The Gompertz-Makeham distribution

by

Fredrik Norström

Master’s thesis in Mathematical Statistics,
Umeå University, 1997

Supervisor: Yuri Belyaev
Abstract

This work is about the Gompertz-Makeham distribution. The distribution has been applied plenty of times. The properties of the Gompertz-Makeham distribution investigated in this work are unimodality of the Gompertz-Makeham distribution and relationship between the median value and the mean residual life time of the Gompertz-Makeham distribution. For most of the realistic set of parameter values the function is unimodal but not for all. The example used in this work gives a set of parameter values for which the function is unimodal. The relationship between the median value and the mean life time, left after some given time, has also been investigated with truncation at fixed age S. A program for simulations of life time with the Gompertz-Makeham distribution has been written in Pascal. For the possibility to have estimators of the unknown parameters of the Gompertz-Makeham distribution the least square estimation has been applied by use of a program written in Pascal. Description of the least square estimation and also the method of Maximum Likelihood and the EM-algorithm are also included in this work. The estimators of parameters can also be obtained by use of the last two methods. For testing the hypothesis that extreme old ages follows the Gompertz-Makeham distribution a goodness of fit test has been applied to real demographic data. For this example the hypothesis that extreme old ages follows the Gompertz-Makeham distribution, with parameters estimated by use of the least square estimation, is rejected.
Contents

1 Introduction. 1

2 Properties and characteristics of the Gompertz-Makeham distribution. 5

3 Programs for simulation and estimation of parameters. 14

3.1 Simulation of Gompertz-Makeham distributed data. 14

3.2 Estimation of the parameters in the Gompertz-Makeham distribution with use of the least square estimation. 16

3.3 Estimation of the parameters in the Gompertz-Makeham distribution with use of the method of Maximum Likelihood. 17

3.4 Estimation of the parameters in the Gompertz-Makeham distribution with use of the EM-algorithm. 18

4 Testing of some hypotheses. 20

4.1 The Goodness of fit test 20

4.2 Kolmogorov test 22

4.3 Likelihood ratio test 22

5 References. 24

Appendix

1 Introduction
An interest of finding a specific function (distribution) that well approximates real life table data has existed through many years. To find a specific distribution function that approximates life table data in a reasonable way has always been a major problem and attempts with different functions haven’t solved the problem. There are several different distribution functions that have been tested for this purpose. The main problem is if those functions describes the real situation good enough. There is always a possibility to estimate parameters using less complicated functions as linear functions and quadratic functions. But these kind of functions aren’t of interest, because they won’t explain life table data good enough. It is simple to prove this by using different tests. The tests will make it possible to reject the hypothesis that life table data are from one of those functions. While many common functions, functions that are often used in other applications according to statistical theory, aren’t possible to use. Another function, which has the possibility to explain life table data satisfactory is necessary to find. Of course, this function must be a distribution function (a distribution has the characteristic that all of the possible outcomes of the distribution has function values between 0 and 1 and the sum of all possible outcomes is 1) otherwise it can’t explain life table data in a natural way. If the function isn’t a distribution function unrealistic situations occurs. The function that explains life table data must have the characteristic that the error (when the unknown parameters of the function are estimated according to life table data) seems to be a “small” random error, or in any sense the error has the characteristic that it doesn’t disturb the real model more than moderate.

Several attempts have been done to find functions that approximates life table data in a satisfactory way. Researches have been done for many different types of functions. All of the functions with different type of theories of mortality behind, as the Brody-Failla theory [Brody, 1923; Failla, 1958], the Simms-Jones theory [Simms, 1942] formed by Simms in 1940 and expounded by Jones later on, the Simms-Sacher theory [Simms, 1942; Sacher, 1956] expanded by Sacher from Simms original theory in 1942. All of these theories have been analysed in the book “Time, cells, and aging” [Strehler, 1962] and the theories are shown that they can’t be useful by Strehler. All of these theories have the outfit that they use the Gompertz function, first introduced by Benjamin Gompertz in 1825, mathematically expressed as $R_m = R_0 e^{at}$, $R_m$ is rate of mortality, a and $R_0$ are constants and $t$ is the time parameter [Gompertz, 1825]. Another theory also using the Gompertz function is the Strehler-Mildvan theory [Strehler, 1960; Lenhoff, 1959]. The theories explained above are all related to a decrease, with aging, in the healthy state. However, the first three of the theories above make certain predictions which are in keeping with observations, although they are not completely consistent with certain other primary observations relating time, physiological function and mortality. The Strehler-Mildvan theory fits the Gompertzian mortality kinetics and assumes a linear decay of physiological function at a rate consistent with observation. Finally, it predicts the quantitatively inverse relationship between Gompertz slope and intercept which has been observed. The theory though hasn’t been used in later years, because of the fact that there are functions that seems to give much better approaches.

In the last decades almost all of the researches that are related to approximations of life tables are made with other functions than the functions explained above. Probably the most important function to apply is the Gompertz-Makeham distribution. This distribution function gives much better approximations for life table data than the approaches described above. The Gompertz-Makeham distribution is also a function that uses the Gompertz function.
The major difference between the Gompertz-Makeham distribution and the functions explained above is that the Gompertz-Makeham function uses more parameters than the simple Gompertz function. The Gompertz-Makeham distribution has the survival function:

\[ \bar{F}_\theta(s) = \exp[-\alpha s - \beta \frac{e^{ys} - 1}{y}] \], \quad \theta = (\alpha, \beta, \gamma), \]

and consequently the (cumulative) hazard function:

\[ H_\theta(s) = \alpha s + \beta \frac{e^{ys} - 1}{y}. \]

The cumulative hazard function is described in section 2.

The Gompertz function has the distribution function: \( F_\theta(s) = B \exp[as] \), \( \theta = (a, B) \), \( a < 0 \), while the Gompertz-Makeham function has the distribution function:

\[ F_\theta(s) = 1 - \exp[-\alpha s - \beta \frac{e^{ys} - 1}{y}]. \]

The Gompertz-Makeham function has three unknown constants while the Gompertz function has only two constants. That’s one of the main reasons why the Gompertz-Makeham function is to prefer for descriptions of real data instead of the Gompertz function.

The Gompertz-Makeham distribution has been investigated in many ways. Still there are things that haven’t been done. In this work different properties will be investigated. One of the investigated properties is, if the probability density function of the Gompertz-Makeham distribution is unimodal, i.e. if there exist only one local maximum of the function (the local maximum can be at the boundary of the function). Another property that will be investigated is the relationship between the median and the mean value of the Gompertz-Makeham distribution. The results of these two investigations are presented in section 2 of this work.

The best way to have the possibility to investigate some of the properties of the Gompertz-Makeham distribution many times is to write a computer program, a program that manages to obtain the wanted results. In this works there has been written programs for different purposes. The programs that has been written are:

1. a program that simulates life times assuming the parameters in the model are known, the program is described in section 3.1, (simulated life times are useful to have for results of different properties), and
2. a program that makes it possible to have the least square estimation of the Gompertz-Makeham distribution related to life table data, a program that is described more in details in section 3.2.
3. a program that are testing hypotheses with the goodness of fit test. Described in section 4.1.

Other estimators than the least square estimator are also possible to have for approximations of life table data. Some of these estimators gives in most cases better approximations than the estimator of the least square estimation. It is necessary to have a good estimation of the unknown parameters, \( \alpha \), \( \beta \) and \( \gamma \), for the opportunity to recognise several different properties of the Gompertz-Makeham distribution. Examples of good estimators are given in section 3.3 and 3.4. In these two sections there is a description of the method of Maximum-Likelihood and also a description of the EM-algorithm.

The least square estimation is a method that minimises the square of the difference between the “real“ value and the value of the function. The method of Maximum-Likelihood is a
method which searches the “most likely“ parameters of a function, the parameters that could have produced known data. The EM-algorithm is almost similar to the method of Maximum-Likelihood but in most cases the EM-algorithm gives simpler expressions and therefore it is an easier way to solve the problems.

Another property that will be investigated is, if the values of the parameters are consistent for every ages. Most of the investigations of different properties of the Gompertz-Makeham distribution uses only estimations of observations that are in the Gompertznian period, the Gompertznian period extends from about the age 35 to the age 90. Still there is a problem, which is if the observations that the Gompertz-Makeham distribution is making use of are estimated for not all ages but only for a period, as the Gompertznian period. There’s a possibility that the estimation reached can’t be used for extreme old ages. The observations for extreme old ages might not follow the Gompertz-Makeham distribution with parameters \( \alpha, \beta \) and \( \gamma \), which gives good approximations for ages between 35 and 90.

In section 4.1, a goodness of fit test has been done for testing the hypothesis that extreme old ages follows the Gompertz-Makeham distribution with the estimated parameters \( \alpha, \beta \) and \( \gamma \) (the parameters are estimated by use of the least square estimation and the period for the estimation is 30-80 years). If the hypothesis that life table data for extreme old ages follows the Gompertz-Makeham distribution, with estimated parameters \( \alpha, \beta \) and \( \gamma \), doesn’t hold, the hypothesis will be rejected. In case the hypothesis explained above will be rejected it is necessary to find another function which give raise to more realistic approximations of life table data for extreme old ages (extreme old ages often is denoted as ages over 90 or 95 years) than the Gompertz-Makeham distribution does. It might be necessary to approximate extreme old ages with a different function than the function that approximates life table data for other ages. An alternative is a density function that has a discontinuity at an extreme old age. This function maybe gives a better approximation for all of the ages than the Gompertz-Makeham distribution does. Attempts to get better approximations by use of a breakpoint in the middle of the period has been done, e.g. Pakin and Hrisanov [1984] has used this kind of approximation.

There are also other tests than the goodness of fit test which are possible to use for testing the hypothesis that extreme old ages follows the Gompertz-Makeham distribution with estimated parameters. For example if the data are explained in a more exact form than years the Kolmogorov test is possible to use. This test is explained in section 4.2. The Goodness of fit test and the Kolmogorov test can also be used for testing whether the Gompertz-Makeham distribution is possible to use for approximations of real life table data at all.

The third and last of the test methods explained in this work is the likelihood ratio test. Based on data \( x \) and the likelihood function \( p(x, \theta), \theta \in \Theta \). The likelihood ratio test consists of testing if
\[
\hat{\lambda}(x) = \frac{\sup\{p(x, \theta); \theta \in \Theta\}}{\sup\{p(x, \theta); \theta \in \Theta_0\}}
\]
is bigger than some test value (known values from tables).

The test value depends on the model assumption and the hypothesis. \( \sup\{p(x, \theta); \theta \in \Theta\} \) and \( \sup\{p(x, \theta); \theta \in \Theta_0\} \) are estimated with the moment of Maximum-Likelihood, the estimators are \( \hat{\theta} \) and \( \hat{\theta}_0 \). The methodology of the likelihood ratio test is possible to find more in details in section 4.3. In that section there is also described how to test \( H_0: \alpha = \alpha_0 \) versus \( H_1: \alpha \neq \alpha_0 \), given \( \beta=\beta_0 \) and \( \gamma=\gamma_0 \).
For the model used in this work it won’t be necessary to test whether the Gompertz-Makeham distribution, with estimated parameters $\alpha$, $\beta$ and $\gamma$, gives a good approximation of real life table data for ages between 30 and 80 or not. The reason for this is the fact that when comparisons between the observations for the life table data used and the values of the Gompertz-Makeham distribution for ages in the period 30 to 80 years has been done with help of a plot. The plot shows that the Gompertz-Makeham distribution gives very good approximation and no essential argument for a test exists.

Many investigations and different attempts have been done to decide if the Gompertz-Makeham distribution has all of the characteristics that are necessary, i.e. if it is possible to use the Gompertz-Makeham distribution according to real data. For example Pakin and Hrisanov [1984] has done an investigation for checking if the parameters have the same characteristics in the complete interval between 35 and 75 years. Investigations of the Gompertz-Makeham distribution have also been done by several others. Most of them gives a good impression of the distribution. This gives us a reason to make the conclusion that the distribution tends to give us good approximations when it is used according to life table data.

An example where real life table data are used is illustrated in section 3.2. The real demographic data are collected from Statistical Abstract [1983] and the data are for Swedish women under the period 1976-1980. A piece of the table in Statistical Abstract, the table used to have the illustration mentioned above, are shown below.

<table>
<thead>
<tr>
<th>Age</th>
<th>Survivors of 100000 born alive</th>
<th>Mean expectation of rested life, years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>45</td>
<td>94264</td>
<td>96750</td>
</tr>
<tr>
<td>46</td>
<td>93928</td>
<td>96555</td>
</tr>
<tr>
<td>47</td>
<td>93547</td>
<td>96360</td>
</tr>
<tr>
<td>48</td>
<td>93136</td>
<td>96136</td>
</tr>
<tr>
<td>49</td>
<td>92721</td>
<td>95891</td>
</tr>
</tbody>
</table>

*Table 1.1.*

Table 1.1. are collected data that shows the numbers of 100000 born alive that are still alive when they are $x$ years of age (the age are counted when a new year begin) and the number are arranged in intervals of length one year. The approximation of real demographic data with the Gompertz-Makeham distribution (the parameters are estimated with help of the least square estimation with estimated parameters obtained in section 3.2) gives very good results. The relation between the Gompertz-Makeham distribution and real demographic data for Swedish women is shown below in Figure 1.1.
As you can see of Figure 1.1 the Gompertz-Makeham distribution gives a very good approximation of real demographic data. This explains why it is of interest to use the Gompertz-Makeham distribution for different approximations relating to life length theory.

The basic reason for making approximations of real demographic data with use of the Gompertz-Makeham distribution is that there are many different professions that have great use of these kind of approximations of life table data. One of the professions that have great use of the Gompertz-Makeham distribution are insurance companies. The Gompertz-Makeham distribution would give them better possibilities to determine insurance’s that better explains the mortality among people for both accidents and natural deaths and it would be very helpful when the fees of the insurance’s are decided. The Gompertz-Makeham distribution isn’t only useful for approximating life lengths for human populations. It might even be of importance to use it in many different biological ways, for example plant biology has great use of the Gompertz-Makeham distribution. Also life lengths for different crops are an application where the Gompertz-Makeham distribution can be of importance to use. The possibility to study life lengths for different crops might even give the possibility to choose a treatment that in some sense raise the quality of these crops.

2 Properties and characteristics of the Gompertz-Makeham distribution

The Gompertz-Makeham distribution is a distribution that gives very good approximations to empirical distributions of life length not only for human populations but also for different biological arts. There will be investigations of some properties of the Gompertz-Makeham distribution in this section.

The Gompertz-Makeham distribution has the survival function:

$$F_0(s) = \exp[-\alpha s - \beta e^{\gamma s} - \gamma].$$

Where $s$ is life time always non-negative ($s \geq 0$) and $\alpha$, $\beta$ and $\gamma$ are known non-negative parameters and $\alpha + \beta \gamma > 0$ and hence $F_0(s) > 0$ for $s > 0$ and $F_0(0) = 1.$
How different values of this three parameters influence to the model will be researched. In general the hazard function of the Gompertz-Makeham distribution often can be used to express the behaviour of the distribution.

The (cumulative) hazard function of a distribution can be defined by the relation

$$ H_\theta(s) := \ln \left( \frac{1}{F_\theta(s)} \right), \quad \theta = (\alpha, \beta, \gamma), $$

(2.3)

and the intensity hazard function is the derivative of the (cumulative) hazard function

$$ h_\theta(s) := \frac{dH_\theta(s)}{ds}. $$

(2.4)

The Gompertz-Makeham distribution has the cumulative hazard function

$$ H_\theta(s) = \alpha s + \beta e^s \gamma - 1 $$

(2.5)

and the intensity hazard function is

$$ h_\theta(s) = \alpha + \beta e^s \gamma. $$

(2.6)

Formulas (2.5) and (2.6) can easily be derived from its survival function (2.1) and the definitions of the hazard functions (2.3) and (2.4). Note that $H_\theta(s) > 0$ if $s > 0$ due to (2.2). Another property that can be derived is the probability density function of the Gompertz-Makeham distribution.

$$ f_\theta(s) = \frac{dF_\theta(s)}{ds} = \frac{d}{ds} \left( 1 - e^{-H(s)} \right) = h_\theta(s) e^{-H(s)} = (\alpha + \beta e^s \gamma) \exp(-((\alpha s + \beta e^s \gamma) \gamma)). $$

(2.7)

It will be of interest if there exists any points of extremum not at the boundary of the probability density function of the Gompertz-Makeham distribution. This knowledge is necessary for the possibility to conclude if the probability density function of the Gompertz-Makeham distribution is unimodal or not. The probability density function is unimodal if the derivative is 0 for at most one value of the derivative. The derivative of the probability density function (2.7) of the Gompertz-Makeham distribution is

$$ \frac{df_\theta(s)}{ds} = \frac{d}{ds} \left( h_\theta(s) e^{-H(s)} \right) = \left( h'_\theta(s) - h_\theta(s) h''_\theta(s) \right) e^{-H(s)} = 
$$

(2.8)

$$ = ((\gamma \beta e^s \gamma) - (\alpha + \beta e^s \gamma)^2) e^{-H(s)}, $$

Because as stated in expression (2.6) the intensity hazard function of the Gompertz-Makeham distribution is $h_\theta(s) = \alpha + \beta e^s \gamma$ and therefore the derivative of it is $h'_\theta(s) = \gamma \beta e^s \gamma$.

Right hand side of equation (2.8) is 0 only when the statements $(\gamma \beta e^s \gamma) = (\alpha + \beta e^s \gamma)^2$ is true, because $0 < H(s) < \infty$ and $e^{-H(s)} > 0$. There exists one or two points of extremum only when $(\gamma \beta e^s \gamma) = (\alpha + \beta e^s \gamma)^2$ and the values of the life length that gives this equality are positive.

Set $z = \beta e^s \gamma$, then (2.8) can be rewritten as

$$ \gamma z = (\alpha + z)^2 = \alpha^2 + 2\alpha z + z^2 \quad \text{or} $$

$$ z^2 = (\gamma - 2\alpha)z - \alpha^2. $$

The value $z$ can be solved by a second grade equation so the roots are

$$ z_1 = \frac{\gamma}{2} (1 - \sqrt{1 - 4\alpha / \gamma}) - \alpha, \quad \text{and} \quad z_2 = \frac{\gamma}{2} (1 + \sqrt{1 - 4\alpha / \gamma}) - \alpha, $$

and in time scale

$$ s_1 = \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( 1 - \sqrt{1 - 4\alpha / \gamma} \right) \right) - \alpha, $$

(2.8)
\[ s_2 = \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right). \] 

(2.9)

As earlier defined \( \alpha, \beta \) and \( \gamma \geq 0 \). We have \( s_1 \leq s_2 \), otherwise the solution is a natural logarithm of a negative value, which is a complex value and for that case \( s_1 \) and \( s_2 \) will have complex time. This fact makes sure that the points of extremum occur at non-negative time if they exist. If \( s_1 \) is positive (\( s_1 \) exists) there are two possibilities. Either there exist both a local minimum and a local maximum of the probability density function or there exists an inflection point and there doesn’t exist any other point of extremum separated from the boundary of the function, occurs if \( s_1 = s_2 \). If \( s_1 \) isn’t positive and \( s_2 \) is positive then there exist only a local maximum of the density probability function. If \( s_2 \) isn’t positive there are no points of extremum for the probability density function of the Gompertz distribution. The function is unimodal if there are no local maximum not at the boundary of the function.

**Corollary** (2.11)
There exists an inflection point of the probability density function of the Gompertz-Makeham distribution if and only if \( 4\alpha = \gamma \) and \( \alpha > \beta \).

\[ \text{If 1-4\alpha/\gamma = 0 \Rightarrow 4\alpha = \gamma. Set 4\alpha = \gamma and it follows that s_1 = s_2,} \]

\[ s_1 = \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right) = \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right) = \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right) = s_2 \]

\[ s_1 \text{ non-negative implies that } \ln \left( \frac{1}{\beta} (2\alpha-\alpha) \right) > 0 \text{ or } (2\alpha - \alpha) > 1 \text{ or } \alpha > \beta. \] There exists an inflection point if and only if \( 4\alpha = \gamma \) and \( \alpha > \beta \). Which was to be proved.

**Corollary** (2.12)
There exist a local minimum not at the boundary of the probability density function of the Gompertz-Makeham distribution if \( 4\alpha < \gamma, 2(\alpha+\beta) < \gamma \) and \( \beta \gamma < (\alpha+\beta)^2 \) are true statements.

**Proof of corollary (2.12)**
Local minimum not at the boundary of the probability density function of the Gompertz-Makeham distribution exists if \( s_1 \) is positive or

\[ \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right) > 0 \Rightarrow \frac{1}{\gamma} \ln \left( \frac{\gamma}{\beta} \left( \frac{1}{2} (1+\sqrt{1-4\alpha/\gamma})-\alpha \right) \right) > 1 \]

\[ \Rightarrow \gamma (1-\sqrt{1-4\alpha/\gamma}) > 2(\alpha+\beta) \Rightarrow 1-\sqrt{1-4\alpha/\gamma} > 2(\alpha+\beta)/\gamma \Rightarrow \]

\[ 1-2(\alpha+\beta)/\gamma > \sqrt{1-4\alpha/\gamma}, \text{ no complex solutions allowed} \Rightarrow \sqrt{1-4\alpha/\gamma} > 0 \Rightarrow 4\alpha < \gamma \] 

(2.13)

(if \( 4\alpha = \gamma \) then there is an inflection point, instead of a local minimum not at the boundary)

\[ 1-2(\alpha+\beta)/\gamma > 0 \text{ or } 2(\alpha+\beta) < \gamma \] 

(2.14)

\[ (\sqrt{1-4\alpha/\gamma})^2 < (1 - 2(\alpha+\beta)/\gamma)^2 \text{ if conditions (2.13) and (2.14) are true then} \]

\[ 1-4\alpha/\gamma < 1+4(\alpha+\beta)^2/\gamma^2 -4(\alpha+\beta)/\gamma \Rightarrow \]

\[ \beta \gamma < (\alpha+\beta)^2 \] 

(2.15)
Due to conditions (2.13), (2.14) and (2.15) there exists a local minimum of the probability density function of the Gompertz-Makeham distribution if $4\alpha < \gamma$, $2(\alpha+\beta) < \gamma$ and $\beta\gamma < (\alpha+\beta)^2$ are true statements. Which has to be proved.

If there’s a local minimum not at the boundary of the probability density function of the Gompertz-Makeham distribution then $\alpha > \beta$, because of the fact that (by use of condition (2.14) and (2.15))

$$4\alpha < \gamma < (\alpha+\beta)^2/\beta \Rightarrow 2\beta(\alpha+\beta) \leq \beta\gamma \leq (\alpha+\beta)^2 \Rightarrow 2\beta \leq \beta\gamma/(\alpha+\beta) \leq (\alpha+\beta).$$

Implies that $2\beta$ never bigger than $(\alpha+\beta)$ if local minimum exists or that $\alpha$ must be bigger or equal to $\beta$ if there exists a local minimum not at the boundary of the probability density function of the Gompertz-Makeham distribution. If $\alpha = \beta$ then $2\alpha = \alpha\gamma/(\alpha+\alpha) = (\alpha+\alpha) \Leftrightarrow \alpha\gamma/2\alpha = 2\alpha \Rightarrow \gamma = 4\alpha$. If the equality $\gamma = 4\alpha$ holds, then there’s an inflection point at the point of extremum not at the boundary and not a local minimum, due to corollary (2.11). Therefore $\alpha$ can’t be equal to $\beta$ if there exists a local minimum not at the boundary, hence $\alpha$ must be bigger than $\beta$.

**Corollary** (2.16)

There exists a local maximum not at the boundary of the probability density function of the Gompertz-Makeham distribution if $4\alpha < \gamma$ and $2(\alpha+\beta) < \gamma$ are true statements or if $4\alpha < \gamma$, $2(\alpha+\beta) \geq \gamma$ and $\beta\gamma > (\alpha+\beta)^2$ are true statements.

**Proof of corollary (2.16)**

Local maximum exists and is not at the boundary if $s_2$ is positive $\Rightarrow$

$$\frac{1}{\gamma} \ln \left( \frac{1}{\beta} \left( \frac{\gamma}{2} \left( 1+\sqrt{1-4\alpha/\gamma} \right) - \alpha \right) \right) > 0 \Rightarrow \frac{1}{\gamma} \left( \frac{\beta}{2} \left( 1+\sqrt{1-4\alpha/\gamma} \right) - \alpha \right) > 1 \Rightarrow$$

$$\gamma \left( 1+\sqrt{1-4\alpha/\gamma} \right) > 2(\alpha+\beta) \Rightarrow 1+\sqrt{1-4\alpha/\gamma} > 2(\alpha+\beta)/\gamma \Rightarrow \sqrt{1-4\alpha/\gamma} > 2(\alpha+\beta)/\gamma - 1$$

and we have (2.13).

The following two conditions are possible, condition (2.14) and

$$2(\alpha+\beta) \geq \gamma.$$  

(2.17)

If conditions (2.13) and (2.14) are true then there exists a local maximum not at the boundary. The case when the condition (2.17) is true leads to following

$$\left( \sqrt{1-4\alpha/\gamma} \right)^2 > (2(\alpha+\beta)/\gamma - 1)^2 \Rightarrow$$

$$1-4\alpha/\gamma > 1+4(\alpha+\beta)^2/\gamma^2 - 4(\alpha+\beta)/\gamma \Rightarrow$$

$$\beta\gamma > (\alpha+\beta)^2.$$  

(2.18)

There exists a local maximum not at the boundary of the probability density function of the Gompertz-Makeham distribution if conditions (2.13), (2.17) and (2.18) are true statements. Due to conditions (2.13), (2.14), (2.17) and (2.18) there exists a local maximum not at the boundary if $4\alpha < \gamma$ and $2(\alpha+\beta) < \gamma$ are true statements or $4\alpha < \gamma$, $2(\alpha+\beta) \geq \gamma$ and $\beta\gamma > (\alpha+\beta)^2$ are true statements. Which was to be proved.

**Definition** (2.19)

A function is unimodal if there exists only one local maximum of the function.

**Theorem** (2.20)
The probability density function of the Gompertz-Makeham distribution function is unimodal if and only if \(4\alpha \geq \gamma\) is a true statement, or if \(\beta\gamma \leq (\alpha+\beta)^2\) and \(2(\alpha+\beta) \geq \gamma\) are true statements or if \(4\alpha < \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements.

**Proof of theorem (2.20)**

By definition (2.19) the probability density function of the Gompertz-Makeham distribution is unimodal if the only local maximum that exists of the function is at the boundary or if the only local maximum of it is not at the boundary of the boundary. The only local maximum is at the boundary (\(s=0\)) of the function when corollary (2.16) isn’t true. If corollary (2.16) isn’t true, the following statements are correct, \(4\alpha \geq \gamma\) or both \(2(\alpha+\beta) \geq \gamma\) and \(\beta\gamma \leq (\alpha+\beta)^2\) are true statements.

The only local maximum isn’t at the boundary if there exists a local maximum not at the boundary and there doesn’t exist a local minimum separated from the boundary of the function. This is the case when corollary (2.16) is true and corollary (2.12) isn’t true. Corollary (2.16) is true if \(4\alpha < \gamma\) and \(2(\alpha+\beta) < \gamma\) are true statements or if \(4\alpha < \gamma\), \(2(\alpha+\beta) \geq \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements and corollary (2.12) isn’t true when \(4\alpha \geq \gamma\) and/or \(2(\alpha+\beta) \geq \gamma\) and/or \(\beta\gamma > (\alpha+\beta)^2\) are true statements. Corollary (2.16) is true and corollary (2.12) isn’t true only if \(4\alpha < \gamma\), \(2(\alpha+\beta) \geq \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements or if \(4\alpha < \gamma\), \(2(\alpha+\beta) < \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements, i.e. corollary (2.16) is true and corollary (2.12) isn’t true when \(4\alpha < \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements.

The probability density function of the Gompertz-Makeham distribution is unimodal if \(4\alpha \geq \gamma\) is a true statement or \(2(\alpha+\beta) \geq \gamma\) and \(\beta\gamma \leq (\alpha+\beta)^2\) are true statements or if \(4\alpha < \gamma\) and \(\beta\gamma > (\alpha+\beta)^2\) are true statements. Which has been proved.

There are four different possibilities of the shape of the probability density function for the Gompertz-Makeham distribution. The different shapes that are possible are illustrated below in Figure 2.1-2.4.

**Figure 2.1**

*Figure 2.1 shows the function for the values of \(\alpha, \beta\) and \(\gamma\) that are estimated in section 4, namely \(\alpha = 5.44*10^{-3}\), \(\beta = 7.12*10^{-6}\) and \(\gamma = 0.113\) and is an example of the probability density function of the Gompertz-Makeham distribution having a local maximum not at the boundary of the function.*
Figure 2.2 shows an example of the probability density function for the Gompertz-Makeham distribution having a local maximum both at the boundary and not at the boundary of the function where the parameter values are $\alpha = 0.2$, $\beta = 0.05$ and $\gamma = 0.85$. Figure 2.3 shows an example of the probability density function of the Gompertz-Makeham distribution where the function has a global maximum at the boundary of the function. The values of the parameters are $\alpha = 0.2$, $\beta = 0.4$ and $\gamma = 0.9$. Figure 2.4 shows an example of the probability density function of the Gompertz-Makeham distribution where the function has a global maximum at the boundary of the Gompertz-Makeham distribution and an inflection point not at the boundary of the function. The values of the three parameters are $\alpha = 0.2$, $\beta = 0.05$ and $\gamma = 0.8$.

Another property that is of interest to research are the relationships between the median value and the mean value of the Gompertz-Makeham distribution. Is the median smaller than the mean value for known $\alpha$, $\beta$ and $\gamma$ for the function or not?

Notable, the median is the value where $F_{\theta}(s) = \frac{1}{2}$.

\[
F_{\theta}(s) = 1 - e^{-H(s)} = \frac{1}{2} \Rightarrow e^{H(s)} = \frac{1}{2} \Rightarrow H_0(s) = \ln(2) \Rightarrow \alpha s + \beta \frac{e^{\gamma s} - 1}{\gamma} = \ln(2) \Rightarrow \\
\alpha s + \beta \frac{e^{\gamma s} - 1}{\gamma} - \ln(2) = 0
\]

(2.21)

$s$ can’t be solved explicitly as a expression of $\alpha$, $\beta$ and $\gamma$.

Therefore the only possibility to have a solution of the median is to have a solution with help of a numerical method. A numerical method that often is used for this kind of approximations is the Newton-Raphson method. With help of the Newton-Raphson the median value will be found with good approximation. The median value will be found as the root of function (2.20) (the root is where the equality of (2.21) holds). The Newton-Raphson is a method that iterates until an approximation has been found with as many digits as desirable of the median value. The main problem with the Newton-Raphson method is that a start value must be chosen for having the possibility to start the iterations that this method is build upon. The start value that is used can’t be chosen too far away from the root of the function. Otherwise the function won’t converge to a value and it won’t give any good solution of the existing problem. To find a useful start value it’s often necessary to use a different method. This method must have the possibility to give a value that’s close enough to the searched root, so that the Newton-Raphson method can be used. The advantage of the method is that the method converges very fast to the searched root. The Newton-Raphson method is faster than most of the other methods. That is the justification of the method.
The Newton-Raphson method has the construction that it uses a start value of \( x_0 \), which will be chosen not too far away from the searched root \( x \). Then the Newton-Raphson method is a method that iterates until the value of \( x_i \) is close enough to the value of \( x_{i-1} \), so that the difference between \( x_i \) and \( x_{i-1} \) is as small as wanted, i.e. the level of the tolerance is desirable.

The Newton-Raphson method has the following design:
\[
 x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad \text{where } f(x) \text{ is a function that is 0 at the root of } x.
\]

More information about the theory of the Newton-Raphson method can be found in different books, e.g. Elden and Wittmeyer-Koch [1987], Flannery et al. [1986] and Adams [1991].

Function (2.21) gives the median value of the Gompertz-Makeham distribution,
\[
f(s) = \frac{\alpha s + \beta}{\gamma} \left( \exp\left(\frac{\gamma s}{\beta} \right) - 1 \right) - \ln(2).
\]

The derivative of function (2.21) is
\[
f'(s) = \frac{\alpha + \beta \exp(\gamma s)}{\gamma}.
\]

The median value of the Gompertz-Makeham distribution has been approximated using the values of \( \alpha \), \( \beta \) and \( \gamma \) estimated with use of the least square estimation in section 3.2 (the unknown parameters \( \alpha \), \( \beta \) and \( \gamma \) are estimated from life tables for Swedish women under the period 1976-1980 [Statistical Abstract, 1983]). The median value is 81.76 years (it is possible to have an approximation with as many digits as wanted).

The mean value of the Gompertz-Makeham distribution is
\[
m_{F,1}(s) = E[S] = \int_0^{\infty} s f(s) ds = \int_0^{\infty} e^{\frac{\alpha s + \beta \exp(\gamma s)}{\gamma}} ds
\]

This integral isn’t a known integral in this manner, so the integral must be simplified to a known integral, otherwise the integral must be solved in a numerical way. This integral is possible to transform to an integral that has a known numerical solution. The result after several transformations shows that it’s possible to rewrite the integral to a well-known integral as shown below.

Set \( u = \frac{\alpha s}{\gamma} \), \( du = \frac{\alpha}{\gamma} ds \Rightarrow \frac{du}{\alpha} = ds \).
\[
\int_0^{\frac{\alpha s + \beta (\exp(\gamma s) - 1)}{\gamma}} e^{\frac{\alpha s + \beta \exp(\gamma s/\gamma)}{\gamma}} ds = e^{\frac{\beta}{\gamma}} \int_0^{\frac{\alpha}{\gamma}} e^{\frac{\beta u + \beta \exp(\gamma u)}{\gamma}} du = \frac{1}{\alpha} e^{\frac{\beta}{\gamma}} \int_0^{\alpha} e^{\frac{\beta u + \beta \exp(\gamma u)}{\gamma}} du
\]

Set \( v = \frac{\beta}{\gamma} e^{\frac{\gamma u}{\alpha}} \), \( dv = \frac{\alpha}{\gamma} du \Rightarrow du = \frac{\alpha}{\gamma} dv \Rightarrow e^{u} = (\frac{\beta}{\gamma})^\frac{\alpha}{\gamma} \) so
\[
\frac{1}{\alpha} e^{\frac{\beta}{\gamma}} \int_0^{\alpha} e^{\frac{\beta u + \beta \exp(\gamma u)}{\gamma}} du = \frac{1}{\alpha} e^{\frac{\beta}{\gamma}} \int_{\beta \frac{\alpha}{\gamma}}^{\alpha} \left( \frac{\beta}{\gamma} \right)^\frac{\alpha}{\gamma} e^{\frac{\alpha}{\gamma} \gamma} dv =
\]
\[
= \frac{1}{\gamma} e^{\frac{\beta}{\gamma}} \left( \frac{\beta}{\gamma} \right)^\frac{\alpha}{\gamma} \int_{\beta \frac{\alpha}{\gamma}}^{\alpha} e^{\gamma^\frac{\alpha}{\gamma} \gamma} dv.
\]

Integrate \( \int_{\beta \frac{\alpha}{\gamma}}^{\alpha} e^{\gamma^\frac{\alpha}{\gamma} \gamma} dv \) by parts \( n \) times so that \( n > \alpha/\gamma \) and \( n < 1 + \alpha/\gamma \) then we have the recurrent relation as below (if \( \alpha < \gamma \) then formula (2.24) is the solution):
\[
\int e^{-v \gamma} \frac{\alpha}{\beta} \frac{1}{\gamma} dv = \gamma \left( e^{\beta \gamma} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) \right) - \int e^{-v \gamma} \frac{1}{\beta} \frac{1}{\gamma} dv = \ldots =
\]

\[
= \frac{\gamma}{\alpha} \left( e^{\beta \gamma} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) \right) + \sum_{j=1}^{n} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) - \sum_{i=1}^{n} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) \int e^{-v \gamma} \frac{1}{\beta} \frac{1}{\gamma} dv.
\]

\[
m_{F_1}(s) = \frac{e^{\beta \gamma}}{\alpha} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) + \sum_{j=1}^{n} \left( \frac{1}{\gamma} - e^{-\frac{\alpha}{\gamma}} \right) \int e^{-v \gamma} \frac{1}{\beta} \frac{1}{\gamma} dv.
\]

If \( \alpha < \gamma \) then

\[
m_{F_1}(s) = \frac{1}{\alpha} \left( 1 - e^{\beta \gamma} \right) \frac{1}{\gamma} \Gamma(1-\alpha/\gamma, \beta/\gamma) = \frac{1}{\alpha} \left( 1 - e^{\beta \gamma} \right) \frac{1}{\gamma} \Gamma(1-\alpha/\gamma, \beta/\gamma)
\]

Where \( \Gamma(\alpha/\gamma, \beta/\gamma) \) denotes an incomplete gamma-function. The incomplete gamma-function has the advantage that it’s a known function that has been evaluated numerical. With help of tables and different packages for computers (for example ”Mathematica”) it’s possible to have any well-approximated value of the function for any combinations of \( \alpha, \beta, \gamma \). By help of ”Mathematica” the mean life time for Swedish women (using life table data for the period 1976-1980) [Statistical abstract, 1983], based on (2.26), is calculated with use of the values of \( \alpha, \beta, \gamma \) estimated in section 3.2 (The values of the parameters are \( \alpha = 5.44 \times 10^3 \), \( \beta = 7.12 \times 10^6 \) and \( \gamma = 0.113 \). The mean life time in this case can be derived from function (2.26), while \( \alpha < \gamma \). It’s very likely that formula (2.26) can be used in all cases when mean life time for human populations are calculated. The mean life time for Swedish women in the given period is 78.72 years. The table value for the mean life time of Swedish women calculated in Statistical Abstract [1983] is 78.51, this value is close to the mean value for the Gompertz-Makeham distribution with estimated parameters. This fact indicates that the mean value for the Gompertz-Makeham distribution is useful.

The relationship between the median and the mean value is researched with the quotient of the median and the mean value for different values of \( \alpha, \beta, \gamma \). The quotient is \( \frac{m_{F_1}(s)}{m_F(s+t)} \) and from (2.21) and (2.25) we can calculate the quotation for given parameters. By use of the estimated values for \( \alpha, \beta, \gamma \) the quotient, estimated for the case of Swedish women described above, the quotation is calculated and the quotation is 81.76/78.72 = 1.039. That means that if the assumption that the life length follows the Gompertz-Makeham distribution is correct more than half of the Swedish women lives at least a little bit longer than the mean life length for Swedish women.

The truncated distributions for the median value and even the mean value, with the truncation at the point \( t = T \) can be of some interest to estimate. The estimations can be useful for different branches, for example for the insurance companies it can be of great use when the life insurance fee is set for different people.
\[ e^{-(\alpha s + \beta \exp(\gamma t)) (\exp(\gamma s) - 1)/\gamma} = \frac{1}{2} \Rightarrow \alpha s + \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma - \ln(2) = 0. \] (2.27)

The truncated distribution of the median value can then, as in the case without having a truncation at \( t = T \) (just set \( T = 0 \) and it will be exactly the case as without truncation, because if \( T = 0 \) then equation (2.27) will be equal to equation (2.21), i.e. \( \alpha s + \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma = \alpha s + \beta \exp(\gamma s)/\gamma = \alpha s + \beta \exp(\gamma s)/\gamma \) ), be used and it will give a approximation of the median value with very good exactness. The approximations of the median value will be done with use of the Newton-Raphson definition (2.22) as described earlier in this section.

The truncated mean value is:

\[
 m_{F,1}(s \mid t \geq T) = E[S \mid t \geq T] = \int_0^\infty s df(s, t) = \int_0^\infty f_0(s) \frac{F_0(s + t)}{F_0(t)} ds = \int_0^\infty \frac{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}}{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}} ds = \int_0^\infty \frac{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}}{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}} ds = \frac{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}}{e^{-\alpha s} \beta \exp(\gamma t) (\exp(\gamma s) - 1)/\gamma}}
\] (2.28)

Set \( \phi = \beta \exp(\gamma t) \), \( \beta \) and \( \exp(\gamma t) \) are constants so that with use of equation (2.24) and changing \( \beta \) with \( \phi \) equation (2.28) will have the same outfit as equation (2.24), but with the difference that \( \beta \) will be changed with \( \phi \) in equation (2.28). If the truncation is at \( T = 0 \) (no truncation at all), the mean life time for the rest of the life time is the same as for new-born babies, namely the equality \( \phi = \beta \exp(\gamma t) = \beta \). The difference between the mean life time left from a truncation at the age \( t \) and the mean life time from day of birth is the constant \( \phi \). The constant \( \phi \) is \( \exp(\gamma t) \) times bigger than the constant for mean life time, the constant \( \beta \), used for the calculation of mean life time from day of birth.

The truncated distribution for the median and the conditional distribution for the mean life time are for the ages 50, 55, 60,..., 85, 90 as shown in table (2.1) below

From the table below the pattern seems to be that with growing age the mean value will begin to decrease less than the median value of life time left at the age \( t \). Somewhere between 65 and 70 mean life time starts to be bigger than the median age left.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( me_F )</th>
<th>( m_{F,1} )</th>
<th>quotient</th>
<th>table value of mean life time</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>32.49</td>
<td>31.04</td>
<td>1.042</td>
<td>30.69</td>
</tr>
<tr>
<td>55</td>
<td>27.70</td>
<td>26.51</td>
<td>1.040</td>
<td>26.22</td>
</tr>
<tr>
<td>60</td>
<td>23.03</td>
<td>22.16</td>
<td>1.032</td>
<td>21.88</td>
</tr>
<tr>
<td>65</td>
<td>18.55</td>
<td>18.05</td>
<td>1.019</td>
<td>17.73</td>
</tr>
<tr>
<td>70</td>
<td>14.38</td>
<td>14.27</td>
<td>0.997</td>
<td>14.11</td>
</tr>
<tr>
<td>75</td>
<td>10.65</td>
<td>10.89</td>
<td>0.964</td>
<td>10.39</td>
</tr>
<tr>
<td>80</td>
<td>7.50</td>
<td>8.00</td>
<td>0.919</td>
<td>7.52</td>
</tr>
<tr>
<td>85</td>
<td>5.02</td>
<td>5.64</td>
<td>0.867</td>
<td>5.30</td>
</tr>
<tr>
<td>90</td>
<td>3.21</td>
<td>3.81</td>
<td>0.816</td>
<td>3.65</td>
</tr>
<tr>
<td>95</td>
<td>1.98</td>
<td>2.48</td>
<td>0.766</td>
<td>2.35</td>
</tr>
</tbody>
</table>

Table (2.1), the median of the life time left and the mean life time left, at age \( t \). From the table for Swedish women earlier mentioned in this section[Statistical Abstract, 1983].
3 Programs for simulation and estimation of parameters

In this section of the work there has been written a program to produce simulated values of life time of the Gompertz-Makeham distribution. This program is explained in section 3.1. A program for the least square estimation has also been written, the program estimates the unknown parameters in the Gompertz-Makeham distribution. Explanation of the program can be found in section 3.2. Descriptions of different methods of estimation can be found in section 3.2, 3.3 and 3.4. The least square estimation is explained in section 3.2. Two other methods of estimation are also described, namely the method of Maximum-Likelihood in section 3.3 and the EM(expectation-maximisation)-algorithm in section 3.4.

3.1 Simulation of the Gompertz-Makeham distribution

In order to find a way of determining different properties of the Gompertz-Makeham distribution there often arises situations where simulation of life times are necessary to do. That’s because of the fact that it can be very difficult to solve some problems without using simulation and in some cases there aren’t even any known methods to solve the problems explicitly without help of simulation. There is also the possibility that the solutions in some cases can be quite complicated to find and the solutions aren’t that much better than the results obtained by simulation. Therefore it is reasonable to simulate the solution instead of solving the problem explicitly. When simulation is used it is most often very important that the number of simulations for life times will be large, otherwise it is possible that too much information will be lost.

The life time will be simulated by splitting the hazard function in two parts. The two parts of the hazard function are $\alpha s$ and $\beta (e^{\gamma s} - 1)/\gamma$. Taking $\alpha$, $\beta$ and $\gamma$ as known parameters, of course when we have real data $\alpha$, $\beta$ and $\gamma$ aren’t known parameters, to make it possible to simulate life times. The unknown parameters must in some way be estimated. By splitting the hazard function in two parts there will be two independent life times simulated. The minimum of this two simulated life times (when the hazard function is split in two parts) is used as the “real” life time. The life times obtained are then useful for investigations of different properties. This simulation study gives possibilities to compare different methods of estimation. For example the expectation and variance of different estimators such as the estimators for the least square estimate, the method of Maximum-Likelihood (ML) and the estimator of the method of moments can be compared and conclusions about the best estimator of those are possible to do.

While the survival function are as stated in (2.1)

$$S(t) = \exp[-\alpha s - \beta].$$

The two parts to be separated are

$$w_1 := \exp[-\alpha s_1]$$

and

$$w_2 := \exp[-\beta \exp((\gamma s_2) - 1)/\gamma].$$

(3.1.1) (3.1.2)

$$H_1(s) = -\alpha s \text{ and } H_2(s) = -\beta$$

From expressions (3.1.1) and (3.1.2), $s$ can be solved explicitly so that $s_1 = \ldots$ and
A normal explanation of the two different parts of the hazard function is that the first part, with the life length $s_1$, describes the risk for death in an accidental way and the other part, with the life length $s_2$, describes the risk for death caused by diseases and other decrements in the healthy state due to the ageing process. This explanation gives us an reason why the hazard function is split in two parts. Therefore an improvement to categories the faces of the different parts in the hazard function is possible to do and it is also possible to give a logical explanation how the deaths are distributed in real life (the percentage of deaths caused by accidents and deaths caused by decrements in the healthy state). Another advantage of splitting the hazard function in two parts is that this simulation technique is much faster than the technique without splitting the hazard function. The faster simulation technique is therefore natural to use while also a logical explanation of the technique exist.

**Theorem (3.1.3)**

If $S = \min(S_1, S_2)$, where $S_1 = -\ln(W_1)$ and $S_2 = \ln(1 - \ln(w_2))$, then

$$H(u) = H_1(u) + H_2(u).$$

$S_1$ and $S_2$ are independent stochastic variables and $W_1$ and $W_2$ are stochastic variables uniformly distributed in the interval $[0, 1]$.

**Note:** $W_1$ and $W_2$ can be thought of as numbers being produced from a random number generator.

**Proof of theorem (3.1.3)**

$$= P(S > u) = P(\min(S_1, S_2) > u) = P((S_1 > u) \cap (S_2 > u)) = P(S_1 > u) P(S_2 > u) = \left[1 - F_1(u)\right] \left[1 - F_2(u)\right] = e^{-H_1(u) - H_2(u)}$$

$H(u) = H_1(u) + H_2(u)$ which was to be proved.

A program has been written in Pascal for simulation of life times. This program is possible to use for different applications according to the Gompertz-Makeham distribution. The program uses the equality $S = \min(U_r, V_r)$. The program has the following construction:

(1) Fix values of $\alpha, \beta$ and $\gamma$ are taken, all of them following statement (2.2), i.e. $\alpha, \beta$ and $\gamma$ are non-negative parameters and $\alpha + \beta \gamma > 0$ and hence $0$ for $s > 0$. For having any set of the data it is necessary to have values of $\alpha, \beta$ and $\gamma$ that are chosen in a realistic way. For example if the value of $\gamma$ is chosen close to 0.2 or bigger, the values of the Gompertz-Makeham distribution for that values will, if even the value of $\beta$ is unsatisfactory, tend to be very large and depending on software the computer won’t have the ability to do all of the calculations that are necessary for a solution.

(2) Set $U_r = -\ln(W_{2r})/\alpha$ and $V_r = \ln(1-\gamma/\beta \ln(W_{2r+1}))/\gamma$. Set $S_r = \min(U_r, V_r)$, $r = 1, \ldots, n$.

Where $U_r$, $V_r$ and $S_r$ denotes the $r$-th value of the simulated life time and $W_r$ is a stochastic uniformly distributed variable in the interval $[0,1]$. $S_r$ will be the simulated value of
life time. The simulated life times are now available to use for different applications.

More information about the program for simulations of life time is possible to find in appendix.

### 3.2 Estimation of the parameters in the Gompertz-Makeham distribution with use of the least square estimation

The least square estimation is used in this section to make it possible to estimate the unknown parameters in the Gompertz-Makeham distribution corresponding to "real" life table data. The least square estimation is a method of estimation that minimises the sum of all squares between known observations and a given function with unknown parameters. The combination of the unknown parameters that has the least sum of squares will then be used as estimated parameters.

In this work a program has been written in the program language Pascal. The program makes it possible to obtain the least square estimation for the unknown parameters (α, β and γ) of the Gompertz-Makeham distribution. The structure of the program is described in the appendix of this work with for example the algorithm of the program. The appendix also contain an example where use of real demographic data has been done. The example makes use of life table data for Swedish women [Statistical Abstract, 1982] under the period 1976-1980. The program that has been written for estimating the least square estimator has got the following overall algorithm (the function that the least square estimation is using is the survival function of the Gompertz-Makeham distribution with unknown parameters α, β and γ):

1. Find values of the three unknown parameters that approximately corresponds to the demographic data that are available, just by testing different values of the unknown parameters.

2. Use the values of the unknown parameters that approximately corresponds to real data and use them, the values are obtained in (1). The values of the parameters will then be tested for different possibilities of the first digit of necessity of the value of the three parameters. The combination of the first number for the three parameters will then be used in next step. Next step will be to test next digit of the parameters and the combination of the parameters that gives the best estimation will be used in next step and this will continue until enough digits of the parameters have been calculated.

3. Repeat (2), but change the "start" value of the parameters to the value of the parameters that was obtained in (2) as the best estimator. Stop this procedure when the least square estimator won’t give more than small enough changes between the two steps. Note, in the program there will be a comparing procedure that changes the least square estimator and the parameters of it if the value of the sum of squares are less than the least value that has been derived earlier.

More about the program is possible to find in the appendix. If the steps in the algorithm is followed it is possible to have the estimated values of α, β and γ with as many correct digits as wanted. The parameters for the example mentioned above are estimated with help of the program and the values of the estimated parameters are α = 5.202*10^{-3}, β = 7.786*10^{-6} and γ = 0.1116. Better values of the parameters is possible to have but the values are good enough for having an acceptable estimation. Figure 1.1 shows the Gompertz-Makeham distribution and real life table data with estimated parameters and it’s easy to understand that the parameters are useful. The plots of the two curves are subjoined in the same plot by use of MatLab. The points for real demographic are in this computer program connected to each other, though the fact that the real demographic data behaves like a discrete function. The ages for the demographic data are 30, 31,....., 79, 80 years.

### 3.3 Estimation of the parameters in the Gompertz-Makeham distribution with use of the method of Maximum-Likelihood
The least square estimation is one of the methods of estimation that can be used to estimate the parameters in the Gompertz-Makeham distribution. Another frequently used method for estimation is the method of Maximum Likelihood.

The method of Maximum Likelihood was first proposed by Gauss in 1821 and is a method that makes sense mostly in parametric models. Suppose \( f(x, \theta) \) is the density function of \( x \) if \( \theta \) is true and that \( \Theta \) is a subset of \( n \) dimensional space. Consider \( f(x, \theta) \) as a function of \( \theta \) for fixed \( x \). This function is normally called the Likelihood function with the notation \( L(\theta|x) \), where \( x \) is thought of as a set of observations. The method of Maximum Likelihood consists of finding the values \( \hat{\theta}(x) \) (\( \hat{\theta}=(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k) \)) which are "most likely" to have produced known data \( L(\hat{\theta}(x), x) \). It is also possible to have the Maximum-Likelihood estimator without assuming that the complete data are known. If \( X=x \), the \( \hat{\theta}(x) \) which satisfies \( L(\hat{\theta}(x), x) = f(x, \hat{\theta}(x)) \) is max \{ \( f(x, \theta) : \theta \in \Theta \) \} is necessary to find.

If such a \( \hat{\theta} \) exists, this value estimates any continuous function \( q(\theta) \) by \( q(\hat{\theta}(x)) \). The estimate \( q(\hat{\theta}(x)) \) is called the Maximum-Likelihood estimation of \( q(\theta) \). The method of Maximum-Likelihood is closer described in several books for example Bickel and Doksum [1977].

The estimation obtained for the Gompertz-Makeham distribution with use of the method of Maximum Likelihood is at the point where the derivative, with respect to all parameters in the model is 0, i.e. the point where \( \frac{\partial f(x, \theta)}{\partial \theta_i} = 0 \) for \( i = 1, 2, 3 \) and \( \theta_1 = \alpha, \theta_2 = \beta \) and \( \theta_3 = \gamma \).

The survival function of the Gompertz-Makeham distribution is, as stated in (2.1),

\[
\overline{F}_0(s) = \exp\left( (\alpha+\beta) \frac{e^{\gamma s} - 1}{\gamma} \right).
\]

The function can be separated in two parts, \( \overline{F}_0(s) = \exp(-\alpha s) \) and \( \overline{F}_0(s) = \exp(-\beta e^{\gamma s} - 1) \), whose have the probability density functions

\[
f_0'(s) = \alpha e^{\gamma s} \exp(-\alpha s) \quad \text{and} \quad f_0''(s) = \beta e^{\gamma s} \exp(-\beta e^{\gamma s} - 1) \]

The probability space of the function are \( x_o = (S_1, \delta_1, S_2, \delta_2, ..., S_n, \delta_n) \), with \( S_i, i=1, ..., n \), known observations and \( \delta_i = I(S_i = S'_i) \), most often there is no knowledge of \( \delta_i \).

The probability of the \( i \)-th outcome is

\[
p_0(S_i, \delta_i) = f_0'(s_i) \overline{F}_0(s_i) \delta_i + f_0''(s_i) \overline{F}_0(s_i)(1-\delta_i) = \alpha \delta_i \exp(-\alpha s_i)(\beta e^{\gamma s_i})^{1-\delta_i},
\]

so the likelihood function of the Gompertz-Makeham distribution is

\[
L(0,x) = \prod_{i=1}^{n} \left( f_0'(s_i) \overline{F}_0(s_i) \delta_i + f_0''(s_i) \overline{F}_0(s_i)(1-\delta_i) \right) = \prod_{i=1}^{n} \alpha \delta_i \exp(-\alpha s_i)(\beta e^{\gamma s_i})^{1-\delta_i}.
\]

From the expression above the loglikelihood function can easily be derived

\[
\ln L(0,x) = -\alpha s_i \ln(\alpha + \beta e^{\gamma s_i}) - \beta \frac{e^{\gamma s_i} - 1}{\gamma}.
\]

The Maximum-Likelihood can be solved with 3 non-linear systems, where \( \frac{\partial \ln L(\theta, x)}{\partial \theta_i} = 0 \) for \( i = 1, 2, 3 \) and \( \theta_1 = \alpha, \theta_2 = \beta \) and \( \theta_3 = \gamma \).

### 3.4 Estimation of the parameters in the Gompertz-Makeham distribution with use of the EM algorithm

Another method that is possible to use for the estimation of parameters of the Gompertz-Makeham distribution is the EM(expectation-maximisation)-algorithm. The EM-algorithm is almost like the method of Maximum Likelihood, but in most cases there will be a simpler way to solve the problems when the EM-algorithm is used instead of the method of Maximum Likelihood.
Consider two statistic models, \( l_0(0,x) \) and \( l_1(0,y) \), where \( x \) isn’t observed and \( y \) is observed and \( y \) is a function of \( x \). With the use of the method of Maximum-Likelihood the systems of equations 

\[
\frac{\partial l_i(\theta, y)}{\partial \theta_i}, \ i=1,2,\ldots,m,
\]

need to be solved. Consider the function 

\[
G(\theta', \theta) = E_{\theta'} l_0(0,X|Y=y).
\]

This function is of course for fixed \( y \) and \( y \) can be observed.

The loglikelihood function can be written in the following form:

\[
l'_1(\theta', Y) = E_{\theta'} l_0(\theta', X) + E_{\theta'} (l_0(\theta', Y) - l_0(\theta', X) - \ln(L_0(\theta', X) / L_1(\theta', X))).
\]

Form the expectation \( E_{\theta'} (l_1(0,Y)|Y=y) = l'_1(\theta', Y) \) then 

\[
E_{\theta'} (l_1(0,Y)|Y=y) = l'_1(\theta', Y) = G(\theta', \theta) - H(\theta', \theta) \text{ with}
\]

\[
H(\theta', \theta) = E_{\theta'} \left( \frac{L_0(\theta', X)}{L_1(\theta', Y)} \right) = E_{\theta'} \left( \ln \left( \frac{L_0(\theta', X)}{L_1(\theta', Y)} \right) \right) = E_{\theta'} \left( \ln(L_0(\theta', X|Y)|Y=y) \right) + E_{\theta'} \ln(L_0(0,X|Y)|Y=y).
\]

It’s easy to proof that this function has maximum at \( \theta' = \theta \), so 

\[
l'_1(\theta', Y) - l_1(0,Y) = G(\theta', \theta) - G(0,0) - (H(\theta', \theta) - H(0,0)) \geq G(\theta', \theta) - G(0,0).
\]

For every \( \theta_m \) \( G(\theta', \theta) \) has the maximum at 

\[
G(0,0) = max G(\theta', \theta(k)) = G(0,0).
\]

Then it is possible to state that 

\[
l'_1(\theta_m,Y) - l_1(0,Y) \geq G(0,0) - G(0,0) > 0.
\]

From the above expression it is possible to state the two-step-approach’s algorithm:

(E): In a neighbourhood of \( \theta(k) \) the function \( G(\theta', \theta(k)) \) is determined.

(M): The neighbourhood \( \theta(k+1) \) is found where 

\[
G(\theta(k+1),\theta(k)) = max_{\theta} G(\theta', \theta(k)) \times G(\theta(k),\theta(k)).
\]

If the maximum can’t be determined in this way the approach \( \theta(k+1) \) are the estimation of the parameters obtained by the EM-algorithm.

Estimators of the parameters for the Gompertz-Makeham distribution can be determined with help of the EM-algorithm.

\[
G(\theta', \theta) = E_{\theta'} (l_0(\theta', x_0)) S_i = s_i, i = 1,\ldots, n = (\sum_{i=1}^{n} \alpha + \beta e^{\gamma s_i}) \ln \alpha' - \alpha' \sum_{i=1}^{n} s_i + \left( \sum_{i=1}^{n} \frac{\beta e^{\gamma s_i}}{\alpha + \beta e^{\gamma s_i}} \right) \ln \beta' + \gamma' \sum_{i=1}^{n} \frac{s_i \beta e^{\gamma s_i}}{\alpha + \beta e^{\gamma s_i}} - \sum_{i=1}^{n} \exp(-\left(\beta^{\gamma s_i} - 1\right)^{-1} \gamma')).
\]

which can be rewritten as 

\[
A \ln \alpha' - B \alpha' + C \ln B' + D \gamma' \cdot E \left( \beta', \gamma' \right), \text{ where } A = \sum_{i=1}^{n} \frac{\alpha}{\alpha + \beta e^{\gamma s_i}}, B = \sum_{i=1}^{n} s_i, \text{ and } E \left( \beta', \gamma' \right) = \sum_{i=1}^{n} \exp(-\left(\beta^{\gamma s_i} - 1\right)^{-1} \gamma')).
\]

Now \( G(\theta', \theta) \) can be written in the form 

\[
G(\theta', \theta) = G_1(\alpha', \alpha, \beta, \gamma') + G_2(\beta', \gamma'; \alpha, \beta, \gamma) \text{ with } G_1(\alpha'; \alpha, \beta, \gamma') = A \ln \alpha' - B, \text{ and } G_2(\beta', \gamma'; \alpha, \beta, \gamma) = C \ln B' + D \gamma' - E \left( \beta', \gamma' \right).
\]

To find the estimations of the parameters with help of the EM-algorithm the functions \( G_1 \) and \( G_2 \) will be maximised. The maximum is at the points where \( G_1 \) and \( G_2 \) are maximised. The functions are maximised where 

\[
\frac{\partial G(\theta', \theta)}{\partial \alpha} = \frac{A}{\alpha'}, B = 0, \text{ and } \frac{\partial G(\theta', \theta)}{\partial \beta} = \frac{C}{\beta'}, \sum_{i=1}^{n} e^{\gamma s_i} - 1 = 0.
\]
\[
\frac{\partial G(\theta', 0)}{\partial \beta'} = D - \beta' \left( \sum_{i=1}^{n} s_i e^{\gamma' s_i} - e^{\gamma' s_i} - 1 \right) = 0.
\]

Set \( \alpha = \alpha(k), \beta = \beta(k) \) and \( \gamma = \gamma(k) \) and with use of the maximisation-step in the algorithm the values of \( \alpha', \beta' \) and \( \gamma' \) are set to \( \alpha' = \alpha(k+1), \beta' = \beta(k+1) \) and \( \gamma' = \gamma(k+1) \).

The solution will then have the following EM-algorithm

\[
\begin{align*}
\alpha(k+1) &= \frac{B(k)}{A(k)}, \\
\beta(k+1) &= \frac{C \gamma'(k+1)}{1 + \sum_{i=1}^{n} \left( e^{\gamma'(k+1)s_i} - 1 \right)}, \\
\gamma(k+1) &= \frac{D(k)}{C(k)} = \frac{\sum_{i=1}^{n} s_i e^{\gamma'(k+1)s_i}}{1 + \sum_{i=1}^{n} \left( e^{\gamma'(k+1)s_i} - 1 \right)}.
\end{align*}
\]

For the first two, \( \alpha(k+1) \) and \( \beta(k+1) \), a solution can be decided explicitly from the formulas while \( \gamma(k+1) \) has to be solved numerically. Now it is possible to estimate the parameters of the Gompertz-Makeham distribution with respect to the EM-algorithm. More about the EM-algorithm is possible to find in Cox and Oakes [1984] and Belyaev and Kahle [1996].

4. Testing of some hypotheses

Of interest it is if the Gompertz-Makeham distribution with estimated parameters (parameters that are mainly estimated for the ages between 30 and 80) is acceptable even for extreme old ages. Is there a possibility to use the estimated parameters of the Gompertz-Makeham distribution even for extreme old ages or not? With different well-known tests it’s possible to test if the estimated parameters of the Gompertz-Makeham distribution are possible to use even for extreme old ages. Three of the tests that are possible to use are the goodness of fit test, the Kolmogorov test and the likelihood ratio test. There is a big disadvantage with the Kolmogorov test, namely that most often there has to be more information about the age than whole years for having the possibility to use the Kolmogorov test.

The hypothesis that the Gompertz-Makeham distribution with estimated parameters can be tested also for extreme old ages is tested versus the hypothesis that the estimated parameters can’t be used for these extreme old ages. If the hypothesis that life table data for extreme old ages follows the Gompertz-Makeham distribution with estimated parameters is rejected, then those parameters can’t be used even for extreme old ages. If that is the case, the best choice maybe is to use another function to have approximation of extreme old ages. If the Gompertz-Makeham distribution describes all but the extreme old ages and another function describes the extreme old ages, there will be a truncated function. The truncated function must have the characteristic of a distribution function, i.e. \( 0 < f(s_i) < 1 \) and \( \sum_{i=1}^{k} f(s_i) = 1 \) for \( i = 1, \ldots, k \) (it’s possible that \( k = \infty \)). Note, two other possibilities are that instead of a truncated function another function than the Gompertz-Makeham distribution is used for the whole age interval or the Gompertz-Makeham distribution is used with a more likely set of the estimated parameters. Another function than the Gompertz-Makeham distribution isn’t of interest in this work while it seems like the Gompertz-Makeham distribution gives the best description of life table data.

4.1 The Goodness of fit test

One of the methods that are possible to use for the test of such a hypothesis is the goodness of fit test. The (multinomial) goodness of fit test is a test that makes use of a test statistic called Pearson’s \( \chi^2 \). This statistic is

\[
\chi^2 = N \sum_{i=j} \frac{(f_0(s_i|s_j) - f_0(s_j|s_i))^2}{f_0(s_i|s_j)}.
\]
Where \( N \) is the total population size of the researched ages. \( f_\theta(s|s_j) \) is real demographic data for death at the ages \( s_j \), \( s_{j+1} \) and death at ages older than \( s_{k-1} \). \( s_k \) denotes the probability to die after the age \( s_{k-1} \), given that alive at age \( s_j \). \( f_\theta(s|s_j) \) is the value of the conditional probability density function of the Gompertz-Makeham distribution at age \( s_j \), given the condition that \( s_j \geq s_j \). Note: \( \sum_{i=j}^{k} f_\theta(s_i|s_j) = 1 \). The test statistic will be compared with \( \chi^2(k-j-1) \) for given significance level. If the value is bigger than \( \chi^2(k-j-1) \) the test will reject the hypothesis.

A goodness of fit test has been done as described in formula (4.1) testing that the demographic data explained in section 3.2 follows the Gompertz-Makeham distribution with estimated parameters. The test is for ages over 95 years and the hypothesis is that ages over 95 years follows the Gompertz-Makeham distribution with the parameters estimated in section 3.2. Figure 4.1 below shows a plot with both the survival function of the Gompertz-Makeham distribution (with the parameters estimated by use of the least square estimation in section 3.2) and values of life time from life table data for women [Statistical Abstract, 1983] for ages between 95 and 100.

By the outfit of Figure 4.1 it’s reasonable to assume that life table data for the ages between 95 and 100 doesn’t follow the Gompertz-Makeham distribution with the estimated parameters. This motivates the fact that there is a need to use different tests to prove this statement. A (multinomial) goodness of fit test for life table data of Swedish women has been done with help of a program written in MatLab. See appendix for more information about the structure of the program. The goodness of fit test has been done assuming that the population of the life table data are of size 4673 women (assuming that the total population of women are 100000, the total population is definitely much bigger than this) and the values of \( f_\theta(s|s_j) \) are calculated with help of the least square estimation of the Gompertz-Makeham distribution, with \( \hat{\Theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (5.202 \times 10^{-4}, 7.786 \times 10^{-6}, 0.1116) \). The ages that are used in the goodness of fit test is 95, 96, 97, 98, 99 and \( \geq 100 \) years. The result of the goodness of fit test is 20.02. This value is compared with the value for the 1 percentage significance level of the \( \chi^2 \)-distribution. The value for the 1 percentage level of the \( \chi^2 \)-distribution is:

\[
\chi^2_{0.05}(k-j-1) = 15.09.
\]

The hypothesis that life table data for extreme old ages follows the Gompertz-Makeham distribution with the parameters \((\hat{\alpha}, \hat{\beta}, \hat{\gamma})\) can be rejected at the 1 percentage level and therefore at all relevant levels. The conclusion is that the Gompertz-Makeham distribution with estimated parameters can not be used for extreme old ages. There has also been tested if the ages 90, 91, 92, 93, 94 and \( \geq 95 \) years corresponds to this estimated parameters and the result of the goodness of fit test for those ages is 56.18.

The probability to die at the ages 95, 96, 97, 98, 99 and \( \geq 100 \) years, given the condition that still alive at 95 years of age are:

\[
\begin{align*}
&f(d = 95 \mid \text{die at age 95}) \mid s \geq 95 = 0.296, \\
&f(d = 96 \mid s \geq 95) = 0.230, \\
&f(d = 97 \mid s \geq 95) = 0.171, \\
&f(d = 98 \mid s \geq 95) = 0.121, \\
&f(d = 99 \mid s \geq 95) = 0.081 \text{ and } f(d \geq 100 \mid s \geq 95) = 0.1025.
\end{align*}
\]
4.2 The Kolmogorov test

Another test that is possible to use is the Kolmogorov test. The Kolmogorov test uses a test statistic that needs more information about the ages (age in days or weeks instead of years) than the goodness of fit test does. Therefore it is more difficult to use the Kolmogorov test, while ages in life tables most often are in whole years. If there’s life table data in more detailed form than whole years the Kolmogorov test is a good test statistic to use.

If there is no knowledge of the distribution function F, the natural estimate of the probability F(x) is defined by 
\[ \hat{F}_n(x) = \text{number of } X_i \leq x \]
Since \[ n \hat{F}_n(x) \] has a binomial, \[ B(n,p) \], distribution with \[ p = F(x) \], \[ \hat{F}_n(x) \] is an unbiased estimate of \[ F(x) \] with variance \[ F(x)[1-F(x)]/n \]. The estimate is consistent, \[ \hat{F}_n(x) \to F(x) \] and is asymptotically normal for each \( x \). The curve \( \hat{F}_n(.) \) is itself a distribution function.

If the sample values are \( X_1 = x_1, \ldots, X_n = x_n \) and \( X \) is a variable with distribution function \( \hat{F}_n(.) \), then
\[ P[X=x_i] = \frac{1}{n}, \text{ for } i = 1, \ldots, n. \]
\( \hat{F}_n(.) \) is called the empirical distribution function of the sample \( X_1, \ldots, X_n \)

Not only is \( \hat{F}_n(x) \) consistent for each \( x \), but by the Glivenko-Cantelli theorem [Fisz, 1963],
\[ \sup_{x \in [0,\infty]} | \hat{F}_n(x) - F_0(x) | \to 0. \]
The Kolmogorov test statistic for testing the hypothesis \( H_0 : F = F_0 \) is defined by \( D_n \)
\[ D_n = \sup_{x \in [0,\infty]} | \hat{F}_n(x) - F(x) | \to 0. \]
It’s reasonable to reject \( H_0 \) in favour of \( H_1 : F \neq F_0 \) for large values of \( D_n \).

For large \( n \) it’s better to use \( \sqrt{n} D_n \), which converges to the so-called Kolmogorov’s distribution \( K(y) \). In the case of real demographic data for Swedish women [Statistical Abstract, 1983] the Kolmogorov test statistic isn’t a good test statistic to use while the data are not as exactly as wanted. The test will be more useful if life times are in days instead of years. In that case it will be more close to the continuity assumption that the Kolmogorov statistic has. It also should be noted that instead of using "exact" unknown parameters we use the estimators of the estimators.

4.3 Likelihood ratio test

Another very useful method of testing parametric hypothesis, such as \( H_0 : \theta_0 \in \Theta \) versus \( H_1 : \theta \in \Theta_1 \), is the likelihood ratio test. The test statistic that is considered for the likelihood ratio test is given by
\[ L(x) = \frac{\sup \{ p(x, \theta) : \theta \in \Theta_1 \} }{\sup \{ p(x, \theta) : \theta \in \Theta_0 \} } \]
Tests that reject \( H_0 \) for large values of \( L(x) \) are called likelihood ratio tests.

The likelihood function \( L(\theta, x) = p(x, \theta) \) can be thought of as a measure of how well \( \theta \) explains the given sample \( x = (x_1, x_2, \ldots, x_n) \). So, if \( \sup \{ p(x, \theta) : \theta \in \Theta_1 \} \) is large compared to \( \sup \{ p(x, \theta) : \theta \in \Theta_0 \} \), the observed sample is better explained by some \( \theta \in \Theta_1 \), and conversely.

When the function that is considered is the Gompertz-Makeham distribution, \( p(x, \theta) \) is a continuous function of \( \theta \) and \( \Theta_0 \) is of smaller dimension than \( \Theta = \Theta_0 \cup \Theta_1 \) so that the likelihood ratio test equals the test statistic
\[ \lambda(x) = \frac{\sup \{ p(x, \theta) : \theta \in \Theta_1 \} }{\sup \{ p(x, \theta) : \theta \in \Theta_0 \} } \]
and this statistic is most often not too difficult to compute.

The four basic steps to evaluate the test statistic is
(1) Calculate the Maximum Likelihood estimate \( \hat{\theta} \) of \( \Theta \);
(2) Calculate the Maximum Likelihood estimate $\hat{\theta}_0$ of $\Theta_0$.

(3) Form $\lambda(x) = p(x, \hat{\theta}) / p(x, \hat{\theta}_0)$.

(4) Find a function $h$ which is strictly increasing on the range of $\lambda$ such that $h(\lambda(x))$ has a simple form or a tabled distribution under $\Theta_0$. Since $h(\lambda(x))$ is equivalent to $\lambda(x)$ the likelihood test with size $\alpha$ is specified with the $(1-\alpha)$:th quantile of $h(\lambda(x))$ and is possible to find in a table. More about the theory behind the likelihood ratio test can be found in e.g. Bickel and Doksum [1977].

For example to test if $\alpha = \alpha_0$, given $\beta = \beta_0$ and $\gamma = \gamma_0$ versus $H_1: \alpha \neq \alpha_0$, given $\beta = \beta_0$ and $\gamma = \gamma_0$, for the case of the Gompertz-Makeham distribution, a simple hypothesis is constructed of the form

$$
\lambda(s) = \sup_{\Theta} \{ p(s, \alpha, \beta_0, \gamma_0): \alpha \in \Theta \} = \sup_{\Theta_0} \{ p(s, \alpha, \beta_0, \gamma_0): \alpha \in \Theta_0 \} = \frac{\sup_{\Theta} \{ p(s, \alpha, \beta_0, \gamma_0): \alpha \in \Theta \}}{p(s, \alpha_0, \beta_0, \gamma_0)}.
$$

While $\sup_{\Theta} \{ p(s, \alpha, \beta_0, \gamma_0): \alpha \in \Theta_0 \}$ contains only constants, besides the time observations $s$, it is possible to rewrite the function to $p(s, \alpha_0, \beta_0, \gamma_0)$. $\sup_{\Theta} \{ p(s, \alpha, \beta_0, \gamma_0): \alpha \in \Theta_0 \} = p(s, \hat{\alpha}, \beta_0, \gamma_0)$, where $\hat{\alpha}$ denotes the Maximum-Likelihood estimate of $\alpha$. The Maximum-Likelihood estimate of the Gompertz-Makeham distribution is possible to have with help of numerical methods.

If $\lambda(s)$ is large the hypothesis $H_0: \alpha = \alpha_0$ is rejected versus $H_1: \alpha \neq \alpha_0$. The value of $\lambda(s)$ is compared with table values of the $(1-\alpha)$:th quantile of $h(\lambda(S))$. Where $S$ is a random variable having the Gompertz-Makeham distribution. Wished significance level can then be tested.
5 References


