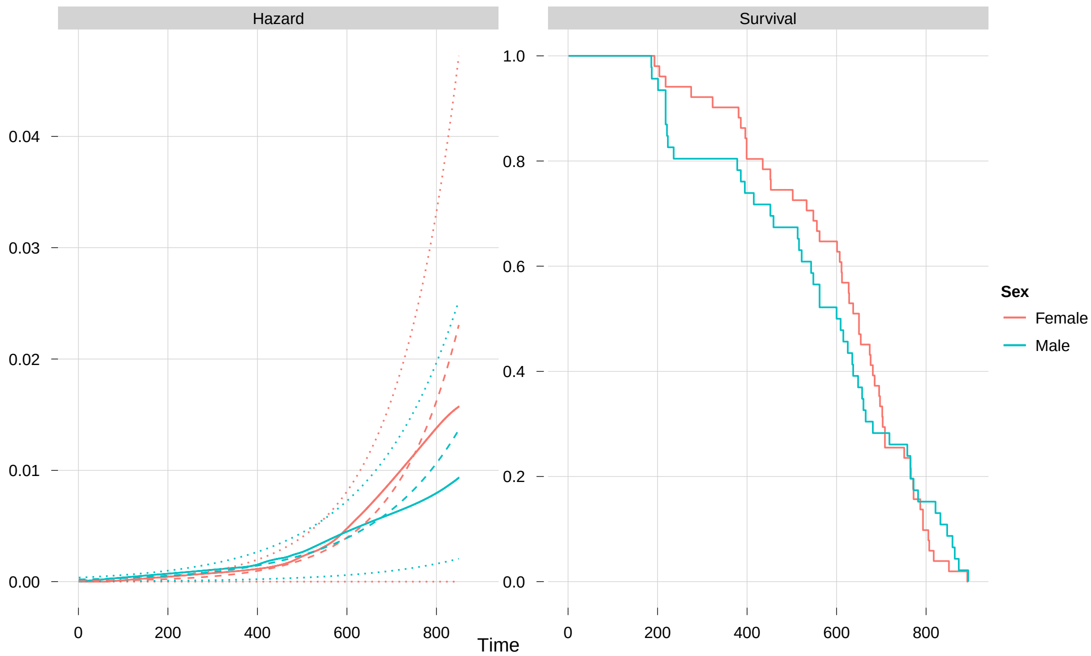

Figure 1: Hazard and survival of the rat data. On the left panel, the solid line corresponds to the non-parametric hazard estimate, the dashed line to the Gompertz fit and the dotted lines are the 95% confidence intervals of the fitted Gompertz hazard.

10 Discussion

The comparison of the power of the tests show that the Anderson-Darling statistic is the most powerful in rejecting the null that the empirical distribution comes from the Gompertz distribution when it was simulated from an alternative distribution. The Anderson-Darling statistic implemented by its computing formula is also the simplest and the quickest to run, and an important advantage of it is that for low values of a, the distribution of the statistic is independent from the Gompertz a and b parameters.

The correlation coefficient test also efficiently refutes other alternative distributions, however, when the alternative distribution is closely related to the Gompertz, such as in the case of Weibull and Gamma-Gompertz distributions, the power of the correlation coefficient test drops. As Legates and McCabe Jr (1999) noted, the tests based on correlation are overly sensitive to outliers and insensitive to proportional differences between the expected and the observed values.

The main problem with testing the mean of the sample hazard lies in the estimation of the sample hazard. Here locally optimal varying kernels were used (Müller and Wang 1994) and as the variance of the kernel hazard estimate was not taken into account, the probability of committing Type I error at $\alpha = 0.05$ reduced to 5% only by $n = 400$. Other sample hazard estimators would necessarily yield different efficiency and critical values.

Juxtaposed with the results for the Gumbel distribution (Pérez-Rodríguez et al. 2009), the Kullback-Leibler test performs unexpectedly poorly relative to the other tests. The main disadvantage of the Kullback-Leibler test lies in the estimation of the sample entropy and choosing the optimal window width for that as it can vary from dataset to dataset with the same sample size and a different window width entails different critical values of the statistic. In the Appendix, the optimal window width is not reported as it was optimized at each draw from the Gompertz distribution, therefore the critical values correspond to an average of the optimal window widths that can be expected when sampling from the Gompertz distribution.

The likelihood ratio test is a powerful test when the alternative distribution is from the generalized extreme value family. A positive externality of the test is that the shape parameter of the generalized extreme value distribution, ξ has to be estimated during the testing procedure. If $\xi < 0$ and the likelihood ratio at the chosen significance level rejects the null hypothesis that $\xi = 0$, than the empirical distribution can be better fitted by a Weibull distribution than by a Gompertz. If $\xi > 0$, the empirical distribution is more likely to be Fréchet-type than Gompertz (Jenkinson 1955).

References

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

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.

Brissaud, J.B. 2005. "The meanings of entropy." *Entropy* 7:68–96.

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.

Gringorten, I.I. 1963. "A plotting rule for extreme probability paper." *Journal of Geophysical Research* 68:813–814.

Harter, H.L. 1961. "Expected values of normal order statistics." *Biometrika* 48:151–165.

Hosking, J.R.M. 1984. "Testing whether the shape parameter is zero in the generalized extreme-value distribution." *Biometrika* 71:367–374.

Jenkinson, A.F. 1955. "The frequency distribution of the annual maximum (or minimum) values of meteorological elements." *Quarterly Journal of the Royal Meteorological Society* 81:158–171.

Kinnison, R. 1989. "Correlation coefficient goodness-of-fit test for the extreme-value distribution." *The American Statistician* 43:98–100.

Kullback, S. and R.A. Leibler. 1951. "On information and sufficiency." *The Annals of Mathematical Statistics* 22:79–86.

Legates, D.R. and G.J. McCabe Jr. 1999. "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation." *Water Resources Research* 35:233–241.

Lenart, A. 2012. "The moments of the Gompertz distribution and maximum likelihood estimation of its parameters."

Li, G. and A. Papadopoulos. 2002. "A note on goodness of fit tests using moments." *STATISTICA* 1:71–86.

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

Müller, H.G. and J.L. Wang. 1994. "Hazard rate estimation under random censoring with varying kernels and bandwidths." *Biometrics* 50:61–76.

Pérez-Rodríguez, P., H. Vaquera-Huerta, and J.A. Villaseñor-Alva. 2009. "A Goodness-of-Fit Test for the Gumbel Distribution Based on Kullback–Leibler Information." *Communications in StatisticsTheory and Methods* 38:842–855.

Sen, P.K. 1959. "On the moments of the sample quantiles." *Calcutta Statistical Association Bulletin* 9:1–19.

Shannon, C.E. 1948. "A mathematical theory of communication." *The Bell Systems Technical Journal* 27:379–423, 623–656.

Shin, H., Y. Jung, C. Jeong, and J.H. Heo. 2011. "Assessment of modified Anderson–Darling test statistics for the generalized extreme value and generalized logistic distributions." *Stochastic Environmental Research and Risk Assessment* 26:105–114.

Sinclair, CD, BD Spurr, and MI Ahmad. 1990. "Modified Anderson darling test." *Communications in Statistics-Theory and Methods* 19:3677–3686.

Song, K.S. 2002. "Goodness-of-fit tests based on Kullback-Leibler discrimination information." *IEEE Transactions on Information Theory* 48:1103–1117.

Stephens, M.A. 1974. "EDF statistics for goodness of fit and some comparisons." *Journal of the American Statistical Association* 69:730–737.

Stephens, M.A. 1977. "Goodness of fit for the extreme value distribution." *Biometrika* 64:583–588.

Vasicek, O. 1976. "A test for normality based on sample entropy." *Journal of the Royal Statistical Society. Series B (Methodological)* 38:54–59.

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

Véron, J. and J.M. Rohrbasser. 2003. "Wilhelm Lexis: The Normal Length of Life as an Expression of the Nature of Things." *Population (english edition)* 58:303–322.

Vogel, R.M. 1986. "The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses." *Water Resources Research* 22:587–590.

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

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*, edited by E. Tabeau, A. van den Berg Jeths, and C. Heathcote, volume 9 of *European Studies of Population*, pp. 105–126. Springer.

Wilson, D.L. 1994. "The analysis of survival (mortality) data: fitting Gompertz, Weibull, and logistic functions." *Mechanisms of Ageing and Development* 74:15–33.

A Variance-covariance matrix of the maximum likelihood estimators of the Gompertz parameters

The variance-covariance matrix, based on the inverse of the observed Fisher information, for the Gompertz distribution is

$$\begin{pmatrix} Var(a) & Cov(a,b) \\ Cov(a,b) & Var(b) \end{pmatrix},$$

where

$$Var(a) = \frac{1}{\frac{n}{a^2} + \frac{\left[\sum_{i=1}^{n}\left(\frac{-1+e^{bX_i}}{b^2} - \frac{e^{bX_i}X_i}{b}\right)\right]^2}{\sum_{i=1}^{n}\left(-\frac{2a\left(-1+e^{bX_i}\right)}{b^3} + \frac{2ae^{bX_i}X_i}{b^2} - \frac{ae^{bX_i}X_i^2}{b}\right)}}$$

$$Cov(a,b) = -\frac{1}{\sum_{i=1}^{n}\left(\frac{-1+e^{bX_i}}{b^2} - \frac{e^{bX_i}X_i}{b}\right) + \frac{n\sum_{i=1}^{n}\left(-\frac{2a\left(-1+e^{bX_i}\right)}{b^3} + \frac{2ae^{bX_i}X_i}{b^2} - \frac{ae^{bX_i}X_i^2}{b}\right)}{a^2\sum_{i=1}^{n}\left(\frac{-1+e^{bX_i}}{b^2} - \frac{e^{bX_i}X_i}{b}\right)}}$$

$$Var(b) = -n \Bigg/ \left[a^2\left(\sum_{i=1}^{n}\left(\frac{-1+e^{bX_i}}{b^2} - \frac{e^{bX_i}X_i}{b}\right)\right)^2 + n\sum_{i=1}^{n}\left(-\frac{2a\left(-1+e^{bX_i}\right)}{b^3} + \frac{2ae^{bX_i}X_i}{b^2} - \frac{ae^{bX_i}X_i^2}{b}\right)\right]$$

B Distribution of test statistics

The Gumbel distribution is a location-scale distribution, therefore the test statistics of it are independent on the location or the scale parameters. However, in the case of the Gompertz distribution the test statistics differ by the a parameter. Therefore, the different a parameters require the definition of different corresponding critical values. To test whether the distribution of the test statistics differ also by the b parameter 50000, $n = 50$ samples were drawn from the Gompertz distribution with $a = \{0.0000015, 0.00015, 0.015, 0.15\}$ and $b = \{0.08, 0.1, 0.12, 0.14, 0.16\}$ and the corresponding Kullbeck-Leibler, correlation coefficient and Anderson-Darling test statistics calculated for them.

B.1 Distribution of the Kullbeck-Leibler discrimination information statistic

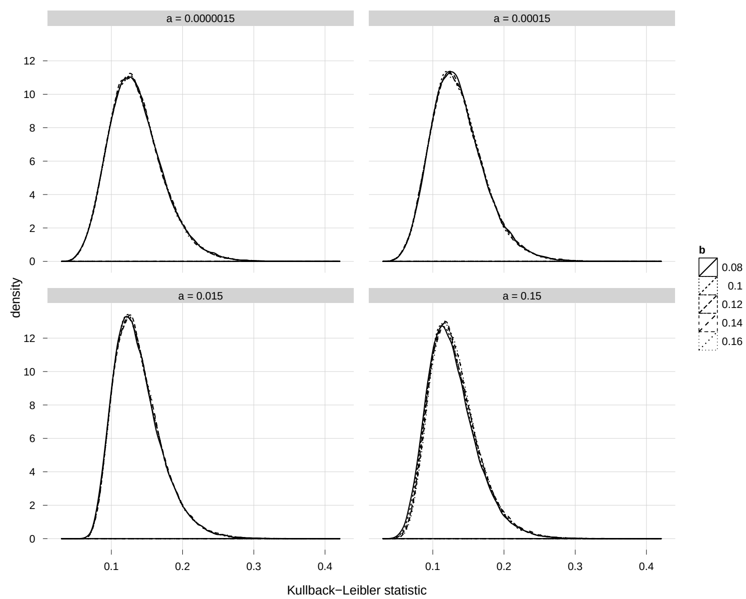

Figure 2: The distribution of the Kullback-Leibler discrimination information statistic for 50000 samples of size 50 drawn from the Gompertz distribution at $m = 5$ for each combination of the Gompertz parameters.

B.2 Distribution of the correlation coefficient statistic

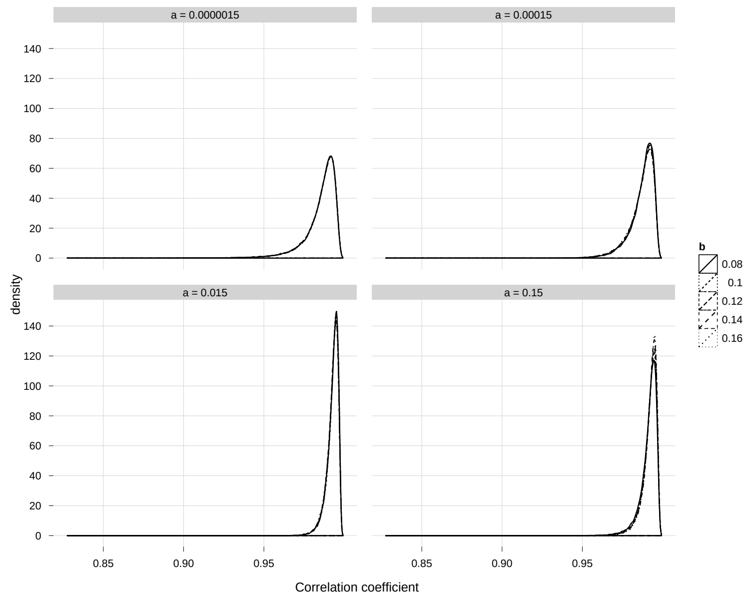

Figure 3: The distribution of the correlation coefficient statistic for 50000 samples of size 50 drawn from the Gompertz distribution for each combination of the Gompertz parameters.

B.3 Distribution of the Anderson-Darling statistic

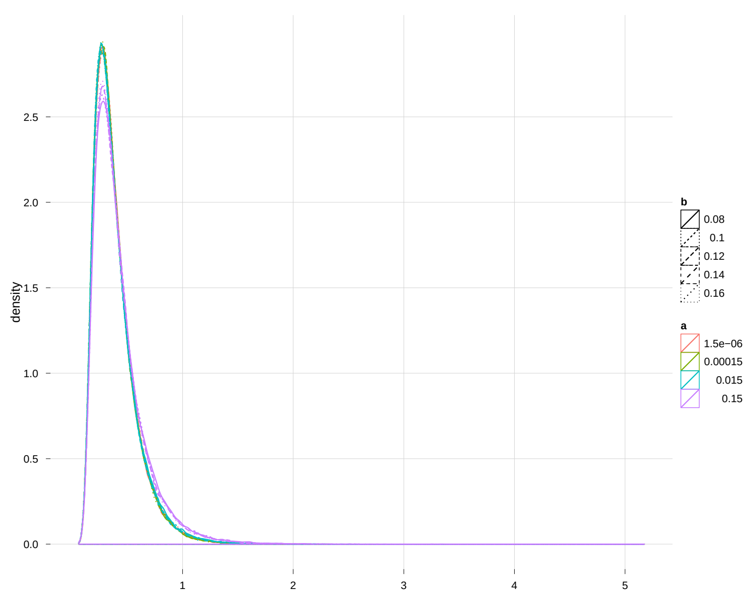

Figure 4: The distribution of the Anderson-Darling statistic for 50000 samples of size 50 drawn from the Gompertz distribution for each combination of the Gompertz parameters .

C Empirical critical values

C.1 Critical values of the Kullback-Leibler discrimination information statistic

				$\hat{a}$		
n	α	0.000001	0.0001	0.01	0.1	0.2
	0.1	0.2189	0.2190	0.2197	0.2207	0.2174
25	0.05	0.2515	0.2507	0.2504	0.2497	0.2465
	0.01	0.3196	0.3202	0.3182	0.3155	0.3111
	0.1	0.1399	0.1392	0.0980	0.1772	0.1863
50	0.05	0.1594	0.1584	0.1359	0.2005	0.2061
	0.01	0.2010	0.2003	0.1943	0.2430	0.2435
	0.1	0.0278	0.0273	0.0323	0.0448	0.0485
75	0.05	0.0432	0.0429	0.0486	0.0617	0.0653
	0.01	0.0763	0.0781	0.0856	0.0971	0.0989
	0.1	0.0506	0.0511	0.0586	0.0651	0.0655
100	0.05	0.0621	0.0624	0.0692	0.0762	0.0766
	0.01	0.0862	0.0872	0.0932	0.0989	0.0990
	0.1	0.0631	0.0639	0.0772	0.0789	0.0813
150	0.05	0.0732	0.0734	0.0839	0.0860	0.0901
	0.01	0.0914	0.0916	0.0983	0.0999	0.1089
	0.1	0.0495	0.0507	0.0825	0.0812	0.0759
200	0.05	0.0598	0.0608	0.0889	0.0877	0.0825
	0.01	0.0799	0.0806	0.1005	0.1003	0.0954
	0.1	0.0392	0.0410	0.0337	0.0434	0.0508
300	0.05	0.0464	0.0474	0.0405	0.0507	0.0573
	0.01	0.0581	0.0586	0.0534	0.0641	0.0694
	0.1	0.0128	0.0150	0.0224	0.0239	0.0243
500	0.05	0.0180	0.0194	0.0258	0.0275	0.0277
	0.01	0.0262	0.0273	0.0322	0.0341	0.0345
	0.1	2E-6	0.0007	0.0231	0.0240	0.0240
1000	0.05	0.0005	0.0062	0.0253	0.0263	0.0264
	0.01	0.0138	0.0177	0.0295	0.0305	0.0306

Table 5: Empirical critical values of the Kullback-Leibler discrimination information statistic

C.2 Critical values of the correlation coefficient statistic

		$\hat{a}$				
n	α	0.000001	0.0001	0.01	0.1	0.2
	0.1	0.9575	0.9596	0.9758	0.9763	0.9743
25	0.05	0.9460	0.9498	0.9705	0.9708	0.9684
	0.01	0.9153	0.9261	0.9576	0.9572	0.9525
	0.1	0.9731	0.9755	0.9875	0.9869	0.9852
50	0.05	0.9650	0.9696	0.9850	0.9840	0.9817
	0.01	0.9405	0.9558	0.9789	0.9768	0.9731
	0.1	0.9798	0.9824	0.9915	0.9909	0.9895
75	0.05	0.9734	0.9783	0.9898	0.9890	0.9870
	0.01	0.9533	0.9689	0.9859	0.9839	0.9807
	0.1	0.9833	0.9861	0.9936	0.9930	0.9917
100	0.05	0.9783	0.9830	0.9923	0.9915	0.9899
	0.01	0.9622	0.9756	0.9895	0.9880	0.9851
	0.1	0.9875	0.9902	0.9957	0.9952	0.9942
150	0.05	0.9835	0.9881	0.9948	0.9942	0.9929
	0.01	0.9706	0.9831	0.9929	0.9917	0.9896
	0.1	0.9895	0.9923	0.9967	0.9963	0.9955
200	0.05	0.9863	0.9907	0.9961	0.9955	0.9945
	0.01	0.9759	0.9870	0.9947	0.9937	0.9920
	0.1	0.9905	0.9942	0.9977	0.9974	0.9969
300	0.05	0.9882	0.9931	0.9973	0.9969	0.9962
	0.01	0.9814	0.9907	0.9963	0.9957	0.9944

Table 6: Empirical critical values of the correlation coefficient statistic

For sample sizes over 300, the critical values of the correlation coefficient statistic was omitted as it the numerical computation of the statistic is not entirely reliable as it requires to compute high values of factorials. In practice, the $n!/b(i-1)!(n-i)!$ term is better to be calculated by $1/\beta(i,(n-i+1))$ as the beta function can be counted until higher values than the factorials separately. For even larger samples, samples can be drawn from the quantile function by (4) by noting that the rank percentiles (rank of the observation divided by sample size + 1) are also bounded by 0 and 1 (see e.g., Kinnison 1989).

C.3 Critical values of the Anderson-Darling statistic

		$\hat{a}$				
n	α	0.000001	0.0001	0.01	0.1	0.2
	0.1	0.6273	0.6276	0.6300	0.6769	0.6920
25	0.05	0.7461	0.7482	0.7487	0.8166	0.8423
	0.01	1.0243	1.0227	1.0291	1.1836	1.2250
	0.1	0.6338	0.6252	0.6238	0.6839	0.7126
50	0.05	0.7545	0.7473	0.7477	0.8262	0.8666
	0.01	1.0380	1.0295	1.0405	1.1804	1.2287
	0.1	0.6299	0.6288	0.6295	0.6830	0.7082
75	0.05	0.7537	0.7483	0.7489	0.8261	0.8560
	0.01	1.0278	1.0341	1.0356	1.1746	1.2155
	0.1	0.6316	0.6278	0.6312	0.6836	0.7113
100	0.05	0.7525	0.7506	0.7521	0.8227	0.8657
	0.01	1.0326	1.0279	1.0453	1.1618	1.2333
	0.1	0.6316	0.6299	0.6263	0.6870	0.7094
150	0.05	0.7533	0.7510	0.7518	0.8303	0.8598
	0.01	1.0389	1.0483	1.0415	1.1833	1.2280
	0.1	0.6347	0.6290	0.6306	0.6851	0.7100
200	0.05	0.7509	0.7526	0.7570	0.8299	0.8621
	0.01	1.0347	1.0427	1.0370	1.1688	1.2292
	0.1	0.6351	0.6319	0.6290	0.6916	0.7128
300	0.05	0.7559	0.7580	0.7496	0.8398	0.8700
	0.01	1.0453	1.0502	1.0352	1.1705	1.2327
	0.1	0.6323	0.6331	0.6334	0.6843	0.7152
500	0.05	0.7516	0.7531	0.7577	0.8308	0.8696
	0.01	1.0331	1.0366	1.0543	1.1642	1.2444
	0.1	0.6405	0.6339	0.6289	0.6843	0.7166
1000	0.05	0.7584	0.7538	0.7530	0.8267	0.8698
	0.01	1.0448	1.0425	1.0498	1.1651	1.2455

Table 7: Empirical critical values of the Anderson-Darling statistic

Please note that the Anderson-Darling statistics are stable over all low values of $\hat{a}$ and increasing by $\hat{a}$. The critical values also increase slightly as the sample size increases. Similar trend was found by Shin et al. (2011) for the modified Anderson-Darling test.