Actuarial Mathematics and Life-Table Statistics

Eric V. Slud
Mathematics Department
University of Maryland, College Park

Chapter 3

More Probability Theory for Life Tables

3.1 Interpreting Force of Mortality

This Section consists of remarks, relating the force of mortality for a continuously distributed lifetime random variable T (with continuous density function f) to conditional probabilities for discrete random variables. Indeed, for m large (e.g. as large as 4 or 12), the discrete random variable $[Tm]/m$ gives a close approximation to T and represents the attained age at death measured in whole-number multiples of fractions $h =$ one m^{th} of a year. (Here $[\cdot]$ denotes the greatest integer less than or equal to its real argument.) Since surviving an additional time $t = nh$ can be viewed as successively surviving to reach times $h, 2h, 3h, \ldots, nh$, and since (by the definition of conditional probability)

$$P(A_1 \cap \cdots \cap A_n) \;=\; P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

we have (with the interpretation $A_k = \{T \geq x + kh\}$)

$$_{nh}p_x \;=\; {}_hp_x \,\cdot\, {}_hp_{x+h} \,\cdot\, {}_hp_{x+2h} \,\cdots\, {}_hp_{x+(n-1)h}$$

The form in which this formula is most often useful is the case $h = 1$: for integers $k \geq 2$,

$$_kp_x \;=\; p_x \,\cdot\, p_{x+1} \,\cdot\, p_{x+2} \,\cdots\, p_{x+k-1} \tag{3.1}$$

Every continuous waiting-time random variable can be approximated by a discrete random variable with possible values which are multiples of a fixed small unit h of time, and therefore the random survival time can be viewed as the (first failure among a) succession of results of a sequence of independent coin-flips with successive probabilities ${}_hp_{kh}$ of heads. By the Mean Value Theorem applied up to second-degree terms on the function $S(x+h)$ expanded about $h = 0$,

$$S(x{+}h) \;=\; S(x) + hS'(x) + \frac{h^2}{2}\,S''(x{+}\tau h) \;=\; S(x) - hf(x) - \frac{h^2}{2}\,f'(x{+}\tau h)$$

for some $0 < \tau < 1$, if f is continuously differentiable. Therefore, using the definition of $\mu(x)$ as $f(x)/S(x)$ given on page 41,

$${}_hp_x = 1 \,-\, h\cdot\Big[\frac{S(x) - S(x+h)}{hS(x)}\Big] \;=\; 1 \,-\, h\,\Big(\mu(x) \,+\, \frac{h}{2}\,\frac{f'(x+\tau h)}{S(x)}\Big)$$

Going in the other direction, the previously derived formula

$${}_hp_x = \exp\left(-\int_x^{x+h} \mu(y)\,dy\right)$$

can be interpreted by considering the fraction of individuals observed to reach age x who thereafter experience hazard of mortality $\mu(y)\,dy$ on successive infinitesimal intervals $[y, y{+}dy]$ within $[x, x{+}h)$. The lives aged x survive to age $x+h$ with probability equal to a limiting product of infinitesimal terms $(1{-}\mu(y)\,dy) \sim \exp(-\mu(y)\,dy)$, yielding an overall conditional survival probability equal to the negative exponential of accumulated hazard over $[x, x+h)$.

3.2 Interpolation Between Integer Ages

There is a Taylor-series justification of "actuarial approximations" for life-table related functions. Let $g(x)$ be a smooth function with small $|g''(x)|$, and let T be the lifetime random variable (for a randomly selected member of the population represented by the life-table), with $[T]$ denoting its integer part, i.e., the largest integer k for which $k \le T$. Then by the Mean Value

Theorem, applied up to second-degree terms for the function $g(t) = g(k+u)$ (with $t = k+u$, $k = [t]$) expanded in $u \in (0,1)$ about 0,

$$E(g(T)) = E(g([T]) \; + \; (T-[T])\, g'([T]) \; + \; \frac{1}{2}\,(T-[T])^2\, g''(T_*)) \tag{3.2}$$

where T_* lies between $[T]$ and T. Now in case the rate-of-change of g' is very small, the third term may be dropped, providing the approximate formula

$$E(g(T)) \approx Eg([T]) \; + \; E\Big((T-[T])\, g'([T])\Big) \tag{3.3}$$

Simplifications will result from this formula especially if the behavior of conditional probabilities concerning $T-[T]$ given $[T]=k$ turns out not to depend upon the value of k. (This property can be expressed by saying that the random integer $[T]$ and random fractional part $T-[T]$ of the age at death are *independent* random variables.) This is true in particular if it is also true that $P(T-[T] \ge s \,|\; k \le T < k+1)$ is approximately $1-s$ for all k and for $0 < s < 1$, as would be the case if the density of T were constant on each interval $[k, k+1)$ (i.e., if the distribution of T were *conditionally uniform* given $[T]$): then $T-[T]$ would be uniformly distributed on $[0,1)$, with density $f(s) = 1$ for $0 \le s < 1$. Then $E((T-[T])\, g'([T])) \; = \; E(g'([T]))/2$, implying by (3.3) that

$$E(g(T)) \;\approx\; E(g([T]) + \frac{1}{2}\, g'([T])) \;\approx E\Big(g([T] + \frac{1}{2}\,)\Big)$$

where the last step follows by the first-order Taylor approximation

$$g(k+1/2) \;\approx\; g(k) \; + \; \frac{1}{2}\, g'(k)$$

One particular application of the ideas of the previous paragraph concern so-called *expected residual lifetimes*. Demographers tabulate, for all integer ages x in a specified population, what is the average number e_x of remaining years of life to individuals who have just attained exact age x. This is a quantity which, when compared across national or generational boundaries, can give some insight into the way societies differ and change over time. In the setting of the previous paragraph, we are considering the function $g(t) = t - x$ for a fixed x, and calculating expectations $E(\cdot)$ conditionally for a life aged x, i.e. conditionally given $T \ge x$. In this setting, the approximation described above says that if we can treat the density of T as constant within each whole year of attained integer age, then

$$\text{Mean (complete) residual lifetime} \;=\; \overset{\circ}{\mathrm{e}}_x \;\approx\; \mathrm{e}_x + \frac{1}{2}$$

where e_x denotes the so-called *curtate mean residual life* which measures the expectation of $[T] - x$ given $T \geq x$, i.e., the expected number of additional birthdays or whole complete years of life to a life aged exactly x.

"Actuarial approximations" often involve an assumption that a life-table time until death is conditionally uniformly distributed, i.e., its density is piecewise-constant, over intervals $[k, k+1)$ of age. The following paragraphs explore this and other possibilities for survival-function interpolation *between* integer ages.

One approach to approximating survival and force-of-mortality functions at non-integer values is to use analytical or what statisticians call *parametric* models $S(x;\vartheta)$ arising in Examples (i)-(v) above, where ϑ denotes in each case the vector of real parameters needed to specify the model. Data on survival at integer ages x can be used to estimate or *fit* the value of the scalar or vector parameter ϑ, after which the model $S(x;\vartheta)$ can be used at *all* real x values. We will see some instances of this in the exercises. The disadvantage of this approach is that actuaries do not really believe that any of the simple models outlined above ought to fit the whole of a human life table. Nevertheless they can and do make assumptions on the shape of $S(x)$ in order to justify interpolation-formulas between integer ages.

Now assume that values $S(k)$ for $k = 0, 1, 2, \ldots$ have been specified or estimated. Approximations to $S(x)$, $f(x)$ and $\mu(x)$ between integers are usually based on one of the following assumptions:

(i) *(Piecewise-uniform density)* $f(k+t)$ is constant for $0 \leq t < 1$;

(ii) *(Piecewise-constant hazard)* $\mu(k+t)$ is constant for $0 \leq t < 1$;

(iii) *(Balducci hypothesis)* $1/S(k+t)$ is linear for $0 \leq t < 1$.

Note that for integers k and $0 \leq t \leq 1$,

$$\left.\begin{array}{c} S(k+t) \\ -\ln S(k+t) \\ 1/S(k+t) \end{array}\right\} \text{ is linear in } \; t \; \text{ under } \left\{\begin{array}{c} \text{assumption (i)} \\ \text{assumption (ii)} \\ \text{assumption (iii)} \end{array}\right. \tag{3.4}$$

Under assumption (i), the slope of the linear function $S(k+t)$ at $t=0$ is $-f(k)$, which implies easily that $S(k+t) = S(k) - tf(k)$, i.e.,

$$f(k) = S(k) - S(k+1) \ , \quad \text{and} \qquad \mu(k+t) = \frac{f(k)}{S(k) - tf(k)}$$

so that under (i),

$$\mu(k + \frac{1}{2}) \ = \ f_T(k + \frac{1}{2}) \Big/ S_T(k + \frac{1}{2}) \tag{3.5}$$

Under (ii), where $\mu(k+t) = \mu(k)$, (3.5) also holds, and

$$S(k+t) = S(k)\, e^{-t\,\mu(k)} \ , \quad \text{and} \qquad p_k = \frac{S(k+1)}{S(k)} = e^{-\mu(k)}$$

Under (iii), for $0 \le t < 1$,

$$\frac{1}{S(k+t)} = \frac{1}{S(k)} + t\left(\frac{1}{S(k+1)} - \frac{1}{S(k)}\right) \tag{3.6}$$

When equation (3.6) is multiplied through by $S(k+1)$ and terms are rearranged, the result is

$$\frac{S(k+1)}{S(k+t)} \ = \ t \ + \ (1-t)\ \frac{S(k+1)}{S(k)} \ = \ 1 \ - \ (1-t)\, q_k \tag{3.7}$$

Recalling that ${}_tq_k = 1 - (S(k+t)/S(k))$, reveals assumption (iii) to be equivalent to

$${}_{1-t}q_{k+t} = 1 - \frac{S(k+1)}{S(k+t)} = (1-t)\Big(1 - \frac{S(k+1)}{S(k)}\Big) \ = \ (1-t)\, q_k \tag{3.8}$$

Next differentiate the logarithm of the formula (3.7) with respect to t, to show (still under (iii)) that

$$\mu(k+t) \ = \ -\frac{\partial}{\partial t}\ \ln S(k+t) \ = \ \frac{q_k}{1-(1-t)q_k} \tag{3.9}$$

The most frequent insurance application for the interpolation assumptions (i)-(iii) and associated survival-probability formulas is to express probabilities of survival for fractional years in terms of probabilities of whole-year survival. In terms of the notations ${}_tp_k$ and q_k for integers k and $0 < t < 1$, the formulas are:

$$ {}_tp_k \;=\; 1 - \frac{(S(k) - t(S(k+1) - S(k))}{S(k)} \;=\; 1 - t\,q_k \qquad \text{under (i)} \tag{3.10} $$

$$ {}_tp_k \;=\; \frac{S(k+t)}{S(x)} \;=\; \left(e^{-\mu(k)}\right)^t \;=\; (1-q_k)^t \qquad \text{under (ii)} \tag{3.11} $$

$$ {}_tp_k \;=\; \frac{S(k+t)}{S(k+1)}\,\frac{S(k+1)}{S(k)} \;=\; \frac{1-q_k}{1-(1-t)q_k} \qquad \text{under (iii)} \tag{3.12} $$

The application of all of these formulas can be understood in terms of the formula for expectation of a function $g(T)$ of the lifetime random variable T. (For a concrete example, think of $g(T) = (1+i)^{-T}$ as the present value to an insurer of the payment of \$1 which it will make instantaneously at the future time T of death of a newborn which it undertakes to insure.) Then assumptions (i), (ii), or (iii) via respective formulas (3.10), (3.11), and (3.12) are used to substitute into the final expression of the following formulas:

$$ \begin{aligned} E\Big(g(T)\Big) \;&=\; \int_0^\infty g(t)\,f(t)\,dt \;=\; \sum_{k=0}^{\omega-1} \int_0^1 g(t+k)\,f(t+k)\,dt \\ &=\; \sum_{k=0}^{\omega-1} S(k)\,\int_0^1 g(t+k)\,\Big(-\frac{\partial}{\partial t}\,{}_tp_k\Big)\,dt \end{aligned} $$

3.3 Binomial Variables & Law of Large Numbers

This Section develops just enough machinery for the student to understand the probability theory for random variables which count numbers of successes in large numbers of independent biased coin-tosses. The motivation is that in large life-table populations, the number l_{x+t} who survive t time-units after age x can be regarded as the number of successes or heads in a large number l_x of independent coin-toss trials corresponding to the further survival of each of the l_x lives aged x , which for each such life has probability ${}_tp_x$. The one preliminary piece of mathematical machinery which the student is assumed to know is the **Binomial Theorem** stating that (for positive integers N and arbitrary real numbers x, y, z),

$$(1+x)^N = \sum_{k=0}^{N} \binom{N}{k} x^k \,, \qquad (y+z)^N = \sum_{k=0}^{N} \binom{N}{k} y^k\, z^{N-k}$$

Recall that the first of these assertions follows by equating the k^{th} derivatives of both sides at $x = 0$, where $k = 0, \ldots, N$. The second assertion follows immediately, in the nontrivial case when $z \neq 0$, by applying the first assertion with $x = y/z$ and multiplying both sides by z^N. This Theorem also has a direct combinatorial consequence. Consider the two-variable polynomial

$$(y+z)^N \;=\; (y+z)\cdot(y+z)\cdots(y+z) \qquad\qquad N \;\; \textit{factors}$$

expanded by making all of the different choices of y or z from each of the N factors $(y+z)$, multiplying each combination of choices out to get a monomial $y^j\, z^{N-j}$, and adding all of the monomials together. Each combined choice of y or z from the N factors $(y+z)$ can be represented as a sequence $(a_1, \ldots, a_n) \in \{0,1\}^N$, where $a_i = 1$ would mean that y is chosen $a_i = 0$ would mean that z is chosen in the i^{th} factor. Now a combinatorial fact can be deduced from the Binomial Theorem: since the coefficient $\binom{N}{k}$ is the total number of monomial terms $y^k\, z^{N-k}$ which are collected when $(y+z)^N$ is expanded as described, and since these monomial terms arise only from the combinations $(a_1, \ldots, a_N)$ of $\{y,\, z\}$ choices in which precisely k of the values a_j are 1's and the rest are 0's,

> The number of symbol-sequences $(a_1, \ldots, a_N) \in \{0,1\}^N$ such that $\sum_{j=1}^N a_j = k$ is given by $\binom{N}{k}$, for $k = 0, 1, \ldots, N$. This number
>
> $$\binom{N}{k} \;=\; \frac{N(N-1)\cdots(N-k+1)}{k!}$$
>
> spoken as 'N choose k', therefore counts all of the ways of choosing k element subsets (the positions j from 1 to N where 1's occur) out of N objects.

The random experiment of interest in this Section consists of a large number N of independent tosses of a coin, with probability p of coming up heads each time. Such coin-tossing experiments — independently replicated two-outcome experiments with probability p of one of the outcomes, designated 'success' — are called *Bernoulli(p) trials*. The space of possible heads-and-tails configurations, or *sample space* for this experiment, consists of the strings of N zeroes and ones, with each string $\underline{\mathbf{a}} = (a_1, \ldots, a_N) \in \{0,1\}^N$ being assigned probability $p^a\,(1-p)^{N-a}$, where $a \equiv \sum_{j=1}^N a_j$. The rule by which probabilities are assigned to sets or *events* A of more than one string $\underline{\mathbf{a}} \in \{0,1\}^N$ is to add the probabilities of all individual strings $\underline{\mathbf{a}} \in A$. We are particularly interested in the event (denoted $[X = k]$) that precisely k of the coin-tosses are heads, i.e., in the subset $[X = k] \subset \{0,1\}^N$ consisting of all strings $\underline{\mathbf{a}}$ such that $\sum_{j=1}^N a_j = k$. Since each such string has the same probability $p^k\,(1-p)^{N-k}$, and since, according to the discussion following the Binomial Theorem above, there are $\binom{N}{k}$ such strings, the probability which is necessarily assigned to the event of k successes is

$$\text{P(k successes in N } \textit{Bernoulli(p)} \text{ trials)} = P(X = k) = \binom{N}{k}\, p^k\,(1-p)^{N-k}$$

By virtue of this result, the random variable X equal to the number of successes in N *Bernoulli(p)* trials, is said to have the **Binomial distribution** with **probability mass function** $p_X(k) = \binom{N}{k}\, p^k\,(1-p)^{N-k}$.

With the notion of Bernoulli trials and the binomial distribution in hand, we now begin to regard the ideal probabilities $S(x+t)/S(x)$ as true but unobservable probabilities ${}_tp_x = p$ with which each of the l_x lives aged x will survive to age $x + t$. Since the mechanisms which cause those lives to survive or die can ordinarily be assumed to be acting independently in a probabilistic sense, we can regard the number l_{x+t} of lives surviving to the (possibly fractional) age $x+t$ as a Binomial random variable with parameters $N = l_x$, $p = {}_tp_x$. From this point of view, the observed life-table counts l_x should be treated as *random data* which reflect but do not define the underlying probabilities ${}_xp_0 = S(x)$ of survival to age x. However, common sense and experience suggest that, when l_0 is large, and therefore the other life-counts l_x for moderate values x are also large, the observed ratios l_{x+t}/l_x should reliably be very close to the 'true' probability ${}_tp_x$. In other words, the ratio l_{x+t}/l_x is a *statistical estimator* of the unknown constant ${}_tp_x$. The good property, called *consistency*, of this estimator to be close with very large probability (based upon large life-table size) to the probability it estimates, is established in the famous **Law of Large Numbers**. The precise quantitative inequality proved here concerning binomial probabilities is called a *Large Deviation Inequality* and is very important in its own right.

Theorem 3.3.1 *Suppose that X is a Binomial(N,p) random variable, denoting the number of successes in N Bernoulli(p) trials.*

(a) Large Deviation Inequalities. If $1 > b > p > c > 0$, then

$$P(X \geq Nb) \leq \exp\Big\{ -N\Big[b \ln\left(\frac{b}{p}\right) \; + \; (1-b)\,\ln\left(\frac{1-b}{1-p}\right)\Big]\Big\}$$

$$P(X \leq Nc) \leq \exp\Big\{ -N\Big[c \ln\left(\frac{c}{p}\right) \; + \; (1-c)\,\ln\left(\frac{1-c}{1-p}\right)\Big]\Big\}$$

(b) Law of Large Numbers. For arbitrarily small fixed $\delta > 0$, not depending upon N, the number N of Bernoulli trials can be chosen so large that

$$P\Big(\Big|\frac{X}{N} - p\Big| \geq \delta\Big) \;\leq\; \delta$$

Proof. After the first inequality in (a) is proved, the second inequality will be derived from it, and part (b) will follow from part (a). Since the event $[X \geq Nb]$ is the union of the disjoint events $[X = k]$ for $k \geq Nb$, which in turn consist of all outcome-strings $(a_1, \ldots, a_N) \in \{0,1\}^N$ for which $\sum_{j=1}^N a_j = k \geq Nb$, a suitable subset of the binomial probability mass function values $p_X(k)$ are summed to provide

$$P(X \geq Nb) = \sum_{k:Nb \leq k \leq N} P(X = k) = \sum_{k \geq Nb} \binom{N}{k} p^k\,(1-p)^{N-k}$$

For every $s > 1$, this probability is

$$\leq \sum_{k \geq Nb} \binom{N}{k} p^k\,(1-p)^{N-k}\, s^{k-Nb} \;=\; s^{-Nb} \sum_{k \geq Nb} \binom{N}{k} (ps)^k\,(1-p)^{N-k}$$

$$\leq\; s^{-Nb} \sum_{k=0}^{N} \binom{N}{k} (ps)^k\,(1-p)^{N-k} \;=\; s^{-Nb}\,(1-p+ps)^N$$

Here extra terms (corresponding to $k < Nb$) have been added in the next-to-last step, and the binomial theorem was applied in the last step. The trick in the proof comes now: since the left-hand side of the inequality does not involve s while the right-hand side does, and since the inequality must be valid for every $s > 1$, it remains valid if the right-hand side is minimized over s. The calculus minimum does exist and is unique, as you can check by calculating that the second derivative in s is always positive. The minimum occurs where the first derivative of the logarithm of the last expression is 0, i.e., at $s = b(1-p)/(p(1-b))$. Substituting this value for s yields

$$\begin{aligned} P(X \geq Nb) \;&\leq\; \left(\frac{b\,(1-p)}{p\,(1-b)}\right)^{-Nb} \left(\frac{1-p}{1-b}\right)^{N} \\ &= \exp\left(-N\Big[b\,\ln\Big(\frac{b}{p}\Big) + (1-b)\,\ln\Big(\frac{1-b}{1-p}\Big)\Big]\right) \end{aligned}$$

as desired.

The second part of assertion (a) follows from the first. Replace X by $Y = N - X$. Since Y also is a count of 'successes' in *Bernoulli*$(1-p)$ trials, where the 'successes' counted by Y are precisely the 'failures' in the *Bernoulli* trials defining X, it follows that Y also has a *Binomial*(N, q) distribution, where $q = 1 - p$. Note also that $c < p$ implies $b = 1 - c > 1 - p = q$. Therefore, the first inequality applied to Y instead of X with $q = 1 - p$ replacing p and $b = 1 - c$, gives the second inequality for $P(Y \geq Nb) = P(X \leq Nc)$.

Note that for all r between $0, 1$, the quantity $r \ln \frac{r}{p} + (1-r) \ln \frac{1-r}{1-p}$ as a function of r is convex and has a unique minimum of 0 at $r = p$. Therefore when $b > p > c$, the upper bound given in part (a) for $N^{-1} \ln P([X \geq bN] \cup [X \leq cN])$ is strictly negative and does not involve N. For part (b), let $\delta \in (0, \min(p, 1-p))$ be arbitrarily small, choose $b = p + \delta$, $c = p - \delta$, and combine the inequalities of part (a) to give the precise estimate (b).

$$P(|\frac{X}{N} - p| \geq \delta) \leq 2 \cdot \exp(-Na) \tag{3.13}$$

where

$$\begin{aligned} a \;&=\; \min\Big((p+\delta)\,\ln(1 + \tfrac{\delta}{p}) + (1-p-\delta)\,\ln(1 - \tfrac{\delta}{1-p})\,, \\ &\qquad (p-\delta)\ln(1 - \tfrac{\delta}{p}) + (1-p+\delta)\ln(1 + \tfrac{\delta}{1-p})\Big) \;>\; 0 \end{aligned} \tag{3.14}$$

This last inequality proves (b), and in fact gives a much stronger and numerically more useful upper bound on the probability with which the so-called *relative frequency of success* X/N differs from the true probability p of success by as much as δ. The probabilities of such *large deviations* between X/N and δ are in fact exponentially small as a function of the number N of repeated *Bernoulli*(p) trials, and the upper bounds given in (a) on the log-probabilities divided by N turn out to be the correct limits for large N of these normalized log-probabilities. □

3.3.1 Exact Probabilities, Bounds & Approximations

Suppose first that you are flipping $20,000$ times a coin which is supposed to be fair (i.e., to have $p = 1/2$) . The probability that the observed number of heads falls outside the range $[9800, 10200]$ is, according to the inequalities above,

$$\leq 2 \cdot \exp\Big[-9800 \, \ln(0.98) \, - \, 10200 \, \ln(1.02) \Big] \; = \; 2\, e^{-4.00} = 0.037$$

The inequalities (3.13)-(3.14) give only an upper bound for the actual binomial probability, and 0.0046 is the exact probability with which the relative frequency of heads based on 20000 fair coin-tosses lies outside the range $(0.98,\, 1.02)$. The ratio of the upper bound to the actual probability is rather large (about 8), but the absolute errors are small.

To give a feeling for the probabilities with which observed life-table ratios reflect the true underlying survival-rates, we have collected in Table 3.3.1 various exact binomial probabilities and their counterparts from the inequalities of Theorem 3.3.1(a). The illustration concerns cohorts of lives aged x of various sizes l_x, together with 'theoretical' probabilities ${}_kp_x$ with which these lives will survive for a period of $k = 1,\, 5,$ or 10 years. The probability experiment determining the size of the surviving cohort l_{x+k} is modelled as the tossing of l_x independent coins with common heads-probability ${}_kp_x$: then the surviving cohort-size l_{x+k} is viewed as the *Binomial*$(l_x,\, {}_kp_x)$ random variable equal to the number of heads in those coin-tosses. In Table 3.3.1 are given various combinations of $x,\ l_x,\ k,\ {}_kp_x$ which might realistically arise in an insurance-company life-table, together, with the true and estimated (from Theorem 3.3.1) probabilities with which the ratios l_{x+k}/l_x agree with ${}_kp_x$

to within a fraction δ of the latter. The formulas used to compute columns 6 and 7 of the table are (for $n = l_x, \; p = {}_kp_x$):

$$\text{True binomial probability} \quad = \sum_{j:j/(np)\in[1-\delta,1+\delta]} \binom{n}{j} \, p^j \, (1-p)^{n-j}$$

$$\text{Lower bound for probability} \quad = \quad 1 \; - \; (1+\delta)^{-np(1+\delta)} \left(1 - \frac{p\delta}{1-p}\right)^{-n(1-p-p\delta)}$$

$$- \; (1-\delta)^{-np(1-\delta)} \left(1 + \frac{p\delta}{1-p}\right)^{-n(1-p+p\delta)}$$

Columns 6 and 7 in the Table show how likely the life-table ratios are to be close to the 'theoretical' values, but also show that the lower bounds, while also often close to 1, are still noticeably smaller than the actual values. .

Much closer approximations to the exact probabilities for Binomial(n, p) random variables given in column 6 of Table 3.3.1 are obtained from the *Normal distribution approximation*

$$P(a \le X \le b) \; \approx \; \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) \; - \; \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right) \tag{3.15}$$

where Φ is the *standard normal distribution function* given explicitly in integral form in formula (3.20) below. This approximation is the **DeMoivre-Laplace Central Limit Theorem** (Feller vol. 1, 1957, pp. 168-73), which says precisely that the difference between the left- and right-hand sides of (3.15) converges to 0 when p remains fixed, $n \to \infty$. Moreover, the refined form of the DeMoivre-Laplace Theorem given in the Feller (1957, p. 172) reference says that each of the *ratios* of probabilities

$$P(X < a) \Big/ \Phi\Big(\frac{a - np}{\sqrt{np(1-p)}}\Big) \quad , \qquad P(X > b) \Big/ \Big[1 \; - \; \Phi\Big(\frac{b - np}{\sqrt{np(1-p)}}\Big)\Big]$$

converges to 1 if the 'deviation' ratios $(b - np)/\sqrt{np(1-p)}$ and $(a - np)/\sqrt{np(1-p)}$ are of smaller order than $n^{-1/6}$ when n gets large. This result suggests the approximation

$$\text{Normal approximation} \quad = \quad \Phi\left(\frac{np\delta}{\sqrt{np(1-p)}}\right) \; - \; \Phi\left(\frac{-np\delta}{\sqrt{np(1-p)}}\right)$$

Table 3.1: Probabilities (in col. 6) with which various Binomial(l_x, ${}_kp_x$) random variables lie within a factor $1 \pm \delta$ of their expectations, together with lower bounds for these probabilities derived from the large-deviation inequalities (3.13)-(3.14). The final column contains the normal-distribution (Central-Limit) approximations to the exact probabilities in column 6.

Cohort $n = l_x$	Age x	Time k	Prob. $p = {}_k p_x$	Toler. frac. δ	Pr. within within $1 \pm \delta$	Lower bound	Normal approx.
10000	40	3	0.99	.003	.9969	.9760	.9972
10000	40	5	0.98	.004	.9952	.9600	.9949
10000	40	10	0.94	.008	.9985	.9866	.9985
1000	40	10	0.94	.020	.9863	.9120	.9877
10000	70	5	0.75	.020	.9995	.9950	.9995
1000	70	5	0.75	.050	.9938	.9531	.9938
10000	70	10	0.50	.030	.9973	.9778	.9973
1000	70	10	0.50	.080	.9886	.9188	.9886

for the true binomial probability $P(|X - np| \le np\delta)$, the formula of which is displayed above. Although the deviation-ratios in this setting are actually close to $n^{-1/6}$, not smaller as they should be for applicability of the cited result of Feller, the normal approximations in the final column of Table 3.3.1 below are sensationally close to the correct binomial probabilities in column 6. A still more refined theorem which justifies this is given by Feller (1972, section XVI.7 leading up to formula 7.28, p. 553).

If the probabilities in Theorem 3.3.1(a) are generally much smaller than the upper bounds given for them, then why are those bounds of interest ? (These are 1 minus the probabilities illustrated in Table 3.3.1.) First, they provide relatively quick hand-calculated estimates showing that large batches of independent coin-tosses are extremely unlikely to yield relative frequencies of heads much different from the true probability or limiting relative frequency of heads. Another, more operational, way to render this conclusion of Theorem 3.3.1(b) is that two very large insured cohorts with the same true survival probabilities are very unlikely to have materially different survival experience. However, as the Table illustrates, for practical purposes the normal approximation to the binomial probabilities of large discrepancies from the expectation is generally much more precise than the large deviation bounds of Theorem 3.3.1(a).

The bounds given in Theorem 3.3.1(a) get small with large N *much* more rapidly than simpler bounds based on Chebychev's inequality (*cf.* Hogg and Tanis 1997, Larsen and Marx 1985, or Larson 1982). We can tolerate the apparent looseness in the bounds because it can be shown that the exponential rate of decay as a function of N in the true tail-probabilities $P_N = P(X \geq Nb)$ or $P(X \leq Nc)$ in Theorem 3.3.1(a) (i.e., the constants appearing in square brackets in the exponents on the right-hand sides of the bounds) are exactly the right ones: no larger constants replacing them could give correct bounds.

3.4 Simulation of Life Table Data

We began by regarding life-table ratios l_x/l_0 in large cohort life-tables as *defining* integer-age survival probabilities $S(x) = {}_xp_0$. We said that if the life-table was representative of a larger population of prospective insureds, then we could imagine a newly presented life aged x as being randomly chosen from the life-table cohort itself. We motivated the conditional probability ratios in this way, and similarly expectations of functions of life-table death-times were averages over the entire cohort. Although we found the calculus-based formulas for life-table conditional probabilities and expectations to be useful, at that stage they were only ideal approximations of the more detailed but still exact life-table ratios and sums. At the next stage of sophistication, we began to describe the (conditional) probabilities ${}_tp_x \equiv S(x+t)/S(x)$ based upon a smooth survival function $S(x)$ as a true but unknown survival distribution, hypothesized to be of one of a number of possible theoretical forms, governing each member of the life-table cohort and of further prospective insureds. Finally, we have come to view the life-table itself as *data*, with each ratio l_{x+t}/l_x equal to the relative frequency of success among a set of l_x *Bernoulli*$({}_tp_x)$ trials which Nature performs upon the set of lives aged x. With the mathematical justification of the Law of Large Numbers, we come full circle: these relative frequencies are random variables which are not very random. That is, they are extremely likely to lie within a very small tolerance of the otherwise unknown probabilities ${}_tp_x$. Accordingly, the life-table ratios are, at least for very large-radix life tables, highly accurate *statistical estimators* of the life-table probabilities which we earlier tried to define by them.

To make this discussion more concrete, we illustrate the difference between the entries in a life-table and the entries one would observe as data in a randomly generated life-table of the same size using the initial life-table ratios as *exact* survival probabilities. We used as a source of life-table counts the Mortality Table for U.S. White Males 1959-61 reproduced as Table 2 on page 11 of C. W. Jordan's (1967) book on *Life Contingencies*. That is, using this Table with radix $l_0 = 10^5$, with counts l_x given for integer ages x from 1 through 80, we treated the probabilities $p_x = l_{x+1}/l_x$ for $x = 0, \ldots, 79$ as the correct one-year survival probabilities for a second, computer-simulated cohort life-table with radix $l_0^* = 10^5$. Using simulated random variables generated in **Splus**, we successively generated, as x runs from 1 to 79, random variables $l_{x+1}^* \sim Binomial(l_x^*, p_x)$. In other words, the mechanism of simulation of the sequence $l_0^*, \ldots, l_{79}^*$ was to make the variable l_{x+1}^* depend on previously generated $l_1^*, \ldots, l_x^*$ only through l_x^*, and then to generate l_{x+1}^* as though it counted the heads in l_x^* independent coin-tosses with heads-probability p_x. A comparison of the actual and simulated life-table counts for ages 9 to 79 in 10-year intervals, is given below. The complete simulated life-table was given earlier as Table 1.1.

The implication of the Table is unsurprising: with radix as high as 10^5, the agreement between the initial and randomly generated life-table counts is quite good. The Law of Large Numbers guarantees good agreement, with very high probability, between the ratios l_{x+10}/l_x (which here play the role of the probability ${}_{10}p_x$ of success in l_x^* *Bernoulli* trials) and the corresponding simulated random relative frequencies of success l_{x+10}^*/l_x^*. For example, with $x = 69$, the final simulated count of 28657 lives aged 79 is the success-count in 56186 *Bernoulli* trials with success-probability $28814/56384 = .51103$. With this success-probability, assertion (a) and the final inequality proved in (b) of the Theorem show that the resulting count will differ from $.51103 \cdot 56186 = 28712.8$ by 300 or more (in either direction) with probability at most 0.08. (Verify this by substituting in the formulas with $300 = \delta \cdot 56186$).

Table 3.2: **Illustrative Real and Simulated Life-Table Data**

Age x	1959-61 Actual Life-Table	Simulated l_x
9	96801	96753
19	96051	95989
29	94542	94428
39	92705	92576
49	88178	87901
59	77083	76793
69	56384	56186
79	28814	28657

3.4.1 Expectation for Discrete Random Variables

The Binomial random variables which have just been discussed are examples of so-called *discrete random variables*, that is, random variables Z with a discrete (usually finite) list of possible outcomes z, with a corresponding list of probabilities or *probability mass function* values $p_Z(z)$ with which each of those possible outcomes occur. (These probabilities $p_Z(z)$ must be positive numbers which summed over all possible values z add to 1.) In an insurance context, think for example of Z as the unforeseeable future damage or liability upon the basis of which an insurer has to pay some scheduled *claim amount* $c(Z)$ to fulfill a specific property or liability insurance policy. The Law of Large Numbers says that we can have a *frequentist* operational interpretation of each of the probabilities $p_Z(z)$ with which a claim of size $c(z)$ is presented. In a large population of N independent policyholders, each governed by the same probabilities $p_Z(\cdot)$ of liability occurrences, for each fixed damage-amount z we can imagine a series of N *Bernoulli*$(p_Z(z))$ trials, in which the j^{th} policyholder is said to result in a 'success' if he sustains a damage amount equal to z, and to result in a 'failure' otherwise. The Law of Large Numbers for these Bernoulli trials says that the number out of these N policyholders who do sustain damage z is for large N extremely likely to differ by no more than δN from $N\, p_Z(z)$.

Returning to a general discussion, suppose that Z is a discrete random variable with a finite list of possible values $z_1, \ldots, z_m$, and let $c(\cdot)$ be a real-valued (nonrandom) cost function such that $c(Z)$ represents an economically meaningful cost incurred when the random variable value Z is given. Suppose that a large number N of independent individuals give rise to respective values Z_j, $j = 1, \ldots, N$ and costs $c(Z_1), \ldots, c(Z_N)$. Here *independent* means that the mechanism causing different individual Z_j values is such that information about the values $Z_1, \ldots, Z_{j-1}$ allows no change in the (conditional) probabilities with which Z_j takes on its values, so that for all j, i, and $b_1, \ldots, b_{j-1}$,

$$P(Z_j = z_i \,|\, Z_1 = b_1, \ldots, Z_{j-1} = b_{j-1}\,) = p_Z(z_i)$$

Then the Law of Large Numbers, applied as above, says that out of the large number N of individuals it is extremely likely that approximately $p_Z(k) \cdot N$ will have their Z variable values equal to k, where k ranges over $\{z_1, \ldots, z_m\}$. It follows that the average costs $c(Z_j)$ over the N independent individuals — which can be expressed exactly as

$$N^{-1} \sum_{j=1}^{N} c(Z_j) \;=\; N^{-1} \sum_{i=1}^{m} c(z_i) \cdot \#\{j = 1, \ldots, N : \; Z_j = z_i\}$$

— is approximately given by

$$N^{-1} \sum_{i=1}^{m} c(z_i) \cdot (N\, p_Z(z_i)) = \sum_{i=1}^{m} c(z_i)\, p_Z(z_i)$$

In other words, the Law of Large Numbers implies that the **average cost per trial** among the N independent trials resulting in random variable values Z_j and corresponding costs $c(Z_j)$ has a well-defined approximate (actually, a limiting) value for very large N

$$\textbf{Expectation of cost} \;=\; E(c(Z)) = \sum_{i=1}^{m} c(z_i)\, p_Z(z_i) \tag{3.16}$$

As an application of the formula for expectation of a discrete random variable, consider the expected value of a cost-function $g(T)$ of a lifetime random variable which is assumed to depend on T only through the function

$g([T])$ of the integer part of T. This expectation was interpreted earlier as the average cost over all members of the specified life-table cohort. Now the expectation can be verified to coincide with the life-table average previously given, if the probabilities $S(j)$ in the following expression are replaced by the life-table estimators l_j/l_0. Since $P([T]=k) = S(k) - S(k+1)$, the general expectation formula (3.16) yields

$$E(g(T)) = E(g([T]) = \sum_{k=0}^{\omega-1} g(k)\,(S(k) - S(k+1))$$

agreeing precisely with formula (1.2).

Just as we did in the context of expectations of functions of the life-table waiting-time random variable T , we can interpret the *Expectation* as a weighted average of values (costs, in this discussion) which can be incurred in each trial, weighted by the probabilities with which they occur. There is an analogy in the continuous-variable case, where Z would be a random variable whose approximate probabilities of falling in tiny intervals $[z, z+dz]$ are given by $f_Z(z)dz$, where $f_Z(z)$ is a nonnegative density function integrating to 1. In this case, the weighted average of cost-function values $c(z)$ which arise when $Z \in [z, z+dz]$, with approximate probability-weights $f_Z(z)dz$, is written as a limit of sums or an integral, namely $\int c(z)\,f(z)\,dz$.

3.4.2 Rules for Manipulating Expectations

We have separately defined *expectation* for continuous and discrete random variables. In the continuous case, we treated the expectation of a specified function $g(T)$ of a lifetime random variable governed by the survival function $S(x)$ of a cohort life-table, as the approximate numerical average of the values $g(T_i)$ over all individuals i with data represented through observed lifetime T_i in the life-table. The discrete case was handled more conventionally, along the lines of a 'frequentist' approach to the mathematical theory of probability. First, we observed that our calculations with $Binomial(n,p)$ random variables justified us in saying that the sum $X = X_n$ of a large number n of independent coin-toss variables $\epsilon_1, \ldots, , \epsilon_n$, each of which is 1 with probability p and 0 otherwise, has a value which with very high probability differs from $n \cdot p$ by an amount smaller than δn, where $\delta > 0$ is an arbitrarily small number not depending upon n. The *Expectation* p of each of the variables ϵ_i is recovered approximately as the numerical average $X/n = n^{-1}\sum_{i=1}^n \epsilon_i$ of the independent outcomes ϵ_i of independent trials. This Law of Large Numbers extends to arbitrary sequences of independent and identical finite-valued discrete random variables, saying that

> if $Z_1, Z_2, \ldots$ are independent random variables, in the sense that for all $k \geq 2$ and all numbers r,
>
> $$P(Z_k \leq r \,|\, Z_1 = z_1, \ldots, Z_{k-1} = z_{k-1}) = P(Z_1 \leq r)$$
>
> regardless of the precise values $z_1, \ldots, z_{k-1}$, then for each $\delta > 0$, as n gets large
>
> $$P\Big(|n^{-1}\sum_{i=1}^n c(Z_i) - E(c(Z_1))| \geq \delta\Big) \longrightarrow 0 \qquad (3.17)$$
>
> where, in terms of the finite set S of possible values of Z ,
>
> $$E(c(Z_1)) = \sum_{z\in S} c(z)\, P(Z_1 = z) \qquad (3.18)$$

Although we do not give any further proof here, it is a fact that the same **Law of Large Numbers** given in equation (3.17) continues to hold if the definition of *independent* sequences of random variables Z_i is suitably generalized, as long as either

> Z_i are discrete with infinitely many possible values defining a set S, and the expectation is as given in equation (3.18) above whenever the function $c(z)$ is such that
>
> $$\sum_{z\in S} |c(z)|\, P(Z_1 = z) \;<\; \infty$$

or

> the independent random variables Z_i are *continuous*, all with the same density $f(t)$ such that $P(q \leq Z_1 \leq r) = \int_q^r f(t)\,dt$, and expectation is defined by
>
> $$E(c(Z_1)) \;=\; \int_{-\infty}^{\infty} c(t)\, f(t)\, dt \qquad (3.19)$$

whenever the function $c(t)$ is such that

$$\int_{-\infty}^{\infty} |c(t)|\, f(t)\, dt \; < \; \infty$$

All of this serves to indicate that there really is no choice in coming up with an appropriate definition of expectations of cost-functions defined in terms of random variables Z, whether discrete or continuous. For the rest of these lectures, and more generally in applications of probability within actuarial science, we are interested in evaluating expectations of various functions of random variables related to the contingencies and uncertain duration of life. Many of these expectations concern superpositions of random amounts to be paid out after random durations. The following rules for the manipulation of expectations arising in such superpositions considerably simplify the calculations. Assume throughout the following that all payments and times which are not certain are functions of a single lifetime random variable T.

(1). If a payment consists of a nonrandom multiple (e.g., face-amount F) times a random amount $c(T)$, then the expectation of the payment is the product of F and the expectation of $c(T)$:

$$\textit{Discrete case:}\quad E(Fc(T)) = \sum_t F\, c(t)\, P(T = t)$$

$$= \; F \sum_t c(t)\, P(T = t) \; = \; F \cdot E(c(T))$$

$$\textit{Continuous case:}\quad E(Fc(T)) = \int F\, c(t) f(t)\, dt = F \int c(t) f(t)\, dt = F\cdot E(c(T))$$

(2). If a payment consists of the sum of two separate random payments $c_1(T)$, $c_2(T)$ (which may occur at different times, taken into account by treating both terms $c_k(T)$ as present values as of the same time), then the overall payment has expectation which is the sum of the expectations of the separate payments:

$$\textit{Discrete case:}\quad E(c_1(T) + c_2(T)) = \sum_t (c_1(t) + c_2(t))\, P(T = t)$$

$$= \sum_t c_1(t)\, P(T = t) + \sum_t c_2(t)\, P(T = t) = E(c_1(T)) + E(c_2(T))$$

$$\textit{Continuous case:} \quad E(c_1(T) + c_2(T)) = \int (c_1(t) + c_2(t))\, f(t)\, dt$$

$$= \int c_1(t)\, f(t)\, dt + \int c_2(t)\, f(t)\, dt = E(c_1(T)) + E(c_2(T))$$

Thus, if an uncertain payment under an insurance-related contract, based upon a continuous lifetime variable T with density f_T, occurs only if $a \le T < b$ and in that case consists of a payment of a fixed amount F occurring at a fixed time h, then the expected present value under a fixed nonrandom interest-rate i with $v = (1+i)^{-1}$, becomes by rule **(1)** above,

$$E(v^h\, F\, I_{[a \le T < b]}) = v^h\, F\, E(I_{[a \le T < b]})$$

where the indicator-notation $I_{[a \le T < b]}$ denotes a random quantity which is 1 when the condition $[a \le T < b]$ is satisfied and is 0 otherwise. Since an indicator random variable has the two possible outcomes $\{0, 1\}$ like the coin-toss variables ϵ_i above, we conclude that $E(I_{[a \le T < b]}) = P(a \le T < b) = \int_a^b f_T(t)\, dt$, and the expected present value above is

$$E(v^h\, F\, I_{[a \le T < b]}) = v^h\, F \int_a^b f_T(t)\, dt$$

3.5 Some Special Integrals

While actuaries ordinarily do not allow themselves to represent real life-table survival distributions by simple finite-parameter families of theoretical distributions (for the good reason that they never approximate the real large-sample life-table data well enough), it is important for the student to be conversant with several integrals which would arise by substituting some of the theoretical models into formulas for various net single premiums and expected lifetimes.

Consider first the *Gamma* functions and integrals arising in connection with Gamma survival distributions. The *Gamma* function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) \;=\; \int_0^\infty x^{\alpha-1}\, e^{-x}\, dx \;, \qquad \alpha > 0$$

This integral is easily checked to be equal to 1 when $\alpha = 1$, giving the total probability for an exponentially distributed random variable, i.e., a lifetime with constant force-of-mortality 1. For $\alpha = 2$, the integral is the expected value of such a unit-exponential random variable, and it is a standard integration-by-parts exercise to check that it too is 1. More generally, integration by parts in the *Gamma* integral with $u = x^\alpha$ and $dv = e^{-x}\,dx$ immediately yields the famous *recursion relation* for the *Gamma* integral, first derived by Euler, and valid for all $\alpha > 0$:

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha\, e^{-x}\, dx = \left(-x^\alpha\, e^{-x}\right)\Big|_0^\infty + \int_0^\infty \alpha\, x^{\alpha-1}\, e^{-x}\, dx \;=\; \alpha\cdot\Gamma(\alpha)$$

This relation, applied inductively, shows that for all positive integers n,

$$\Gamma(n+1) \;=\; n\cdot(n-1)\cdots 2\cdot\Gamma(1) \;=\; n!$$

The only other simple-to-derive formula explicitly giving values for (non-integer) values of the *Gamma* function is $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, obtained as follows:

$$\Gamma(\frac{1}{2}) \;=\; \int_0^\infty x^{-1/2}\, e^{-x} dx \;=\; \int_0^\infty e^{-z^2/2}\,\sqrt{2}\, dz$$

Here we have made the integral substitution $x = z^2/2$, $x^{-1/2}\, dx = \sqrt{2}\, dz$. The last integral can be given by symmetry as

$$\frac{1}{\sqrt{2}}\int_{-\infty}^\infty e^{-z^2/2}\, dz \;=\; \sqrt{\pi}$$

where the last equality is equivalent to the fact (proved in most calculus texts as an exercise in double integration using change of variable to polar coordinates) that the *standard normal distribution*

$$\Phi(x) \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-z^2/2}\, dz \tag{3.20}$$

is a *bona-fide* distribution function with limit equal to 1 as $x \to \infty$.

One of the integrals which arises in calculating expected remaining lifetimes for Weibull-distributed variables is a *Gamma* integral, after integration-by-parts and a change-of-variable. Recall that the *Weibull* density with parameters λ, γ is

$$f(t) \;=\; \lambda\,\gamma\, t^{\gamma-1}\, e^{-\lambda\, t^\gamma} \;, \qquad t > 0$$

so that $S(x) = \exp(-\lambda\, x^\gamma)$. The expected remaining life for a Weibull-distributed life aged x is calculated, via an integration by parts with $u = t - x$ and $dv = f(t)dt = -S'(t)dt$, as

$$\int_x^\infty (t-x)\,\frac{f(t)}{S(x)}\,dt = \frac{1}{S(x)}\Big[-(t-x)\,e^{-\lambda t^\gamma}\Big|_x^\infty + \int_x^\infty e^{-\lambda t^\gamma}\,dt \Big]$$

The first term in square brackets evaluates to 0 at the endpoints, and the second term can be re-expressed via the change-of-variable $w = \lambda\, t^\gamma$, to give, in the Weibull example,

$$E(T - x\,|\,T \geq x) = e^{\lambda x^\gamma}\,\frac{1}{\gamma}\,\lambda^{-1/\gamma}\int_{\lambda\, x^\gamma}^\infty w^{(1/\gamma)-1}\,e^{-w}\,dw$$

$$= \Gamma(\frac{1}{\gamma})\,e^{\lambda\, x^\gamma}\,\frac{1}{\gamma}\,\lambda^{-1/\gamma}\Big(1 - G_{1/\gamma}(\lambda\, x^\gamma)\Big)$$

where we denote by $G_\alpha(z)$ the *Gamma distribution function* with shape parameter α,

$$G_\alpha(z) \;=\; \frac{1}{\Gamma(\alpha)}\int_0^z v^{\alpha-1}\,e^{-v}\,dv$$

and the integral on the right-hand side is called the *incomplete Gamma function*. Values of $G_\alpha(z)$ can be found either in published tables which are now quite dated, or among the standard functions of many mathematical/statistical computer packages, such as **Mathematica**, **Matlab**, or **Splus**. One particular case of these integrals, the case $\alpha = 1/2$, can be recast in terms of the standard normal distribution function $\Phi(\cdot)$. We change variables by $v = y^2/2$ to obtain for $z \geq 0$,

$$G_{1/2}(z) \;=\; \frac{1}{\Gamma(1/2)}\int_0^z v^{-1/2}\,e^{-v}\,dv \;=\; \frac{1}{\sqrt{\pi}}\int_0^{\sqrt{2z}} \sqrt{2}\,e^{-y^2/2}\,dy$$

$$= \sqrt{\frac{2}{\pi}}\cdot\sqrt{2\pi}\cdot(\Phi(\sqrt{2z}) - \Phi(0)) \;=\; 2\Phi(\sqrt{2z}) - 1$$

One further expected-lifetime calculation with a common type of distribution gives results which simplify dramatically and become amenable to numerical calculation. Suppose that the lifetime random variable T is assumed *lognormally distributed* with parameters m, σ^2. Then the expected remaining lifetime of a life aged x is

$$E(\,T - x\,|\,T \geq x\,) = \frac{1}{S(x)}\int_x^\infty t\,\frac{d}{dt}\Phi(\frac{\log(t)-\log(m)}{\sigma})\,dt \,-\, x$$

Now change variables by $y = (\log(t) - \log(m))/\sigma = \log(t/m)/\sigma$, so that $t = m\,e^{\sigma y}$, and define in particular

$$x' = \frac{\log(x) - \log(m)}{\sigma}$$

Recalling that $\Phi'(z) = \exp(-z^2/2)/\sqrt{2\pi}$, we find

$$E(\,T - x\,|\,T \ge x\,) = \frac{1}{1 - \Phi(x')} \int_{x'}^{\infty} \frac{m}{\sqrt{2\pi}}\, e^{\sigma y - y^2/2}\, dy$$

The integral simplifies after completing the square $\sigma y - y^2/2 = \sigma^2/2 - (y - \sigma)^2/2$ in the exponent of the integrand and changing variables by $z = y - \sigma$. The result is:

$$E(\,T - x\,|\,T \ge x\,) = \frac{m e^{\sigma^2/2}}{1 - \Phi(x')}\,(1 - \Phi(x' - \sigma))$$

3.6 *Exercise Set 3*

(1). Show that: $\frac{\partial}{\partial x}\, {}_tp_x = {}_tp_x \cdot (\mu_x - \mu_{x+t})$.

(2). For a certain value of x, it is known that ${}_tq_x = kt$ over the time-interval $t \in [0, 3]$, where k is a constant. Express μ_{x+2} as a function of k.

(3). Suppose that an individual aged 20 has random lifetime Z with continuous density function

$$f_Z(t) = 0.02\,(t - 20)\, e^{-(t-20)^2/100} \quad , \qquad t > 20$$

(a) If this individual has a contract with your company that you must pay his heirs $\$10^6 \cdot (1.4 - Z/50)$ on the date of his death between ages 20 and 70, then what is the expected payment ?

(b) If the value of the death-payment described in (a) should properly be discounted by the factor $\exp(-0.08(Z - 20))$ (i.e. by the effective interest rate of $e^{.08} - 1$ per year) to calculate the present value of the payment, then what is the expected present value of the insurance contract ?

Hint for both parts: After a change of variables, the integral in (a) can be evaluated in terms of incomplete Gamma integrals $\int_c^\infty s^{\alpha-1}\, e^{-s}\, ds$, where the complete Gamma integrals (for c=0) are known to yield the **Gamma function** $\Gamma(\alpha) = (\alpha-1)!$, for integer $\alpha > 0$. **Also:** $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for all real > 0, and $\Gamma(1/2) = \sqrt{\pi}$.

(4). Suppose that a life-table mortality pattern is this: from ages 20 through 60, twice as many lives die in each 5-year period as in the previous five-year period. Find the probability that a life aged 20 will die between exact ages 40 and 50. If the force of mortality can be assumed constant over each five-year age period (20-24, 25-29, etc.), and if you are told that $l_{60}/l_{20} = 0.8$, then find the probability that a life aged 20 will survive at least until exact age 48.0 .

(5). Obtain an expression for μ_x if $l_x = k\, s^x\, w^{x^2}\, g^{c^x}$, where k, s, w, g, c are positive constants.

(6). Show that: $\int_0^\infty l_{x+t}\, \mu_{x+t}\, dt \;=\; l_x$.

(7). A man wishes to accumulate $\$50,000$ in a fund at the end of 20 years. If he deposits $\$1000$ in the fund at the end of each of the first 10 years and $\$1000 + x$ in the fund at the end of each of the second 10 years, then find x to the nearest dollar, where the fund earns an effective interest rate of 6% .

(8). Express in terms of annuity-functions $a^{(m)}_{\overline{N}|}$ the present value of an annuity of \$100 per month paid the first year, \$200 per month for the second year, up to \$1000 per month the tenth year. Find the numerical value of the present value if the effective annual interest rate is 7% .

(9). Find upper bounds for the following Binomial probabilities, and compare them with the exact values calculated via computer (e.g., using a spreadsheet or exact mathematical function such as **pbinom** in **Splus**) :

(a). The probability that in *Bernoulli* trials with success-probability 0.4, the number of successes lies outside the (inclusive) range [364, 446].

(b). The probability that of 1650 lives aged exactly 45, for whom ${}_{20}p_{45} = 0.72$, no more than 1075 survive to retire at age 65.

(10). If the force of mortality governing a cohort life-table is such that

$$\mu_t \;=\; \frac{2}{1+t} + \frac{2}{100-t} \quad \text{for real } t\,, \qquad 0 < t < 100$$

then find the number of deaths which will be expected to occur between ages 1 and 4, given that the radix l_0 of the life-table is $10,000$.

(11). Find the expected present value at 5% APR of an investment whose proceeds will with probability $1/2$ be a payment of \$10,000 in exactly 5 years, and with the remaining probability $1/2$ will be a payment of \$20,000 in exactly 10 years.
Hint: calculate the respective present values V_1, V_2 of the payments in each of the two events with probability 0.5, *and find the expected value of a discrete random variable which has values V_1 or V_2 with probabilities* 0.5 *each.*

3.7 Worked Examples

Example 1. Assume that a cohort life-table population satisfies $l_0 = 10^4$ *and*

$$d_x \;=\; \begin{cases} 200 & \textit{for} \quad 0 \le x \le 14 \\ 100 & \textit{for} \;\; 15 \le x \le 48 \\ 240 & \textit{for} \;\; 49 \le x \le 63 \end{cases}$$

(a) Suppose that an insurer is to pay an amount $\$100 \cdot (64-X)$ *(without regard to interest or present values related to the time-deferral of the payment) for a newborn in the life-table population, if* X *denotes the attained integer age at death. What is the expected amount to be paid ?*

(b) Find the expectation requested in (a) if the insurance is purchased for a life currently aged exactly 10 *.*

(c) Find the expected present value at 4% *interest of a payment of* $1000 *to be made at the end of the year of death of a life currently aged exactly* 20.

The first task is to develop an expression for survival function and density governing the cohort life-table population. Since the numbers of deaths are constant over intervals of years, the survival function is piecewise linear, and the life-distribution is *piecewise uniform* because the the density is piecewise constant. Specifically for this example, at integer values y,

$$l_y \;=\; \begin{cases} 10000 - 200y & \text{for} \;\; 0 \le y \le 15 \\ 7000 - 100(y-15) & \text{for} \;\; 16 \le y \le 49 \\ 3600 - 240(y-49) & \text{for} \;\; 50 \le y \le 64 \end{cases}$$

It follows that the terminal age for this population is $\omega = 64$ for this population, and $S(y) = 1 - 0.02\,y$ for $0 \le y \le 15$, $0.85 - 0.01\,y$ for $15 \le y \le 49$, and $1.536 - .024\,y$ for $49 \le y \le 64$. Alternatively, extending the function S linearly, we have the survival density $f(y) = -S'(y) = 0.02$ on $[0, 15)$, $= 0.01$ on $[15, 49)$, and $= 0.024$ on $[49, 64]$.

Now the expectation in (a) can be written in terms of the random lifetime variable with density f as

$$\int_0^{15} 0.02 \cdot 100 \cdot (64 - [y])\, dy \;+\; \int_{15}^{49} 0.01 \cdot 100 \cdot (64 - [y])\, dy$$

$$+ \int_{49}^{64} 0.024 \cdot 100 \cdot (64 - [y])\, dy$$

The integral has been written as a sum of three integrals over different ranges because the analytical form of the density f in the expectation-formula $\int g(y) f(y) dy$ is different on the three different intervals. In addition, observe that the integrand (the function $g(y) = 100(64 - [y])$ of the random lifetime Y whose expectation we are seeking) itself takes a different analytical form on successive one-year age intervals. Therefore the integral just displayed can immediately be seen to agree with the summation formula for the expectation of the function $100(64 - X)$ for the integer-valued random variable X whose probability mass function is given by

$$P(X = k) \;=\; d_k / l_0$$

The formula is

$$E(g(Y)) \;=\; E(100(64 - X)) \;=\; \sum_{k=0}^{14} 0.02 \cdot 100 \cdot (64 - k) \,+$$

$$\sum_{k=15}^{48} 0.01 \cdot 100 \cdot (64 - k) \,+\, \sum_{k=49}^{63} 0.024 \cdot 100 \cdot (64 - k)$$

Thus the solution to (a) is given (after the change-of-variable $j = 64 - k$), by

$$2.4 \sum_{j=1}^{15} j \;+\; \sum_{j=16}^{49} j \;+\; 2 \sum_{j=50}^{64} j$$

The displayed expressions can be summed either by a calculator program or by means of the easily-checked formula $\sum_{j=1}^{n} j \;=\; j(j+1)/2$ to give the numerical answer \$3103 .

The method in part (b) is very similar to that in part (a), except that we are dealing with conditional probabilities of lifetimes given to be at least 10 years long. So the summations now begin with $k = 10$, or alternatively end with $j = 64 - k = 54$, and the denominators of the conditional probabilities $P(X = k | X \geq 10)$ are $l_{10} = 8000$. The expectation in (b) then becomes

$$\sum_{k=10}^{14} \frac{200}{8000} \cdot 100 \cdot (64 - k) + \sum_{k=15}^{48} \frac{100}{8000} \cdot 100 \cdot (64 - k) + \sum_{k=49}^{63} \frac{240}{8000} \cdot 100 \cdot (64 - k)$$

which works out to the numerical value

$$3.0 \sum_{1}^{15} j \;+\; 1.25 \sum_{16}^{49} j \;+\; 2.5 \sum_{50}^{54} j \;=\; \$2391.25$$

Finally, we find the expectation in (c) as a summation beginning at $k = 20$ for a function $1000 \cdot (1.04)^{-X+19}$ of the random variable X with conditional probability distribution $P(X = k | X \geq 20) = d_k / l_{20}$ for $k \geq 20$. (Note that the function 1.04^{-X+19} is the present value of a payment of 1 at the end of the year of death, because the end of the age- X year for an individual currently at the 20^{th} birthday is $X - 19$ years away.) Since $l_{20} = 6500$, the answer to part (c) is

$$1000 \Big\{ \sum_{k=20}^{48} \frac{100}{6500} (1.04)^{19-k} \;+\; \sum_{k=49}^{63} \frac{240}{6500} (1.04)^{19-k} \Big\}$$

$$=\; 1000 \Big(\frac{1}{65} \, \frac{1 - 1.04^{-29}}{0.04} \;+\; \frac{24}{650} \, 1.04^{-29} \, \frac{1 - (1.04)^{-15}}{0.04} \Big) \;=\; 392.92$$

Example 2. Find the change in the expected lifetime of a cohort life-table population governed by survival function $S(x) = 1 - (x/\omega)$ *for* $0 \leq x \leq \omega$ *if* $\omega = 80$ *and*

(a) the force of mortality $\mu(y)$ *is multiplied by* 0.9 *at all exact ages* $y \geq 40$, *or*

(b) the force of mortality $\mu(y)$ *is decreased by the constant amount* 0.1 *at all ages* $y \geq 40$.

The force of mortality here is

$$\mu(y) \;=\; -\frac{d}{dy} \ln(1 - y/80) \;=\; \frac{1}{80 - y}$$

So multiplying it by 0.9 at ages over 40 changes leaves unaffected the density of $1/80$ for ages less than 40, and for ages y over 40 changes the density from $f(y) = 1/80$ to

$$f^*(y) \;=\; -\frac{d}{dy} \Big(S(40) \, \exp(-0.9 \int_{40}^{y} (80 - z)^{-1} \, dz) \Big)$$

$$= -\frac{d}{dy}\Big(0.5\, e^{0.9\,\ln((80-y)/40)}\Big) \;=\; -0.5\,\frac{d}{dy}\left(\frac{80-y}{40}\right)^{0.9}$$

$$= \frac{0.9}{80}\,(2 - y/40)^{-0.1}$$

Thus the expected lifetime changes from $\int_0^{80} (y/80)\, dy = 40$ to

$$\int_0^{40} (y/80)\, dy \;+\; \int_{40}^{80} y\,\frac{0.9}{80}\,(2-y/40)^{-0.1}\, dy$$

Using the change of variable $z = 2 - y/40$ in the last integral gives the expected lifetime $= 10 + .45(80/.9 - 40/1.9) = 40.53$.

Example 3. Suppose that you have available to you two investment possibilities, into each of which you are being asked to commit \$5000*. The first investment is a risk-free bond (or bank savings-account) which returns compound interest of* 5% *for a* 10*-year period. The second is a 'junk bond' which has probability* 0.6 *of paying* 11% *compound interest and returning your principal after 10 years, probability* 0.3 *of paying yearly interest at* 11% *for* 5 *years and then returning your principal of* \$5000 *at the end of the* 10^{th} *year with no further interest payments, and probability* 0.1 *of paying yearly interest for* 3 *years at* 11% *and then defaulting, paying no more interest and not returning the principal. Suppose further that the going rate of interest with respect to which present values should properly be calculated for the next* 10 *years will either be* 4.5% *or* 7.5%*, each with probability* 0.5*. Also assume that the events governing the junk bond's paying or defaulting are independent of the true interest rate's being* 4.5% *versus* 7.5% *for the next* 10 *years. Which investment provides the better expected return in terms of current (time-*0*) dollars ?*

There are six relevant events, named and displayed along with their probabilities in the following table, corresponding to the possible combinations of true interest rate (Low versus High) and payment scenarios for the junk bond (Full payment, Partial interest payments with return of principal, and Default after 3 years' interest payments):

Event Name	Description	Probability
A_1	Low $\cap$ Full	0.30
A_2	Low $\cap$ Partial	0.15
A_3	Low $\cap$ Default	0.05
A_4	High $\cap$ Full	0.30
A_5	High $\cap$ Partial	0.15
A_6	High $\cap$ Default	0.05

Note that because of independence (first defined in Section 1.1), the probabilities of intersected events are calculated as the products of the separate probabilities, e.g.,

$$P(A_2) \;=\; P(Low)\,\cdot\,P(Partial) \;=\; (0.5)\,\cdot\,(0.30) \;=\; 0.15$$

Now, under each of the events $A_1, A_2, A_3,$ the present value of the first investment (the risk-free bond) is

$$5000 \left\{ \sum_{k=1}^{10} 0.05\,(1.045)^{-k} \;+\; (1.045)^{-10} \right\} \;=\; 5197.82$$

On each of the events $A_4, A_5, A_6,$ the present value of the first investment is

$$5000 \left\{ \sum_{k=1}^{10} 0.05\,(1.075)^{-k} \;+\; (1.075)^{-10} \right\} \;=\; 4141.99$$

Thus, since

$$P(Low) \;=\; P(A_1 \cup A_2 \cup A_3) \;=\; P(A_1) + P(A_2) + P(A_3) \;=\; 0.5$$

the overall expected present value of the first investment is

$$0.5\,\cdot\,(5197.82 + 4141.99) \;=\; 4669.90$$

Turning to the second investment (the junk bond), denoting by PV the present value considered as a random variable, we have

$$E(PV \,|\, A_1)/5000 \;\;=\;\; 0.11 \sum_{k=1}^{10} (1.045)^{-k} \;+\; (1.045)^{-10} \;=\; 1.51433$$

$$
\begin{aligned}
E(PV \mid A_4)/5000 &= 0.11 \sum_{k=1}^{10} (1.075)^{-k} + (1.075)^{-10} = 1.24024 \\
E(PV \mid A_2)/5000 &= 0.11 \sum_{k=1}^{5} (1.045)^{-k} + (1.045)^{-10} = 1.12683 \\
E(PV \mid A_5)/5000 &= 0.11 \sum_{k=1}^{5} (1.075)^{-k} + (1.075)^{-10} = 0.93024 \\
E(PV \mid A_3)/5000 &= 0.11 \sum_{k=1}^{3} (1.045)^{-k} = 0.302386 \\
E(PV \mid A_6)/5000 &= 0.11 \sum_{k=1}^{3} (1.075)^{-k} = 0.286058
\end{aligned}
$$

Therefore, we conclude that the overall expected present value $E(PV)$ of the second investment is

$$
\sum_{i=1}^{6} E(PV \cdot I_{A_i}) = \sum_{i=1}^{6} E(PV|A_i)\, P(A_i) = 5000 \cdot (1.16435) = 5821.77
$$

So, although the first-investment is 'risk-free', it does not keep up with inflation in the sense that its present value is not even as large as its starting value. The second investment, risky as it is, nevertheless beats inflation (i.e., the expected present value of the accumulation after 10 years is greater than the initial face value of \$5000) although with probability $P(\mathit{Default}) = 0.10$ the investor may be so unfortunate as to emerge (in present value terms) with only 30% of his initial capital.

3.8 Useful Formulas from Chapter 3

$$ {}_kp_x \;=\; p_x \, p_{x+1} \, p_{x+2} \cdots p_{x+k-1} \quad , \qquad k \ge 1 \quad \text{integer} $$

p. 63

$$ {}_{k/m}p_x \;=\; \prod_{j=0}^{k-1} {}_{1/m}p_{x+j/m} \quad , \qquad k \ge 1 \quad \text{integer} $$

p. 63

(i) Piecewise Unif.. $S(k+t) \;=\; tS(k+1)+(1-t)S(k)$, k integer , $t \in [0,1]$

p. 66

(ii) Piecewise Const. $\mu(y)$ $\ln S(k+t) \;=\; t \ln S(k+1)+(1-t) \ln S(k)$, k integer

p. 66

(iii) Balducci assump. $\dfrac{1}{S(k+t)} \;=\; \dfrac{t}{S(k+1)} + \dfrac{1-t}{S(k)}$, k integer

p. 66

$$ {}_tp_k \;=\; \frac{S(k) - t(S(k+1) - S(k))}{S(k)} \;=\; 1 - t\, q_k \qquad \text{under (i)} $$

p. 68

$$ {}_tp_k \;=\; \frac{S(k+t)}{S(x)} \;=\; \left(e^{-\mu(k)}\right)^t \;=\; (1-q_k)^t \qquad \text{under (ii)} $$

p. 68

$$ {}_tp_k \;=\; \frac{S(k+t)}{S(k+1)} \, \frac{S(k+1)}{S(k)} \;=\; \frac{1-q_k}{1-(1-t)q_k} \qquad \text{under (iii)} $$

p. 68

Binomial(N,p) probability $\quad P(X=k) = \binom{N}{k} p^k \, (1-p)^{N-k}$

p. 70

Discrete r.v. Expectation $\quad E(c(Z)) = \sum_{i=1}^{m} c(z_i) \, p_Z(z_i)$

p. 79

$$\Gamma(\alpha) \;=\; \int_0^\infty x^{\alpha-1} \, e^{-x} \, dx$$

p. 84

$$\Phi(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} \, dz$$

p. 84

Bibliography

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

[2] Cox, D. R. and Oakes, D. **Analysis of Survival Data**, Chapman and Hall, London 1984

[3] Feller, W. **An Introduction to Probability Theory and its Applications**, vol. I, 2nd ed. Wiley, New York, 1957

[4] Feller, W. **An Introduction to Probability Theory and its Applications**, vol. II, 2nd ed. Wiley, New York, 1971

[5] Gerber, H. **Life Insurance Mathematics**, 3rd ed. Springer-Verlag, New York, 1997

[6] Hogg, R. V. and Tanis, E. **Probability and Statistical Inference**, *5th ed.* Prentice-Hall Simon & Schuster, New York, 1997

[7] Jordan, C. W. **Life Contingencies**, 2nd ed. Society of Actuaries, Chicago, 1967

[8] Kalbfleisch, J. and Prentice, R. **The Statistical Analysis of Failure Time Data**, Wiley, New York 1980

[9] Kellison, S. **The Theory of Interest**. Irwin, Homewood, Ill. 1970

[10] Larsen, R. and Marx, M. **An Introduction to Probability and its Applications**. Prentice-Hall, Englewood Cliffs, NJ 1985

[11] Larson, H. **Introduction to Probability Theory and Statistical Inference**, 3rd ed. Wiley, New York, 1982

[12] Lee, E. T. **Statistical Models for Survival Data Analysis**, Lifetime Learning, Belmont Calif. 1980

[13] The R Development Core Team, **R: a Language and Environment**.

[14] **Splus**, version 7.0. MathSoft Inc., 1988, 1996, 2005

[15] Spiegelman, M. *Introduction to Demography*, Revised ed. Univ. of Chicago, Chicago, 1968

[16] Venables, W. and Ripley, B. **Modern Applied Statistics with S**, 4th ed. Springer-Verlag, New York, 2002