ACTUARIAL RESEARCH CLEARING HOUSE
1997 VOL. 1

Stochastic Simulations using Spreadsheet Software

Margie Rosenberg
University of Wisconsin-Madison

Howard Young
University of Michigan

Abstract

The combination of mathematical theory, statistical theory, and computers will enhance the educational process of actuaries and equip actuaries to enter industry better prepared to analyze data. Some actuarial models can be studied analytically; now with the ease of working with computers, these models and others can also be studied as stochastic processes by using simulated data. Actuaries in the workplace use spreadsheets on a daily basis. Classroom preparation on model building using spreadsheet packages would provide valuable experience. This paper suggests ways of presenting ideas from the theory of compound interest and actuarial mathematics using simulation on spreadsheets.

1 Introduction

Actuaries have been historically educated in the fundamentals of mathematical and statistical theory applied to risk settings in business. With an increasing prevalence of computers in industry, in education, and at home, the educational process of actuaries can be enhanced by combining the theory with the use of computers. Some actuarial models can be studied analytically; now with the ease of working with computers, these models and others can be studied using simulated data.

Simulation is powerful tool that is widely used in business. Almost every topic in business can be approached through simulation. Complex models often can be simulated and the results analyzed more quickly than from the mathematical analytical principles. The main benefits of simulations are (i) building a computer model of a business process forces one to think carefully about the process; (ii) selecting inputs to the model forces one to think carefully about hypothetical conditions; (iii) analyzing the output of the simulation provides insight into the operational characteristics of the business process; and (iv) obtaining insight results in better decision making.

The increased statistical power of spreadsheets now permit easier use of simulation. Actuaries in the workplace use spreadsheets on a daily basis. University education of actuaries includes the use of spreadsheets as part of the learning process. Greater exposure of university students to the practical use of computers to help solve problems would be invaluable and lead to a greater understanding of business processes.

There are many examples of problems that could be better solved using statistical packages and using programming languages. The use of spreadsheets as a first step in the problem solution or as an initial exposure to the problem solution can be a worthwhile tool. The spreadsheet clearly indicates where there are problems in formulas, such as dividing by zero, or referencing a cell that does not exist, such as in a table. Spreadsheet packages are not as versatile in their selection of random number generation tools, such as generation of an exponential deviate, nor their choice of types of plots, such as boxplots, but one can accomplish most tasks with some creativity. For example, generation of exponential random numbers can be accomplished using a simple formula covered in most statistics texts and included in the Society of Actuaries Course 130 exam and Casualty Actuarial Society's 4B exam. (See the example in Section 2.)

This paper presents examples from the theory of compound interest and actuarial mathematics using simulation on spreadsheets. These examples are extensions of examples or exercises presented in the Actuarial Mathematics [1] and The Theory of Interest [2] textbooks. The examples illustrate how students of the material can gain a better understanding of the material with hands-on computer work: the theory presented in the textbooks can *come alive* and the ideas presented will be shown with numbers and pictures rather than formulas.

The outline of the paper is as follows. Section 2 briefly introduces simulation in spreadsheets using the software EXCEL 5.0.[1] Section 3 discusses possible ways of presenting the results. Section 4 provides the classroom examples and Section 5 discusses how one may incorporate the spreadsheet simulation technique into the classroom. Spreadsheets for the six examples presented in this paper are available upon request.

2 Spreadsheet Simulation

Early spreadsheet packages contained only a function RAND(), or some version of this function, which would generate a uniformly distributed deviate between 0 and 1. The inverse transform technique is one way of generating deviates from other distributions. This technique uses the cumulative probability distribution function to generate a value from a particular distribution. Briefly the approach is to generate random numbers from an uniform distribution on the interval $(0,1)$ which is denoted as $\mathcal{U}(0,1)$. These deviates are then included in another formula to generate values from the desired distribution.

Define F as the cumulative distribution function, with density function, f, for the random variable V. The distribution function is continuous and positive on an interval (a,b) such that $F(b) - F(a) = 1$ for $-\infty \le a < b \le \infty$. If V has density f, then $U = F(V)$ is uniformly distributed on $(0,1)$. Thus if $U \sim \mathcal{U}(0,1)$, then $F^{-1}(U)$ has density f. That is, if we generate a uniform deviate, then we can find a deviate from our desired c.d.f. F.

[1] Other spreadsheet software could have been used. Some of the methods presented in the paper are specified to EXCEL, but the functions could be translated for use in other software packages.

For example, suppose that $V \sim \mathcal{E}(\lambda)$, i.e. an exponential random variable with parameter λ. The c.d.f. of V is: $F_V(v) = 1 - e^{-\lambda \cdot v}$ for $v \geq 0$ The process to generate a deviate from an exponential distribution with parameter λ is as follows:

Step 1: Generate $U \sim \mathcal{U}(0, 1)$ and call the result, u.

Step 2: Set $u = 1 - e^{-\lambda \cdot v}$.

Step 3: Solve for v. The value, v, is a deviate from an exponential distribution with parameter λ.

The RAND() function can be dynamic. By entering "=RAND()" in a cell and pressing enter, the contents of the cell will change when changes are made to the spreadsheet. If one desires that the contents of the cell not change, then instead of pressing enter, one would press the F9 key and a fixed value of an uniform random number will be generated. The use of the dynamic version allows multiple trials of a simulation to be easily completed. This dynamic version, coupled with a macro to copy values of successive runs to another sheet, facilitates further analysis of the problem under study. For large spreadsheets, the use of the dynamic random function may be very time consuming to re-calculate a worksheet. One may prefer to set the calculation to 'manual' as opposed to 'automatic' to facilitate spreadsheet design and re-calculate the spreadsheet only periodically.

More recent versions of spreadsheet packages have increased the number of available functions. In EXCEL 5.0 two functions LOGINV and NORMINV have been added that calculate the inverse of the lognormal cumulative distribution function and the inverse of the normal distribution function. These two functions will be used in Example 2. Other techniques to simulate deviates exist, but are beyond the scope of this paper.

The Analysis Tools included in EXCEL 5.0 is another approach to generate random numbers. These tools include many statistical functions, among them random number generation. The types of random numbers quickly generated include uniform, normal, bernoulli, binomial, patterned, and discrete. Multiple numbers of variables can be calculated with multiple numbers of deviates, but they must be generated from the same distribution with one dialogue box. A random seed can be included so that the deviates can be reproduced at later times. The numbers produced are fixed and not dynamic. For more trials, more numbers will need to be generated.

Traditional statistics packages such as SAS and MINITAB are more versatile in their choice of random number functions and may provide better statistical routines to generate those values. If a complex distribution is desired, EXCEL would not be the choice of package to do the simulation.

3 Presentation of Results

Spreadsheet information is displayed in a tabular format. For a large simulation, the amount of information displayed may be overwhelming. Thus ways of summarizing the information to provide an easy means of analyzing the results is critical.

Summary statistics, such as averages, standard deviations, ranges, and percentiles, provide some information. These functions can be coded right into the spreadsheet. The advantage of using these functions is with the dynamic approach. New samples of deviates from the desired distribution would be recalculated by pressing the F9 key and new summary statistics would be generated at the same time.[2] The summary of all the simulation trials can be saved on a separate worksheet with the use of a macro and these statistics summarized as well. Some of the functions available include: AVERAGE, CORREL, FREQUENCY, KURT, LARGE, MAX, MEDIAN, MIN, MODE, PERCENTILE, QUARTILE, SMALL, STDEV, and VAR. All of these functions are described in EXCEL's HELP.

The Analysis Tools also provides a descriptive statistics option. The advantage of using the Analysis Tools is in its ease of use, but the numbers are fixed for the range of information given in the dialogue box for the descriptive statistics menu. If the number of variables and trials is known in advance, then this approach is much faster than the dynamic approach. However, if multiple simulations are desired, then this

[2] Note that this is a different use of the F9 key as described earlier. Here the formula RAND() is entered into the cell with the enter key. Then, every change of the spreadsheet will change the value in the cell. Or the cell's contents will change by pressing the F9 key.

approach is not as efficient as the dynamic approach. The summary statistics provided by the Analysis Tools are shown in Example 3.

In addition to summary statistics, graphs of defined items may be beneficial. EXCEL includes many plot options such as histograms, scatterplots, and timeplots. Noticeably missing is a boxplot option.[3] The above distinction regarding the dynamic vs. Analysis Tools summary statistics can be made here as well. The use of a frequency function or lookup function would be valuable to help display summary statistics and provide the input for a column chart to pictorially display the results. The histogram function as part of the Analysis Tools allows a chart to be created as well, but recall that the results are fixed, and if the numbers are updated, the histogram will need to created again.

4 Classroom Examples

Example 1: Theory of Interest Example 10.12: GIC binomial lattice

This example is an extension of Example 10.12 from The Theory of Interest [2]; the substance of which is as follows:

> The issuer of a Guaranteed Investment Contract (GIC), with a guarantee of 8.5% annual effective rate, will invest the proceeds in three-month instruments. The interest rate for the first quarter is 8.4% convertible quarterly. It is assumed that future interest rates move according to the random walk model; the probability of an upward movement is 0.4 and of a downward movement is 0.6. The amount of upward and downward movements each quarter are 0.5% and 0.4% convertible quarterly, respectively. What is the probability that the issuer will lose money on a one year GIC?

The solution given for the example shows an enumeration of the eight possible interest rate paths for three quarters. The probability of each path is calculated along with the quarterly interest rates and the annual accumulated value. Two paths will result in a rate less than the guarantee: an annual rate of 8.3% has probability of 0.144, and an annual rate of 8.0% has probability of 0.216.

While this technique theoretically applies to any duration GIC, the calculation procedure increases greatly for even relatively short durations: a 3-year GIC would have 2,048 possible interest rate paths (since there are 11 unknown quarters), while a 5-year GIC would have more than half-a-million paths. Therefore, simulation is a useful alternative approach.

Table 1 shows the input data and Figure 1 shows the results from simulating 1,000 runs for a 3-year GIC. For each run, a numerical interest rate path is generated, the resulting numerical values are retained on the worksheet and used in subsequent calculations. To simplify the printout, the paths are converted to indicate Up (U) or Down (D) rate changes between quarters, by using the Excel formula: IF(Current Quarterly Rate - Previous Quarterly Rate > 0 , "U" , "D") .

Determination of the interest rate, for any quarter, is simulated by comparing the $\mathcal{U}(0, 1)$ deviate with 0.4, the probability of an upward movement; if less, there is an increase, otherwise a decrease in the quarterly interest rate. This comparison is accomplished using the IF function, and including the RAND() function directly in the IF function. The cells for each quarter, after the initial period, have the formula: IF(RAND()< 0.4, prior quarter rate + increase, prior quarter rate - decrease).

The PRODUCT function gives the 3-year accumulation, and the equivalent annual rate is calculated by a formula for each of the 1,000 runs. For each run, the cumulative number of runs (from run 1 through the then current run) with Annual Average Factor less than the Target Factor is determined, and divided by the total number of runs up to that point. That is done by using the formula:

$$\frac{COUNTIF(range, criteria)}{COUNT(range)}$$

where the range covers the Annual Average Factor cells from run 1 through the run currently being processed, and the criteria is "< 1.085".

[3] Note: There are macro packages developed by outside vendors for EXCEL that can generate boxplots.

Table 1: **Input Data**

First quarterly interest rate	0.021
Possible increase in quarterly rate	0.00125
Probability of increase	0.4
Possible decrease in quarterly rate (use absolute value)	0.001
Total number of quarters	12
Target Annual Accumulation Factor (1 + Annual Effective Rate)	1.085

Figure 1: Simulated Interest Paths, Average Annual Accumulation Factors, and Percentage of Losses below Target Rate

Quarter->	2	3	4	5	6	7	8	9	10	11	12	AnnAvg	Ratio of
Trial												Factor	Runs<1.085
100	U	D	D	U	D	D	D	D	U	D	U	1.082	0.500
200	D	U	U	U	U	D	U	U	U	D	D	1.100	0.500
300	U	D	D	U	D	D	D	D	D	U	D	1.080	0.527
400	D	D	D	U	U	U	D	D	U	D	U	1.083	0.538
500	D	U	D	D	D	U	U	D	D	U	D	1.082	0.554
600	U	D	D	D	U	D	D	D	D	U	U	1.080	0.553
700	D	D	D	D	D	U	D	U	D	D	U	1.072	0.550
800	U	U	D	U	U	D	D	U	D	D	D	1.095	0.541
900	D	D	D	D	D	D	U	D	D	D	U	1.068	0.552
1000	U	U	D	U	D	D	D	U	U	D	D	1.092	0.550
Increase Ratio	0.42	0.42	0.40	0.40	0.40	0.39	0.38	0.40	0.42	0.39	0.42	**Mean= 0.404**	

The above summarizes 1000 simulation runs; print-out shows each 100th run.
For each run, an Interest Rate Path for Quarters 2-12 is generated as follows:

1) Each cell, for those Quarters, uses the function
   IF(RAND()<Prob, Previous Q rate + Inc, Previous Q rate - Dec)
2) The resulting numerical values (not shown above), and the 1st Q rate,
   produce the AvgAnn factor = PRODUCT(range) ^(4/12)
3) To simplify the printout, the paths are converted to Up or Down notation
   by using IF(CurrQ-PrevQ>0, "U", "D")

The Increase Ratio, by Quarter, for all 1000 runs is calculated using
COUNTIF(range,"U") / 1000
For all Quarters combined, the Mean = AVERAGE (range)

Also the proportion of runs with AvgAnn Factor < 1.085 is calculated on a cumulative basis (from Run 1 through the Current Run) using the following:
COUNTIF(range, "<1.085") / COUNT(range)

For example, Figure 1 shows an interest rate path for run 300, and the resulting Annual Average Factor for that specific run. On the same line, the ratio of the number of runs below target of 8.5% is shown for the first 300 runs.

As a check on the overall process, the ratio of increases in each quarter, and for all quarters combined, is shown for the entire 1,000 run sample. In Figure 1, each of those items is close to the expected value of 0.4. The COUNTIF function was also used for this, by applying it to the array of Up and Down indicators.

In order to show a printout of a reasonable amount of data, the rows with the intermediate runs (as well as some other portions of the worksheet) are hidden. Recalculation of the entire worksheet takes a considerable period of time to generate another 1,000 run simulation. However, a single press of the F9 key accomplishes this without opening the hidden portions of the worksheet.

The worksheet format can easily be modified to consider a GIC with a different duration. The primary change is to insert the necessary number of columns and copy the IF function where needed. Care must be given to see that all needed recopying is done, and that the PRODUCT function captures the appropriate range of cells. The array used to determine the ratio of increases also needs to be modified. Finally, since the number of quarters is a parameter in the Input Data section it must be changed; the other Input Data items shown also are parameters and can similarly be changed.

Example 2: Actuarial Mathematics Example 4.2: Actual versus expected accumulated fund for a group of 100 lives

Simulation runs can provide additional insights into the analytic results developed in Example 4.2 of Actuarial Mathematics [1]. The substance of that example is as follows:

> One hundred independent lives are subject to a constant force of mortality, $\mu = 0.04$; each life is insured for death benefit of 10 units; benefits will be paid from an investment fund expected to earn $\delta = 0.06$.

The text shows that $E[Z] = 4$ and $Var[Z] = 9$. Then, using a normal approximation, it is determined that an initial amount of 449.35 would be required for a 0.95 probability that the fund will be sufficient. Also, a single illustration is shown of the run-out of the fund, using a force of interest of 6%, and selected times-at-death for the first two years.

A single simulation run involves a generation of 100 uniformly distributed random numbers, and those are converted into time-at-death deviates (as described in Section 2 of this paper) using $\mu = 0.04$. For each of the time-at-death deviates, the present value of the 10 unit death benefit is calculated using $\delta = 0.06$. Once a single run has been set-up, new deviates can be generated by hitting F9; also the formulas can be copied to show multiple runs on a single worksheet.

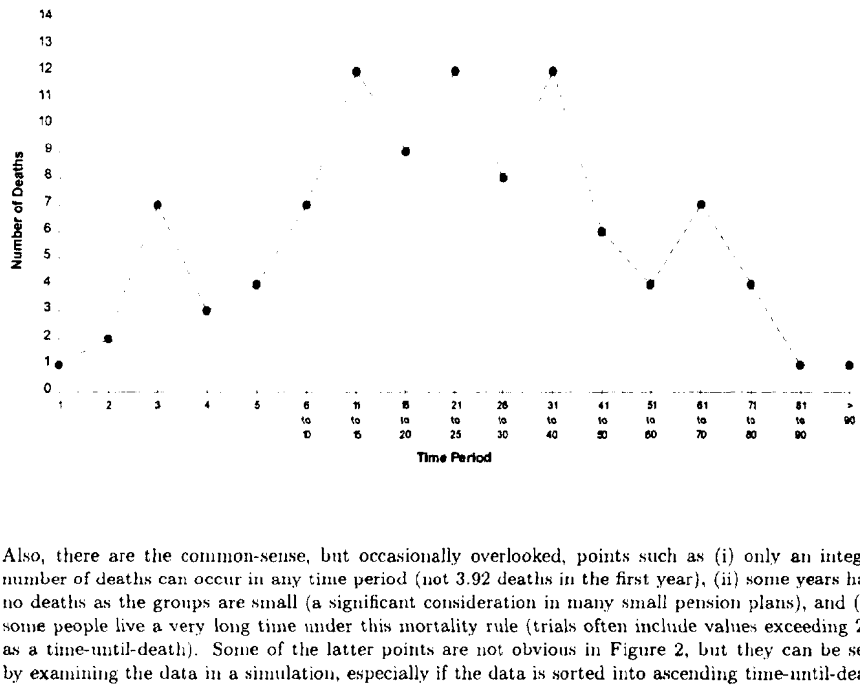
The AVERAGE and VAR functions provide approximate values of $E[Z]$ and $Var[Z]$; Table 2 shows such values for ten simulation runs. One can see how close these results are to the analytically determined values. Note that 100 times each mean equals the initial fund needed for the sample in that run. In Table 2, only one of the runs (with 4.6 mean) would require a fund in excess of the 95% minimum which was calculated at 449.35. Of course, ten runs is not sufficient to draw any conclusions. By merely hitting F9, the means of many runs can be computed quickly. Using this procedure, 100 runs were generated, and only 6 of those would require an excessive fund. This example shows the sampling variation using simulation.

Table 2: Simulation Output for 10 Runs

	Run Number									
Statistic	1	2	3	4	5	6	7	8	9	10
Mean	3.4	4.0	4.1	4.6	3.8	4.2	4.3	3.7	4.0	4.1
Variance	8.4	9.9	8.5	9.1	8.4	9.4	9.2	9.8	8.4	8.9

The FREQUENCY function yields the number of deaths in specific time intervals when applied to the data in each simulation run. The results are easily converted into a chart; for example, Figure 2 shows the results for one run. Deaths are grouped for later time periods. Unlike the smoothly decreasing number of deaths per year implied by analytic use of the constant force of mortality, this run has 1 death in the first year, 2 in the second, 7 in the third and exhibits a jagged curve.

Figure 2: Frequency Output for 1 Run of 100 Deaths



Also, there are the common-sense, but occasionally overlooked, points such as (i) only an integral number of deaths can occur in any time period (not 3.92 deaths in the first year), (ii) some years have no deaths as the groups are small (a significant consideration in many small pension plans), and (iii) some people live a very long time under this mortality rule (trials often include values exceeding 200 as a time-until-death). Some of the latter points are not obvious in Figure 2, but they can be seen by examining the data in a simulation, especially if the data is sorted into ascending time-until-death sequences. An extract of that sorted data is shown as run 1 in Figure 3; note that the FREQUENCY function can process sorted or unsorted data.

To sort data, the spreadsheet software in EXCEL requires that cells have fixed values, rather than formulas; otherwise the recalculation that occurs during the sorting process generates new random deviates and the result is just a new unsorted sample. While cell contents can be directly converted from formulas to values, a better approach is to copy them using the Edit-PasteSpecial command, so that the original formula cells can continue to be used to generate new samples.

Sorted data for any run can be used to simulate the run-out of an initial fund, with stochastic time-at-death patterns. Furthermore, the run-out can be done with various interest processes, including some stochastic concepts.[4]

[4]The first edition of Actuarial Mathematics does not include analysis of stochastic interest processes; it is expected that will be covered in the next edition. Therefore, even though it is somewhat inconsistent to use stochastic interest in the run-out of an initial amount that was based on constant interest, the extension is easy to understand.

Figure 3: Run-out of initial fund of 400 and death benefit of 10 with stochastic mortality and three interest rate processes

Time-at-Death		Stochastic Rate of Interest		Time-at-Death using Run 1 Fund Value, with Interest using			Time-at-Death using Run 4 Fund Value, with Interest using		
Run 1	Run 4	Log Normal	Random Walk	Force at 6% (fixed)	Log Normal	Random Walk	Force at 6% (fixed)	Log Normal	Random Walk
				400	400	400	400	400	400
0.524	0.525	1.062	1.062	403	403	403	403	403	403
1.503	0.592	1.062	1.062	417	417	417	394	394	394
1.618	0.727	1.054	1.054	410	410	410	388	387	387
2.070	0.849	1.053	1.045	411	409	408	380	380	379
2.181	1.103	1.074	1.057	404	402	400	376	377	375
2.300	2.357	1.067	1.061	397	396	393	396	399	394
2.549	2.729	1.077	1.075	393	393	390	395	400	395
2.565	2.818	1.054	1.068	383	383	381	387	391	387
2.619	2.831	1.073	1.079	375	375	372	377	382	377
2.643	3.105	1.055	1.073	365	365	363	373	378	375
3.536	3.233	1.059	1.070	375	374	376	366	370	368
3.658	3.757	1.063	1.072	368	367	369	368	372	371
3.814	4.223	1.057	1.068	361	360	363	368	372	373
4.279	4.502	1.059	1.065	362	360	364	365	368	370
		HIDDEN ROWS HERE							
58.366	55.088	1.075	0.927	1913	1341	-74	341	182	-82
59.043	55.657	1.040	0.906	1982	1367	-79	343	176	-88
60.476	57.739	1.067	0.911	2150	1491	-79	379	192	-82
60.880	59.343	1.061	0.911	2193	1517	-86	407	201	-81
61.359	59.813	1.057	0.906	2247	1548	-92	409	196	-87
64.546	62.445	1.065	0.909	2711	1880	-78	469	221	-78
64.551	64.835	1.064	0.910	2701	1871	-88	531	247	-72
66.497	66.940	1.064	0.912	3026	2100	-83	593	271	-69
69.073	68.318	1.065	0.915	3522	2460	-76	634	285	-71
70.555	68.725	1.047	0.901	3839	2625	-76	639	281	-78
70.829	70.254	1.061	0.901	3893	2658	-83	691	298	-77
71.228	80.627	1.060	0.900	3977	2711	-90	1277	536	-36
79.662	87.419	1.059	0.897	6587	4374	-46	1910	779	-27
81.893	95.102	1.071	0.905	7521	5088	-47	3018	1310	-23
104.810	203.078	1.066	0.909	29735	21926	-15	1964955	1280796	-10

Two simulations using RAND() and inverse transform. Values frozen and sorted.

Fund runouts, using interest rate (from specified process) for period between time-at-death (for the appropriate Run).

One set of new RAND() values is used as *Probability* for:

Force of interest generated by LOGINV(*Probability*, 0.06, 0.01).

Interest rate generated by: NORMINV(*Probability*, PriorIntRate, 0.01)

Figure 3 shows data from 6 run-outs of the fund for two mortality runs using interest rates generated from three different interest processes. The two mortality runs correspond to run 1 and run 4 of Table 2. They were chosen because their results were the extreme mean values in that table, with approximate initial fund requirements of 340 and 460 respectively.

The three interest rate processes all start with $\delta = 0.06$; therafter they are determined by either:

1. The 6% constant force of interest.
2. A lognormal generation of the force of interest, using the function LOGINV, with a mean of 6% and a standard deviation of 1%.
3. A random walk generation of the interest rate, using the function NORMINV, with a mean for each period equal to the prior period interest rate and a 1% standard deviation.

For processes (2) and (3), a single new run of uniformly distributed random numbers is generated. These serve as the probability input items for the LOGINV and NORMINV functions. A new simulation is generated each time the worksheet is recalculated; however, as noted above, the mortality run deviates are fixed, so they do not change with recalculations of the worksheet.

For convenience, interest rates are assumed to change immediately after **each benefit payment**, rather than annually or at other fixed time intervals. Thus, the length of time a particular interest rate is used depends on the mortality of the group. The three interest rate processes have the same apriori expected means but different variances, so it is not surprising that the simulated run-outs produce quite different residual amounts.

The simulated run-outs use an initial fund of 400 rather than the 95% minimum initial amount of 449.35. The final expected value of a large number of mortality simulations would be zero for the constant force of interest process.

Rather than comparing the final fund amounts which are distorted by the large difference in times for the last death (about 105 years for trial 1, and 203 years for trial 4), a comparison of values immediately after payment of the penultimate death benefits (which are respectively about 82 and 95 years in the future) is shown in Table 3.

Table 3: Comparison of Fund Value

Run Number	$\delta = 0.06$	LogNormal	RandomWalk
1	7,521	5,088	- 47
4	3,018	1,310	- 23

It is not suggested that these are estimates of averages for the possible population of run-outs. These reflect only two samples from the joint distribution of mortality and interest rate variables. The goal here, aside from learning the simulation concepts and related spreadsheet techniques, is to provide concrete illustrations which can be dynamically revised of the actuarial material.

Example 3: Actuarial Mathematics Section 3.6: Assumptions for fractional ages

Actuaries in their work generally assume expected mortality probabilities through the use of a life table which are displayed by integer year of age last birthday. To complete calculations at fractional ages and fractional durations, such as for year-end reserves or to calculate functions such as the complete expectation of life from the life tables, an assumption for the time of death within the year of age is made.

Three possible assumptions are the uniform distribution of deaths (UDD), constant force (CF), and Balducci (BAL). Let q_x be the annual mortality rate for a life aged (x). Table 3.5 of Actuarial Mathematics [1] describes ${}_tq_x$, for $0 \le t \le 1$ for each of the assumptions: $t \cdot q_x$ for UDD, $1 - \exp[-\mu \cdot t]$ for CF, and $\frac{t \cdot q_x}{1-(1-t)q_x}$ for BAL. Note that t is restricted between 0 and 1 and that ${}_tq_x$ does not integrate to 1, as it is not conditional on the death of (x).

We can use the above to define the cumulative conditional distribution function for death of (x), given death prior to age $(x+1)$. For $0 \leq t \leq 1$, $Pr\,(T(x) \leq t \mid T(x) \leq 1) = \frac{tq_x}{q_x}$. Since $Pr\,(T(x) \leq t) = tq_x$, then $Pr\,(T(x) \leq t) = q_x \cdot Pr\,(T(x) \leq t \,|\, T(x) \leq 1)$. For $0 \leq t \leq 1$, $Pr\,(T(x) \leq t \,|\, T(x) \leq 1) \sim \mathcal{U}(0,1)$, as it is a cumulative distribution function.

One way to illustrate the difference between the assumptions is to graph the cumulative conditional distribution function for death of (x), given death prior to age $(x+1)$. Two mortalitiy rates are illustrated in Figure 4. The first graph uses a very high mortality rate of 0.50 which clearly shows the impact of choosing of the three assumptions. For example, the probability that someone will die within the first quarter of the year, given that they die during that year of age is 0.25 under the UDD assumption, 0.32 under CF, and 0.40 under BAL. The next graph showing a 0.01 probability of dying in a given age, demonstates indistinguishable differences between the three assumptions.

The moments for these conditional distributions can be calculated as well. For example, the mean can be found using the equation $\int_0^1 (1 - \frac{tq_x}{q_x})dt$, where $\frac{tq_x}{q_x}$ is the cumulative distribution function. The mean is 0.5 for UDD, independent of the mortality rate, $1 - \frac{1}{q_x} - \frac{1}{\log(1-q_x)}$ for CF, and $\frac{q_x - 1}{q_x}\left(1 + \frac{\log(1-q_x)}{q_x}\right)$ for BAL. With the mortality rate at 0.5, the theoretical means under UDD, CF, and BAL are 0.5, 0.4427, and 0.3863. Other moments can be found as well. All of these calculations and graphs can be easily completed on a spreadsheet.

These same results can be illustrated by simulation. Using the inverse transform technique, uniform deviates between 0 and 1 are generated. Each of these values are multiplied by the one-year mortality rate to solve for the time of death during the year under each of the three assumptions. The steps are: (i) Generate $U \sim \mathcal{U}(0,1)$; (ii) Set $U = Pr\,(T(x) \leq t \,|\, T(x) < 1)$; (iii) For UDD, set $t \cdot q_x = q_x \cdot U$. Thus $t = U$; (iv) For CF, set $1 - (1-q_x)^t = q_x \cdot U$. Thus $t = \frac{\log(1-q_x \cdot U)}{\log(1-q_x)}$; (v) For BAL, set $\frac{t \cdot q_x}{1-(1-t) \cdot q_x} = q_x \cdot U$. Thus $t = \frac{U \cdot (1-q_x)}{1-q_x \cdot U}$.

One thousand values are generated for U and time until death values for each of the three assumptions are calculated assuming that the mortality rate is 0.5. Descriptive statistics for the generated values are shown in Table 4.

Table 4: Descriptive Statistics by Fractional Age Assumption Given $q_x = 0.50$

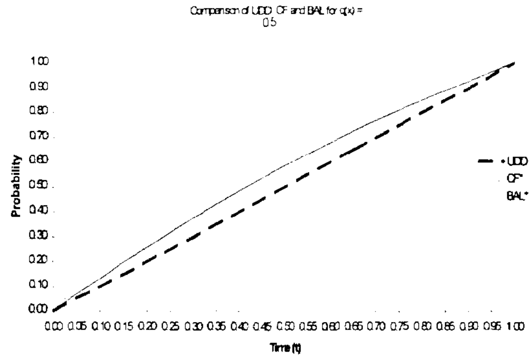
	UDD	CF*	BAL*
Mean	0.5038	0.4468	0.3905
Standard Error	0.0092	0.0091	0.0089
Median	0.5012	0.4162	0.3344
Mode	0.1591	0.1196	0.0864
Standard Deviation	0.2903	0.2872	0.2819
Sample Variance	0.0843	0.0825	0.0795
Kurtosis	-1.2289	-1.1625	-0.9724
Skewness	-0.0090	0.2230	0.4604
Range	0.9935	0.9944	0.9948
Minimum	0.0052	0.0037	0.0026
Maximum	0.9987	0.9981	0.9974
Sum	503.8358	446.7720	390.5416
Count	1000	1000	1000
Confidence Level(95.000%)	0.0180	0.0178	0.0175

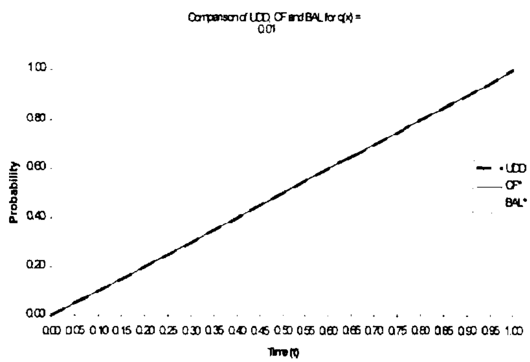
*Depends on q(x) = 0.5

Note how closely the sample means compare to the theoretical means: all are within one standard error. As a side project, one could illustrate the concept of sampling distributions, by repeatedly drawing samples of 1,000 and recording the sample mean for a number of trials. The distribution of these sample means could be described. The mean should correspond to the theoretical mean and the standard deviation to the theoretical standard error.

Histograms based on these simulated values are shown in Figure 5. The charts show a level number of values by cell for UDD, some skewness for the CF, and greater shifting to the left for BAL.

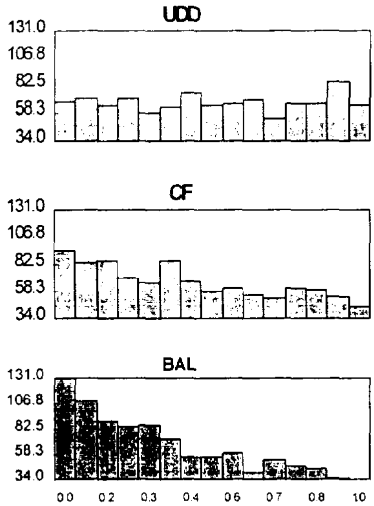
Figure 4: Comparison of UDD, CF, and BAL Assumptions for Two Mortality Rates 0.5 and 0.01





The student has gained an understanding of the three assumptions through this exercise. The conclusion can be reached that the choice of the fractional age assumption is not critical for low mortality rates.

Figure 5: Histogram of 1000 Deaths Simulated using UDD, CF, and BAL Assumptions for Mortality Rate of 0.5



Example 4: Actuarial Mathematics Exercise 3.24: Fractional age assumption in the calculation of the median age at death

Exercise 3.24 asks the student to use the U.S. Life Table and an assumption of uniform distribution of deaths in each year of age to find the median of T, where T is the future lifetime of a person aged (0). This problem is one that is easier to do theoretically, but the simulation of the answer requires a good working knowledge of the concepts. To do the problem without simulation, one could just look for where 50,000 people are alive, somewhere between ages 77 and 78. Using the uniform assumption, one could just do a linear interpolation to solve for the answer of 77.59 years.

Figure 6 graphs the l_x and d_x expected values from age 0 to 111. The deaths at age 0 last birthday are high at 12.6 deaths per 1,000, but then drop to around 0.5 deaths per 1,000. One can calculate the q_x's at each age and see that they are all very small except at the very highest ages. Thus the choice of the fractional age assumption will not be significant.

The simulation of this exercise uses the LOOKUP function in EXCEL. A probability density function and then the cumulative distribution function is created for each life aged (0). One hundred $\mathcal{U}(0,1)$ values are generated to represent 100 hundred newborns. The uniform deviate is compared against the cumulative distribution function to determine the integer age at death. The LOOKUP function takes the value less than or equal to the lookup value, one must be careful to code the spreadsheet formula correctly. It may be, depending on the design of the spreadsheet, that one must add 1 to the value returned by the LOOKUP function. Care must also be employed for treatment at age 0 last birthday.

This in particular highlights the advantage of using a spreadsheet, as an error for looking up a value will be clearly displayed in the spreadsheet. The fractional probability piece is found by subtracting the uniform deviate by the lookup value and dividing by the difference between the c.d.f.'s for the ages surrounding the deviate. Then UDD deviates, or other fractional age assumption, can be found as shown in Example 3. The total time until death is found by adding the integer and fractional pieces.

The simulation of 100 lives shows the distribution of the time until death rather than just a number of 77.59 years for the median. The minimum and maximum times until death can be found by using the MIN and MAX functions. The differences in time of death between individuals, even though they are identically distributed, can be dramatic. For example, for one trial of 100 lives assuming UDD, the median time until death is 77.94, the minimum is 0.88, and the maximum is 107.18. For another sample of 100 lives, the median time until death is 78.14, the minimum is 18.61, and the maximum is 101.89.

This is a challenging exercise to complete on the computer, but very practical, as actuaries commonly work with data that is discrete and must make assumptions for interim time periods. The exercise could be taken further to examine the difference in the results for the choice of fractional age assumption. Since the annual mortality rates are small for almost all ages, the spreadsheet results confirm those in the prior example, showing little differences in the median age at death and other statistics.

Example 5: Actuarial Mathematics Exercise 8.7: Joint life probability and expectations using DeMoivre's Law with $\omega = 100$.

The cumulative distribution function and density function of order statistics has always been a difficult concept for students to understand. This example and the next will show that the use of simulation makes a difficult concept easy and allows one to quickly get reliable answers.

Exercise 8.7 of Actuarial Mathematics assumes deMoivre's law with $\omega = 100$. The problem asks the student to calculate $\mathring{e}_{40:50}$, $\mathring{e}_{\overline{40:50}}$ and variances of the associated time until death random variables, and the correlation. It is a good problem, as the student needs to pay attention to the upper limits of the integral because of ω and the joint life assumption.

A spreadsheet is designed to separately input ω and the two desired issue ages, in this case (40) and (50). For sake of comparison to the book answers, these are kept as 100, 40, and 50 respectively. The RAND() function is used separately for each life to simulate the time until death, which implies that the two lives are independent. The actual formula coded is (ω − issue age) × RAND(), which provides the flexibility for the input noted above. For each life, 100 deviates are generated. For each pair of deviates, the MIN function is used to choose the smaller of the two values and the MAX function to choose the maximum of the two values. The 100 values for the minimum is an approximation for the density of the first death, while the other 100 values for the maximum is an approximation for the density of the second death. Table 5 shows the theoretically correct values for the average, variance, standard deviation for the time of death for each single life, the time of death for the first death, and the time of death for the second death. Also included in Table 5 are the results of two trials for the modeling of the distribution. Comparing each statistic for each trial to the theoretically true value, one can see that each trial will produce variation from the truth. This is one way for the student to understand how a realization of a process may differ from the expected average. Also this concept can be reinforced by just examining each realization from the 100, instead of the average of the 100 realizations.

Corresponding to the results in trials one and two, a frequency histogram is dynamically generated using the FREQUENCY function in EXCEL. See Figure 7 for a chart of results. Of importance to note is that the formula is entered as an array formula, which means that CTRL+SHIFT+ENTER must be entered with all the appropriate cells highlighted. The *bins* of how to split up the counts into groups can be defined separately or a default is used. The counts in each bin is the number of cells that are less than the number for the bin and above the immediate prior bin. So the first bin is the number of times of death less than or equal to 5, while the next bin is the number between 5 and 10. The chart gives a feel for the difference between the distributions of the min and the max of the time until death. This method shows how the distribution varies depending on the samples, as the chart is updated on each recalculation of the random numbers.

Table 5: Comparison of Theoretical and Simulation Results for Exercise 8.7

Theoretical Results

Statistics	T(40)	T(50)	min(T(40),T(50))	max(T(40),T(50))
Average	30.00	25.00	18.06	36.94
Variance	300.00	208.33	160.11	182.33
Std.dev.	17.32	14.43	12.65	13.50
Corr(T(40),T(50)) = 0.00				
Corr(min,max) = 0.49				

Trial One Simulation Results

Statistics	T(40)	T(50)	min(T(40),T(50))	max(T(40),T(50))
Average	28.92	24.72	17.54	36.10
Variance	288.47	199.01	155.36	166.95
Std.dev.	16.98	14.11	12.46	12.92
Corr(T(40),T(50)) = 0.00				
Corr(min,max) = 0.51				

Trial Two Simulation Results

Statistics	T(40)	T(50)	min(T(40),T(50))	max(T(40),T(50))
Average	29.74	26.66	18.23	38.17
Variance	330.76	189.51	141.46	182.94
Std.dev.	18.19	13.77	11.89	13.53
Corr(T(40),T(50)) = −0.11				
Corr(min,max) = 0.44				

Example 6: Actuarial Mathematics Exercise 8.27: Joint life fractional age assumption

Exercise 8.27 combines both concepts of joint life functions and the fractional age assumption. The problem states: " Let $T(x)$ and $T(y)$ be independent and uniformly distributed in the next year of age. Given that both (x) and (y) die within the next year, demonstrate that the time-of-failure of (xy) is not uniformly distributed over the year. [Hint: Show that $Pr\,[T(xy) \le t \mid (T(x) \le 1) \cap (T(y) \le 1)] = 2t - t^2$]" Students have difficulties with calculating the density function of the minimum of two random variables. *This example will show how easy it is to use simulation to illustrate the concepts of this exercise.*

First, to solve this problem theoretically, one calculates the cumulative density function as:

$$\begin{aligned}
Pr\left[T(xy) \le t \mid (T(x) \le 1) \bigcap (T(y) \le 1)\right] &= 1 - Pr\,[T(x) > t, T(y) > t \mid (T(x) \le 1) \cap (T(y) \le 1)] \\
&= 1 - Pr\,[T(x) > t \mid (T(x) \le 1)] \cdot Pr\,[T(y) > t \mid (T(y) \le 1)] \\
&= 1 - (1-t)(1-t) \\
&= 2t - t^2
\end{aligned}$$

The density function of the minimum is $2(1-t)$, which is not uniform for $0 \le t \le 1$. Using the density function we can find the expected value equal to $\frac{1}{3}$ and the variance equal to $\frac{1}{18}$. In addition, one can note that the joint survival probability of surviving past time 0.1 is 0.81 (0.9×0.9), while if uniform, the value should be 0.9.

Figure 6: U.S. Life Table for Total Population

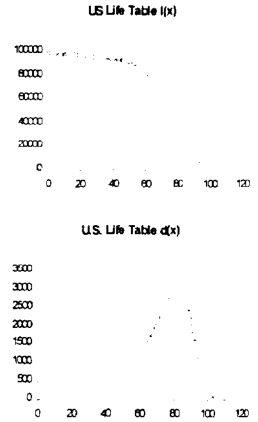
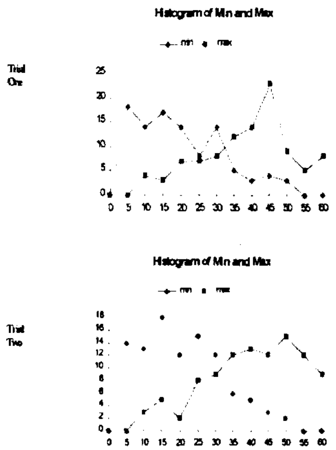


Figure 7: Histogram for Exercise 8.7

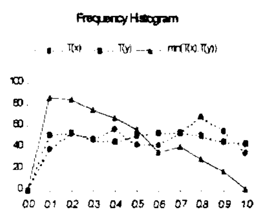


The simulation exercise is as easy as creating a column of $U(0,1)$ deviates for $T(x)$ and a separate set for $T(y)$, as the distribution is conditional on each of their deaths in the next year and the lives are independent. The smaller of the two numbers is a deviate generated from the minimum. Summary statistics, a histogram, and a chart confirm the answer in the exercise. Table 6 shows the summary statistics of one trial to be nearly identical to the theoretical answers. Also, a histogram of the results shown in Figure 8, shows how the individual times until death are uniform, while the frequency of the minimum is not.

Table 6: Summary Statistics for Exercise 8.27

Statistics	min(T(x),T(y))
average	0.331
variance	0.053

Figure 8: Histogram for Exercise 8.27



5 Classroom Incorporation

The simulation examples described in this paper range from easy to complex. The classroom may be a place to demonstrate the results, but is not a place to create the raw spreadsheets. The incorporation of simulation is most appropriate in homework assignments and more lengthy projects. This section describes two experiences that have been used.

Example 1 involving the GIC, can be used to introduce the concept and technique of simulation to a theory of interest class, as well as to cover various uses of spreadsheet software. Even though the example is based on material at the very end of textbook, the substance of the problem and its solution can be explained in general terms very early in the course. The advantage of this is that simulation of interest rates can be continued throughout the course, such as in examining variable rate mortgages versus fixed rate mortgages. Students may or may not have some background in probability distributions when taking interest theory, so the level of explanation regarding simulation will need to be tailored to their background.

Most actuarial students are likely to have had some spreadsheet experience, but many may not have used the logic functions or other specialized capabilities described in this paper. In order to accommodate the range of prior knowledge, the following approach has been useful.

First, simply reproduce the one-year analysis in the textbook over the eight possible interest rate paths. This can be done with the type of spreadsheet formulas and functions that students are likely to know already or can learn quite quickly. Interestingly, even in this very simple activity, there may be some surprises; e.g.

while students are very likely to know that Excel has a SUM function, they may not know there is a similar PRODUCT function.

Second, again reproduce the eight path analysis. However, instead of tailoring the formula in each cell based on whether the path involves an increase or decrease, use the IF function to produce that result. This approach introduces the student to more sophisticated programming skills.

Third, introduce a simulation analysis of the 1-year GIC. Instead of specifying the interest rate paths in advance, build on the IF statement practiced in step 2. A large number of rows (e.g. 100 or 1000) can be set up by copying the formula, and a new run of random paths can be generated by triggering a recalculation of the worksheet.

The COUNTIF function can be used to determine how many of the interest rate paths produce a loss: that is, a full year accumulation less than the guarantee. The proportion of the total number of paths is an estimate of the probability of loss, and can be compared with the value determined by steps 1 and 2. Students see that there is variation from run to run, but that the overall result is about as expected.

Extending this simulation approach to a longer time period is quite easy. However, it is very important to emphasize that a significant change occurs: in the 1-year case, the correct answer based on the eight paths is already known and unreasonable simulation results can be the basis for reexamining the spreadsheet logic and details. For the longer period, the correct answer is not known; of course, that is more realistic since a major reason for using simulation is when complete or otherwise accurate calculations are not feasible. Other ways of checking the feasibility of the results will be needed. For example, simply checking whether the proportion of quarterly increases are near the 'true' probability of an increase has revealed errors. Students need to learn that they cannot trust complex spreadsheets to produce reliable results unless they put a lot of effort into the art of developing check-values.

If students have done some spreadsheet simulations in prior courses, even greater emphasis can be put on sophisticated techniques and their implications. Thus the complexities of the LOOKUP function should not be a barrier, and the ideas discussed in Example 4 regarding alternative fractional age rules can be included.

While Example 4 showed a simulation of Exercise 3.24 in Actuarial Mathematics which refers to the U.S. Life Table shown in the text, the procedure is applicable to other life tables. Rather than having students type in the Illustrative Life Table values (Appendix 2A), a preliminary project can be constructed which is also instructive in its own right. For ages 13 and above, that construction involves application of the Makeham formula as described on pages 71-72 of the text. The 113 value shown in Appendix 2A should be used as the starting point for the construction; then it is a small task to type in values for younger ages.

Example 4 can then be applied to the Illustrative Life Table. Of course, the simulation results can be checked against the table by simply determining the UDD estimate for $l_x = 50,000$.

Example 4 simulations can also be used to illustrate the relationship between curtate and complete time-until-death variables as discussed on pages 70-71 of the text, and in Exercise 3.40. For UDD it is easy to demonstrate that average of the curtate expectation plus 0.5 provides a reasonable estimate of the complete expectation. That half-year adjustment can be justified by analytic reasoning, or justified by simulations as discussed in Example 3 of this paper.

References

[1] Bowers, N., Gerber, H., Hickman, J., Jones, D., and Nesbitt, C., **Actuarial Mathematics**, Society of Actuaries:Itasca, Illinois. 1986.

[2] Kellison, S., **The Theory of Interest, 2nd edition**, Irwin:Illinois, 1991.