Introducing Actuarial Science through Simulation

Kevin L. Shirley, FSA
email: shirleykl1@appstate.edu
Department of Mathematical Sciences
Appalachian State University, Boone, NC

Abstract

Mathematics departments often tout a career as an actuary as a path that their students may pursue after graduation. However, preparation for a career as an actuary has rapidly evolved since the late 1990s. Actuarial science is solidifying itself as an academic field of study. Actuarial organizations, such as the Society of Actuaries (SOA), classify qualified undergraduate programs as Centers of Actuarial Excellence. Traditional mathematics and statistics departments may feel they need additional resources for their students who pursue this path. The author hopes that papers such as this one will begin to fill this need. The purpose of this paper is to provide an introductory experience in actuarial modeling for undergraduate mathematics students. Using simulation as a method to explore the underlying distribution of an insurance product with financial risk allows the student to study problems that would otherwise be inaccessible without first taking an introductory course in contingencies.

1. Introduction

We are all exposed to risk on a daily basis. As children, we are taught to avoid certain risks or to develop habits to reduce our exposure to risks. Teenagers are given a curfew because being out past a certain hour is considered risky. We are taught to wash our hands before eating to reduce the risk of contracting contagious diseases. The consequences of an event related to these risks are often thought of in physical terms, but almost always have a financial component as well. Actuaries specialize in modeling risks that have financial consequences. The death of a family's primary financial provider, the loss of income due to disability, the expenses resulting from being involved in an automobile accident, or even dropping one's cell phone are all examples of risks with financial consequences. Actuaries have traditionally self-studied such models through a rigorous professional examination process. Today, future actuaries are receiving an increasing amount of their training at the undergraduate level before entering the workforce [10]. Students interested in reading about the expected preparation of the actuary will find information in [10]. In this paper, the author introduces risk modeling through simulation. The simulation method is increasingly utilized in the workplace. It also provides a way for a student with a less advanced background or a student attending a university with no actuarial science courses to experience the work of an actuary at an earlier stage in their academic development. An early exposure to risk models not only broadens the student's mathematics background, but also adds to the overall knowledge necessary to make an informed decision to pursue a career as an actuary. With this goal in mind, we will develop a model often used by actuaries to solve a broad range of problems.

2. The life table

Actuaries use life tables in a number of applications involving mortality risk. These tables contain information on the mortality of groups of lives. The information in the tables includes the number of lives in the group surviving to a given age and the number of deaths each year. Actuaries use life tables relevant to the populations involved in their risk calculation. For example, the life table used when determining mortality rates for a group of truck drivers will differ from a table used in determining mortality rates for fitness coaches. Finding information on life tables applicable to specific groups is normally quite difficult but is included in the duties of an actuary. Life tables for the general population can be found at [1]. To illustrate the use of a life table, we consider a sample of the total US population life table as shown in Table 1.

Table 1: a portion of US population life table 2010

Age	q(x)	l(x)	d(x)
0-1	0.006123	100,000	612
1-2	0.000428	99,388	43
2-3	0.000275	99,345	27
3-4	0.000211	99,318	21
⋮	⋮	⋮	⋮
34-35	0.001180	97,587	115
35-36	0.001235	97,472	120
36-37	0.001302	97,352	127
37-38	0.001377	97,225	134
38-39	0.001461	97,091	142
39-40	0.001557	96,949	151

In Table 1, a hypothetical cohort of 100,000 newborns is assumed. The life functions are shown as the column labels and defined as follows:

- $q(x)$, the probability of dying between the ages of x and $x+1$;
- $l(x)$, the number of lives surviving to age x;
- $d(x) = 100{,}000 \cdot q(x)$, the number dying between the ages of x and $x+1$.

We refer to an individual between the ages of x and $x+1$ as (x). From Table 1, relative frequency probabilities for the population are calculated. For example, the probability a newborn from this cohort survives to age 3 is $\frac{l(3)}{l(0)} = \frac{99318}{100{,}000}$, or the probability (35) survives to age 36 is $\frac{l(36)}{l(35)} = \frac{97352}{97472}$. One can construct the probabilities that (35) survives to age 35 + n by forming the ratios $\frac{l(35+n)}{l(35)}$. Hence, to calculate the probability of (35) dying before attaining age 35 + n we use the complement rule of probabilities $1 - \frac{l(35+n)}{l(35)}$. By constructing the probabilities that (35) dies before age 35 + n, we are constructing the cumulative probability distribution for the future lifetime random variable for (35). Since the future lifetime variable for (35) is a continuous random variable, we obtain a discrete approximation to this random variable. In a more general context, the future lifetime distribution is called the time to failure distribution and is the foundation for many actuarial calculations. A formal introduction to this random variable follows.

3. The time to failure random variable

The first step in modeling events contingent upon the occurrence of some failure is defining the "time to failure" random variable. For a life or product currently age x, this random variable will be denoted T_x. Understanding the distribution for T_x is foundational in modeling problems related to life insurance. In this paper we model life insurance for a group of insureds, and T_x is referred to as the future lifetime variable for (x). One of the more common future lifetime variables to use for human mortality is called the Gompertz model [3], developed by Benjamin Gompertz in 1825. This model uses two parameters, denoted by B and c. These parameters need to be estimated for the population of which our group of insureds is a subpopulation. For example, if we insure drivers from a trucking company, then we would estimate the parameters for the larger population of truck drivers. For the population used in this paper, we will use the US life table 2010 without modification. The cumulative distribution function for this distribution is

$$F_{T_x}(t) = 1 - \exp\left[-\frac{B}{\ln(c)} c^x (c^t - 1)\right].$$

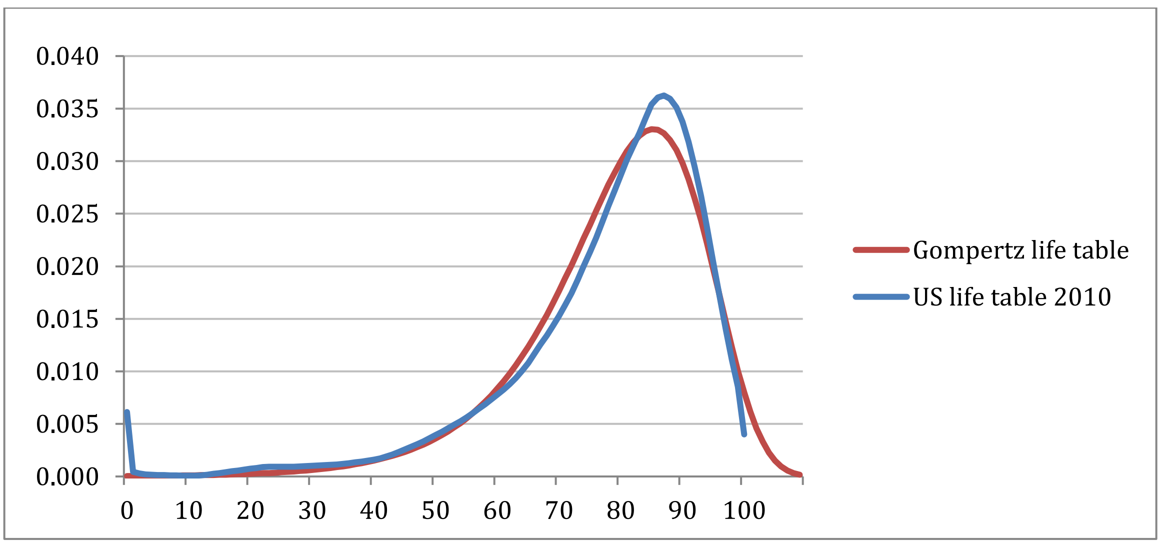

Figure 1: US life table 2010 distribution compared to the Gompertz model. (Produced by [Excel].)

Consider the future lifetime variable for a newborn, (0). The probability density functions for the US life table 2010 [1] and the Gompertz model with parameter values $c = 1.094$ and $B = 0.00004$ are shown in Figure 1. Notice the spike in the distribution from birth to about 1 year of age. This spike illustrates that a high risk of mortality still exists among newborns today. One should also note the local maximum occurring at approximately age 23. The tendency for risky behavior, accidents, and suicide in the late teens through the early twenties results in a mortality rate similar to that of an individual in their early thirties. The global maximum in this US life table occurs around age 87. The median age at death in the table is about age 82. The Gompertz distribution shown cannot capture the high mortality at birth or the local maximum occurring around age 23. However, it does capture both the general shape and mode of the distribution.

We use the matching quartile method from mathematical statistics to determine the parameter values for c and B [5]. To obtain the first quartile from the US life table, Table 3 in the appendix, we observe that $l(72) = 74{,}913$ or about 25% of the cohort has died by age 72. Since $l(89) = 27{,}131$ and $l(90) = 23{,}619$, the third quartile is about 89.5. We obtain the values for the parameters above by solving the system of equations

$$F_{T_0}(72) = 0.25$$
$$F_{T_0}(89.5) = 0.75.$$

A notable difference in the two models is the differing lifespans. Due to a lack of data for individuals living beyond age 100, the US life table ends at this age. A lifespan of age 110 is used in the Gompertz model shown in Figure 1. For comparison purposes, the probability that a newborn die in their forties using the Gompertz model is $F_{T_0}(50) - F_{T_0}(40) = 0.0230$ whereas the US life table 2010 gives 0.0250.

In order to provide a closer fit to an actual human population future lifetime distribution, more parameters could be used. One such model is the Makeham distribution [3], [4]. In the Makeham distribution a parameter is added to account for deaths that occur independent of age. For example, deaths due to accident are largely independent of age. The cumulative distribution function for the Makeham distribution is

$$F_{T_x}(t) = 1 - \exp\left[-At - \frac{B}{\ln(c)} c^x (c^t - 1)\right].$$

Also, using a different parameter fitting method such as the method of moments, a least likelihood estimation [5], or least squares [9] may result in a better fit.

4. A life insurance model

With a future lifetime distribution specified, payments contingent upon the death of the individual can be modeled. The details and terms of the insurance product will be written in a contract called the life insurance policy. These policies can be quite complicated. Actuaries play a critical role in how they are written. The contracts specify the event or events that must occur in order for the insured to receive payment. These specified events range from a single event in a term life insurance policy, to multiple events in a whole life policy with riders attached. Claim triggering events include death, disability, critical illness, long term care, and others. In this section and the next, we will model whole life policies with a death benefit paid at the time of death.

Since death may occur many years in the future, the value of the death benefit in today's dollars must take into account the time value of money. Knowing the value of the death benefit at the time the policy is written is important in determining the correct premium to charge. Using a level annual effective interest rate i, and the future lifetime random variable T_x, the present value of a \$1 death benefit paid at the instant of death for an insured currently age x is

$$Z_x = \frac{1}{(1+i)^{T_x}}$$

which is itself a random variable . To define Z_x in a standard way, let δ be the continuously compounded annual interest rate. Then $\delta = \ln(1+i)$. To take the present value (PV) of \$1 paid at time T_x in the future we write

$$Z_x = e^{-\delta T_x} = \frac{1}{(1+i)^{T_x}}.$$

If the death benefit is R, the present value is written

$$R \cdot Z_x = Re^{-\delta T_x}.$$

This random variable is well studied in actuarial modeling textbooks [3] and [4]. In introducing this random variable, textbooks commonly use an abundance of actuarial notation to develop a general theory of contingencies. We avoid the lengthy introduction and the development of actuarial notation. Instead, we use simulation and technology to gain insight into the distribution of Z_x for human populations. Furthermore, we use simulation to solve a problem that life pricing actuaries often encounter.

5. A group insurance model

Having specified a model for a single insured age x, we now consider a group of n insureds with current ages $x_1, x_2, \ldots, x_n$. Assuming that each life is independent, we have the following independent future lifetime random variables for the members in this group, $T_{x_1}, T_{x_2}, \ldots, T_{x_n}$. Let R_i be the death benefit paid at the time of death for the i^{th} insured. Then the present value of aggregate death benefit amount for the group is

$$Z_{G(n)} = \sum R_i \cdot Z_{x_i} = R_1 e^{-\delta T_{x_1}} + R_2 e^{-\delta T_{x_2}} + \cdots + R_n e^{-\delta T_{x_n}}.$$

We use information about the distribution of $Z_{G(n)}$ to compute the solution for the actuarial problem presented in the next section. However, the analytical determination of this distribution, as with any sum of independent non-identically distributed random variables, requires the use of convolution, and convolution is not tractable for $Z_{G(n)}$. To overcome this problem, we use simulation. Still, some analytical information about the aggregate distribution, such as $E[Z_{G(n)}]$ and $Var[Z_{G(n)}]$, is readily calculated from the single life distributions. By the linear property of the expected value we have,

$$E[Z_{G(n)}] = \sum R_i \cdot E[Z_{x_i}].$$

From the assumption that the lives are independent, the formula for the variance is written,

$$Var[Z_{G(n)}] = \sum {R_i}^2 \cdot Var[Z_{x_i}] .$$

We also use the assumption of independent lives in the method used to approximate this distribution through simulation. This assumption is so central to our analysis that if it fails to be reflected in the group of lives for which the analysis is intended, the results of our calculations will likely be misleading. We revisit the independence assumption in the discussion on risk in the next section.

6. The actuarial problem

Working as a life pricing and product development actuary for TransAmerica, the author was often presented with cases similar to the following. A small-sized employer desires to provide a worksite benefit for its employees. The employer has provided our life insurance company with a table of information that includes ages and proposed benefits. See Table 2. The employer requests a single premium for the entire group to provide this one-time life insurance benefit. An employer chooses to provide such a benefit for several reasons. If the benefit amount is based upon the number of years that the employee has been with the company, then it may be considered a longevity bonus. If it is based upon the current salary of the employee, then the employer may be compensating its employees for past wages that were perceived to be too low. Sometimes an employer uses worksite benefits as part of a total compensation package to attract the highest quality workers.

Table 2: a sample table from an employer showing employee information

Insured age	Benefit Amount	Sex	Tobacco Use
25	35,000	M	N
38	55,000	F	Y
45	70,000	F	N
61	65,000	M	Y
19	25,000	F	N
29	40,000	M	N
33	45,000	M	N
42	60,000	F	N
26	35,000	M	Y
35	50,000	M	N

The job of the actuary is to calculate a competitive premium that provides enough value to cover the death benefits as they are paid, the expenses the insurance company incurs as a result of this business, and an acceptable amount of profit. For a single premium group plan, the actuary begins by applying a premium principle for the benefit and risk and proceeds to incorporate expenses and other items into the model. For this paper, we apply one or more premium principles suitable for the death benefit and mortality risk the company is accepting.

A premium principle is one method used to determine the charge for a particular risk [8]. We should charge the employer a premium P that is at least the expected present value of the aggregate death benefit,

$$P \geq E\left[Z_{G(n)}\right],$$

called the actuarial present value of the benefit. Notice, if the deaths do not occur at the times expected, a premium amount equal to $E\left[Z_{G(n)}\right]$ may prove to be insufficient to pay the benefits. In most traditional long-term life insurance plans, the risk to the insurance company is the difference between the expected claims and the actual claims. If a number of deaths occur earlier than predicted on the life policies, then the claims in a particular year can be substantially larger than expected. This variation from expected is the risk that the insurance company accepts. It is measured by the standard deviation of the distribution of the present value of the death benefit,

$$STDEV\left[Z_{G(n)}\right] = \sqrt{\mathrm{Var}\left[Z_{G(n)}\right]}.$$

Since a premium principle includes a charge for risk, the standard deviation appears in some premium principle formulas. Also, claims due to death are more predictable when we insure more independent lives. Hence, the mortality risk of independent lives can be viewed as diversifiable [4].

If the lives considered are not independent, then the risk may actually increase with an increase in the number of lives insured. Consider a group of soldiers in the same platoon. Many deaths could occur from this group due to a single event, such as a particular battle or accident. An accident involving the platoon causing a large number of deaths would result in a much greater claim than we would expect if the lives were independent. In this case, the mortality risk is not diversifiable.

Another source of risk is the long-term interest rate we choose for calculating the present value of the death benefit. If the invested premium yields a return substantially less than this interest rate, we may fail to collect enough premium to cover the claim at the time of death. This interest rate and investment risk is another example of a non-diversifiable risk. The more policies we issue with this fixed interest rate, the more the effect an insufficient investment return will have on our ability to pay the claims. However, in this paper, we only model the diversifiable mortality risk of independent lives.

To formalize a small sample of premium principles, we use a loss variable X. In the case of a single life currently age x, with death benefit R, $X = R \cdot Z_x$. In the case of the group model $X = Z_{G(n)}$.

The premium principles we use are,

a) Expected Value Principle:
$$P = E[X](1+\alpha), \quad \alpha > 0;$$
b) Standard Deviation Principle:
$$P = E[X] + \alpha \cdot STDEV[X], \quad \alpha > 0;$$
c) Percentile Principle:
$$P = Q_\alpha \text{ where } \Pr[X < Q_\alpha\,] = \alpha, \quad 0 < \alpha < 1$$
for a continuous random variable X.

An actuary may use a premium principle as a starting point to develop an economically sound premium or as a way to determine how much of the premium charged is projected to cover the benefit and its corresponding risk. We apply these principles to show net premiums that cover the benefit and mortality risk.

7. Sampling from a distribution

To apply these premium principles, we must understand the distribution for $Z_{G(n)}$. To gain this understanding, we approximate the distribution using a Monte Carlo method. For each insured life in the group, we take a random sample from their future lifetime distribution. Using the sample value, we calculate the present value of the benefit and sum the results to obtain the aggregate present value for the group. This process is repeated 10,000 times. In this section, we discuss the method used to sample the distribution of T_x and apply it to estimate the present value of benefit distribution for a single insured.

One of the simplest ways to sample from a distribution is to use the inverse function method [5], [6]. Suppose we need to sample from a continuous probability distribution for a random variable, Q, with a cumulative distribution function $F(q)$. Let u be a randomly selected number from the interval (0,1). Define the random variable $Y = F^{-1}(u)$. Y is well-defined since $F(q)$ is an increasing function. Let F_Y be the cumulative distribution function for Y. We can identify F_Y by doing a short calculation.

$$\begin{aligned} F_Y(q) &= \Pr[Y \le q] \\ &= \Pr\,[F^{-1}(u) \le q] \\ &= \Pr\,[F(F^{-1}(u)) \le F(q)] \\ &= \Pr[u \le F(q)] \\ &= F(q) \end{aligned}$$

since u is in $U[0,1]$. Hence, Y has the same distribution as Q. We summarize the inverse function method algorithm to generate a random variable Q that has cdf $F(q)$,

1. Generate $u \in U[0,1]$;
2. Set $Q = F^{-1}(u)$.

To simulate the present value of benefit random variable, $R \cdot Z_x = Re^{-\delta T_x}$, we sample from the future lifetime distribution for (x), T_x, and then use the sample value to calculate the simulated value for Z_x. For human mortality, we use the Gompertz model previously described. Let $u = F_{T_x}(t)$ and solve for t to obtain the inverse function,

$$F_{T_x}{}^{-1}(u) = \frac{1}{\ln(c)} \ln\left(1 - \ln(1-u)\frac{\ln(c)}{Bc^x}\right).$$

When using the Makeham distribution for a human population, this algebraic process is not possible. In this case, we can use a numerical solver, such as Goal Seek in EXCEL to solve for particular inverse function values.

To demonstrate the simulation process, consider a single insured (35). We use the previously defined parameter values for B and c, $B = 0.00004$, $c = 1.094$, in the Gompertz distribution. Suppose the death benefit for this insured is 100,000, and the continuously compounded interest rate is $\delta = 0.04$. If the random number 0.67 is chosen from the uniform distribution, then the sample future lifetime for this individual is $F_{T_{35}}{}^{-1}(0.67) = 52.06$ years. The simulated value is

$$R \cdot Z_{35} = 100{,}000e^{-0.04 \times 52.06} = 12{,}463.07.$$

We simulate the distribution for $R \cdot Z_{35}$ by repeating this process 10,000 times and constructing a histogram of the results. One histogram resulting from this simulation is shown in Figure 2.

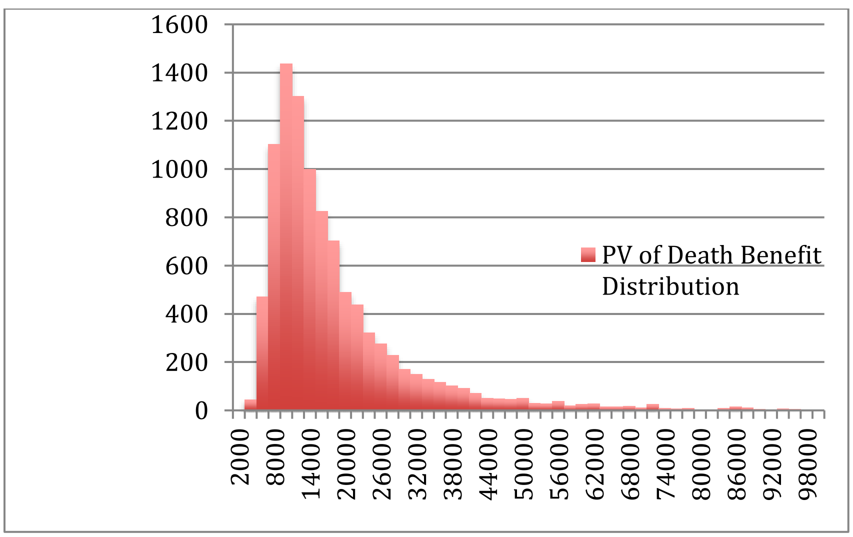

Figure 2: a histogram for the distribution of $R \cdot Z_{35}$ created using simulation. (Produced by [Excel].)

From this simulation we obtain information about the shape of the distribution, an estimate of the distribution mean, 19,305, an estimate of the standard deviation, 12,951, and estimates for the percentiles. If we apply the expected value principle with $\alpha = 0.5$ to this insured life, then we find the net premium $1.5 \cdot E[X] = 28{,}957.50$. If we apply the standard deviation principle with $\alpha = 1$, then we get $E[X] + \sigma_x = 32{,}256$. Using the 95th percentile, about 42,000, the percentile principle yields the net premium that has a 0.95 probability of covering the death benefit at the time of death. This probability assumes that we make at least a 4% continuously compounded return on the investments funded by the 42,000.

8. Solving the group insurance problem

Now that we can simulate the present value of the death benefit for a single insured, it is straightforward to extend our simulation to the group problem. There are four reasons simulation is worth considering when solving the group insurance problem. First, with simulation, we can use commonly accessible software, such as Microsoft EXCEL. Second, simulation allows students earlier access to studying the type of problem introduced in this paper. By contrast, using a traditional analytic approach requires a significant amount of knowledge and notation from a life contingencies textbook. Third, simulation provides enough information about the percentiles of the distribution that the percentile principle can be applied. It also allows us to estimate the shape of the distribution and its tail. Finally, the use of simulation lets us explore net premiums of other insurances and incorporate other measures of risk into our calculation as discussed in Section 9. A description of the algorithm follows:

1. Let n be the number of insured lives in the group;
2. Enter a Loop $k = 1$ to n ‘initialize age and benefit amounts
    a. Age of insured I, x_I.
    b. Benefit Amount for insured I, B_I
3. Next k
4. Enter Loop $J = 1$ to 10,000 ‘ Number of simulations
    a. Initialize total present value of death benefit for the group $Z_T = 0$
    b. Enter Loop $I = 1$ to n
        i. Sample the future lifetime of insured I
            1. Generate $u \in U[0,1]$;
            2. $T_I = {F_{T_{x_I}}}^{-1}(u)$
        ii. Simulate the Present value of the death benefit for insured I, $Z_I = B_I e^{-\delta T_I}$
        iii. Add to simulated group total $Z_T = Z_T + Z_I$
    c. Next I
    d. Record simulated group total.
5. Next J

A visual basic macro that uses Microsoft EXCEL to carry out the simulation is included in the appendix.

We apply the algorithm to the insurance model for the group of ten employees whose ages and benefit amounts are shown in Table 2. Using the Gompertz distribution with values $B = 0.00004$, and $c = 1.094$ for the future lifetime distribution results in the histogram in Figure 3. This histogram shows an estimate for the distribution of $Z_{G(10)}$.

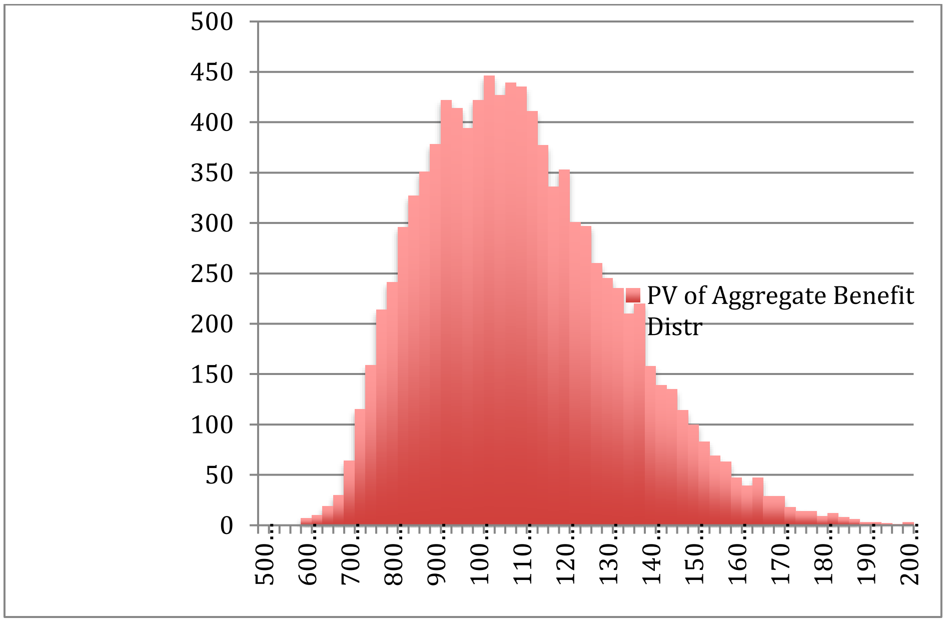

Figure 3: a histogram for the distribution of $Z_{G(n)}$ created using simulation. (Produced by [Excel].)

The horizontal axis in Figure 3 has units × 1000 dollars. From this simulation we obtain estimates of the distribution mean, 111,387, the standard deviation, 22,742, and the 95th percentile, 152,500. If we apply the expected value principle to $X = Z_{G(10)}$ with $\alpha = 0.5$, then we calculate $1.5 \cdot E[X] = 167{,}080.50$. Applying the standard deviation principle with $\alpha = 1$, $E[X] + \sigma_X = 134{,}129$. Using the 95th percentile, we find 152,500 is the net premium.

9. Conclusion and further study

A natural extension of the previous model is to consider term insurance. An n-year term life insurance policy pays the benefit if death occurs within n-years and pays nothing otherwise. Hence the random variable for the present value of $1 of benefit for (x) is defined by

$$Z_{x,term-n} = \begin{cases} e^{-\delta t}, & T_x < n \\ 0 \quad , & T_x \geq n \end{cases}.$$

For example, consider 20-year term insurance on (45) with a 100,000 benefit amount. We simulate the distribution for $X = 100{,}000 \cdot Z_{45,term-20}$ using the Gompertz distribution previously defined and $\delta = 0.04$. The EXCEL macro for this simulation is contained in the accompanying spreadsheet. The simulation results in an approximate mean of 7,420 and an approximate standard deviation of 20,685. In the resulting histogram, Figure 4, there are no scenarios resulting in a present value of benefit between 0 and 44,000. This observation is explained by the 20-year term of the policy. No claim is paid if (45) lives 20 years or longer leading to the lower bound for the present value of benefit $100{,}000 \cdot e^{-.04\times 20} = 44{,}933$. If we apply the standard deviation principle with $\alpha = 1, E[X] + \sigma_X = 28{,}105$. Because $\sigma_X > E[X]$, most of this net premium is paying for risk, which is economically unsound. In order to address this problem, the charge for risk must be reduced. We can reduce the risk charge per policy by issuing a large number of policies. If N identical policies are issued to independent lives, the resulting expected aggregate present value and aggregate standard deviation are $N \cdot 7{,}420$ and $\sqrt{N} \cdot 20{,}685$, respectively. Hence, the net premium per policy is

$$7{,}420 + \frac{20{,}685}{\sqrt{N}}.$$

Therefore, if 10,000 identical policies are issued, the net premium per policy is reduced to 7,626.85. The reduction in the risk charge from issuing a large number of policies illustrates the role that diversifiable risk plays in explaining the economics of the insurance industry.

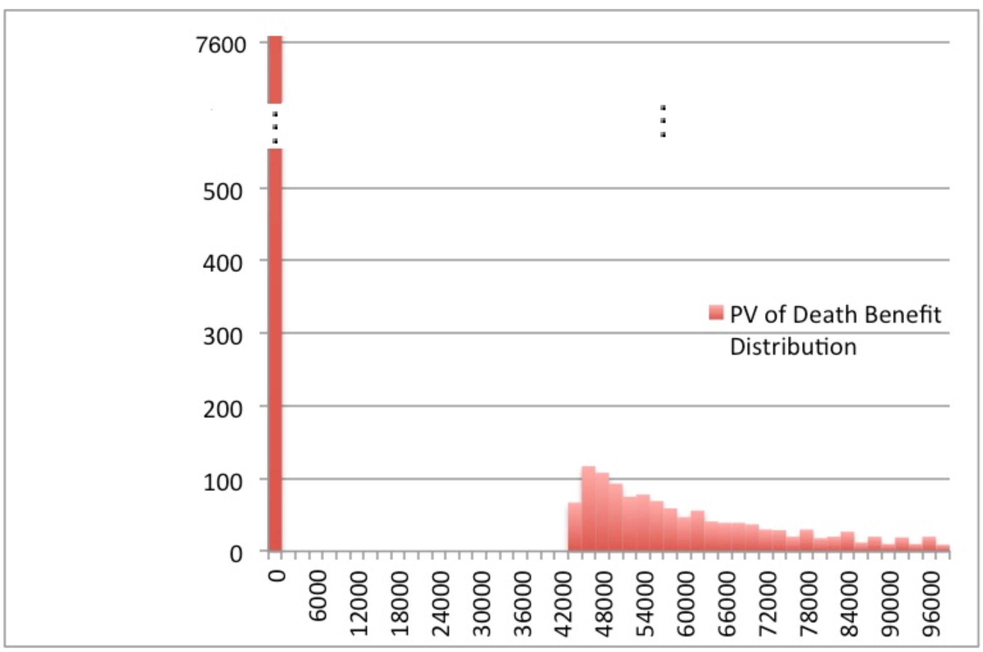

Figure 4: a histogram for the distribution of $100{,}000 \cdot Z_{45,term-20}$ created using simulation. (Produced by [Excel].)

It is instructive to compare the net premiums obtained at different ages and for different terms in order to understand the premium differences one observes in sample term life insurance rate sheets. Sample term life insurance rate sheets and quotes can be found online. Term insurance premium rates are normally given as a level monthly payment and will include a loading for profit and expenses. To determine a level premium paid periodically, it is necessary to model life annuities. The interested reader can find detailed discussions of life annuities in actuarial modeling textbooks [3], [4]. Also, the premium quote from a sample rate sheet will vary according to risk class. The risk classes used by insurance companies differ, but most include the designations preferred and non-preferred, male and female, tobacco and non-tobacco. A preferred, male, non-tobacco user is an example of a risk class using these designations. The actuary will use a different life table to calculate the premiums for each risk class. To make the group life net premiums as determined in Section 8 more applicable, the reader can use life tables that reflect the sex and tobacco use of each life in Table 2.

One issue with term insurance is that the insurance contract ends at the end of the term, and the insured's premium seems to have been lost. Of course, it actually goes to pay claims on other term policies whose unfortunate owners do not survive their term. In the late 1990s, in order to help alleviate the feeling of having lost the premium, companies introduced return of premium (ROP) term policies. In ROP term, if a claim is not made during the term, the entire premium is returned to the policy owner. Naturally, there is a cost to this additional benefit. Using simulation and the models developed thus far, this additional cost and the effect this benefit has on the present value distribution can be explored. We begin with the present value equation where R is the death benefit and P is the net premium,

$$Z_{x,ROP-n} = R \cdot Z_{x,term-n} + \begin{cases} 0, & T_x < n \\ Pe^{-\delta n}, & T_x \geq n \end{cases}$$

Since P is included in the random variable for the benefit, circularity results if the distribution for $Z_{x,ROP-n}$ is used in determining P. However, one can use the expected value principle as a starting point when calculating this premium. For example, if we use the expected value of the present value of the ROP benefit for the net premium, we obtain the equation,

$$P = R \cdot E\left[Z_{x,term-n}\right] + Pe^{-\delta n} \cdot prob[T_x \geq n],$$

and solving for P

$$P = \frac{R \cdot E[Z_{x,term-n}]}{1 - e^{-\delta n} \cdot prob[T_x \geq n]}.$$

To illustrate the determination of P, consider 20-year ROP term insurance issued to (45) with a 100,000 benefit amount. From the previous example,

$$100{,}000 \cdot E\left[Z_{45,term-20}\right] \approx 7{,}420.$$

Using the Gompertz distribution,

$$prob[T_{45} \geq 20] = 1 - F_{T_{45}}(20) = 0.8793.$$

Thus

$$P \approx \frac{7{,}420}{1 - e^{-.04\times 20} \cdot 0.8793} = 12{,}266.39.$$

Having a value for P, we can use simulation to estimate the distribution of $Z_{x,ROP-n}$. Beginning with the estimate of the distribution, the reader may develop a method or algorithm that can be used to determine P using the percentile principle for the present value of the ROP benefit.

It is also interesting to consider the use of distributions other than the Gompertz distribution for T_x. When using other parameterized future lifetime distributions, such as the Makeham distribution, it may be more appropriate to use another technique, such as the rejection method, for sampling. Useful and well-described methods of sampling and simulation are found in [6] and [7].

Another natural extension of the group problem is to explore the cost of interest rate risk. In this case, both the future lifetime variable and an interest rate path must be simulated. Suppose $N + f$, where N is a positive integer and $0 < f < 1$, is a sample value from the distribution of T_x. Let δ_i, $i = 1, \ldots, N + 1$, be a continuously compounded annual interest rate path where δ_i is the rate effective in year i. Then the present value of the benefit random variable becomes,

$$R \cdot Z_x = Re^{-\delta T_x} = Re^{-\left(\sum_{i=1}^{N} \delta_i + f \cdot \delta_{N+1}\right)}.$$

With this addition, our model captures not only mortality risk, but also interest rate risk. A very accessible introduction to modeling interest rates is found in [2].

Finally, the student may research the failure rates for commonly purchased items such as cell phones or laptops. From the failure rates, life table functions as illustrated in Table 1 can be calculated, and a curve fit to the data. Since a warranty is just term life insurance with a benefit amount equal to the cost or a prorated cost of the item, the term insurance model with some modification can be applied to determine net warranty prices.

Acknowledgements

The author would like to thank the reviewers and proofreaders for their many valuable suggestions.

Supplementary Electronic Material

Excel worksheet with data and macros:

Intro to ActSci Simulation.xlsm

References

[1] Arias, E., *United States Life Tables, 2010*, Centers for Disease Control (CDC), National Vital Statistics Report. Vol 63 No 7 (2014, November). Retrieved from http://www.cdc.gov/nchs/data/nvsr/nvsr63/nvsr63_07.pdf

[2] Babbel, D.F., Merrill, C.B. *Valuation of Interest Sensitive Financial Instruments.* SOA Monograph M-FI96-1. Schaumburg, Illinois, The Society of Actuaries, 1996.

[3] Bowers, N., Gerber, H., Hickman, J., Jones, D., Nesbitt, C. *Actuarial Mathematics, 2nd Edition.* Schaumburg, Illinois, The Society of Actuaries, 1997.

[4] Dickson, D.C.M., Hardy, M.R., & Waters, H.R. . *Actuarial Mathematics for Life Contingent Risks.* (International Series on Actuarial Science). Cambridge University Press, 2009.

[5] Hogg, R.V., Tanis, E.A. . *Probability and Statistical Inference, 5th Edition.* Upper Saddle River, NJ. Prentice Hall, 1997.

[6] Ross, S.M. . *Simulation, 3rd Edition.* San Diego, CA. Academic Press, 2002.

[7] Scollnik, D.P.M. . *Simulating Random Variates from Makeham's Distribution and from others with exact or nearly log-concave densities*. Transactions of Society of Actuaries, Vol 47. 1995.

[8] Young, V., *Premium Principles.* Encyclopedia of Actuarial Science, Vol 3., John Wiley and Sons, 2006

[9] Xinlong F., Guoliang H., Abdurishit, Estimation of parameters of the Makeham distribution using the least squares method, Mathematics and Computers in Simulation, Vol 77 No. 1, p. 34-44, February 2008.

[10] "*Education FAQs*," found at Beanactuary.org, 2016, http://www.beanactuary.org.

Software

[Excel] A product of Microsoft Corporation
https://www.microsoft.com/en-us/

Appendix

Visual Basic Macro for simulating the distribution for Group Life policies.

```
Sub GroupLife()
' GroupLife Macro
Sheets("Gompertz GL").Select

Let n = Excel.Cells(1, 16)
Let Delta = Excel.Cells(2, 16)  ' Continuously compounded interest rate

Let A = Excel.Cells(2, 3)
Let B = Excel.Cells(4, 3)
Let C = Excel.Cells(3, 3)
Application.Calculation = xlCalculationManual 'Speeds up macro
For i = 1 To 10

Let X = Excel.Cells(i + 3, 15) 'Age of insured
Let Benefit = Excel.Cells(i + 3, 16) ' Benefit Amount
Excel.Cells(3, i + 18) = X

   For j = 1 To n

    'This generates a randomly chosen value from U[0,1].
      Randomize Timer
      u = Rnd
      T = 1 / Log(C) * Log(1 - Log(1 - u) * Log(C) / (B * (C ^ X)))
      Z = Benefit * Exp(-Delta * T)
      'Excel.Cells(3 + j, 26) = T
      Excel.Cells(3 + j, i + 18) = Z

   Next j
Next i
Application.Calculation = xlCalculationAutomatic
End Sub
```

Table 3. Life table for the total population: United States, 2010, [1]

Age	Prob of dying between x and x+1 q(x)	Number surviving to age x l(x)	Number dying between x and x+1 d(x)	Person-years lived between age x and x+1 L(x)	Total Person-years lived above age x. T(x)	Expectation of life at age x. e(x)
0-1	0.006123	100,000	612	99,465	7,866,027	78.7
1-2	0.000428	99,388	43	99,366	7,766,561	78.1
2-3	0.000275	99,345	27	99,331	7,667,195	77.2
3-4	0.000211	99,318	21	99,307	7,567,864	76.2
4-5	0.000158	99,297	16	99,289	7,468,556	75.2
5-6	0.000145	99,281	14	99,274	7,369,267	74.2
6-7	0.000128	99,267	13	99,260	7,269,993	73.2
7-8	0.000114	99,254	11	99,249	7,170,733	72.2
8-9	0.000100	99,243	10	99,238	7,071,484	71.3
9-10	0.000087	99,233	9	99,229	6,972,246	70.3
10-11	0.000079	99,224	8	99,220	6,873,017	69.3
11-12	0.000086	99,216	9	99,212	6,773,797	68.3
12-13	0.000116	99,208	12	99,202	6,674,585	67.3
13-14	0.000175	99,196	17	99,188	6,575,383	66.3
14-15	0.000252	99,179	25	99,167	6,476,195	65.3
15-16	0.000333	99,154	33	99,138	6,377,028	64.3
16-17	0.000412	99,121	41	99,101	6,277,891	63.3
17-18	0.000492	99,080	49	99,056	6,178,790	62.4
18-19	0.000573	99,032	57	99,003	6,079,734	61.4
19-20	0.000655	98,975	65	98,942	5,980,731	60.4
20-21	0.000744	98,910	74	98,873	5,881,789	59.5
21-22	0.000829	98,836	82	98,795	5,782,916	58.5
22-23	0.000892	98,754	88	98,710	5,684,120	57.6
23-24	0.000925	98,666	91	98,621	5,585,410	56.6
24-25	0.000934	98,575	92	98,529	5,486,789	55.7
25-26	0.000936	98,483	92	98,437	5,388,260	54.7
26-27	0.000943	98,391	93	98,344	5,289,824	53.8
27-28	0.000953	98,298	94	98,251	5,191,479	52.8
28-29	0.000971	98,204	95	98,157	5,093,228	51.9
29-30	0.000998	98,109	98	98,060	4,995,071	50.9
30-31	0.001029	98,011	101	97,961	4,897,011	50.0
31-32	0.001063	97,910	104	97,858	4,799,051	49.0
32-33	0.001099	97,806	108	97,752	4,701,193	48.1
33-34	0.001137	97,699	111	97,643	4,603,440	47.1

Table 3. (Cont.) Life table for the total population: United States, 2010, [1]

Age	Prob of dying between x and x+1 q(x)	Number surviving to age x l(x)	Number dying between x and x+1 d(x)	Person-years lived between age x and x+1 L(x)	Total Person-years lived above age x. T(x)	Expectation of life at age x. e(x)
34-35	0.001180	97,587	115	97,530	4,505,797	46.2
35-36	0.001235	97,472	120	97,412	4,408,267	45.2
36-37	0.001302	97,352	127	97,289	4,310,855	44.3
37-38	0.001377	97,225	134	97,158	4,213,567	43.3
38-39	0.001461	97,091	142	97,020	4,116,408	42.4
39-40	0.001557	96,949	151	96,874	4,019,388	41.5
40-41	0.001663	96,798	161	96,718	3,922,514	40.5
41-42	0.001793	96,637	173	96,551	3,825,796	39.6
42-43	0.001962	96,464	189	96,370	3,729,245	38.7
43-44	0.002177	96,275	210	96,170	3,632,875	37.7
44-45	0.002423	96,065	233	95,949	3,536,705	36.8
45-46	0.002676	95,833	256	95,704	3,440,756	35.9
46-47	0.002931	95,576	280	95,436	3,345,052	35.0
47-48	0.003205	95,296	305	95,143	3,249,616	34.1
48-49	0.003505	94,990	333	94,824	3,154,473	33.2
49-50	0.003830	94,658	363	94,476	3,059,649	32.3
50-51	0.004177	94,295	394	94,098	2,965,173	31.4
51-52	0.004535	93,901	426	93,688	2,871,075	30.6
52-53	0.004903	93,475	458	93,246	2,777,386	29.7
53-54	0.005284	93,017	491	92,771	2,684,140	28.9
54-55	0.005684	92,526	526	92,263	2,591,369	28.0
55-56	0.006117	92,000	563	91,718	2,499,106	27.2
56-57	0.006589	91,437	603	91,136	2,407,388	26.3
57-58	0.007095	90,834	644	90,512	2,316,253	25.5
58-59	0.007626	90,190	688	89,846	2,225,741	24.7
59-60	0.008180	89,502	732	89,136	2,135,895	23.9
60-61	0.008767	88,770	778	88,381	2,046,759	23.1
61-62	0.009397	87,992	827	87,578	1,958,378	22.3
62-63	0.010085	87,165	879	86,725	1,870,800	21.5
63-64	0.010863	86,286	937	85,817	1,784,075	20.7
64-65	0.011758	85,348	1,004	84,847	1,698,258	19.9
65-66	0.012810	84,345	1,080	83,805	1,613,411	19.1
66-67	0.014011	83,264	1,167	82,681	1,529,606	18.4

Table 3(Cont). Life table for the total population: United States, 2010, [1]

	Prob of dying between x and x+1	Number surviving to age x	Number dying between x and x+1	Person-years lived between age x and x+1	Total Person-years lived above age x.	Expectation of life at age x.
Age	q(x)	l(x)	d(x)	L(x)	T(x)	e(x)
67-68	0.015290	82,098	1,255	81,470	1,446,925	17.6
68-69	0.016601	80,843	1,342	80,172	1,365,455	16.9
69-70	0.018005	79,501	1,431	78,785	1,285,283	16.2
70-71	0.019548	78,069	1,526	77,306	1,206,499	15.5
71-72	0.021294	76,543	1,630	75,728	1,129,192	14.8
72-73	0.023275	74,913	1,744	74,041	1,053,464	14.1
73-74	0.025528	73,169	1,868	72,236	979,423	13.4
74-75	0.028061	71,302	2,001	70,301	907,188	12.7
75-76	0.030820	69,301	2,136	68,233	836,886	12.1
76-77	0.033775	67,165	2,268	66,031	768,654	11.4
77-78	0.037252	64,896	2,418	63,688	702,623	10.8
78-79	0.041136	62,479	2,570	61,194	638,935	10.2
79-80	0.045411	59,909	2,721	58,549	577,741	9.6
80-81	0.050146	57,188	2,868	55,754	519,193	9.1
81-82	0.055445	54,321	3,012	52,815	463,438	8.5
82-83	0.061272	51,309	3,144	49,737	410,624	8.0
83-84	0.067764	48,165	3,264	46,533	360,887	7.5
84-85	0.075818	44,901	3,404	43,199	314,354	7.0
85-86	0.085319	41,497	3,540	39,727	271,155	6.5
86-87	0.094975	37,956	3,605	36,154	231,429	6.1
87-88	0.105525	34,351	3,625	32,539	195,275	5.7
88-89	0.117007	30,726	3,595	28,929	162,736	5.3
89-90	0.129450	27,131	3,512	25,375	133,807	4.9
90-91	0.142873	23,619	3,375	21,932	108,432	4.6
91-92	0.157280	20,245	3,184	18,653	86,500	4.3
92-93	0.172661	17,061	2,946	15,588	67,847	4.0
93-94	0.188988	14,115	2,668	12,781	52,259	3.7
94-95	0.206214	11,447	2,361	10,267	39,478	3.4
95-96	0.224274	9,087	2,038	8,068	29,211	3.2
98-99	0.282492	3,935	1,112	3,379	10,316	2.6
99-100	0.302838	2,823	855	2,396	6,937	2.5
>100	1.000000	1,968	1,968	4,542	4,542	2.3