

A Review of the Methodologies Used in the Derivation of Formulas for Parametric Survival Functions with Illustrative Numerical Examples

Charles J. Mode¹ & Candace K. Sleeman²

¹ Professor Emeritus, Department of Mathematics, Drexel University, Philadelphia, PA 19104, USA

² Director of Academic Analytics, Southern New Hampshire University, 33 S Commercial Street, Manchester, NH 03101, USA

Correspondence: Charles J. Mode, Professor Emeritus, Department of Mathematics, Drexel University, Philadelphia, PA 19104, USA

Received: December 11, 2018 Accepted: January 17, 2019 Online Published: January 28, 2019

doi:10.5539/ijsp.v8n2p99 URL: <https://doi.org/10.5539/ijsp.v8n2p99>

Abstract

In this review paper formulas for survival functions are derived that take into account risks of deaths in early life including infancy, mid life, a random component of deaths due to accidents and deaths in older ages. The basic ideas used in the derivation of survival function for each of the components just mentioned are risk functions. Given a formula for a risk function, it is possible to derive a formula for the corresponding survival function. By using the theory of competing risks, a formula for survival function that takes into account the risks of deaths in various stages of life expressed as a product of survival functions for the risks of deaths under consideration. For many applications information on the numerical values of parameters in survival functions is not available. Consequently, rationales are developed for assigning plausible values to parameters that take into account personal ideas of an investigator may have for each stage of life. For every assignment of parameter values in the paper, a numerical version of survival functions are plotted in graphs so that an assessment of the plausibility of the chosen parameter values may be made. Also included in the paper is an application of survival functions in an experiment to make an assessment as to whether a small population of chimpanzees, or some other endangered species of animals, will have descendants that make up a surviving population 200 years into the future.

Keywords: risks functions, survival functions, Gompertz, Makeham, competing risks, survival of age structured populations

1. Introduction

In computer simulation experiments designed to study the evolution of age structured animal or human populations with respect to two or more genetic autosomal loci in deep time, such as hundreds or thousands of years before the present, it is almost always the case that no mortality data exists. Consequently, an investigator will need to rely on model parametric survival functions. As is well known, efforts have been made to construct parametric survival functions for nearly the past two centuries. Among the first such efforts was that of Gompertz B. Gompertz 1825. As Gompertz examined human survival data at ages 70 or more, he came to the realization that the risk of death increased exponentially for the older ages. In 1869 Makeham published a paper indicating that a positive constant should be added to Gompertz's observation to take into account deaths by accidents. In 1872, Thiele proposed a formula for the risk of death for the whole life. For further details on the proposals of Gompertz and Makeham, the book by Smith and Keyfitz *The work of Thiele* is presented In the book and the formula that Thiele proposed is presented on page 72.

In 1982 Mode and Busby published a paper on an eight parameter model of human mortality. In this model attention was focused on three stages of life in humans. The first stage was that for early life for ages x in the interval $[0, 10)$ expressed in years. The second stage was for midlife was ages in the interval $(10, 30]$, and that for the older ages was the interval $(30, k]$, where $k > 0$ was the greatest age considered in a model. For humans, this age is often chosen as $k = 100$. A second paper on formulation of a of parametric survival function was that of Mode and Jacobson (1984) in which age intervals mentioned above were also used. Although the formula for the parametric function used in the two papers for the same three age intervals under consideration were different, the same method of splicing the survival functions together was the same for both papers, which provided a parametric survival function for the entire age interval $[0, k]$. Although both models fit survival functions estimated from many period life tables, they are also mildly flawed, because of small discontinuities at the splice points. The two papers just mentioned were included in the references in the paper by Gage and Mode (1993) along with several other parametric survival functions, which were tested for goodness of fit to data.

It should be emphasized, however, that in this paper the focus of attention will be based on rationals for constructing theoretical parametric survival functions that will be used in computer simulation experiments in evolutionary genetics and other fields in biology when no survival data are available.

In the paper Mode et al. 2013 a survival function related to that of Mode and Busby was used in a paper on a stochastic model of a two sex age structured population with a genetic component. A four parameter parametric survival function that was used to produce the content of the paper was based on partitioning the life span of each individual into two age intervals $[0, 30)$ and $[30, r)$. For this model, age $x = 30$ was a splice point and there was a noticeable discontinuity at that age point that appeared in age distribution estimated from data generated in Monte Carlo simulation experiments as well as those computed using the embedded deterministic model. Although it seems unlikely that the conclusions of the paper based on computer experiments would change significantly if another survival model was used to remove such discontinuities, in future experiments it would be advantageous to use a parametric survival function that did not have noticeable discontinuities. Moreover, it seems likely that it would also be beneficial to readers of the paper if these annoying discontinuities were removed.

Even though it was realized that there was need for parametric survival functions that did not have any noticeable discontinuities, no sustained efforts were made to formulate parametric survival functions with no discontinuities throughout the ages considered in future computer simulation experiments. However, some time in 2015 in a search for APL functions stored in folders on flash memory sticks and other storage devices, by chance an old folder was found containing APL software designed to implement parametric survival functions with no discontinuities. As it turned out the software had been written in connection with research on aging funded by the National Institute of Aging during the late 1980s. By reading the APL code, it was realized that some potentially significant issues were not considered in the software so it was rewritten to make the output more informative as to the significance of the ideas contained in the written code. An inspection of the numerical output of the revised code led to a need to further develop the software to fix potentially flawed ideas. In the remaining sections of this paper, these potential flaws will be investigated and alternative procedures will also be investigated.

2. A Review of Fundamentals

There is a large literature on survival analysis, but it beyond the scope of this paper to provide review of this literature. There is also a rather large number of books on the subject, and among them are Elandt-Johnson and Johnson (1980) If a reader is interested in a more in depth account of the ideas presented in this section, it is recommended that this book be consulted. To make this paper self contained, in this section attention will be focused on the basic fundamentals and definitions that are necessary to provide a background for the contents of this paper.

Let X denote a random variable representing the life span of any individual in some biological population. Because age is measured in terms of a number in the interval $R^+ = [0, \infty]$, the range of X is this interval. By definition the distribution function of the random variable X is

$$P[X \leq x] = F(x) \tag{2.1}$$

for all $x \in R^+$, and is the probability that an individual dies in the age interval $[0, x)$. It will be assumed that $F(0) = 0$ and

$$\lim_{x \uparrow \infty} F(x) = 1. \tag{2.2}$$

Any new born individual in the population will be assigned the age $x = 0$. Let $S(x)$ denote the probability that any individual is alive at age $x > 0$. Then

$$S(x) = P[X > x] = 1 - F(x) \tag{2.3}$$

is called the survival function, and it follows that $S(x) \rightarrow 0$ as $x \rightarrow \infty$.

If $F(x)$ is differentiable, then there is a function $f(x)$ such that

$$\frac{dF(x)}{dx} = f(x) \tag{2.4}$$

for all $x \in R^+$. The function $f(x)$ is called the probability density function *p.d.f.* It also follows that

$$F(x) = \int_0^x f(s) ds \tag{2.5}$$

for all $x \in R^+$. From this equation, it follows that if one constructs or defines a probability density function $f(x)$ that is defined for all $x \in R^+$ and has the property

$$\int_0^\infty f(x) dx = 1, \tag{2.6}$$

then

$$\lim_{x \uparrow \infty} F(x) = 1. \tag{2.7}$$

A well known example of a *pdf* is that of the exponential distribution which has the form

$$f(x) = \beta e^{-\beta x} \tag{2.8}$$

for $x \in R^+$, where $\beta > 0$ is a positive parameter. The distribution function of this distribution is

$$F(x) = \int_0^x \beta e^{-\beta s} ds = 1 - e^{-\beta x}, \tag{2.9}$$

and the survival function is

$$S(x) = e^{-\beta x} \tag{2.10}$$

for all $x \in R^+$. From these equations, it can be seen that $F(0) = 0$, $S(0) = 1$. Moreover, $F(x) \rightarrow 1$ and $S(x) \rightarrow 0$ as $x \rightarrow \infty$.

A central concept in the formulation of parametric survival functions is the idea of a risk function. As a step toward defining this function, consider the conditional probability

$$P[x \leq X \leq x + h \mid X > x] \tag{2.11}$$

and let $F(x)$ denote the distribution of the random variable X and let $f(x)$ denote $f(x)$ the probability density function of the distribution of X , where $x > 0$ and $h > 0$. Then

$$P[x \leq X \leq x + h \mid X > x] = \frac{F(x + h) - F(x)}{1 - F(x)}. \tag{2.12}$$

By definition, the risk function $\theta(x)$ is defined by the limit

$$\lim_{h \downarrow 0} \frac{F(x + h) - F(x)}{1 - F(x)} = \frac{f(x)}{1 - F(x)} = \theta(x) \tag{2.13}$$

for $x > 0$. An alternative definition of the risk function is

$$-\frac{d \ln(1 - F(x))}{dx} = \frac{f(x)}{1 - F(x)} = \theta(x). \tag{2.14}$$

From this definition, it follows that

$$S(x) = 1 - F(x) = \exp \left[- \int_0^x \theta(s) ds \right]. \tag{2.15}$$

By taking the derivative of both sides of this equation, it can be seen that the *pdf* has the form

$$\frac{dF(x)}{dx} = f(x) = S(x)\theta(x) \tag{2.16}$$

for $x > 0$.

From the definition of the risk function, it can be seen that if $f(x)$ is the *pdf* of some distribution that if a formula for $f(x)$ is given, then a formula for the risk function may be derived. Consider, for example, the exponential distribution described above. Then

$$\theta(x) = \frac{f(x)}{S(x)} = \frac{\beta e^{-\beta x}}{e^{-\beta x}} = \beta > 0 \tag{2.17}$$

for all $x > 0$. From this general formula connecting the survival function and the risk function, it can be seen that the survival function of the exponential distribution is

$$S(x) = e^{-\beta x} \tag{2.18}$$

as given above.

There are a number of points of view that may be implemented in connection with formulating models of parametric survival functions. The following example from reliable theory, which is used in connection with quality control in

manufacturing, will provide a basis for formulating the parametric survival function presented in this paper. Suppose some device has four components, and let the random variables X_k for $k = 1, 2, 3, 4$ taking values in the interval $0 \leq x < \infty$ denote that random life spans of the four components. Suppose these random variables are independently distributed with survival functions $S_k(x)$ for $k = 1, 2, 3, 4$. Then the survival function of the system is

$$S(x) = \prod_{k=1}^4 \lim_{s \rightarrow \infty} S_k(s) \tag{2.19}$$

When any component of the system fails, the system fails. Consequently, the life span of the system is given by the random variable

$$Y = \min \{X_1, X_2, X_3, X_4\} \tag{2.20}$$

Observe that $Y > y$, if and only if, $X_k > y$ for every $k = 1, 2, 3, 4$. Therefore, the survival function of the system is

$$S(y) = \prod_{k=1}^4 S_k(y) \tag{2.21}$$

for $y > 0$.

Let $\theta_k(x)$ denote the risk function for component $k = 1, 2, 3, 4$. Then the survival function of the system has the form

$$S(y) = \prod_{k=1}^4 S_k(y) = \prod_{k=1}^4 \exp \left(- \int_0^y \theta_k(s) ds \right) = \exp \left(- \sum_{k=1}^4 \int_0^y \theta_k(s) ds \right) \tag{2.22}$$

Therefore, from the right most formula in 2.22 it can be seen that

$$\frac{d \ln S(y)}{dy} = - \sum_{k=1}^4 \theta_k(y) \tag{2.23}$$

A procedure that has been widely used in the formulation of parametric survival functions is to define the risk function under consideration and then use expression 2.23 in deriving a symbolic form of a survival function.

The survival function of the Gompertz distribution has the form

$$S_G(y) = \exp[-\alpha (\exp \beta y - 1)] \tag{2.24}$$

where α and β are positive parameters. As can be seen from this formula, this function has the property

$$\lim_{y \rightarrow \infty} S_G(y) = 0 \tag{2.25}$$

It is well known that the Gompertz survival function fits the survival function that is estimated in many period life tables of ages $y \geq 30$. Throughout this paper, the Gompertz survival function will be used as a basis for formulating parametric survival functions for all ages $y > 0$.

Because it fits empirical survival functions for ages $y \geq 30$, it follows that the function needs to be modified only for ages $0 < y < 30$. As will be shown in subsequent sections of this paper, the modifying functions may not converge to zero as $y \rightarrow \infty$. But as can be seen from the right most expression in equation 2.22, if the Gompertz survival function is one of the survival functions for the system, then

$$\lim_{y \rightarrow \infty} S(y) = \prod_{k=1}^4 S_k(y) = 0 \tag{2.26}$$

In the following sections of this paper, several examples of the situation just described will be presented.

3. A Brief History of Proposed Risk Functions for Early Life

The form of the risk function for early life model proposed by Thiele in 1871 was

$$\mu_T(x) = \alpha_0 e^{-\beta_0 x} \tag{3.1}$$

where the parameters $\alpha_0 > 0$ and $\beta_0 > 0$ are positive. As can be seen from this formula $\mu_T(x) \rightarrow 0$ as $x \rightarrow \infty$. The integral of this risk function is

$$H_T(x) = \int_0^x \mu_T(s) ds = \frac{\alpha_0}{\beta_0} (1 - \exp[-\beta_0 x]) \tag{3.2}$$

for $x > 0$. The survival function in this case is by definition

$$S_T(x) = \exp[-H_T(x)] \tag{3.3}$$

As can be seen from 3.2, it follows that

$$\lim_{x \rightarrow \infty} S_T(x) = \exp\left[-\frac{\alpha_0}{\beta_0}\right] \tag{3.4}$$

Technically, $S_T(x)$ is not a survival function, because $S_T(x)$ does not converge to 0 as $x \rightarrow \infty$. However, it could be used to adjust the Gompertz survival function for the ages $0 < x < 30$ as indicated in 2.24 in section 2. Siler (1970). also used the risk function in 3.1 in his model of animal mortality.

The formula for risk of death in early life in this paper is the modified version that proposed by Thiele and has the form

$$\mu_0(x) = \alpha_0 \beta_0 e^{-\beta_0 x} \tag{3.5}$$

for $x > 0$. The integral of this risk function is

$$H_0(x) = \int_0^x \mu_0(s) ds = \alpha_0 (1 - \exp[-\beta_0 x]) \tag{3.6}$$

for $x > 0$. Just as in 3.3, the survival function in this formulation has the definition $S_0(x) = \exp[-H_0(x)]$. As can be seen in this case

$$\lim_{x \rightarrow \infty} H_0(x) = \alpha_0 \tag{3.7}$$

When data are not available, then an investigator needs to explain the rationale used to justify his procedure for choosing numerical values of the parameters. Let p_0 denote the probability that an infant born as age $x = 0$ survives to age 1 year, and suppose that

$$e^{-\alpha_0} = p_0 \tag{3.8}$$

Then

$$\alpha_0 = -\ln p_0 \tag{3.9}$$

For the case of $p_0 = 0.8$,

$$\ln .8 = -0.223 14 \tag{3.10}$$

Hence, for this choice of p_0 , the parameter α_0 has the value

$$\alpha_0 = 0.223 14 \tag{3.11}$$

If an investigator is working with human data and if some period life table are available for a few past centuries, then it would be possible to use this data as a guide for the choice of the numerical value for the parameter p_0 .

The next step is to consider a rationale for choosing a value for the parameter β_0 . As can be seen from equation 3.6, the expression $(1 - \exp[-\beta_0 x])$, decreases the value of the parameter α_0 by a fraction for each $x > 0$. Suppose, for example, that at age 3 years the equation

$$(1 - \exp[-\beta_0 3]) = 0.5 \tag{3.12}$$

holds. Then

$$\exp[-\beta_0 3] = 1 - 0.5 = 0.5, \tag{3.13}$$

and β_0 has the value

$$\beta_0 = -\frac{\ln 0.5}{3} = 0.23105, \tag{3.14}$$

so that the integral of the risk function has the value at $x = 3$.

$$H_0(3) = 0.223 14 \times 0.23105 = 5.1556 \times 10^{-2} \tag{3.15}$$

Observe that this value is smaller than the value chosen for $\alpha_0 = 0.22314$.

As can be seen from 3.7 and 3.11,

$$\lim_{x \rightarrow \infty} H_0(x) = \alpha_0 = 0.223 14 > 5.1556 \times 10^{-2} \tag{3.16}$$

From formula 3.6, it can be seen that $H_0(x)$ increases for any chosen value of the parameter β_0 , but will eventually converge to the limit α_0 .

In studies of human mortality in computer simulation experiments, ages from 1 to 100 may be under consideration. In such experiments, it has been observed that the value of $H_0(100)$ is the limit α_0 . For the illustrative numerical example under consideration this value is

$$H_0(100) = 0.22314 . \tag{3.17}$$

Consequently,

$$S_0(100) = \exp[-H_0(100)] = \exp[-0.22314] = 0.8 . \tag{3.18}$$

From equation 2.21, it can be seen that if the survival function $S_0(x)$ were included in this product, then it may reduce the values of the Gompertz survival more than intended for ages $x > 30$. If an investigator, thinks this it is valid for the situation under consideration, then the survival function $S_0(x)$ would be included in the product in 2.21 for all ages $x \geq 1$. For example in such a formulation of a model on evolutionary genetics, the function $S_0(x)$ could represent a genotype such that individuals of this genotype are more at risk of dying than those of other genotypes so that this survival function would be considered be valid.

An alternative to formulating a survival function is presented in Mode and Busby . A brief history of risk functions just outlined is incomplete. In this connection, it is suggested that an interested reader have a look at the paper of Mode and Jacobson (1984) and the references cited therein. Briefly, the risk of death function used in that paper for early life was related to the probability density function of the Weibull distribution, which was much more complicated than the one in 3.5. In the early planning stages for this paper, a decision was made to use risks functions that have only a few parameters in order to minimize the number of parameters that must be assigned values for computer experimenters connected with evolutionary genetics.

4. A Brief History of Risk Functions That Have Been Proposed for Midlife

In his proposal for a risk function for the whole life, Thiele (1872) **thei** offered the risk function of the from

$$\mu_1(x) = \alpha_1 \exp \left[- \left(\frac{1}{2b_1^2} (x - c_1)^2 \right) \right] \tag{4.1}$$

for mid-life. As can be seen from this formula is nearly that of the probability density function of a normal distribution with mean c_1 and variance b_1^2 . If the formula below form is chosen at the risk function

$$\mu_1(x) = \alpha_1 \frac{1}{b_1 \sqrt{2\pi}} \exp \left[- \left(\frac{1}{2b_1^2} (x - c_1)^2 \right) \right] , \tag{4.2}$$

then it would be a multiple the probability density function of a normal distribution with mean c_1 and variance b_1^2 . Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left[- \frac{z^2}{2} \right] dz \tag{4.3}$$

for $x \in (-\infty, \infty)$. By definition the function $\Phi(x)$ is the distribution function of the standard normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. It can be shown that the integral of the risk function in 4.2 is

$$H_1(x) = \alpha_1 \Phi \left(\frac{x - c_1}{b_1} \right) . \tag{4.4}$$

Therefore, the survival function $S_1(x)$ has the form

$$S_1(x) = \exp[-H_1(x)] \tag{4.5}$$

for $x \in (-\infty, \infty)$. It is well known that

$$\lim_{x \rightarrow \infty} \Phi(x) = 1 . \tag{4.6}$$

From this observation, it follows from 4.4 and 4.5 that

$$\lim_{x \rightarrow \infty} S_1(x) = \exp[-\alpha_1] . \tag{4.7}$$

Because $\Phi(x) \rightarrow 0$ as $x \rightarrow -\infty$, it follows that $S_1(x)$ may not $\rightarrow 1$ as $x \rightarrow 0$. For example, from 4.4 and 4.5, it can be seen that

$$S_1(0) = \exp\left[-\alpha_1 \Phi\left(\frac{-c_1}{b_1}\right)\right] \neq 1. \tag{4.8}$$

Consequently, it seems desirable to search for a distribution such that the range of the random variable X representing age takes on only non-negative values $x \geq 0$. Among the distributions with this property is the log-normal distribution, which is defined as follows. Suppose the random variable W has a normal distribution with mean μ and variance σ^2 . Then the random variable

$$X = \exp[W], \tag{4.9}$$

then the random variable X takes on values x such that $x > 0$. From 4.9 it follows that

$$\ln X = W. \tag{4.10}$$

Thus, the random variable X is said to have a log-normal distribution. In what follows, it will be helpful to represent the random variable W in the form $W = \mu + \sigma Z$, where Z has a standard normal distribution with mean 0 and variance 1. Given this representation of W , the distribution function of X may be easily derived. By definition

$$F(x) = P[X \leq x]. \tag{4.11}$$

is the distribution function of the random variable X . But, because the $\ln x$ is a monotone increasing function for $x > 0$, it follows that

$$\begin{aligned} F(x) &= P[X \leq x] = P[W \leq \ln x] = P[\mu + \sigma Z \leq \ln x] \\ &= P\left[Z \leq \frac{\ln x - \mu}{\sigma}\right] = \Phi\left(\frac{\ln x - \mu}{\sigma}\right). \end{aligned} \tag{1}$$

From this equation, it can be seen from 4.3 that the probability density function of the random variable X has the form

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \tag{4.13}$$

From now on the risk function for mid-life will be chosen as $\mu_1(x) = \alpha_1 f(x)$ in 4.13. It can also be seen that from 4.12, it can also be shown that

$$\lim_{x \downarrow 0} F(x) = 0, \tag{4.14}$$

which is a desirable property.

To connect the parameters μ and σ with more useful statistical interpretations, it can be shown that the mode of the log-normal density is

$$m_1 = \exp\left[\mu - \sigma^2\right]. \tag{4.15}$$

And if X_1 is a log-normal random variable, then its expectation is

$$E[X_1] = \mu_1 = \exp\left[\mu + \frac{\sigma^2}{2}\right] \tag{4.16}$$

Equation 4.15 may be verified by finding the derivative of the log-normal density function in 4.13 and setting it to 0, and then solve that resulting equation for mode in 4.15. It would also be wise to check that m_1 is indeed the maximum of the density function in 4.13. As a first step in showing the μ_1 in 4.16 is the correct formula for the expectation in 4.16, it will be helpful to have a look at the formula for the moment generating function of normal random variable with mean μ and variance σ^2 . It is well known that the moment generating function of the normal distribution is

$$G_N(s) = E[\exp X_1 s] = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[xs] \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] dx = \exp\left[\mu s + \frac{s^2 \sigma^2}{2}\right]. \tag{4.17}$$

By definition, because $X_1 = \exp W$

$$\begin{aligned} E[X_1] &= E[\exp W] = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[w] \exp\left[-\frac{(w - \mu)^2}{2\sigma^2}\right] dw = G_N(1) \\ &= \exp\left[\mu + \frac{\sigma^2}{2}\right], \end{aligned} \tag{2}$$

which demonstrates the validity of 4.16.

The integral of the risk function $\mu_1(x) = \alpha_1 f(x)$, where $f(x)$ is the log-normal density is

$$H_1(x) = \alpha_1 \int_0^x f(x) dx = \frac{\alpha_1}{x\sigma\sqrt{2\pi}} \int_0^x \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] dx = \alpha_1 \Phi\left(\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \tag{4.19}$$

Therefore, the survival function for mid-life has the form

$$S_1(x) = \exp[-H_1(x)] \tag{4.20}$$

from 4.19, it can be seen that

$$\lim_{x \rightarrow \infty} S_1(x) = \exp[-\alpha_1] \tag{4.21}$$

but the rate of convergence may be slow.

Given the formulas for the mode m_1 in 4.15 and that in 4.16 for μ_1 , the expectation of a random variable with a log-normal distribution, the next step in the formulation is to derive formulas for computing the parameters μ and σ as functions of m_1 and μ_1 , which in a computer simulation experiment would be assigned values based on some rational used in designing a computer experiment. From 4.15 and 4.16, it follows that

$$\ln m_1 = \mu - \sigma^2 \tag{4.22}$$

and

$$\ln \mu_1 = \mu + \frac{\sigma^2}{2}. \tag{4.23}$$

Let v_1 and v_2 denote two 2×1 vectors defined as follows

$$v_1 = \begin{pmatrix} \ln m_1 \\ \ln \mu_1 \end{pmatrix} \tag{4.24}$$

and

$$v_2 = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}. \tag{4.25}$$

Then consider the 2×2 matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & \frac{1}{2} \end{pmatrix}. \tag{4.26}$$

From 4.22 and 4, 23, it can be seen that

$$Av_2 = v_1. \tag{4.27}$$

If the matrix A is non-singular, then

$$v_2 = A^{-1}v_1. \tag{4.28}$$

It can be shown that the inverse of the matrix A is

$$A^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{pmatrix}. \tag{4.29}$$

It follows that

$$\mu = \frac{1}{3} \ln m_1 + \frac{2}{3} \ln \mu_1 \tag{4.30}$$

and

$$\sigma^2 = -\frac{2}{3} \ln m_1 + \frac{2}{3} \ln \mu_1 = \frac{2}{3} (\ln \mu_1 - \ln m_1). \tag{4.31}$$

If $\ln \mu_1 - \ln m_1 > 0$, then $\sigma^2 > 0$, and

$$\sigma = \sqrt{\frac{2}{3} (\ln \mu_1 - \ln m_1)}. \tag{4.32}$$

Thus, if μ , σ and α_1 are known, then function $H_1(x)$ in (4, 19) could be computed for chosen values of $x > 0$. Included in the scientific word processor used to write this paper are symbolic and numerical computation engines. The inverse matrix in 4.29 was calculated using the symbolic computation engine, and the numerical computation engine was used,

was used to produce the the numerical computation engine was used to produce the numerical results presented in the rest of this section.

By way of an illustrative example, suppose that in some computer simulation experiment, it was assumed that ages of greatest risk of death in mid-life were $x \in [15, 35]$. If it assumed that the risk of death is greatest as age 25, then the value of the mode of the log-normal density function could be chosen as $\mu_1 = 25$. From formula 4.32 in order to compute a value of $\sigma > 0$, the mean μ_1 of the log-normal distribution should be chosen as a number such that $\mu_1 > m_1$. Suppose, for example, the value chosen for $m_1 = 22$. Then, $\ln 25 = 3.2189$ and $\ln 22 = 3.091$. Then $\ln 25 - \ln 22 = 0.12783$, and

$$\sigma = \sqrt{\frac{2}{3} \times 0.12783} = 0.29192 . \tag{4.33}$$

In this case

$$\mu = \frac{1}{3} \times 3.091 + \frac{2}{3} \times 3.2189 = 3.1763 . \tag{4.34}$$

Given these values of μ and σ , an investigator may wish to graph the log-normal density function in 4.13 for $x \in [15, 35]$. to provide insights into plausibility of the values the chosen values of μ and σ and their implications for predicting the potential results of the computer simulation experiments being considered. By assigning a value to the promoter α_1 , it would also be possible to compute the values of the survival function

$$S_1(x) = \exp[-H_1(x)] \tag{4.35}$$

for any $x > 0$.

In the Mode and Sleeman 2000 , the risk function based on the log-normal distribution as indicated in section was suggested as a risk function for mid-life in section in chapter 13 section 2 of the book. But the account of this risk function is much more complete than in the book. Furthermore, one of the formulas presented in this chapter for finding μ and σ may contain an error, but in this section the formula was derived using a different method based on matrix theory and its derivation is more transparent. The review presented in this section is incomplete. For example, the risk function for mid-life suggested by Heligman and Pollard (1980) has not been displayed in this section It was not included, because of its complexity. But, if a reader is interested in an application of this risk function the paper by Mode and Jacobson (1984) may be consulted.

5. Makeham and Gompertz Risk Functions

The Makeham risk function is included in the formulation to accommodate the risks of accidents throughout the life span of an individual. It is assumed that this risk function is constant and is given by

$$\theta_2(x) = \alpha_2 \tag{5.1}$$

,where α_2 is a positive constant. In this case, the integral of the risk function is

$$H_2(x) = \int_0^x \theta_2(s)ds = \alpha_2x \tag{5.2}$$

for $x \geq 0$. Therefore, for $x \geq 0$, the survival function has the simple form

$$S_2(x) = \exp[-\alpha_2x] . \tag{5.3}$$

It has been suggested that a useful trial value of α_2 , based on period studies of human mortality, is about 0.001. The two-parameter latent risk function, due to Gompertz (19-th Century) deals with risks of deaths at the older ages. Let α_3 and β_3 be positive parameters. Then, it will be assumed that for $x \geq 0$ the latent risk function $\theta_3(x)$ has the form

$$\theta_3(x) = \alpha_3\beta_3 \exp[\beta_3x] . \tag{5.4}$$

Observe that, as it should, this risk function increases as age x of an individual increases, and, by assumption, the risk of death increases exponentially with increasing age. The integral of this risk function has the form

$$H_3(x) = \alpha_3 \exp(\exp[\beta_3x] - 1) \tag{5.5}$$

for $x \geq 0$. Therefore, the latent survival function for this component is

$$S_3(x) = \exp[-\alpha_3(\exp[\beta_3x] - 1)] \tag{5.6}$$

for $x \geq 0$.

By applying a general but standard formula that a density is the risk function times the survival function, it can be seen that the probability density function of the Gompertz distribution has the form

$$f_3(x) = \theta_3(x)S_3(x) = \alpha_3\beta_3 \exp[\beta_3x] \exp[-\alpha_3(\exp[\beta_3x] - 1)] \tag{5.7}$$

for $x \geq 0$. Although this distribution may be derived from intuitively appealing assumptions, it is more difficult to handle from a mathematical point of view than some other distributions that arise in probability and statistics. Nevertheless, because many advanced mathematical functions are now available to research workers in such software packages as MAPLE, MATHEMATICA, and MATLAB, an outline of the mathematics used in analyzing the Gompertz distribution seems appropriate.

Because the parameters α_3 and β_3 do not have obvious statistical interpretations, such as an expectation or variance, it is difficult to assign tentative values to them. Quite often, however, there is some feeling about the modal age of death for those who survive to old age. Let m_3 denote the mode of the Gompertz distribution. Then by using elementary calculus to find the maximum of the density $f_3(x)$, it can be shown that the equation

$$\alpha_3 = \exp[-\beta_3m_3] \tag{5.8}$$

supplies a connection among the parameters α_3, β_3 and m_3 . In particular, if m_3 is assigned a value and β_3 is known, then α_3 is determined. But, to find a plausible value of β_3 , more input is needed. To this end, one may also have some idea about plausible values for σ_3 , the standard deviation of the age of death among those who survive to old age. Thus, it would be helpful to express σ_3 in terms of the parameters of the Gompertz distribution.

In this connection, it will be useful to consider the moment generating function of the Gompertz distribution, which is defined by

$$M_3(s) = \int_0^\infty e^{sx} f_3(x) dx \tag{5.9}$$

for those values of s for which the integral converges. By using some advanced calculus, it can be shown that the moment generating function of the Gompertz distribution has the form

$$M_3(s) = e^{\alpha_3} \left[\frac{\Gamma(\frac{s}{\beta_3} + 1)}{\alpha_3^{\frac{s}{\beta_3}}} - \sum_{v=0}^\infty \frac{(-1)^v \alpha_3^{v+1}}{v! (\frac{s}{\beta_3} + v + 1)} \right], \tag{5.10}$$

where $\Gamma(\cdot)$ is the famous gamma function and $s \in (-\epsilon, \infty)$ with $\epsilon > 0$ near zero.

Recall that the gamma function is defined by the integral

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \tag{5.11}$$

which converges for all $z > 0$. As is well known, this function corresponds with the factorial function on the positive integers. In fact, if $z = n$, a positive integer, then $\Gamma(n) = (n - 1)!$. In a word, this function fills the gaps among the integers, and, in particular, it can be shown that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

If the random variable X_3 has a Gompertz distribution, then its expectation is

$$E[X_3] = M_3^{(1)}(0) = \frac{dM_3(s)}{ds} \Big|_{s=0}, \tag{5.12}$$

the second moment is

$$E[X_3^2] = M_3^{(2)}(0) = \frac{d^2M_3(s)}{d^2s} \Big|_{s=0}. \tag{5.13}$$

and the variance of the distribution is given by the well known formula

$$\sigma_3^2 = E[X_3^2] - (E[X_3])^2. \tag{5.14}$$

After considerable analysis, it can be shown that the exact formula for the expectation is

$$E[X_3] = e^{\alpha_3} \left[m_3 - \frac{C}{\beta_3} + \frac{1}{\beta_3} \sum_{v=0}^\infty \frac{(-1)^v \alpha_3^{v+1}}{v!(v+1)^2} \right], \tag{5.15}$$

where $C \approx 0.57721 \dots$ is Euler's constant. Furthermore, the exact formula for the second moment is

$$E[X_3^2] = e^{\alpha_3} \left[\left(m_3 - \frac{C}{\beta_3} \right)^2 + \frac{\pi^2}{\beta_3^2 6} - \frac{2}{\beta_3^2} \sum_{v=0}^{\infty} \frac{(-1)^v \alpha_3^{v+1}}{v!(v+1)^3} \right]. \tag{5.16}$$

The following two results

$$\Gamma^{(1)}(1) = -C \tag{5.17}$$

and

$$\Gamma^{(2)}(1) = \frac{\pi^2}{6} + C^2, \tag{5.18}$$

which were used in deriving the above formulas, are given in several books on special functions and advanced calculus. The symbol $\Gamma^{(k)}(x)$ is, by definition, the k -th derivative of the Gamma function.

When $\alpha_3 > 0$ is small, then $\exp[\alpha_3] \approx 1$ and the above infinite series may be neglected. Thus, the approximations

$$E[X_3] \approx m_3 - \frac{C}{\beta_3} \tag{5.19}$$

and

$$E[X_3^2] \approx \left(m_3 - \frac{C}{\beta_3} \right)^2 + \frac{\pi^2}{\beta_3^2 6} \tag{5.20}$$

hold for small α_3 . Therefore, a formula for the approximate variance of the Gompertz distribution is

$$\sigma_3^2 \approx \frac{\pi^2}{\beta_3^2 6}. \tag{5.21}$$

Equivalently,

$$\beta_3 \approx \frac{\pi}{\sigma_3 \sqrt{6}} \tag{5.22}$$

is an approximate expression connecting the parameter β_3 with the standard deviations σ_3 . However, if an investigator is working with MAPLE, MATHEMATICA, MATLAB or other software packages, programs may be written to compute the variance σ_3^2 more accurately, using the above infinite series. Thus, the accuracy of these approximations may be assessed. Among the classic books that contain information on the above mathematics are those of Artin and Widder.

6. Introducing Randomness Into the Gompertz Survival Function

In a previous study using Monte Carlo simulation methods, it was found that an age structured stochastic process converged in distribution to a parametrized survival function, see Mode and Sleeman 2012 . In previous work the idea of a stochastic environment has a also been considered see Mode and Jacobson (1987) . In this section. some techniques for incorporating randomness into the Gompertz distribution will be considered. The Gompertz distribution is based on two parameters α and β . Consequently, in this section the focus of attention will be some suggestions for introducing randomness into these parameters. In equation 5.8 there is a formula connecting and parameters α and β as well as the mode m of the Gompertz distribution.

Among those who study and apply Monte Carlo simulation methods there is a well known equation. Let

$$F(x) = P[X \leq x] \tag{6.1}$$

denote the distribution function of the random variable X . It will be assumed that the domain of the random variable X is a continuum of real numbers. Then

$$F(X) = U, \tag{6.2}$$

where U is a random variable on the interval $[0, 1]$. Whenever it is possible to solve this equation to express X as a function of U , then it is possible to compute realizations of the random variable X by computing realizations of the random variable U .

In order to view the mode of the Gompertz distribution as a random variable, it will be necessary to define positive numbers that represent ages. Let a and b be positive numbers representing ages such that $a < b$, and let V denote a random variable with a uniform distribution on the interval $[a, b]$. Its probability density function is

$$g(v) = \frac{1}{b - a} \tag{6.3}$$

for all $v \in [a, b]$, The distribution function of the random variable V is

$$G(v) = \int_a^v g(s) ds = \frac{v-a}{b-a} \quad (6.4)$$

for all $v \in [a, b]$. From the comments stated above, it follows that $G(V) = U$. Thus,

$$\frac{V-a}{b-a} = U. \quad (6.5)$$

By solving this equation for V , it can be seen that

$$V = a + (b-a)U. \quad (6.6)$$

From this equation, it becomes transparent that realizations of the random variable V reduces to computing realizations of the random variable U . If $U = 0$, then $V = a$, and if $U = 1$, then $V = b$. From these observations, it follows that that realizations of the random variable V will always be in the interval $[a, b]$.

For example, if the mode of the Gompertz distribution is $m = 60$, $b = 70$ and $a = 50$, then $b - a = 20$, then the above equation takes the form

$$V = 50 + 20U. \quad (6.7)$$

In this case, in a Monte Carlo simulation experiment, the mode of the Gompertz distribution would be somewhere in the interval $[50, 70]$ for any realization of the process.

Similarly, if the standard deviation σ of the Gompertz distribution is viewed as a random variable such that all realizations of σ belong to the interval $[c, d]$, where c and d are positive numbers such that $c < d$, then

$$\sigma = c + (d-c)U. \quad (6.8)$$

From 5.8, it follows that if the mode m is a random variable, then so is the parameter α . It is of interest to note that in this case β may be constant or a random variable.

When designing a computer simulation experiment, it is of interest to know the expectation and variance of random variables given in 6.6 and 6.8. It is well known that if the random variable U is uniformly distributed on the interval $[0, 1]$, then its expectation and variance are given by the formulas

$$E[U] = \frac{1}{2} \quad (6.9)$$

and

$$\text{var}[U] = \frac{1}{12}. \quad (6.10)$$

From these observations, it follows that the expectation and variance of the random variable V in 6.6 are

$$E[V] = \frac{a+b}{2} \quad (6.11)$$

and

$$\text{var}[V] = \frac{(b-a)^2}{12} \quad (6.12)$$

For example, if $a = 50$ and $b = 70$ as suggested above, then

$$E[V] = \frac{50+70}{2} = 60.0 \quad (6.13)$$

and

$$\text{var}[V] = \frac{(20)^2}{12} = 33.333 \quad (6.14)$$

From these observations, it can be seen that the variation around the mean 60.0 will be relatively small in samples of realizations of the random variable V . In the next section, the concepts described in the preceding sections will be examined numerically and illustrated by using graphs.

7. A Method for Computing Numerical Versions of Parametric Survival Functions

In this numerical version of a survival function, the parameters values for early life, see section 3, were chosen as $\alpha_0 = 0.953$, which is the probability of an infant surviving one year. The chosen value of the expectation of the exponential distribution was $\beta_0 = 1.361$. For midlife the parameter α_1 , which is the probability of surviving midlife, was chosen as $\alpha_1 = 1$. The values of the mode and mean for the lognormal distribution were, see section 4 for details, were chosen as $m_1 = 25$ and $\mu_1 = 33$. The value of the Makeham parameter was chosen as $\alpha_2 = 0.001$. The value of the standard deviation σ_3 for the Gompertz distribution was chosen $\sigma_3 = 13.255$ and the value of the mode of the Gompertz was chosen as age $m_3 = 75$. A detailed account of these parameters is given in section 5. It is also shown in section 5 that the values of parameters determine the values of the Gompertz parameters α_3 and β_3 . The value of the greatest age was chosen as $GA = 100$. It should be emphasized that the values of the parameters stated above, which were chosen to produce a plausible numerical realization of a survival function for humans that may have long lives, are arbitrary. Any user of the ideas may chose any values of the parameters that are, in his mind, plausible.

Let $S_0(x)$ denote the function designed to adjust the Gompertz survival function for deaths in early life. Then, $S_0(0) = 1$ and $S_0(x)$ was evaluated numerically for $x = 1, 2, 3, \dots, 100$. The function $S_1(x)$ was designed to adjust the Gompertz survival function for midlife deaths was also evaluated numerical for the integers $x = 1, 2, 3, \dots, 100$. Similarly, the survival functions for the Makeham distribution $S_2(x)$ and Gompertz distribution $S_3(x)$ were also evaluated numerically for the positive integers listed above. In this section two additional survival functions will be considered

$$S_{0123}(x) = S_0(x) \times S_1(x) \times S_2(x) \times S_3(x) \tag{7.1}$$

and

$$S_{023}(x) = S_0(x) \times S_2(x) \times S_3(x) \tag{7.2}$$

and evaluated for $x = 1, 2, 3, \dots, 100$. Both these survival function satisfy the condition $S_{0123}(0) = 1$ and $S_{023}(0) = 1$. Observe that the survival function $S_{023}(x)$ differs from $S_{0123}(x)$ because the function $S_1(x)$ for midlife has been excluded.

If $S(x)$ is the survival function for the random variable X with values in the interval $0 \leq x \leq 100$, then it is well known that the expected value of X is

$$E[X] = \sum_{x=0}^{100} S(x) . \tag{7.3}$$

This expectation is known as the expectation of the life span for any newly born infant in the population. For the case of the survival function S_{0123} this expectation has the value

$$E[X | S_{0123}] = 40.95676571 . \tag{7.4}$$

The expectation of S_{023} is

$$E[X | S_{023}] = 64.084919 . \tag{7.5}$$

For the case of the Gompertz survival function, this expectation has the value

$$E[X | S_3] = 69.592489 . \tag{7.6}$$

Presented in Figure 7.1 is a graph comparing the Gompertz survival function with S_{0123} for the ages $x = 0, 1, 2, \dots, 100$. As can seen from the graph of the survival function S_{0123} , it appears that this function over adjusts the Gompertz so it is doubtful whether this function could be applied in a computer simulation experiment, unless the objective of the experiment was to project the evolution of a population with high mortality rates. The expectation of the life of an infant for this survival function is about 41 years, see 7.4 and is much smaller then that of the Gompertz survival function which is about 70 years, see 7.6. The reason why the function $S_{0123}(x)$ over adjust the Gompertz survival function is that the numerical value of this function at age 100 is less than one. If it were one, then its affects would be neutral and the numerical value of the Gompertz function would not be changed.

Figure 7.2 contains the graphs of the survival function S_{023} and the Gompertz survival function. From these graphs it can be seen that the high rate of infant immortality my be easily seen by viewing the graph of the survival function S_{023} in the early years of life. Given this survival function, the expectation of life is about 64 years, see 7.5. It seems more likely that this survival function would be more suitable for using in a computer simulation experiment than the function S_{0123} when high mortality rates for infants are under consideration.

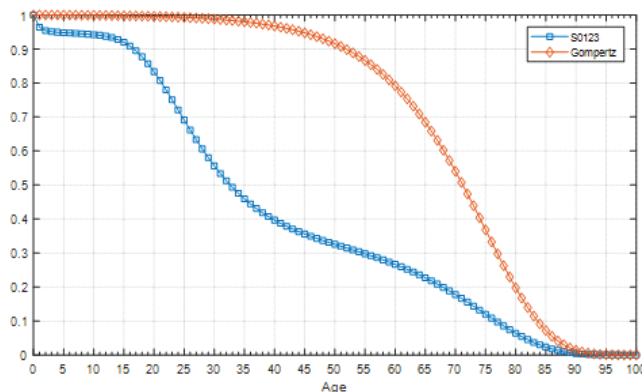


Figure 7.1. A Comparison of the Graphs of the Survival Function S_{0123} and the Gompertz Survival Function

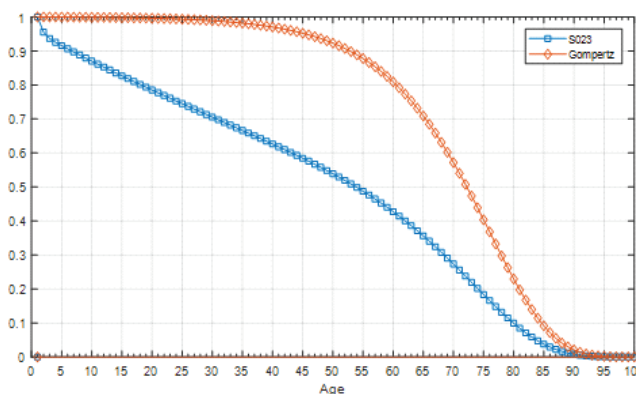


Figure 7.2. A Comparison of the Graphs of the Survival Function S_{023} and the Gompertz Survival Function

In experiments not reported in this paper, an attempt was made to use the assumptions that the function $S_0(x)$ was less than one only for the ages $x = 1, 2, \dots, 10$ and $S_0(x) = 1$ for ages $x = 11, 12, \dots, 100$. Similarly, the function for mid-life $S_1(x)$ was less than one only for the ages $x = 15, 16, \dots, 30$, representing mid-life. For all ages not in the mid-life range, it was assumed that $S_1(x) = 1$. According to theory, when these functions were one, the Gompertz function would not be changed in the products 7.1 and 7.2. However, when these products were used at some ages $x = 1, 2, \dots, 100$, the Gompertz function was not monotone decreasing as the age x increased. Consequently, the adjusted Gompertz function was no longer useful as a survival function for all ages. A decision was made, therefore, to abandon the approach just described.

In previous work, Mode and Busby 1984, as mentioned in a previous section, survival function for infancy, mid-life and older ages were spliced together. When the parameters of this mode were estimated from a life table it was found that there were small bumps at the splice points, but the over all the fit to data was very good. Consequently, this splicing method remains an alternative to the methods described in this paper based on the numerical evaluation of all functions in a product of functions for all ages $x = 0, 1, 2, \dots, 100$. But, it should be mentioned that if all the parameters of the were assigned plausible numerical values not based on data, the bumps in the splice points may be very noticeable.

In section 6, a method was developed to treat the mode of the Gompertz as a random variable. In an illustrative example, the bounds of variation in the mode of the Gompertz function were 50 and 70. For the case of a mode of 50, the expectation of life would be

$$E[X | S_3, m_3 = 50] = 44.968, \tag{7.9}$$

and the case $m_3 = 70$, this expectation is

$$E[X | S_3, m_3 = 70] = 64.621 \tag{7.10}$$

Thus, in any Monte Carlo simulation experiment based on these bounds, the expectation of the Gompertz distribution would lie in the interval $[44.968, 64.621]$, which shows that in such a population the number of deaths would vary sig-

nificantly among years. Because the computer implementation of these ideas would entail major a change to existing software, the ideas just mentioned will not be implemented in the software used in this paper.

Natural disasters such as earthquakes, floods and volcanic eruptions are often the cause of deaths in human and animal population in the regions of such disasters. Such disasters are difficult to predict, but it would be interest to design and implement some software based on the idea of random waiting times among disasters. An example of such random waiting times are those among earthquakes in some geological area area. For example in discrete time formulations, it may be assumed that the waiting times T among earthquakes are random variables in some geological area that follow a geometric distribution with the probability density function

$$P [T = t] = p (1 - p)^{t-1} \tag{7.11}$$

where $0 < p < 1$ and $t = 1, 2, \dots$.

In any Monte Carlo simulation experiment, it would be straight forward to simulate a sequence T_1, T_2, \dots of the random variables that are realizations of the random variable T . In such an experiment, it would be required that some methods for simulating deaths would be needed. A very simple way of simulating such deaths across all ages is to assume that the parameter α_2 in the Makeham distribution is a random variable. To illustrate these ideas, suppose the survival function under consideration is

$$S_{023}(x) = S_0(x) \times S_2(x) \times S_3(x) , \tag{7.12}$$

where

$$S_2(x) = \exp(-\alpha_2 x) \tag{7.13}$$

is the Makeham survival function. The value of this parameter in the formulation under consideration was chosen as $\alpha_2 = 0.001$. If, for example, if this value is increased to $\alpha_2 = 0.01$, then the survival function in 7.12 would be significantly decreased for all ages $x = 0, 1, 2, \dots, 100$, and in any in any Monte Carlo simulation experiment the simulated number of death using the survival function in 7.11, would be much greater than if the parameter value $\alpha_2 = 0.001$ were used. In the next section, an algorithm for simulating deaths across all ages will be described, and will be used if all parameters are constant and not random variables.

8. Algorithms for Projecting an Age Structured Population With Stochastic Laws of Evolution Forward in Time

The objective of this section is to define as set of algorithms to simulate the stochastic evolution of an age structured population forward in time. It will be assumed that the greatest age any individual may live is 50 years. At time t let $X(x, t)$ denote the number of females in a population for age $x = 0, 1, 2, \dots, 50$ for $t = 0, 1, 2, \dots, N$, where $N > 1$ is the number of years considered in a projection experiment. Let $X(0, t)$ denote the number of female infants in the population at time t . Similarly, let $Y(0, t)$ denote the number of male infants in the population at time t , and let $Y(x, t)$ denote the number of males in the population at time t for ages $x = 1, 2, \dots, 50$.

Let $S_f(x)$ and $S_m(x)$ denote, respectively, the survival functions for females and males. By definition $S_f(0) = S_m(0) = 1$. Let $p_f(x)$ denote the conditional probability that a female of age $x - 1$, survives to age x for $x = 1, 2, \dots, 50$. Then, it follows that

$$p_f(x) = \frac{S_f(x)}{S_f(x-1)} \tag{8.1}$$

for $x = 1, 2, \dots, 50$. Observe that $p_f(1)$ is the conditional probability that an infant female survives to age 1. Similarly, let $p_m(x)$ denote the conditional probability that a male of age $x - 1$ survives to age x for $x = 1, 2, \dots, 50$. The evolution of the population will be viewed as stochastic, because it will be assumed that the number of females $X(x, t)$ is a random variable who's realizations may be simulated.

A random variable X is said to have a binomial distribution with parameters N and p if its pdf has the formula

$$P [X = x] = \binom{N}{x} p^x (1 - p)^{n-x} \tag{8.2}$$

for $x = 0, 1, 2, \dots, N$. In this formula N is a positive integer and p is a probability such that $(0 < p < 1)$ In what follows, it will be convenient to use a kind of shorthand such as

$$X \sim B(N, p) \tag{8.3}$$

to indicate that a random variable X has a binomial distribution with parameters N and p .

The next step in the formulation of the evolutionary model is to take into account the births of infants and their entry into an evolving population. From now on the focus of attention will be the evolution of a population of wild chimpanzees.

Suppose that the ages female chimps may give birth to an infant are $x = 15, 16, \dots, 40$. There are several distributions depending on one or more parameters that may be viewed as candidates for the number of infants a fertile female chimp may give birth during every year of her fertile ages. Let the random variable $Y(x)$ denote the number of offspring a female chimp may have during any year for ages $x = 1, 2, \dots, 40$. For the sake of simplicity, it will be assumed that the random variable $Y(x)$ has a Poisson distribution with probability density function

$$P[Y(x) = y] = \exp[-\lambda(x)] \frac{(\lambda(x))^y}{y!} \tag{8.4}$$

with parameter $\lambda(x) > 0$ for each age fertile age x . Let the random variable B denote the total number infants the females of a chimp population have during any year of a population projection. Then, if it is assumed that the random variables $Y(x)$ for $x = 15, 16, \dots, 40$ are independent, then the random variable

$$B = \sum_{x=15}^{40} Y(x) \tag{8.5}$$

has a Poisson distribution with the parameter

$$\lambda = \sum_{x=15}^{40} \lambda(x) . \tag{8.6}$$

It is well known that the expectation and variance of B is λ . Let Z be a standard normal random variable with expectation 0 and variance 1. Then it is also well known that the distribution of the random variable

$$\frac{B - \lambda}{\sqrt{\lambda}} \simeq Z \tag{8.7}$$

when the parameter λ is large, Z has a standard normal distribution with mean 0 and variance 1. From 8, 7 it can be seen that

$$B \simeq \lambda + \sqrt{\lambda}Z . \tag{8.8}$$

Observe that if $\lambda(x) = 1$ for $x = 15, 16, \dots, 40$, then $\lambda = 26$, which is sufficiently large to use the approximation 8.7. The number B in 8, 8 may not be positive. Thus, B will be chosen as

$$B = |\lambda + \sqrt{\lambda}Z| . \tag{8.9}$$

Therefore, the distribution of the random variable B is the absolute normal distribution with parameter λ . In general, the number B in 8.9 is not a positive integer. To remedy this situation, for any positive real number X , let $[X]$ denote the greatest integer in X . Thus, if B is chosen as

$$B = [B] , \tag{8.10}$$

then B is non-negative integer.

The next step in the formulation is to define an algorithm for computing the number female and male infant chimps that are added to the population at any time t . Let p_f denote the probability that an infant chimp is female, and let p_m denote the probability an infant is male. Then, $p_m = 1 - p_f$. Let the random variable $X(0, t)$ the number of female infant chimps added to the population at time t , and let $B(t)$ the total number of infants that will become members of the population at time t . Then, it will be assumed that

$$X(0, t) \sim B(B(t), p_f) . \tag{8.11}$$

From this result it follows that the number of male infant chimps added to the population at time t is

$$Y(0, t) = B(t) - X(0, t) . \tag{8.12}$$

At time t , the number of females in the population by age may be represented by the array

$$X(0, t), X(1, t), X(2, t), \dots, X(50, t) \tag{8.13}$$

Let

$$X((1, t + 1)), X(2, t + 1), \dots, X(50, t + 1) \tag{8.14}$$

denote the array of females by age in the population who are the survivors of the population depicted in 8.13. Then, the survivors in 8.14 are simulated using the formula

$$X(x + 1, t + 1) \sim B(X(x, t), p_f(x)) \tag{8.15}$$

for $x = 0, 1, \dots, 49$. Note if there are individuals of age 50 in the population, then they will no longer be viewed as members of an evolving population.

Observe that the age structured population represented by the array in 8.14 does not contain female infants. Therefore, the number $X(0, t + 1)$ of female infants at time $t + 1$ could be simulated using the algorithm described above and added to the array 8.14 to simulate the array

$$X(0, t + 1), X(1, t + 1), \dots, X(50, t + 1) \tag{8.16}$$

By using the algorithm just described, an array analogous to that in 8.16 could also be simulated for males. In this paper, Monte Carlo simulation methods will be used to simulate a realizations of the stochastic process under consideration, using the algorithms described above. The numerical values of the parameters of the model used in this experiment will be given in the next section.

At any time during a population projection, it is of interest to display an estimates of total population size and the age distribution. At time t , the total size of the female population is

$$T(t) = \sum_{k=0}^{50} X(k, t) \tag{8.17}$$

Therefore, at any time t , the age distribution is

$$A(x, t) = \frac{X(x, t)}{T(t)} \tag{8.18}$$

for $x = 0, 1, \dots, 50$. If both females and males are being considered in a population projection experiment, then the same formulas can be used to estimate the total population size for males as well as their age distribution.

When dealing with a stochastic processes evolving in time, it is often of interest to estimate the value of some random variable at time $t + 1$, given the value of the random variable at time t . For the stochastic processes under consideration, it is well known that the minimum mean square estimate $\widehat{X}(x, t + 1)$ of the random variable $X(x, t + 1)$ is the conditional expectation

$$\widehat{X}(x, t + 1) = E[X(x, t + 1) | X(x - 1, t)] = X(x - 1, t) p(x - 1), \tag{8.19}$$

for $x = 1, 2, \dots, 50$. Observe that this is the expectation of a random variable with a binomial distribution as indicated in 8.3. Estimates of the random function $Y(x, t + 1)$ for males could also be computed using a formula of the type given in 8.19.

The formula in 8.19 may also be written in the form

$$\widehat{X}(x, t + 1) = \widehat{X}(x - 1, t) p(x - 1) \tag{8.20}$$

which provides a recursive formula for computing estimates $\widehat{X}(x, t)$ of the random function $X(x, t)$ for $t = 1, 2, \dots$, given $X(x, 0)$ for all ages $x = 0, 1, 2, \dots, 50$. A similar recursive equation may be used to compute $\widehat{Y}(x, t)$ as an estimate of the random function $Y(x, t)$ for $t = 1, 2, \dots$ and $x = 0, 1, 2, \dots, 50$. At each t in the recursive procedure, the function $X(0, t)$ will be assigned the value $\widehat{X}(0, t) = \lambda$, the expected total number of births, see 8.6.

The recursive equations for the females and males just described will be referred to as the deterministic model embedded in the stochastic process. For the case of the embedded deterministic model only one trajectory of the process can be computed, Whereas, for the stochastic model in a computer simulation experiment, each trajectory of the process will differ from all other trajectories in a sample of $R > 1$ realizations of the process. It should also be mentioned that when the embedded deterministic model is used, it will be possible to compute age distributions as well as other sets of values that may be of interest to an investigator.

During the preliminary testing of the ideas underlying stochastic process, it was observed that if it is assumed that λ , the expected total number of offspring born in any year was constant, then the number of births in any realization of the process was nearly constant, i.e., it was not random. To correct this flaw, it was decided to let λ be a random variable defined as follows. Let a denote a positive number less than b . Then it was assumed that λ was a random variable defined by the equation

$$\lambda = a + (b - a)U, \tag{8.21}$$

where U is a uniform random variable on the interval $[0, 1]$. At this point, it may be helpful to consult section 6 that contains details on the properties of random variables defined in equation 8.21. For example, the expectation of the random variable λ is

$$E[\lambda] = \frac{a + b}{2}. \tag{8.22}$$

In computer simulation experiments the numbers a and b are usually chosen such that the expression on the right is a preassigned positive number.

9. Assigned Values of Parameters Used in an Illustrative Monte Carlo Simulation Experiment

The parameter values chosen in the simulation experiment considered in this paper were designed to follow the short term evolution a population of animals living in the wild with short lives. It was assumed that the greatest age of any individual in the population as it evolved was 50 years. Differences in the survival on males and female were taken into account by selecting two survival functions with different parameters values. The survival functions for both sexes were chosen using the early life survival function in section 3 and Makeham Gompertz survival functions described in sections 3 and 5.

The survival function, S_{023F} was chosen for both females and males in the experiment under consideration. The reason for choosing this type of survival function for both females and males is that there is no information from chimp populations in the wild that there is a midlife effect with regarding mortality. as there is in humans, particularly for males in midlife. .Consequently the midlife component S_1 was omitted.

Observe that to get a numerical version of the survival function S_0 for females, it is necessary to assign values for two parameters in S_0 , the fraction that survive infancy and a mean of an exponential distribution. The fraction that survived infancy was chosen as $\alpha_0 = 0,80$ and the mean of the exponential distribution was assigned the value $\beta_0 = 030030$. The parameter for the Makeham distribution was assigned the value $\alpha_2 = 0.02$. Finally to get a numerical version of the Gompertz survival function numerical values must be assigned for the standard deviation σ_G and the mode m_G . The numbers chosen for these two parameters were $\sigma_G = 6$ and $m_G = 33$. All the parameters for the male survival function were the same as those for the females except that the mode of the Gompertz distribution for males was chosen as $m_G = 30$. Numerical values for the female S_{023F} and male the survival functions S_{023M} were computed by using the procedures set forth in section 7.

Very little about survivorship in populations of wild Chimpanzees is known. Consequently every numerical version of a survival function must be checked for plausibility. By using the formulas for the expectation of life given in section 7, the numerical values for the expectation of life, expressed in terms of years, for females was

$$E[X | S_{923F}] = 18.684 \tag{9.1}$$

and that for males was

$$E[X | S_{023M}] = 17.289. \tag{9.2}$$

The difference in these two numbers is due to different modes assigned to the Gompertz survival function. For females the mode of the Gompertz distribution was chosen as $m_G = 33$ and that for males was chosen as $m_G = 30$. These two expectations of life seem to be plausible for wild Chimpanzee populations.

It is often informative to include graph forms of the two survival functions under consideration. Presented in Figure 9.11 is a graph of the survival functions for females and males.

In the simulation experiment under consideration, a realization of the embedded deterministic model as well a sample of realizations of the stochastic model were computed. The parameter assignments for the deterministic model were as follows. The number of years of evolution of both the female and male populations was set as 200. The expected number of the total infants born in any year was assigned the value $\lambda = 20$. The initial number of females and males in the population was chosen as 300. The probability that an offspring was female was chosen as $p_f = 100/205$, and the probability an offspring was male was the complementary probability $p_m = 1 - p_f$.

For the case of the stochastic model, the number of years of evolution of the population was 200, and the number of replications of the experiment was chosen as 100. The initial number for both females and males was 300 and was the same as those used in the experiment with the deterministic model. In the experiment with the stochastic model λ was a random variable, and the numbers in equation 8.21 were chosen as $a = 15$ and $b = 25$. It is interesting to note that the expected value of the random variable λ in this case is

$$E[\lambda] = 20, \tag{9.3}$$

which is the same value for λ used in the experiment with the deterministic model.

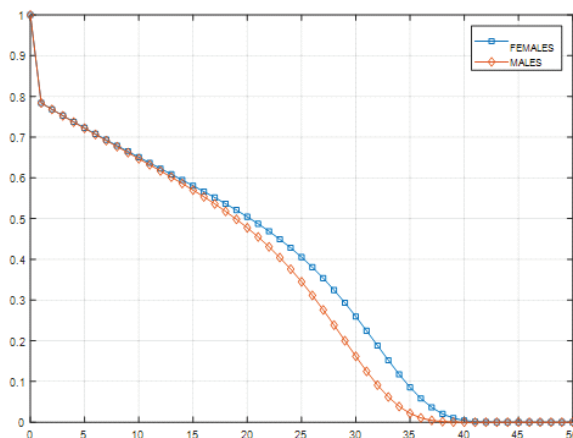


Figure 9.1. A Graph of the Female and Male Survival Functions

An important part of the procedure in setting up a computer simulation experiment is that of specifying the initial age structure for the female and male populations. Let $X(0, 0)$ denote the initial age structure of the female population and let $Y(0, 0)$ denote the initial age structure for the male population at time $t = 0$. Observe that $X(0, 0)$ and $Y(0, 0)$ are vectors with 1 row and 51 columns, and, as indicated above, the survival function used studying the 200 years of evolution of the female population was S_{023F} and that for males it was S_{023M} . The initial vector for the female population was chosen as

$$X(0, 0) = 300 \times S_{023F} , \tag{9.4}$$

and similarly that for the males was chosen as

$$Y(0, 0) = 300 \times S_{023M} \tag{9.5}$$

These assignments will work well in experiments with the deterministic model, but when using the stochastic model the vectors of the female and male populations must have values in the set $\mathbb{S} = (x \mid x = 0, 1, 2, 3, \dots)$. This set is often referred to as the set of non-negative integers. Thus if the initial vectors for the age structure of the female and male populations are as

$$X(0, 0) = [X(0, 0)] \tag{9.6}$$

and

$$Y(0, 0) = [Y(0, 0)] , \tag{9.7}$$

where the symbols $[X(0, 0)]$ and $[Y(0, 0)]$ indicate that the greatest integer of each element in the vectors $X(0, 0)$ and $Y(0, 0)$ have been selected. The operation defined in equations 9.7 and 9.8 ensures that in the Monte Carlo simulation experiment the elements of all vector belong to the set \mathbb{S} of non-negative integers. The software used in the Monte Carlo simulation experiment makes it necessary that the elements of all vectors are in the set \mathbb{S} .

10. Analysis and Interpretation of Monte Carlo Simulated Data

The results presented in this section were based data that was simulated by what is known by Monte Carlo Methods. A review of Monte Carlo Simulation Methods as they apply to mutation and selection formulated in Wright-Fisher models of evolution was published by Mode and Gallop (2008). In this paper, the formula used to compute random numbers in the experiments described in this section is also presented.

By definition the age structure of a population for females and males is the number of individuals in each age classification: $0, 1, 2, \dots, 51$. As expected in the experiment with the embedded deterministic model, the age structure of the population converged to a constant for each age classification for both females and males. By a little more than 50 years into the population projection experiment, convergence to constant age structure had occurred for both sexes. Various terms are used to describe such a population. Among these terms is to say that the projected population has converged to an equilibrium. Another term that is used that the population has converged to a stable population.

It is of interest to display a graph of the age structure that the deterministic process converged to for females and males after about 50 years into the projection of 200 years as presented in Figure 10.1.

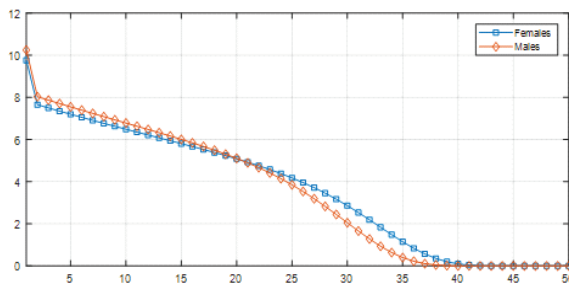


Figure 10.1. Age Structure of a Population in Equilibrium for the Deterministic Model

According to the prediction of the deterministic model, after a population reaches an equilibrium in the age structure for both females and males, this age structure would remain constant year after year. It seems very unlikely that if one were able to observe a wild population of Chimps, that the age structure of the female and male populations would remain constant.

For the experiment with the stochastic model, 200 years of evolution were computed and replicated 100 times. Let \mathbb{R} denote the array of simulated data. The first 200 rows of this array contain the simulated data for the first 200 years of the Monte Carlo simulation experiment, the next 200 rows contain the simulated data for the second 200 years of evolution and so on. It follows that the array consists of $200 \times 100 = 20,000$ rows and 51 columns, representing the set of ages under consideration. There are many ways to statistically summarize such simulated data.

The method chosen to statistically summarize the simulated data was to order each column of the array from the smallest number to the largest. In this procedure the array is expressed in terms of the order statistics for each of the 51 ages considered in the experiment. Row 1 of this array contains the minimum age *Min* for each age classification. And similarly, row 20,000 contains the maximum age *Max* for each age classification. Let Q_{25} denote the 0.25 quantile, which is the fraction of all columns such that the inequality is valid $x_{ij} \leq Q_{25}$ for $i = 1, 2, \dots, 25$ and $j = 0, 1, 2, \dots, 50$. Note Q_{25} is the row

$$Q_{25} = 0.25 \times 20000 = 5000 \tag{10.1}$$

of the ordered data array. Similarly, the Q_{50} quantile is row

$$Q_{50} = 0.5 \times 20000 = 10000 \tag{10.2}$$

of the ordered data array. Q_{50} is also known as the median of the data. It is easy to see that the quantile Q_{75} may also be defined. For the case under consideration *Max* is defined as row 20,000 of the ordered data array. It is clear that the notation just described could be extended to the case N , some large number rows and that software could be written to take into account extended the case just outlined

Presented in Figure 10.2 are the graphs of the Min, Max and Quantiles for females.

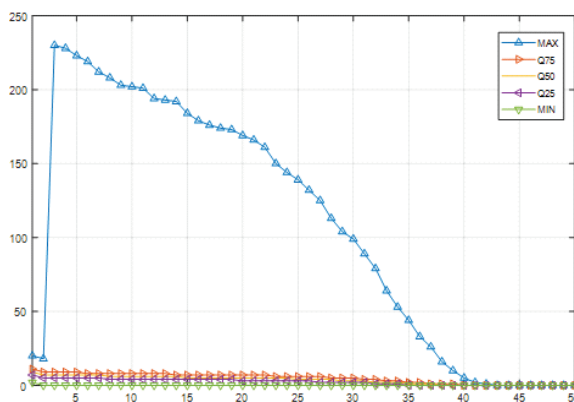


Figure 10.2. Graphs of the MIN, MAX and Quantiles for Females

It can be seen by inspecting the above graphs for females, that Min, Q_{25}, Q_{50} and Q_{75} are the x - axis, which indicates that they are small for each age group. It will be informative, therefore, to include a graph of Min, Q_{25}, Q_{50} and Q_{75} for females. Figure 10.3 contains the graph of Min, Q_{25}, Q_{50} and Q_{75} for females.

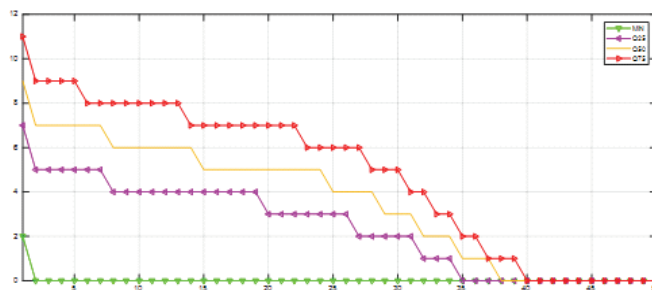


Figure 10.3. Graphs of Min, Q_{25}, Q_{50} and Q_{75} for Females

From graphs 10.2 and 10.3, it can be seen that the values of Min and the quantiles range from 0 to about 12. These numbers indicate that the numbers in the quantiles and Min are small, indicating that the Max for all ages is truly an outlier.

Another statistic that was included in the simulated data was total population size for each year of the projection of 200 years. This number was computed for each year of the projection by summing over all ages and sexes. In the experiment with the stochastic model, a sample of total population size was computed for each year the projection, which was replicated 100 times. Presented in Table 10.1 are the summary statistics for this experiment.

Table 10.1. Summary Statistics for Total Population Size

Min	293
Q_{25}	325
Q_{50}	344
Q_{75}	360
Max	9744

From this table it can be seen that there is also an outlier 9744 for total population size. It is also interesting to note that 0.75 of the realizations of the process are less than or equal to 360. It is of compelling interest to attempt to find an explanation as to why there are outliers in the simulated data. It is well known that for the type of age structured model under consideration that the initial age structure of the population has a profound effect on the subsequent evolution of the population that can be seen in the experiments with the deterministic and stochastic models. For the case of the deterministic model the expectation of the total number of offspring per year was assigned as constant number. Given this number and a numerical version of the survival function, the population will evolve to a fixed age structure. If the initial age structure differs from the fixed age structure, then unexpected results may be seen in the simulated data such as outliers that may appear in the first years of a projection. The initial age structure used in the experiment with the deterministic model was also used in the experiment with the stochastic model. In the experiment with the stochastic model, outliers were also observed during the initial years of an experiment.

It is of interest to compare, the evolution of total population size that was observed when using the deterministic model with that of the stochastic model. In the early years of the projection using the deterministic model, total population size was in the nine thousands but in about fifty years it had converged to 358.912, which is close to the median for the stochastic model displayed in table 10.1. Presented in Figure 10.4 is a comparison of total population size with the Q_{50} quantile for the stochastic model.

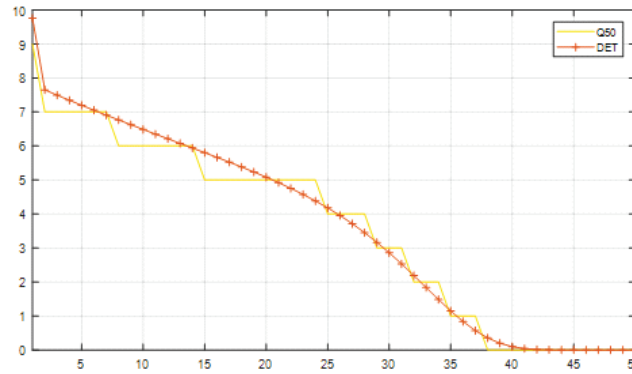


Figure 10.4. A Comparison of Age Structure According to the Deterministic Model in Year 100 with the Median for the Stochastic Model

As can be seen from Figure 10.4, the trajectory for total population size is close to that of the *Q50*, the median trajectory for females.

In the experiments with the deterministic and stochastic models, a survival function was computed for females and males in each year of the projection. In the experiment with the stochastic model each projection of 200 years was replicated 100 times and the mean survival function were computed for both females and males In Figure 10.5 a comparison plot of the deterministic survival function in year 100 of the projection with the mean survival function for females computed from the 100 realizations of the stochastic process.

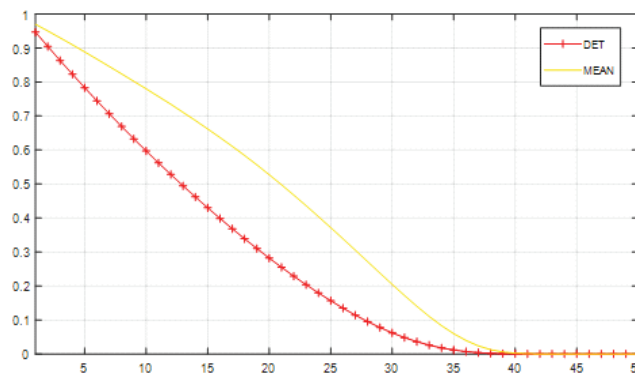


Figure 10.5. A Comparison of the Computed Survival Functions for the Deterministic and Stochastic Models

By inspecting Figure 10.5, it can be seen that the survival function for the deterministic model lies below the mean survival function for females. This graph illustrates that the output of the deterministic model may differ significantly from that of the stochastic model. In the graphs displayed above, the focus of attention has been the female population. Similar graphs for the male population were observed, but they have not been displayed here because they were similar to the graphs for the female population.

When attempting to forecast whether a small population of chimpanzee or that of another species would still exist 200 years into the future, the presence of outliers in a computer simulation experiment is troublesome, because it would not be possible for a small population of chimps to grow to a total population size of over 9,000 in a few years. Such outliers occur when the initial age structure for females and males differs from that determined by the expected number of offspring each year and the conditional probabilities that individuals of each age survive another year. In the experiment with the deterministic model, the convergence to a constant age structure for females and males occurred between 50 and 100 years into the projection. Given this information, a decision was made to do an experiment with the stochastic model using a slight modification of the age structures for females and males that arose as limits in the experiment with the deterministic model. Presented in Table 10.2 are the summary statistics computed from the simulated data for total

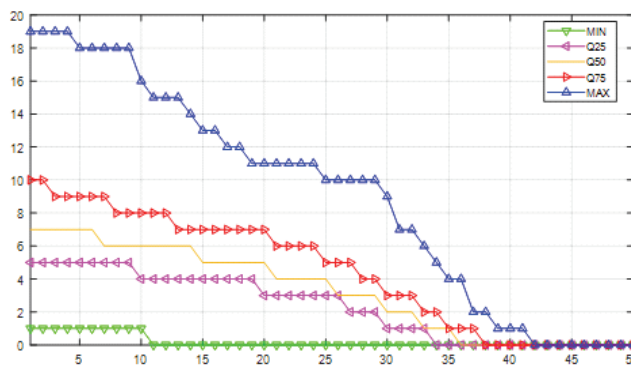


Figure 10.6. Graph of Selected Quantiles and the *Min* and *Max* for Females

population size in the experiment with the stochastic model.

Table 10.2. Summary Statistics for Total Population Size for Experiment 2 with the Stochastic Model

<i>Min</i>	314
<i>Q</i> ₂₅	334
<i>Q</i> ₅₀	354
<i>Q</i> ₇₅	369
<i>Max</i>	406

Unlike experiment 1 on total population size presented in Table 10.1, where the *Max* statistic is 9744, the *Max* statistic in Table 10.2 is only 406 and would not be classified as an outlier when compared with the other statistics in Table 10.2. This second experiment provides evidence that when the initial age structures for females and males as chosen as the greatest integers in the stationary age structure of the deterministic are used as the initial age structures in experiments with the stochastic model, then the stochastic model performs well in the sense that no outliers appeared in the simulated data on total population size.

It is also of interest to display a graph of the of selected quantiles as well as the *Min* and *Max* age structures. Figure 10.6 contains the graph of interest for the female population.

The information conveyed in Figure 10.6 may be viewed as a forecast as to whether an existing population of chimps will be the ancestors of a population of chimps in 200 years into the future. It is not possible to state with certainty that an ancestral population will exist 200 years into the future. But, the statistical information in Figure 10.6 will give conservationists ideas about the uncertainties of the existence of a future population. As can be seen from Figure 10.6 the age structure of the quantiles *Q*₂₅, *Q*₅₀ and *Q*₇₅ are zero only at the higher ages as one would expect. The age structure of the *Max* is also well above zero except for the higher ages. It is very interesting to note that age structure of the *Min* is zero beyond age 15. This result suggest that in some realizations of the process the population became extinct. But, on the other hand, the statistical information in Table 10.2 on the total population size provides evidence that the population would not become extinct during 200 years of evolution. The graph for males was very similar to that for females, and has, therefore, been omitted.

It is well known that the class of stochastic models under consideration will converge to what is called a quasi-stationary distribution, given that extinction does not occur. In the experiment just discussed, it is not clear as to whether the data generated by the stochastic mode did converge to the quasi-stationary distribution, because 200 years of evolution is too short a time for convergence to occur. A knowledge of the existence of a quasi-stationary distribution will, however, provide a structure for dealing with the uncertainty that arises in Monte Carlo simulation experiments with stochastic models.

The results just described provide convincing evidence that after all the parameter values have been assigned numerical values in the deterministic and stochastic models, then the first step of an experiment would be that of using the deterministic model to get estimates of the initial age structures for the female and male populations for the stochastic model. In the experiments reported in this section, a few seconds of running time of the deterministic were sufficient to compute

200 years of evolution. The next step in the experiment would that of using the adjusted age structures that were estimated in the experiment with the determinist model as initial age structures in an experiment with the stochastic model. Even though outliers may be observed in the experiment with the deterministic model, it seems very unlikely that outliers would be observed in an experiment with the stochastic model as suggested by observed results of experiment 2. The running time to the completion of the experiment 2 with the stochastic model was in the range of 2 to 4 minutes. These relatively short running times make it feasible to do several exploratory experiments before selecting an experiment with an intention of publishing the results.

References

- Artin, E. (1964). *THE GAMMA FUNCTION*. New York, Chicago, San Francisco, Toronto, London, Holt, Rinehart, and Winston, 1964.
- Elandt-Johnson, & Johnson, N. L. (1980). *Survival Models and Data Analysis*. John Wiley and Sons.
- Gage, T. B., & Mode, C. J. (1993). Some Laws of Mortality: How Well do They Fit. *Humanbiology*, 65, 445-461.
- Gompertz, B. (1825). On the Nature of the Function expression of the Law of Human Mortality, and a New Mode of Determining the Value of Life Contingencies. *Trans R. Soc., user A*, 115, 513 - 585.
- Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality, *J. Inst. Actuaries*, 107, 49-80. <https://doi.org/10.1017/S0020268100040257>
- Makeham, W. M. (1860). On the Law of Mortality. *J. Inst. Actuaries*, 13, 325-358. <https://doi.org/10.1017/S2046166600003238>
- Mode, C. J., & Busby, R. C. (1982). An Eight Parameter Model of Human Mortality - The Single Decrement Case. *Bulletin of Mathematical Biology*, 44(5), 647-659.
- Mode, C. J., & Jacobson, M. E. (1984). A parametric algorithm for computing model period and cohort human survival functions. *International journal of bio-medical computing*, 15(5), 341-356. [https://doi.org/10.1016/0020-7101\(84\)90040-0](https://doi.org/10.1016/0020-7101(84)90040-0)
- MODE, C. J. (1988). On statistically assessing critical population size of an endangered species in a random environment. *Mathematical Medicine and Biology: A Journal of the IMA*, 5(2), 147-166.
- Mode, C. J., & Jacobson, M. E. (1987). A study of the impact of environmental stochasticity on extinction probabilities by Monte Carlo integration. *Mathematical Biosciences*, 83(1), 105-125. [https://doi.org/10.1016/0025-5564\(87\)90006-X](https://doi.org/10.1016/0025-5564(87)90006-X)
- Mode, C. J., & Jacobson, M. E. (1987). On estimating critical population size for an endangered species in the presence of environmental stochasticity. *Mathematical Biosciences*, 85(2), 185-209. [https://doi.org/10.1016/0025-5564\(87\)90052-6](https://doi.org/10.1016/0025-5564(87)90052-6)
- Mode, C. J., Pickens, G. T., & Ewbank, D. C. (1987). Demographic Heterogeneity and Uncertainty in Population Projections. *Mathematical Medicine and Biology: A Journal of the IMA*, 4(3), 223-236.
- Mode, C. J., & Root, T. (1988). Projecting age-structured populations in a random environment. *Mathematical Biosciences*, 88(2), 223-245. [https://doi.org/10.1016/0025-5564\(88\)90043-0](https://doi.org/10.1016/0025-5564(88)90043-0)
- Mode, C. J. (1985). *Stochastic Processes in Demography and Their Computer Implementation*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo. <https://doi.org/10.1007/978-3-642-82322-0>
- Mode, C. J. (1991). A stochastic model for the development of an AIDS epidemic in a heterosexual population. *Mathematical biosciences*, 107(2), 491-520. [https://doi.org/10.1016/0025-5564\(91\)90020-J](https://doi.org/10.1016/0025-5564(91)90020-J)
- Mode, C. J., & Gallop, R. J. (2008). A review on Monte Carlo simulation methods as they apply to mutation and selection as formulated in Wright-Fisher models of evolutionary genetics. *Mathematical Biosciences*, 211(2), 205-225. <https://doi.org/10.1016/j.mbs.2007.05.015>
- Mode, C. J., & Sleeman, C. K. (2000). *Stochastic processes in epidemiology: HIV/AIDS, other infectious diseases, and computers*. World Scientific. <https://doi.org/10.1142/4243>
- Mode, C. J., Sleeman, C. K., & Raj, T. (2013). Simulating the emergence of mutations and their subsequent evolution in an age-structured stochastic self-regulating process with two sexes. *International Journal of Stochastic Analysis*, 2013. <https://doi.org/10.1155/2013/826321>
- Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60(4), 750-757. <https://doi.org/10.2307/1936612>

Smith, D., & Keyfitz, N. (2012). *Mathematical demography: selected papers* (Vol. 6). Springer Science & Business Media.

Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the Danish Life Insurance Company of 1871. *Journal of the Institute of Actuaries*, 16(5), 313-329. <https://doi.org/10.1017/S2046167400043688>

Widder, D. (1947). *ADVANCED CALCULUS*. New York, Prentice Hall.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).