TRANSACTIONS OF SOCIETY OF ACTUARIES
1984 VOL. 36

MAXIMUM LIKELIHOOD ALTERNATIVES TO ACTUARIAL ESTIMATORS OF MORTALITY RATES

JAMES D. BROFFITT

ABSTRACT*

This paper will investigate the problem in estimating q_x using data obtained from insurance company records. For the ith life, let $[x+a_i,x+b_i]$ be the age interval for which observation is possible ($0 \leq a_i < b_i \leq 1$). This life is observed from age $x+a_i$ to either death, withdrawal, or age $x+b_i$ where observation terminates. For data of this form, maximum likelihood estimators are derived under both the assumptions of a uniform distribution of deaths and a constant force of mortality within the unit age interval $[x,x+1]$. The estimators for partial data (frequency counts of deaths and withdrawals) and full data (ages at death and withdrawal) are developed. These estimators are compared with the usual actuarial estimators and the product limit estimator, using a Monte Carlo study and asymptotic results.

1. INTRODUCTION

The Problem

Consider the problem of estimating q_x. Since q_x is a population parameter, its true value can never be known with complete accuracy, but it may be estimated from sample data. The simplest approach is to run a binomial experiment, where we start with a random sample of N lives (all aged x, selected from the population of interest) and observe them for a one-year period. If D denotes the random number of observed deaths among these N lives, then D/N provides the appropriate estimator of q_x.

In studies based on insured lives, this ideal solution is not applicable. Rather than actually conducting an experiment, the data used in mortality estimation is taken from records that insurance companies maintain on the lives they insure. Some insureds who are under observation at age x may let their policies lapse prior to attaining age $x + 1$, and, consequently, we lose track of what happens to them. These lives prematurely withdraw from

*My thanks to Stuart Klugman for suggesting this project and to Jim Hickman and Cecil Nesbitt for their useful comments on an earlier version of the manuscript.

observation, and any deaths among them that occur after withdrawal but before age $x + 1$ are unobservable and cannot be counted in D. In this case, D will be less than the actual number of deaths resulting from the original N lives, and D/N will underestimate q_x.

This problem arises because there are two causes of decrement (death and withdrawal) which are competing for lives. Observation on a life can terminate by only one of these two causes, which are operating simultaneously. The ratio D/N is estimating the probability of an *observed* death, but q_x is the probability of an *actual* death, whether observed or unobserved.

One remedy is to use the so-called actuarial estimator

$$\hat{q}_x = \frac{D}{N - \frac{1}{2}W}, \tag{1.1}$$

where W equals the random number of lives that withdraw. Notice that in the ratio D/N, each life contributes a count of one to the denominator or exposure. When a life withdraws, it loses the potential for contributing to D; consequently, it should not contribute a full unit to the exposure. Assuming that, on the average, withdrawals occur at age $x + \frac{1}{2}$, we deduct a half unit of exposure for each withdrawal, which produces the estimator given by (1.1).

An alternate explanation of (1.1) is obtained by assuming that half the withdrawals occur at age x and half at age $x + 1$. This is equivalent to running a binomial experiment where we begin with a sample of size $N - \frac{1}{2}W$. Since D deaths are observed, the estimator of q_x is (1.1).

This paper contains a unified treatment of maximum likelihood estimators for q_x, derived from models which make allowance for withdrawals. A summary of these estimators is contained in Appendix A. The major reason for considering maximum likelihood estimators is that under mild conditions they are asymptotically efficient. That is, among a large class of consistent estimators that have asymptotic normal distributions, there is none that has a variance any smaller than that of the maximum likelihood estimator when N is large. Thus, in the case of large sample sizes, we expect the maximum likelihood estimator to be better than other consistent asymptotically normal estimators.

Most of the maximum likelihood estimators in sections 2 and 3 have been previously derived by other writers, for example, Steelman [13], Elveback [8], Chiang [5] and [6], and Elandt-Johnson and Johnson [7]. In particular, the development of the full data maximum likelihood estimators in section 2 is similar to that given by Steelman [13].

The Basic Random Withdrawal Model

With respect to the age interval $[x,x+1]$, each life in the sample is observed from age x until it either dies, withdraws, or attains age $x+1$. For each life, we observe the method by which observation terminates and the time (T), measured from age x, at which this termination occurs. (If observation terminates on a life at age s, then $T = s-x$.) In order to model T, we associate with each life a pair of independent random variables Y and Z which are the times to death and withdrawal, respectively, each being measured from age x. Then $T = \min(Y,Z,1)$. For example, if $Z < Y$ and $Z < 1$, then $T = Z$ and withdrawal will be observed at age $x+Z$.

The independence of Y and Z is necessary for the development of the likelihood function in the random withdrawal model. In practice, the validity of this assumption deserves careful scrutiny.

We assume throughout that Y is a continuous random varible, and, unless otherwise stated, we also assume that Z is continuous. We shall call this the random withdrawal model since Z is considered a random variable. The case where Z is considered nonrandom is called the fixed withdrawal model and is treated in section 3.

The parameter of interest is $q_x = P[Y \leq 1] = 1 - p_x$. Let $Q_x = P[Y \leq Z, Y \leq 1]$, which is the probability of an observed death. Clearly, $q_x \geq Q_x$. Following Jordan [9], we shall call q_x the rate of death and Q_x the probability of death. Corresponding to withdrawal, we analogously define $r_x = P[Z \leq 1]$ and $R_x = P[Z \leq Y, Z \leq 1]$ as the rate and probability of withdrawal, respectively. The distribution functions of Y and Z are denoted respectively by $F(\cdot)$ and $H(\cdot)$; for example, ${}_tq_x = P[Y \leq t] = F(t)$. The density functions, when they exist, will be denoted by the corresponding lower case letters $f(\cdot)$ and $h(\cdot)$. Finally, the forces of death and withdrawal at age x will be denoted by μ_x and ν_x, respectively. In order to simplify notation, we may delete the subscript x, unless needed for clarity, and merely write q, r, Q, and R.

To see how q and Q are mathematically related, we condition on $Y = t$ and obtain

$$\begin{aligned} Q &= P[Y \leq Z, Y \leq 1] \\ &= \int_0^1 [1 - H(t)]\, dF(t) \\ &= q - \int_0^1 H(t) dF(t) \end{aligned} \tag{1.2}$$

$$= q - qr + \int_0^1 F(t)dH(t)$$

$$= q[1 - r(1-c)], \tag{1.3}$$

$$c = \int_0^1 F(t)dH(t)/\, qr = P[Y \leq Z \mid Y \leq 1, Z \leq 1],$$

which is the conditional probability of an observed death given that death and withdrawal are due to occur before age $x + 1$.
Similarly,

$$R = P[Z \leq Y, Z \leq 1]$$

$$= \int_0^1 [1 - F(t)]dH(t)$$

$$= r[1-qc]. \tag{1.4}$$

Using traditional actuarial notation, we note the alternate form for Q:

$$Q = \int_0^1 (1 - {}_tr_x)(1 - {}_tq_x)\mu_{x+t}\, dt. \tag{1.5}$$

Since ${}_tq_x = F(t)$ and ${}_tr_x = H(t)$, the equality of (1.2) and (1.5) is clear.

Let D_i be an indicator random variable that equals 1 if the ith life is observed to die and 0 otherwise. Then $D = \sum_1^N D_i$. Similarly W_i will be the indicator variable for withdrawal, and $W = \sum_1^N W_i$ is the total number of withdrawals. Finally $S = N-D-W$ denotes the number of lives surviving under observation to age $x+1$.

Unidentifiability of q_x

Generally our data consists of observations on T_i, D_i, W_i, $i = 1, \ldots, N$. However, suppose we have only the summary statistics D and W. How can we estimate q? We know that $D/N \xrightarrow{p} Q$ and $W/N \xrightarrow{p} R$, where $\xrightarrow{p}$ denotes convergence in probability. As $N \to \infty$, D/N and W/N will grow closer to Q and R. From observing D and W, the most we could ever hope to learn about the model are the values of Q and R and, consequently, any functions of them. Suppose we consider the limiting case ($N = \infty$) and assume we have the exact values of Q and R. What, then, is q? The problem is to solve

$$Q = q[1 - r(1 - c)] \tag{1.3}$$

and

$$R = r[1 - qc] \tag{1.4}$$

for q. We have two equations in three unknowns (q, r and c), and unfortunately there is not a unique solution. So even if we are given the values of Q and R, we cannot uniquely determine q. It is in this sense that we say q is not identifiable. This undesirable result was previously observed by Lindley [11]. Based only on D and W, it is impossible to construct a consistent estimator of q, unless we are willing to impose additional assumptions on the model.

In particular, the actuarial estimator

$$\hat{q} = \frac{D}{N - \frac{1}{2}W} \tag{1.1}$$

is generally not a consistent estimator of q. Since

$$\frac{D}{N - \frac{1}{2}W} \xrightarrow{p} \frac{Q}{1 - \frac{1}{2}R},$$

the actuarial esitmator will be consistent if and only if $Q/(1 - \frac{1}{2}R) = q$. This is equivalent to

$$\int_0^1 F(t)dH(t) = \frac{F(1)H(1)}{2-F(1)},$$

which is satisfied if

$$F(x) = 1 - 1/(1+kH(x))^{1/2}, \quad 0 \leq x \leq 1 \tag{1.6}$$

where k is any positive constant. Thus if (1.6) holds, the actuarial estimator (1.1) does converge in probability to q. This result is a special case of Theorem 1 in Breslow and Crowley [4].

Date-to-Date Studies

In anniversary-to-anniversary studies, each life comes under observation at an integer (insuring) age, and if it is neither a death nor a withdrawal, it leaves observation at an integer age. In this case the basic random withdrawal model is appropriate since each life observed within the unit age interval $[x, x+1]$ must begin observation at age x (assuming x is integral), and continue under observation to age $x+1$, unless previously removed by death or withdrawal.

However, complications arise in date-to-date studies, since within the age interval $[x, x+1]$, some lives may begin observation after age x, and some may be "forced" to withdraw if the study ends before they can attain age $x+1$. These possibilities are illustrated in Figure 1. The history of each life is represented by a 45° line which connects the points where observation began and ended. With respect to the age interval [35,36], both lives (1) and (2) entered into observation at age 35¼ and terminated observation at age 35¾. Life (1) was a withdrawal, whereas life (2) was a forced withdrawal.

It is important to distinguish between withdrawals and forced withdrawals since they can play different roles in the likelihood function. For the age interval $[x, x+1]$, a withdrawal is a life that terminates observation after age x and before age $x+1$, while alive, and before the study ends. This type of withdrawal corresponds to a lapsing policy-holder. A forced withdrawal occurs in $[x, x+1]$ when a life is under observation and between ages x and $x+1$ when the study ends. Within a unit age interval, any life whose 45° line is to the upper right of the dashed line (see Figure 1) is a potential forced withdrawal. For example, at the moment life (3) attained age 36 it became a potential forced withdrawal for the interval [36,37], but it was never a potential forced withdrawal for [35,36].

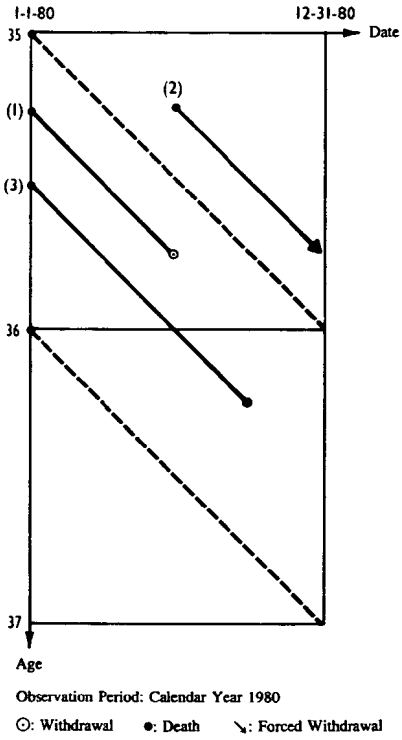


Fig. 1—Age-date diagram for mortality data.

Typically in mortality studies a single life may be observed over several consecutive unit age intervals. In this paper, all definitions and terminology will be with respect to a single age interval, $[x, x+1]$. For example, life (3) in figure 1 was a survivor in [35,36] but a death in [36,37]. When reference is made to the age that a life entered into observation, it will be with respect to a particular age interval. Thus, it is permissible to say that life (3) entered into observation at age 35½ and at age 36, since each is correct for the appropriate age interval. N will denote the number of lives observed between ages x and $x+1$. For the data in figure 1, $N = 3$ for the interval [35,36], and $N = 1$ for the interval [36,37].

We should note the difference between the data used in actuarial studies and the data used in medical studies. In the medical setting, the variable of interest is usually not age but rather duration since diagnosis or treatment for a certain disease. In this case, we may think of ${}_tq_x$ as the probability that a patient who has survived x time units since treatment will die within the

next t time units. Since all patients start observation at duration 0, the problem of late entry is generally nonexistent in medical studies of survival analysis. In this sense, there is an added complexity in actuarial data.

2. MAXIMUM LIKELIHOOD ESTIMATES (MLEs) IN THE RANDOM WITHDRAWAL MODEL

A Preliminary Example

Suppose $x_1, \ldots, x_N$ are observed values of the independent discrete random variables $X_1, \ldots, X_N$, respectively. If $P[X_i = x] = f_i(x;\theta)$ for some θ in the set Ω, then the likelihood function is

$$L(\theta) = \prod_{i=1}^{N} f_i(x_i;\theta) \ , \ \theta \in \Omega. \tag{2.1}$$

Thus $L(\theta)$ is the probability of observing, in repeated samples, the values actually obtained when the parameter value is θ. The maximum likelihood estimate (MLE) of θ is $\hat{\theta}$ if

$$L(\hat{\theta}) \geq L(\theta) \text{ for all } \theta \in \Omega.$$

That is, $\hat{\theta}$ is the parameter value that maximizes the probability of the sample.

If the random variables are continuous, and $f_i(x;\theta)$ denotes the probability density function (pdf) of X_i, then the likelihood function is again given by (2.1). However, in this case $L(\theta)$ is not the probability of the sample. If we abuse the terminology a bit and interpret

$$P[x \leq X_i \leq x + dx] \doteq f_i(x;\theta)dx$$

as the probability that x is the observed value of X_i, then the "probability" of the sample is essentially

$$\prod_{i=1}^{N} f_i(x_i;\theta)dx_i = L(\theta) \prod_{i=1}^{N} dx_i.$$

The likelihood function is then obtained by simply removing the differential element Πdx_i. This method of obtaining the likelihood function will prove useful. The idea is to write the "probability" of the sample and then remove any differential elements.

As an example, consider the simplified problem where each life is ob-

served from age x to either death or age $x+1$, whichever occurs first. Thus, withdrawal is impossible and $T = \min(Y,1)$. We may view this as a special case of the basic random withdrawal model given by $Z = \infty$. In the terminology of reliability engineering, this is a life testing model with type I censoring.

Our data consist of observations on the random variables D_i, T_i, $i = 1, \ldots, N$, and, accordingly, we denote the observed values by d_i and t_i. Now the contribution to the likelihood function by the ith life is

$$P[t_i \le T_i \le t_i + dt_i, D_i = d_i] \doteq \begin{cases} 1 - q_x & , \text{if } d_i = 0,\ t_i = 1 \\ (1 - {}_{t_i}q_x)\mu_{x+t_i}\, dt_i & , \text{if } d_i = 1,\ t_i < 1. \end{cases}$$

Letting $\mathfrak{D}$ be the subset of subscripts corresponding to deaths and using $\prod_{\mathfrak{D}}$ to denote $\prod_{i\in\mathfrak{D}}$, the probability of the sample is

$$\prod_{i=1}^{N} P[t_i \le T_i < t_i + dt_i, D_i = d_i] \doteq (1-q_x)^{N-d} \prod_{\mathfrak{D}} [(1 - {}_{t_i}q_x)\mu_{x+t_i} dt_i]$$

where $d = \sum_{1}^{N} d_i$ is the observed value of the random variable D. To obtain the likelihood function, we remove the differential elements so that

$$L = (1 - q_x)^{N-d} \prod_{\mathfrak{D}} [(1 - {}_{t_i}q_x)\mu_{x+t_i}]. \tag{2.2}$$

To find the value of q_x that maximizes (2.2), we must first express ${}_t q_x$ in terms of q_x. This may be done by placing an additional restriction on the model. Three common assumptions (see Batten [1]) are uniform distribution of deaths, the Balducci hypothesis, and constant force of mortality. Using Batten's notation, we label these assumptions A, B, and C, respectively. Batten discusses the implications and merits of each assumption. While assumptions A and C seem generally acceptable, assumption B leads to some unreasonable results and is considered only because it leads to a simple estimator of q_x when the derivation of the estimator is based on the actuarial approach (see section 3). For these reasons we will restrict our attention to assumptions A and C, and throughout this paper maximum likelihood estimators will be derived under both of these assumptions. Table 1 summarizes some properties of the three assumptions when $0 \le a \le 1$, $0 \le t \le 1$, and $0 \le t+a \le 1$.

TABLE 1

PROPERTIES OF THE THREE MORTALITY ASSUMPTIONS

	A Uniform Distribution of Deaths	B Balducci Hypothesis	C Constant Force of Mortality
${}_{t-a}q_{x+a}$	$\frac{(t-a)q_x}{1-aq_x}$	$\frac{(t-a)q_x}{1-(1-t)q_x}$	$1-e^{-\mu(t-a)}$
μ_{x+t}	$\frac{q_x}{1-tq_x}$	$\frac{q_x}{1-(1-t)q_x}$	μ

Under A, (2.2) reduces to

$$L(q) = q^d(1-q)^{N-d} \tag{2.3}$$

and the maximum likelihood estimator of q is $\hat{q} = D/N$. (Although we shall be careful to distinguish between the random variables T, D, W, and their observed values t, d, w, we shall not make a notational distinction between $\hat{q}$ the estimator and $\hat{q}$ the estimate. Thus, in the previous example, $\hat{q}$ could denote either the estimator D/N or the estimate d/N.)

There are two levels at which data may be recorded. We may record only the indicator variables $D_1, \ldots, D_N$, or, in addition, we may record the times $T_1, \ldots, T_N$. In the first case, we shall refer to the data as partial, and in the second as full. Both cases will be considered in the sequel.

For the problem under immediate discussion, it is interesting to note that the partial data case corresponds to the binomial experiment discussed in Section 1. The corresponding likelihood function is (2.3), and this result does not require additional assumptions such as A, B, or C. That is, the MLE of q is the same whether we use partial data without additional assumptions or full data with assumption A.

Under C, it is convenient to express the likelihood function in terms of μ rather than q. This is permissible since there is a one-to-one correspondence between these two parameters, which is given by $q = 1 - e^{-\mu}$. Then (2.2) becomes

$$L(\mu) = \mu^d \exp[-\mu(N-d) - \mu\Sigma t_i]$$

$$= \mu^d \exp[-\mu \sum_1^N t_i].$$

Hence, $\hat{\mu} = D/\sum_{1}^{N} T_i$, and, accordingly,

$$\hat{q} = 1 - \exp(-D/\sum_{1}^{N} T_i).$$

Notice that $\hat{\mu}$ is the number of deaths per total time the N lives were under observation within the age interval $[x,x+1]$. As we shall see, this will always be the result for the full data case with assumption C.

The General Random Withdrawal Model

We are now ready to extend the basic random withdrawal model to include cases where lives may enter observation after age x and may be forced to withdraw before age $x+1$.

With respect to the age interval $[x,x+1]$, let $x+a_i$ be the age where observation begins for the ith life. This life is observed until it dies, withdraws, or attains age $x+b_i$, at which point observation ceases, either by forced withdrawal ($b_i < 1$) or the attainment of age $x+1$ ($b_i = 1$). Clearly, $0 \leq a_i < b_i \leq 1$. If the age and date at entry are such that the life would be at least $x+1$ years of age at the close of the study, then $b_i = 1$. However, if the life enters at a point where it does not have the potential to attain age $x+1$ by the close of the study, then $b_i < 1$.

We emphasize that the definitions of a_i and b_i are with respect to a particular age interval. For the interval [35,36] in figure 1, life (1) has $a = 1/4$, $b = 1$, (2) has $a = 1/4$, $b = 3/4$, and for (3) $a = 1/2$, $b = 1$. For the interval [36,37], life (3) has $a = 0$, $b = 1/2$, and neither (1) nor (2) are members of the sample.

Let Y_a and Z_a be the random times to death and withdrawal respectively, measured from age $x+a$. We assume Y_a and Z_a are independent. The time at which observation terminates is

$$T = \min(a+Y_a, a+Z_a, b).$$

Notice that T is measured from age x rather than age $x+a$, so the age that observation terminates is $x+T$.

For the generalized model, "survivors" will refer to those lives which are forced to withdraw at the closing date or are observed to reach age $x+1$ prior to the closing date. These are the lives which attain age $x+b$ while under observation. In figure 1, for example, lives (2) and (3) are survivors for the age interval [35,36]. The term "withdrawals" shall refer to nonforced withdrawals only. Accordingly, the random variable W_i will be the indicator only for withdrawals that are not forced.

Let the subsets $\mathcal{D}$, $\mathcal{W}$, and $\mathcal{S}$ contain the subscripts corresponding to lives that are observed to die ($T=a+Y_a$), withdraw ($T=a+Z_a$), and survive to age $x+b$ ($T=b$), respectively. Clearly $\mathcal{D}\cup\mathcal{W}\cup\mathcal{S} = \{1,\ldots,N\}$. Then, for example, we shall use $\sum_{\mathcal{D}}$ and $\prod_{\mathcal{W}}$ to denote $\sum_{i\in\mathcal{D}}$ and $\prod_{i\in\mathcal{W}}$.

Full Data Problem

We now analyze the full data problem and construct the likelihood function by considering the probability of the sample.

$$P[t\leq T_i\leq t+dt, D_i=d, W_i=w] \doteq (1-{}_{b-a}q_{x+a})(1-{}_{b-a}r_{x+a}),$$
$$\text{if } d=w=0,\ t=b$$
$$\doteq (1-{}_{t-a}q_{x+a})(1-{}_{t-a}r_{x+a})\mu_{x+t}\,dt,$$
$$\text{if } d=1,\ w=0,\ t<b$$
$$\doteq (1-{}_{t-a}q_{x+a})(1-{}_{t-a}r_{x+a})\,v_{x+t}\,dt,$$
$$\text{if } d=0,\ w=1,\ t<b.$$

To get the likelihood we form $\prod_1^N P[t_i\leq T_i\leq t_i+dt_i, D_i=d_i, W_i=w_i]$ where t_i,d_i,w_i, $i=1,\ldots,N$ are the observed values and remove the differential elements. Thus,

$$L = \prod_{i=1}^{N}(1-{}_{t_i-a_i}q_{x+a_i})(1-{}_{t_i-a_i}r_{x+a_i})\cdot\prod_{\mathcal{D}}\mu_{x+t_i}\cdot\prod_{\mathcal{W}}v_{x+t_i}. \tag{2.4}$$

Since the factors in (2.4) containing r and v do not depend on q or μ, they may be absorbed into a constant, and (2.4) becomes

$$L\propto\prod_{i=1}^{N}(1-{}_{t_i-a_i}q_{x+a_i})\cdot\prod_{\mathcal{D}}\mu_{x+t_i}. \tag{2.5}$$

Full Data—Assumption C

Under assumption C, (2.5) reduces to

$$L\propto\mu^d e^{-\mu\sum_1^N(t_i-a_i)}$$

The MLE of μ is readily found to be

$$\hat{\mu} = D/\sum_{1}^{N}(T_i - a_i),$$

and, consequently,

$$\hat{q} = 1 - e^{-D/\sum_{1}^{N}(T_i - a_i)} \tag{2.6}$$

Again we comment that $\hat{\mu}$ is the number of deaths per total time the N lives were under observation within the age interval $[x,x+1]$.

This estimator may be extended to the case where x is very large, and it is unreasonable to assume μ_{x+t} is constant for $0 < t < 1$. That is, suppose we partition $[x,x+1]$ into subintervals $[x_0,x_1],(x_1,x_2],\cdots,(x_k,x_{k+1}]$ where $x = x_0$, $x+1 = x_{k+1}$, and assume $\mu_{x+t} = \mu_i$ for $x_i < t < x_{i+1}$, $i = 0, \cdots,k$. Then the MLE of q is

$$\hat{q} = 1 - e^{\sum_{i=0}^{k}(x_{i+1} - x_i)\hat{\mu}_i}$$

where $\hat{\mu}_i$ is the ratio of the number of deaths observed between ages x_i and x_{i+1} to the total time the N lives were under observation within the interval $[x_i,x_{i+1}]$.

Full Data—Assumption A

Under assumption A, (2.5) becomes

$$L \propto \prod_{1}^{N}\left(\frac{1-t_i \cdot q}{1-a_i \cdot q}\right) \cdot \prod_{\mathcal{D}}\left(\frac{q}{1-t_i \cdot q}\right)$$

$$= q^{d} \cdot \prod_{1}^{N}(1-a_i \cdot q)^{-1} \prod_{\mathcal{W} \cup \mathcal{S}}(1-t_i \cdot q). \tag{2.7}$$

Generally, the MLE will be the solution to $\dfrac{\partial}{\partial q} \ln L = 0$, which may be written as

$$\frac{d}{\hat{q}} + \sum_{i=1}^{N}\frac{a_i}{1-a_i \cdot \hat{q}} - \sum_{\mathcal{W} \cup \mathcal{S}}\frac{t_i}{1-t_i \cdot \hat{q}} = 0, \tag{2.8}$$

and must be solved by iteration.

NOTE: It is tempting to find the MLE by solving $\frac{\partial}{\partial q} \ln L(q) = 0$ or $\frac{\partial}{\partial q} L(q) = 0$. However, this may not always work. For example, if $d = 0$, (2.7) is a decreasing function of q; therefore, $\frac{\partial}{\partial q} L(q) < 0$ and $\hat{q} = 0$. On the other hand, if $d = N$, $\frac{\partial}{\partial q} L(q) > 0$ and $\hat{q} = 1$. In general, care should be exercised so that these special cases are detected before blindly trying to solve $\frac{\partial}{\partial q} \ln L(q) = 0$.

It is interesting to notice that under certain conditions (2.8) yields explicit solutions for $\hat{q}$. Suppose $a_i \equiv 0$, $b_i \equiv 1$ (which is the case for anniversary-to-anniversary studies based on insuring age), and for each withdrawal $t_i \equiv \frac{1}{2}$; then (2.8) stimplifies to

$$\frac{d}{\hat{q}} - \frac{N-d-w}{1-\hat{q}} - \frac{\frac{1}{2}w}{1-\frac{1}{2}\hat{q}} = 0.$$

The MLE obtained by solving the above equation is

$$\hat{q} = \frac{(2N+D-W) - \sqrt{(2N+D-W)^2 - 8ND}}{2N}. \tag{2.9}$$

Suppose $a_i \equiv 0$, $b_i \equiv 1$, and half the withdrawals occur at 0 with the other half at 1; then (2.8) becomes

$$\frac{d}{\hat{q}} - \frac{N-d-w}{1-\hat{q}} - \frac{\frac{1}{2}w}{1-\hat{q}} = 0,$$

and the solution is

$$\hat{q} = \frac{D}{N-\frac{1}{2}W}, \tag{1.1}$$

the actuarial estimator discussed in section 1.

Partial Data Problem

We now turn to the partial data problem where our data consists of observations only on D_i, W_i, $i=1,\cdots,N$. The likelihood function is

$$L = \prod_{i=1}^{N} P[D_i = d_i, W_i = w_i]$$

where d_i and w_i are the observed values. There are three possibilities for (d_i,w_i). An observed death, withdrawal, and survival to age $x+b_i$, correspond repectively to $(d_i,w_i) = (1,0)$, $(0,1)$, and $(0,0)$. The probabilities of these events are denoted by

$$\begin{aligned} P[D_i = d_i, W_i = w_i] &= Q_i \text{ for } d_i = 1,\ w_i = 0 \\ &= R_i \text{ for } d_i = 0,\ w_i = 1 \\ &= 1 - Q_i - R_i \text{ for } d_i = w_i = 0. \end{aligned}$$

Then

$$L = \prod_{i=1}^{N} Q_i^{d_i} R_i^{w_i} (1 - Q_i - R_i)^{1-d_i-w_i}. \tag{2.10}$$

The probabilities Q_i,R_i have been subscripted since they depend on the values of a and b, which may be different for different lives. Temporarily suppressing the subscripts, Q_i and R_i may be written as

$$Q = \int_a^b (1 - {}_{t-a}q_{x+a})(1 - {}_{t-a}r_{x+a})\mu_{x+t}\, dt \tag{2.11}$$

and

$$R = \int_a^b (1 - {}_{t-a}q_{x+a})(1 - {}_{t-a}r_{x+a})\nu_{x+t}\, dt. \tag{2.12}$$

In order to express L in terms of q, we must make a simplifying assumption on the distribution of time to withdrawal as well as the time to death. We shall consider two cases.

Partial Data—Assumption C

Corresponding to assumption C, we shall assume that $\mu_{x+t} = \mu$ and $\nu_{x+t} = \nu$ for $0 \leq t \leq 1$. Thus, the forces of death and withdrawal are both constant over the unit age interval. Then (2.11) becomes

$$Q = \int_a^b \mu e^{-(t-a)(\mu+\nu)}\, dt$$

$$= \frac{\mu}{\mu+\nu}(1-e^{-(\mu+\nu)h}),$$

where $h = b-a$. Similarly,

$$R = \frac{\nu}{\mu+\nu}(1-e^{-(\mu+\nu)h}).$$

Substituting these expressions into (2.10) yields,

$$L = \frac{\mu^d \nu^w}{(\mu+\nu)^{d+w}} \cdot \prod_{\mathcal{D}\cup\mathcal{W}} [1-e^{-(\mu+\nu)h_i}] \cdot \prod_{\mathcal{S}} e^{-(\mu+\nu)h_i}.$$

Consequently,

$$\frac{\partial\, lnL}{\partial\mu} = \frac{d}{\mu} - \frac{d+w}{\mu+\nu} + \sum_{\mathcal{D}\cup\mathcal{W}} \frac{h_i e^{-(\mu+\nu)h_i}}{1-e^{-(\mu+\nu)h_i}} - \sum_{\mathcal{S}} h_i$$

$$\frac{\partial\, ln\, L}{\partial\nu} = \frac{\partial\, ln\, L}{\partial\mu} - \frac{d}{\mu} + \frac{w}{\nu}. \tag{2.13}$$

The MLE's of μ and ν are obtained by solving $\frac{\partial\, ln\, L}{\partial\mu} = 0$ and $\frac{\partial\, ln\, L}{\partial\nu} = 0$ simultaneously. From (2.13), $\hat{\nu} = \frac{w}{d}\hat{\mu}$. Substituting $\frac{w}{d}\hat{\mu}$ for $\hat{\nu}$ in $\frac{\partial\, ln\, L}{\partial\mu} = 0$ provides the following equation which may be solved by iteration

$$\sum_{\mathcal{D}\cup\mathcal{W}} \frac{h_i \nu^{h_i}}{1-\nu^{h_i}} = \sum_{\mathcal{S}} h_i, \tag{2.14}$$

where $h_i = b_i - a_i$, and $v = e^{-\hat{\mu}(1+\frac{w}{d})}$. After solving (2.14) for v, we may compute $\hat{q}$ by

$$\hat{q} = 1 - v^{\frac{D}{D+W}}. \tag{2.15}$$

An explicit solution occurs when $a_i \equiv 0$, $b_i \equiv 1$. In this case $v = s/N$ and

$$\hat{q} = 1 - (\frac{S}{N})^{\frac{D}{D+W}}, \text{ where } S = N - D - W. \tag{2.16}$$

Again assume $a_i \equiv 0$ and let N_1 lives have $b_i \equiv 1$, and N_2 lives have $b_i \equiv \frac{1}{2}$ with $N = N_1 + N_2$. That is, N_2 of the lives are potential forced withdrawals, all at age $x + \frac{1}{2}$. Let $D_{\cdot j}$, $W_{\cdot j}$, $S_{\cdot j}$ be the number of observed deaths, withdrawals, and survivors among the N_j lives, $j = 1,2$. Thus $N_j = D_{\cdot j} + W_{\cdot j} + S_{\cdot j}$. Using this notation, (2.14) may be written as

$$\frac{v}{1-v}(d_{\cdot 1} + w_{\cdot 1}) + \frac{\frac{1}{2}v^{1/2}(d_{\cdot 2} + w_{\cdot 2})}{1 - v^{1/2}} = s_{\cdot 1} + 1/2\, s_{\cdot 2}. \tag{2.17}$$

The solution for v in (2.17) is

$$v = \left[\frac{-(d_{\cdot 2} + w_{\cdot 2}) + \sqrt{(d_{\cdot 2} + w_{\cdot 2})^2 + 4(2N_1 + N_2)(2s_{\cdot 1} + s_{\cdot 2})}}{2(2N_1 + N_2)}\right]^2,$$

which together with (2.15) provides a second explicit solution for $\hat{q}$. This estimator was previously developed by Chiang [5],[6] for use in medical follow-up studies.

Partial Data—Assumption A

Finally we consider assumption A. In addition to assuming a uniform distribution of deaths, we shall also assume a uniform distribution of withdrawals within the unit age interval $[x, x+1]$. Thus ${}_tq_x = t \cdot q_x$ and ${}_tr_x = t \cdot r_x$ for $0 \leq t \leq 1$. From (2.11)

$$Q = \int_a^b \frac{1-t\cdot q}{1-a\cdot q}\cdot\frac{1-t\cdot r}{1-a\cdot r}\cdot\frac{q}{1-t\cdot q}\,dt$$

$$= \frac{hq(1-kr)}{(1-aq)(1-ar)} \tag{2.18}$$

where $h = b-a$, $k = (a+b)/2$. Similarly,

$$R = \frac{hr(1-kq)}{(1-aq)(1-ar)}\cdot \tag{2.19}$$

Substituting (2.18) and (2.19) into (2.10) produces

$$L = \prod_{\mathcal{D}} \frac{h_i q(1-k_i r)}{(1-a_i q)(1-a_i r)}\cdot \prod_{\mathcal{W}} \frac{h_i r(1-k_i q)}{(1-a_i q)(1-a_i r)}\cdot \prod_{\mathcal{S}} \frac{(1-b_i q)(1-b_i r)}{(1-a_i q)(1-a_i r)}$$

$$\propto q^d \prod_{\mathcal{W}} (1-k_i q)\cdot \prod_{\mathcal{S}} (1-b_i q)/\prod_1^N (1-a_i q).$$

Thus, $\dfrac{\partial \ln L}{\partial q} = 0$ is

$$\frac{d}{\hat{q}} + \sum_1^N \frac{a_i}{1-a_i\cdot\hat{q}} - \sum_{\mathcal{W}} \frac{k_i}{1-k_i\cdot\hat{q}} - \sum_{\mathcal{S}} \frac{b_i}{1-b_i\cdot\hat{q}} = 0. \tag{2.20}$$

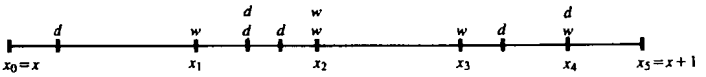
Again (2.20) may be solved by iteration to obtain $\hat{q}$. It is interesting to compare (2.20) with (2.8), which was obtained under assumption A for the full data problem. If t_i is replaced by k_i for withdrawals in (2.8), a reasonable approximation, then (2.20) is obtained. Accordingly, if $a_i \equiv 0$, $b_i \equiv 1$, and consequently $k_i \equiv \frac{1}{2}$, then the explicit solution to (2.20) is as given in equation (2.9).

The Product Limit Estimator

We conclude this section with a discussion of the product limit estimate (PLE) proposed by Böhmer [2] and later developed by Kaplan and Meier [10]. The PLE is the full data maximum likelihood estimator of q_x when the distributions of Y_a and Z_a are discrete. Its derivation does not require addi-

tional assumptions such as A,B, or C. Under the fixed withdrawal model (section 3), it is again the full data MLE when Y_a is discrete.

We shall not provide a derivation of the PLE, but we shall explain how it is computed. We begin with a simple example involving fifteen lives for which $a_i \equiv 0$, $b_i \equiv 1$. The data is displayed in figure 2. There are six deaths and five withdrawals which occurred at the ages indicated in the diagram.



Subinterval sample size: 15 13 8 7 4

$N = 15, D = 6, W = 5 S = 4$

Fig. 2—Data for PLE example.

First we partition the unit age interval $[x,x+1]$ into subintervals where the end points are given by $x,x+1$, and any age at which a withdrawal occurred. The subintervals are $[x,x_1],(x_1,x_2],(x_2,x_3],(x_3,x_4],(x_4,x+1]$. For the unit age interval the survival probability is

$$1-q_x = p_x = \prod_{i=0}^{4} ({}_{x_{i+1}-x_i}p_{x_i}),$$

where $x_0 = x$ and $x_5 = x+1$. Since withdrawals do not occur within a subinterval, we may estimate the survival probability ${}_{x_{i+1}-x_i}p_{x_i}$ using the concept of a binomial experiment. For example, ${}_{x_2-x_1}\hat{p}_{x_1} = 10/13$, since there are thirteen lives under observation after the withdrawal at age x_1 and ten of these lives survived to age x_2. The PLE of q_x is then

$$\hat{q}_x = 1 - \prod_{i=0}^{4} ({}_{x_{i+1}-x_i}\hat{p}_{x_i})$$

$$= 1 - \frac{14}{15}\cdot\frac{10}{13}\cdot\frac{8}{8}\cdot\frac{5}{7}\cdot\frac{4}{4}.$$

Notice that any deaths that occurred at the same age as a withdrawal are counted as deaths in the subinterval immediately preceding their occurrence. Thus the death at age x_4 is assumed to have occurred in the interval $(x_3,x_4]$.

An alternate method of computing the PLE is given by the formula

$$\hat{q} = 1 - \prod_{j\in J} \frac{N-j}{N-j+1}. \tag{2.21}$$

$$\hat{q} = 1 - \prod_{j \in J} \frac{N-j}{N-j+1}. \tag{2.21}$$

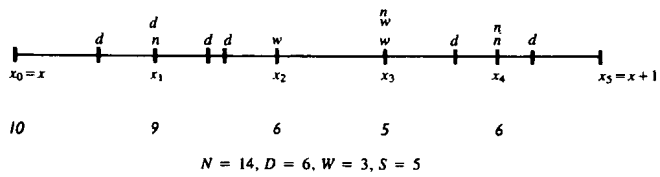
To find the set J we first order the t_i's, that is, $t_{(1)} \leq \cdots \leq t_{(N)}$. If there are any ties between times of deaths and withdrawals, the deaths receive the smaller subscripts. Then J is the set of ordered subscripts corresponding to times of deaths. In the example, $J = \{1,3,4,5,9,10\}$, and (2.21) is

$$1 - \frac{14}{15} \cdot \frac{12}{13} \cdot \frac{11}{12} \cdot \frac{10}{11} \cdot \frac{6}{7} \cdot \frac{5}{6},$$

which agrees with the previous result.

The PLE of q_x may not always exist. To see this, suppose we alter the previous example so that the four survivors become withdrawals at age x_4. Then $D = 6$, $W = 9$, $S = 0$. Since we do not observe any lives from x_4 to x_5, there is no way to estimate ${}_{x_5-x_4}p_{x_4}$ and, consequently, q_x; unless, of course, we are willing to impose some additional assumptions on the model.

Computation of the PLE may be extended to the generalized model where $a_i \neq 0$ or $b_i \neq 1$. Actually the inclusion of forced withdrawals does not create a problem since they are treated exactly as withdrawals were in the previous computation. That is, in computing the PLE no distinction is made between withdrawals and forced withdrawals. New entrants, that is, observations for which $a_i > 0$, are also treated like withdrawals, except the subinterval sample size is increased by one for each new entrant. To illustrate this, we consider the data displayed in figure 3 (new entrants are denoted by n). The number of starters, that is, observations for which $a_i = 0$), is ten, and there are four new entrants, so $N = 14$.



$N = 14, D = 6, W = 3, S = 5$

Fig. 3—Example of the PLE with new entrants.

There are five subintervals and

$$\hat{q} = 1 - \prod_{i=0}^{4} {}_{x_{i+1}-x_i}\hat{p}_{x_i}$$
$$= 1 - \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{6} \cdot \frac{4}{5} \cdot \frac{5}{6}.$$

The PLE will not exist if one of the subinterval sample sizes is zero, unless all of the lives observed over some other subinterval died. In this case $\hat{q} = 1$. We also note that formula (2.21) is not applicable when the data include new entrants.

Consider the set containing the ages x, $x+1$, and all ages where new entries or withdrawals (including forced withdrawals) occur. Let $x = x_0 < x_1 < \cdots < x_{k+1} = x+1$ be the distinct ordered points in this set. We may formally define the PLE as

$$\hat{q}_x = 1 - \prod_{i=0}^{k} {}_{x_{i+1}-x_i}\hat{p}_{x_i}$$
$$= 1 - \prod_{i=0}^{k} \frac{N(i)-D(i)}{N(i)},$$

where

$D(i)$ = the number of deaths at ages in the interval $(x_i, x_{i+1}]$,

$M(i)$ = (the number of new entrants) − (the number of withdrawals) observed at age x_i,

$N(0)$ = the number of starters, and

$N(i) = N(i-1) - D(i-1) + M(i)$, $i = 1, \cdots, k$.

In effect, $N(i)$ is the sample size for the interval $(x_i, x_{i+1}]$, of which $N(i)-D(i)$ survive to age x_{i+1}.

3. ESTIMATION IN THE FIXED WITHDRAWAL MODEL

Fixed Withdrawal Model

The fixed withdrawal model arises when the time to withdrawal (Z_a) is treated as fixed but unknown. Since this time is no longer subject to random variation, we shall use the notation z_a rather than Z_a. Alternatively, we may arrive at the fixed model by starting with the random model and conditioning on $Z_a = z_a$.

The time to termination of observation, measured from age x, is then

$$\begin{aligned} T &= \min(a+Y_a,\ a+z_a,\ b) \\ &= \min(a+Y_a,\ c), \end{aligned}$$

where $c = \min(a+z_a,\ b)$. For each life we know that $a \leqslant c \leqslant b$, but the precise value of c is unknown. However, for lives that are withdrawals or survivors, observation ceases at age $x+c$, so the value of c is observed. For lives that are deaths, observation ceases prior to age $x+c$, and c remains unknown. Thus, $\{c_i;\ i \in \mathfrak{D}\}$ must be treated as unknown parameters.

In the fixed withdrawal model, a life is observed from age $x+a$ to either death or age $x+c$ where observation ceases. The only relevant data is T_i, D_i, $i=1, \ldots, N$. Withdrawals and survivors may be combined and treated as a single group; however, to maintain consistency, we shall continue to keep them separate.

Full Data Problem

From

$$\begin{aligned} P[t \leqslant T_i \leqslant t+dt,\ D_i=d] &\doteq (1-{}_{t-a}q_{x+a})\mu_{x+t}\,dt,\ d=1,\ t<c \\ &= 1-{}_{t-a}q_{x+a} \qquad ,\ d=0,\ t=c, \end{aligned}$$

we obtain the likelihood

$$L = \prod_1^N (1-{}_{t_i-a_i}q_{x+a_i}) \cdot \prod_{\mathfrak{D}} \mu_{x+t_i}.$$

Since this is identical to the likelihood given in (2.5), the estimators for the full data problem are the same for both the random and fixed withdrawal models. In particular,

$$\hat{q} = 1 - e^{-D/\sum_1^N (T_i-a_i)} \tag{2.6}$$

is the full data MLE for both models under assumption C.

Partial Data Problem

The likelihood is

$$L = \prod_{1}^{N} P[D_i = d_i]$$

$$= \prod_{1}^{N} Q_i^{d_i} (1-Q_i)^{1-d_i},$$

where Q_i (with subscripts deleted) is

$$Q = \int_a^c (1 - {}_{t-a}q_{x+a})\mu_{x+t}\, dt$$

$$= {}_{c-a}q_{x+a}. \tag{3.1}$$

Partial Data-Assumption C

Under assumption C (3.1) is

$$Q = 1 - e^{-\mu(c-a)},$$

and the likelihood is

$$L(\mu, c_i, i \in \mathcal{D}) = \prod_{\mathcal{D}} [1 - e^{-\mu(c_i - a_i)}] \cdot \prod_{\mathcal{W} \cup \mathcal{S}} e^{-\mu(c_i - a_i)}. \tag{3.2}$$

To maximize $L(\mu, c_i, i \in \mathcal{D})$, we first fix μ and maximize with respect to c_i, $i \in \mathcal{D}$. Since L increases with c_i ($i \in \mathcal{D}$) and $a_i \leq c_i \leq b_i$, then $\hat{c}_i = b_i$. Thus,

$$L(\mu, \hat{c}_i, i \in \mathcal{D}) = \prod_{\mathcal{D}} [1 - e^{-\mu(b_i - a_i)}] \cdot \prod_{\mathcal{W} \cup \mathcal{S}} e^{-\mu(c_i - a_i)}.$$

Then $\dfrac{\partial}{\partial \mu} \ln L(\mu, \hat{c}_i, i \in \mathcal{D}) = 0$ may be written as

$$\sum_{\mathcal{D}} \frac{h_i e^{-\hat{\mu} h_i}}{1-e^{-\hat{\mu} h_i}} = \sum_{\mathcal{W} \cup \mathcal{S}} (c_i - a_i) , \tag{3.3}$$

where $h_i = b_i - a_i$. An iterative solution will provide $\hat{\mu}$ and, consequently, $\hat{q} = 1 - e^{-\hat{\mu}}$.

For the special case $a_i \equiv 0$, $b_i \equiv 1$, we obtain from (3.3)

$$\hat{q} = 1 - e^{-\hat{\mu}}$$

$$= \frac{D}{D + \sum_{\mathcal{W} \cup \mathcal{S}} c_i} \tag{3.4}$$

$$= \frac{D}{N - W + \sum_{\mathcal{W}} c_i} \tag{3.5}$$

since $c_i = b_i = 1$ for $i \in \mathcal{S}$.

If $\sum_{\mathcal{W}} c_i = \frac{1}{2} W$, then (3.5) reduces to the partial data actuarial estimator

$$\hat{q} = \frac{D}{N - \frac{1}{2} W} . \tag{1.1}$$

Also notice that (3.4) agrees with the exposure concept of the Balducci-based actuarial estimator. Deaths receive exposure from entry to age $x+1$. All others receive exposure equal to the duration of their observation.

Partial Data—Assumption A

Under assumption A, (3.1) becomes

$$Q = \frac{(c-a)q}{1-a \cdot q} ,$$

and

$$L = \prod_{\mathcal{D}} \frac{(c_i - a_i)q}{1 - a_i \cdot q} \cdot \prod_{\mathcal{W} \cup \mathcal{S}} \frac{1 - c_i \cdot q}{1 - a_i \cdot q} .$$

Thus $\frac{\partial \ ln\ L}{\partial q} = 0$ is

$$\frac{d}{\hat{q}} + \sum_{1}^{N} \frac{a_i}{1-a_i \cdot \hat{q}} - \sum_{\mathcal{WU}_{\mathcal{S}}} \frac{c_i}{1-c_i \cdot \hat{q}} = 0. \tag{3.6}$$

Notice the similarity between (2.20) and (3.6). If the c's for withdrawals are replaced by $(a+b)/2$ in (3.6), the result is (2.20).

The Actuarial Estimator

We shall end our discussion of estimators with a consideration of the full data actuarial estimator

$$\hat{q}_x = \frac{D}{\sum_{1}^{N} (c_i - a_i) + \sum_{\mathcal{D}} (1-c_i)}. \tag{3.7}$$

We have previously noted that under certain conditions, including $a_i \equiv 0$, $b_i \equiv 1$, some of the MLE's agree with the actuarial estimator. Although we have not found a model for which (3.7) is the MLE of q_x in the general case, (3.7) may be derived under assumption B as a modified method of moments estimator (see Broffitt and Klugman [3]).

We have previously noted that under certain conditions, including $a_i \equiv 0$, $b_i \equiv 1$, some of the MLE's agree with the actuarial estimator. Although we have not found a model for which (3.7) is the MLE of q_x in the general case, (3.7) may be derived under assumption B as a modified method of moments estimator (see Broffitt and Klugman [3]).
The method of moments technique begins with

$$E(D) = \sum_{1}^{N} {}_{c_i - a_i} q_{x+a_i}. \tag{3.8}$$

Then the right-hand side is expressed in terms of q_x, and the left-hand side is replaced by D. The solution for q_x in the resulting equation is the method of moments estimator (MME).

Using assumption B, we have

$$E(D_i) = {}_{c_i - a_i}q_{x+a_i}$$

$$= {}_{1-a_i}q_{x+a_i} - {}_{c_i - a_i}p_{x+a_i} \cdot {}_{1-c_i}q_{x+c_i}$$

$$= {}_{1-a_i}q_{x+a_i} - E(1-D_i)\,{}_{1-c_i}q_{x+c_i}$$

$$= [(1-a_i) - E(1-D_i)(1-c_i)]q_x.$$

Hence (3.8) becomes

$$E(D) = [\sum(1-a_i) - \sum(1-c_i)E(1-D_i)]q_x \,. \tag{3.9}$$

The modification which produces the actuarial estimator is to replace $E(1-D_i)$ by $1-D_i$ in (3.9). The resulting equation for computing the modified MME is

$$D = [\sum(1-a_i) - \sum(1-c_i)(1-D_i)]\hat{q}_x$$

and, thus,

$$\hat{q}_x = \frac{D}{\sum(1-a_i) - \sum(1-c_i)(1-D_i)}\,, \tag{3.10}$$

which is the same as (3.7).

4. COMPARISONS

Preliminaries

Up to this point we have been concerned with determining plausible estimators for q_x. In this section we shall compare some of these estimators

to see which ones do the best job in estimating q_x. Four estimators will be compared:

$$BP = \frac{D}{N - \frac{1}{2}W}$$

$$BF = \frac{D}{\sum_{1}^{N}(T_i - a_i) + \sum_{\mathcal{D}}(1 - T_i)}$$

$$CF = 1 - e^{-D/\sum_{1}^{N}(T_i - a_i)}$$

$$PL = 1 - \prod_{i=0}^{k} x_{i+1} - x_i \hat{p}_{x_i}.$$

BP and *BF* are the partial data and full data actuarial estimators; *CF* is the full data MLE under assumption *C*; and *PL* is the product limit estimator. For the three estimators *BP*, *BF*, and *CF*, the first letter in the two letter designations denotes the assumption under which the estimator was derived, *B* for Balducci and *C* for constant force of mortality. The second letter denotes the data type, *P* for partial and *F* for full. These four estimators seem to be the logical ones for comparison. While *BP* and *BF* are the currently used actuarial estimators, *CF* and *PL* were the only MLE's applicable in the $a \neq 0$, $b \neq 1$ case that did not require an iterative solution. This should not be construed as an indictment against MLE's, whose calculation requires iteration. In fact, Steelman [13] has indicated that the convergence is very rapid.

It is interesting to notice that when $W = 0$, $a \equiv 0$, and $b \equiv 1$, then $BP = BF = PL = D/N$. Thus, if there are few withdrawals, new entrants, and forced withdrawals, these three estimators should give nearly equal results.

Since q_x is generally quite small (except for extreme ages), it is of interest to compare the estimators in this case. Let $\hat{\mu} = D/\sum_{1}^{N}(T_i - a_i)$,

then

$$\frac{\hat{\mu}}{1+\hat{\mu}} \leq BF \leq \hat{\mu}$$

and

$$\frac{\hat{\mu}}{1+\hat{\mu}} \leq CF \leq \hat{\mu}.$$

Thus,

$$|BF\text{-}CF| \leq \hat{\mu} - \frac{\hat{\mu}}{1+\hat{\mu}} \leq \hat{\mu}^2,$$

which will be small when q_x is small.

The upper bound of $\hat{\mu}^2$ will generally be crude. A more representative comparison can be made by assuming the average age of observed deaths is $x+\frac{1}{2}$; then

$$BF = \frac{\hat{\mu}}{1+\frac{1}{2}\hat{\mu}} = \hat{\mu} - \frac{1}{2}\hat{\mu}^2 + \frac{1}{4}\hat{\mu}^3 - \frac{1}{8}\hat{\mu}^4 + \cdots$$

and

$$CF = 1 - e^{-\hat{\mu}} = \hat{\mu} - \frac{1}{2!}\hat{\mu}^2 + \frac{1}{3!}\hat{\mu}^3 - \frac{1}{4!}\hat{\mu}^4 + \cdots.$$

Ignoring terms of degree 4 and higher,

$$|BF - CF| \doteq \frac{1}{12}\hat{\mu}^3.$$

Thus for small q_x, BF and CF should be quite close. If, in addition, there are few withdrawals, new entrants, and forced withdrawals, all four estimators will give nearly equal results.

It is rather interesting to notice that the estimates BP, CF, and PL must each be between 0 and 1, whereas BF may take on any nonnegative value.

To see this, suppose we observe only one life which began observation at age $x+a$ and died before age $x+1$. Then $BF = 1/(1-a)$, and as $a \to 1$, $BF \to +\infty$. This undesirable result must surely stem from the unreasonableness of the Balducci assumption.

It is also of interest to note that PL is an unbiased estimator of q_x. To see this, suppose we condition on the ages of withdrawal, new entry, and forced withdrawal. Within this conditional setting,

$$\begin{aligned} E\left(\prod_{i=0}^{k} {}_{x_{i+1}-x_i}\hat{p}_{x_i}\right) &= E\left[E\left(\prod_{i=0}^{k} {}_{x_{i+1}-x_i}\hat{p}_{x_i} \middle| N(k)\right)\right] \\ &= E\left[E\left(\prod_{i=0}^{k-1} {}_{x_{i+1}-x_i}p_{x_i} \middle| N(k)\right) \cdot E\left({}_{x_{k+1}-x_k}\hat{p}_{x_k} \middle| N(k)\right)\right] \\ &= {}_{x_{k+1}-x_k}p_{x_k}\, E\left(\prod_{i=0}^{k-1} {}_{x_{i+1}-x_i}\hat{p}_{x_i}\right). \end{aligned}$$

The last two equations follow, since given $N(k)$, ${}_{x_{k+1}-x_k}\hat{p}_{x_k}$ is clearly independent of the previous $\hat{p}$'s and is an unbiased estimator of ${}_{x_{k+1}-x_k}p_{x_k}$. In general,

$$E\left(\prod_{i=0}^{j} {}_{x_{i+1}-x_i}\hat{p}_{x_i}\right) = {}_{x_{j+1}-x_j}p_{x_j}\, E\left(\prod_{i=0}^{j-1} {}_{x_{i+1}-x_i}\hat{p}_{x_i}\right), \quad j = 1, \cdots, k.$$

Consequently, we obtain

$$\begin{aligned} E\left(\prod_{i=0}^{k} {}_{x_{i+1}-x_i}\hat{p}_{x_i}\right) &= \prod_{i=1}^{k} {}_{x_{i+1}-x_i}p_{x_i} \cdot E\left({}_{x_1-x_0}\hat{p}_{x_0}\right) \\ &= p_x. \end{aligned}$$

Thus, since the conditional expectation (given the ages of withdrawal, new entry, and forced withdrawal) of PL is q_x, unconditionally, we have $E(PL) = q_x$.

Asymptotic Distributions

Our comparisons will be based mainly on the probabilities $P[|\hat{q}_x - q_x| \leq \epsilon]$. For each estimator these probabilities will be computed for various values of N and ϵ. The distributions of BP and PL are discrete (PL must be a rational number), while the distributions of BF and CF are mixtures of discrete and continuous distributions ($P[BF = 0] = P[CF = 0] = P[D = 0] > 0$). Rather than attempting to derive the exact distributions of these estimators, probabilities will be approximated using Monte Carlo techniques and asymptotic distributions. The asymptotic distributions will be determined under the random withdrawal model and only for the $a \equiv 0$, $b \equiv 1$ case.

Derivations of the asymptotic distributions are provided in Appendix B. Each estimator has a normal asymptotic distribution. We denote this by

$$\sqrt{N}\,(\hat{q}-m_{\hat{q}}) \xrightarrow{D} N(0,\tau^2_{\hat{q}}),$$

where $m_{\hat{q}}$ and $\tau^2_{\hat{q}}$ are the asymptotic mean and variance of $\hat{q}$. Appendix B contains formulas for m_{BP}, m_{BF}, m_{CF}, ($m_{PL} = q$), τ^2_{BP}, τ^2_{BF}, τ^2_{CF}, and τ^2_{PL}. These formulas are in terms of $F(\cdot)$ and $H(\cdot)$, the distribution functions of Y and Z, respectively.

Our comparisons will be based on two different sets of assumptions for $F(\cdot)$ and $H(\cdot)$.

Distribution Assumption A: $F(t) = qt$, $0 \le t \le 1$

$$H(t) = rt,\ 0 \le t \le 1$$

Distribution Assumption C: $F(t) = 1-e^{-\mu t}$, $0 \le t \le 1$

$$H(t) = 1-e^{-\nu t},\ 0 \le t \le 1.$$

Under assumption A, the times to death and withdrawal are uniformly distributed within the unit age interval $[x,x+1]$, and under C, the forces of mortality and withdrawal are constant in $[x,x+1]$. Appendix B contains the formulas needed for computing the asymptotic means and variances under assumptions A and C.

Under assumption C, $m_{CF} = q$ and τ^2_{CF} simplifies to

$$\tau^2_{CF} = \frac{\mu(\mu+\nu)e^{-2\mu}}{1-e^{-(\mu+\nu)}}.$$

Since under C both *CF* and *PL* are consistent estimators of q, and since *CF* is the MLE in this model, we know that $\tau^2_{CF} \le \tau^2_{PL}$. However, it is still of interest to compare their magnitudes. The asymptotic efficient of *PL* relative to *CF* is

$$ARE(PL,CF) = \frac{\tau^2_{CF}}{\tau^2_{PL}} = \left[\frac{\mu+\nu}{e^{1/2(\mu+\nu)}-e^{-1/2(\mu+\nu)}}\right]^2, \text{ under C.}$$

TABLE 2

VALUES OF *ARE(PL,CF)* UNDER ASSUMPTION C.

$\mu+\nu$	*ARE(PL,CF)*
2.0	0.724
1.0	0.921
0.5	0.979
0.1	0.999

Since $\lim_{\mu+\nu\to\infty} ARE(PL,CF) = 0$, for large values of $\mu+\nu$, *CF* will be vastly superior to *PL*. On the other hand, $\lim_{\mu+\nu\to 0} ARE(PL,CF) = 1$, so for small values of $\mu+\nu$, *CF* will be better than *PL* but not by very much. Table 2 provides a few values for *ARE(PL,CF)*. When $\mu+\nu = 0.5$, for example, *PL*, based on 1000 observations, would have the same variance as *CF*, based on 979 observations. Again we see that for small values of q and r, there is little difference in the performaces of *CF* and *PL*.

Numerical Results

For selected values of q and r, Table 3 displays the means and variances of the asymptotic distributions of the four estimators under the assumptions A and C. As a check on these asymptotic values, we have included the means and variances when the sample size is 100. These were obtained from a Monte Carlo Study where 1,000 replications were generated. It appears that the agreement between $N = 100$ and $N = \infty$ is quite good.

In all but one case, *PL* has the largest τ^2 and smallest $|m-q|$, and *BP* has the smallest τ^2 and largest $|m-q|$. The one exception is distribution assumption A, $q = 0.3$, $r = 0.1$, where *CF* has the smallest τ^2 and largest $|m-q|$. The estimators *BP* and *BF* always tend to underestimate q. This is also true for *CF* under assumption A. Under C, *CF* is asymptotically unbiased, and so is *PL* under any assumption. Finally, we note that (for the range of q and r values reported) increasing either q or r produces the predictable result that τ^2 increases and $|m-q|$ increases. Results similar to those noted here have been reported in other Monte Carlo studies; for examples see Steelman [13] and Miller and Hickman [12].

TABLE 3

ASYMPTOTIC MEANS AND VARIANCES FOR SELECTED VALUES OF q AND r.

q	r	Distribution	N	BP		BF		CF		PL	
				m	τ^2	m	τ^2	m	τ^2	m	τ^2
0.3	0.6	A	∞	0.2819	0.2535	0.2877	0.2675	0.2913	0.2782	0.3000	0.3384
			100	0.2822	0.2535	0.2882	0.2639	0.2917	0.2742	0.3010	0.3362
		C	∞	0.2723	0.2500	0.2935	0.2864	0.3000	0.3090	0.3000	0.3530
			100	0.2711	0.2423	0.2924	0.2800	0.2989	0.3024	0.2992	0.3505
0.3	0.3	A	∞	0.2923	0.2299	0.2948	0.2354	0.2948	0.2332	0.3000	0.2550
			100	0.2896	0.2296	0.2923	0.2368	0.2924	0.2346	0.2974	0.2574
		C	∞	0.2923	0.2299	0.2974	0.2386	0.3000	0.2445	0.3000	0.2550
			100	0.2897	0.2409	0.2949	0.2500	0.2978	0.2558	0.2972	0.2691
0.3	0.1	A	∞	0.2977	0.2163	0.2984	0.2179	0.2966	0.2105	0.3000	0.2226
			100	0.2978	0.2207	0.2986	0.2210	0.2968	0.2126	0.3003	0.2265
		C	∞	0.2982	0.2165	0.2992	0.2183	0.3000	0.2182	0.3000	0.2222
			100	0.2968	0.2098	0.2978	0.2116	0.2988	0.2136	0.2986	0.2145
0.1	0.3	A	∞	0.0991	0.1032	0.0994	0.1041	0.0996	0.1048	0.1000	0.1077
			100	0.0985	0.1092	0.0988	0.1097	0.0991	0.1108	0.0995	0.1130
		C	∞	0.0984	0.1026	0.0997	0.1054	0.1000	0.1066	0.1000	0.1085
			100	0.0979	0.1002	0.0992	0.1029	0.0995	0.1039	0.0997	0.1062
0.05	0.3	A	∞	0.0498	0.0552	0.0499	0.0554	0.0499	0.0557	0.0500	0.0566
			100	0.0490	0.0536	0.0491	0.0538	0.0492	0.0543	0.0491	0.0547
		C	∞	0.0493	0.0547	0.0499	0.0561	0.0500	0.0564	0.0500	0.0572
			100	0.0495	0.0562	0.0501	0.0574	0.0502	0.0580	0.0500	0.0578
0.01	0.3	A	∞	0.0100	0.0116	0.0100	0.0116	0.0100	0.0116	0.0100	0.0118
			100	0.0100	0.0113	0.0100	0.0113	0.0100	0.0113	0.0100	0.0116
		C	∞	0.0099	0.0115	0.0100	0.0118	0.0100	0.0118	0.0100	0.0119
			100	0.0100	0.0117	0.0101	0.0120	0.0101	0.0120	0.0101	0.0121

For the case $a \equiv 0$, $b \equiv 1$, figures 4 through 9 display plots of $P[|\hat{q}-q| \leq \epsilon]$ against sample size, for each estimator and for selected values of q, r, and ϵ. The values in figures 4, 5, 6, and 8 were computed under distribution assumption A. Figure 7 provides values computed under assumption C, and figure 9 applies to both assumptions A and C.

For each combination of q, r, ϵ, and N, $P[|\hat{q}-q| \leq \epsilon]$ was computed for each estimator by using the asymptotic normal approximation. The choices for N were 10, 32, 100, 316, 1,000, 3,162, 10,000, and 31,623. These values were selected so that $\log_{10}(N)$ would be evenly spaced.

Each curve in figure 9 represents all four estimators and both distribution assumptions A and C. For each combination of q, r, N, and ϵ in figure 9, eight values (since there are four estimators and two distribution assumptions) of $P[|\hat{q}-q| \leq \epsilon]$ were computed. In nearly every case, these eight values were within 0.01 of each other. This allowed the use of a single curve to represent all four estimators and both distributions. The one exception was $q = 0.1$, $\epsilon = 0.005$, and assumption C, where BP fell off slightly at the larger values of N.

The smaller we take q, the larger N must be for the normal approximation to be accurate. To check the asymptotic values, empirical distributions were generated for each estimator. For each combination of q, r, N, and distribution assumption, 1,000 samples of size N were generated. For each of these samples, BP, BF, CF, and PL were computed. This provided empirical distributions from which $P[|\hat{q}-q| \leq \epsilon]$ was calculated. For the smaller values of N, if there was significant disagreement between the asymptotic and empirical calculations, $P[|\hat{q}-q| \leq \epsilon]$ was not plotted. This accounts for the missing values in figures 4 through 9.

Let $\hat{q}$ represent BP, BF, or CF, and let m be the mean of the corresponding asymptotic distribution. Then, since $\hat{q} \xrightarrow{p} m$,

$$\lim_{N\to\infty} P[|\hat{q}-q| \leq \epsilon] = 1, \text{ if } |m-q| \leq \epsilon$$

$$= 0, \text{ if } |m-q| > \epsilon.$$

That is, $P[|\hat{q}-q| \leq \epsilon]$ approaches 1 or 0, depending on whether the asymptotic bias is less than or equal to ϵ or greater than ϵ. This explains why some of the curves bend back toward the horizontal axis as $N \to \infty$.

A general inspection of figures 4 through 8 reveals that for $N \leq 316$, all estimators have essentially equal values of $P[|\hat{q}-q| \leq \epsilon]$. (Of course, there was some variability in these values, usually 0.02 or less, but this was not deemed large enough to warrant plotting multiple points.) For large values

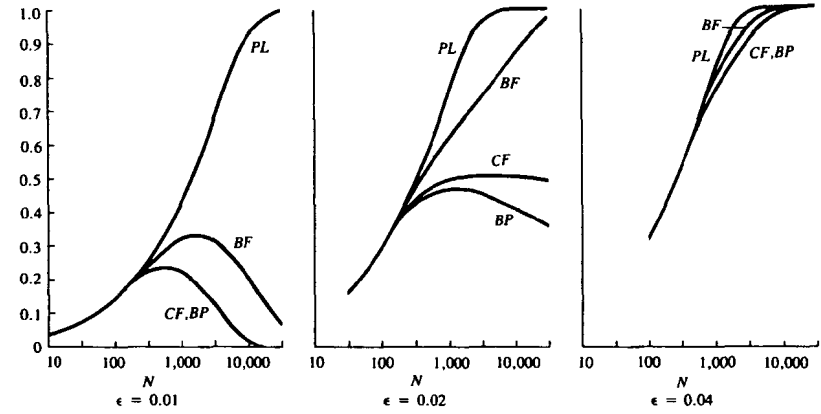


Fig. 4—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.5$, $r = 0.3$, Distribution A.

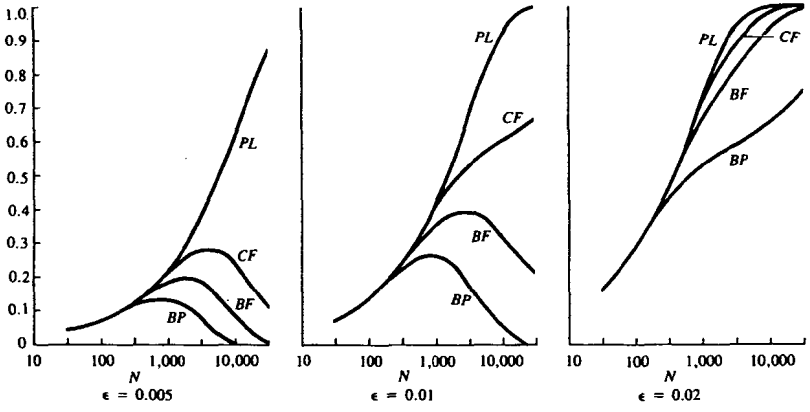


Fig. 5—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.3$, $r = 0.6$, Distribution A.

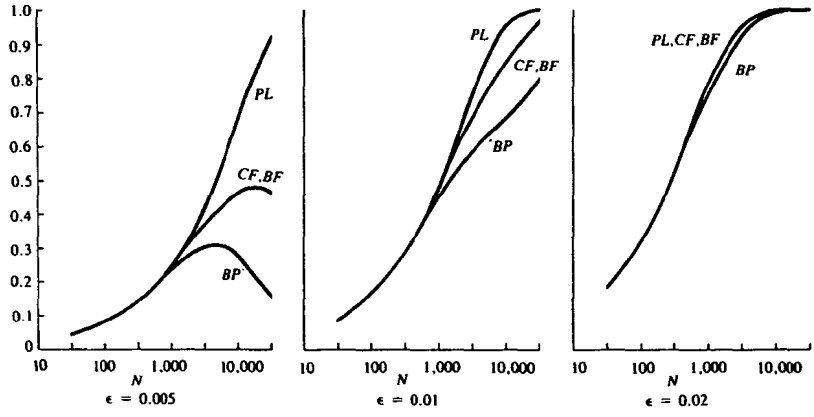


Fig. 6—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.3$, $r = 0.3$, Distribution A.

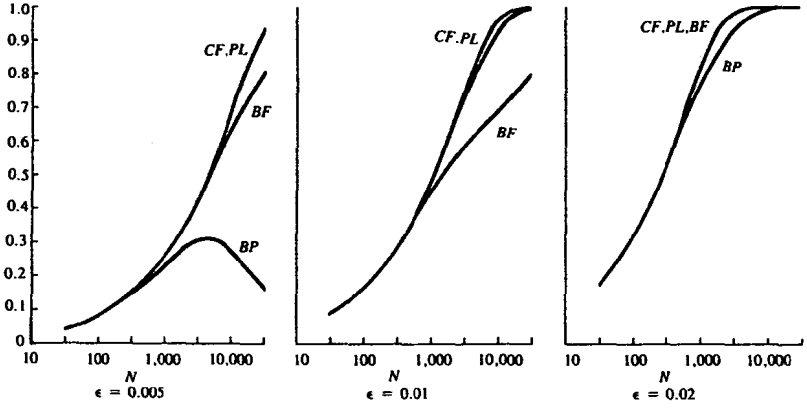


Fig. 7—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.3$, $r = 0.3$, Distribution C.

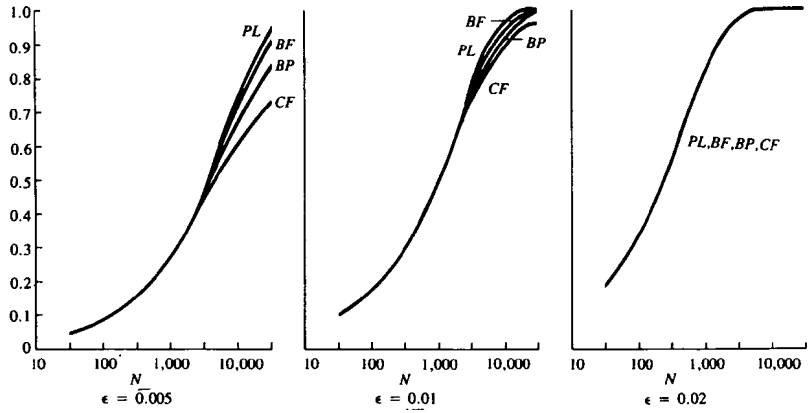


Fig. 8—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.3$, $r = 0.1$, Distribution A.

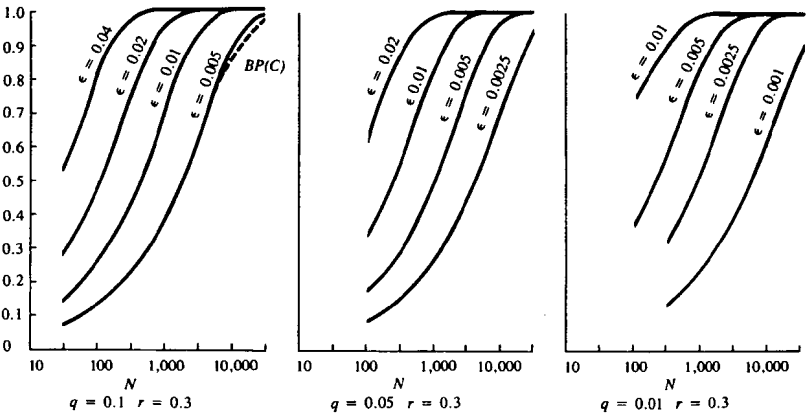


Fig. 9—$P[|\hat{q}-q| \leq \epsilon]$ versus N, Distributions A and C.

TABLE 4

COMPARISONS AMONG BP, BF, CF, PL FOR $a \equiv 0$, $b \equiv 1$

q	r	Distribution	Comparisons
0.5	0.3	A	$BP < CF < BF < PL$
0.3	0.6	A	$BP < BF < CF < PL$
0.3	0.3	A	$BP < BF = CF < PL$
0.3	0.3	C	$BP < BF < PL = CF$
0.3	0.1	A	$CF < BP < BF < PL$
0.1	0.3	A,C	$BP = BF = CF = PL$
0.05	0.3	A,C	
0.01	0.3	A,C	

of N, PL is by far the best estimator when the distributions are A, and essentially as good as CF when the distributions are C. (Clearly, for large enough N, a consistent estimator will always be better than an inconsistent one.)

Using $\hat{q}_1 < \hat{q}_2$ to denote that $\hat{q}_2$ is preferred to $\hat{q}_1$, table 4 summarizes the comparisons displayed in figures 4 through 9.

To complete the analysis, a few comparisons were made in the $a \not\equiv 0$, $b \not\equiv 1$ case. The calculations of $P[|\hat{q}-q| \leq \epsilon]$ were based on Monte Carlo experiments with uniform distributions of death and withdrawal, $(q,r) = (0.3,0.3)$, $(0.1,0.3)$, and $N = 10,32,100,316$, and $1,000$.

Values of a and b were generated at random as follows: specify probabilities α and β, and generate X_1,X_2, independent and identically distributed uniform random variables on the interval (0,1). Then let

$$\begin{aligned} a &= 0 && \text{with probability } \alpha \\ &= \min(X_1,X_2) && \text{with probability } 1\text{-}\alpha \\ b &= 1 && \text{with probability } \beta \\ &= \max(X_1,X_2) && \text{with probability } 1\text{-}\beta. \end{aligned}$$

After computing a, we must generate Y_a and Z_a. Assumption A implies

$$\begin{aligned} P(Y_a \leq t) &= \frac{tq}{1-aq} && , 0 \leq t \leq 1-a \\ &= \text{unspecified} && , t > 1-a \end{aligned}$$

and

$$\begin{aligned} P(Z_a \leq t) &= \frac{tr}{1-ar} && , 0 \leq t \leq 1-a \\ &= \text{unspecified} && , t > 1-a. \end{aligned}$$

A convenient way to compute Y_a is

$$Y_a = I[U_1 \leq \frac{(1-a)q}{1-aq}] \cdot U_2 + I[U_1 > \frac{(1-a)q}{1-aq}] \cdot 2,$$

where $I[\cdot]$ is the indicator function, and U_1 and U_2 are independent random variables; U_1 is uniform on (0,1), and U_2 is uniform on $(0,1-a)$. That is, $Y_a = U_2$ with probability $\frac{(1-a)q}{1-aq}$, and $Y_a = 2$ (any value greater than $1-a$ will do) with probability $1 - \frac{(1-a)q}{1-aq}$. Z_a may be generated in an analogous manner.

For each combination of q, r, N, α, and β, 1,000 samples were generated, and the estimators were computed. The resulting empirical distributions were used to calculate $P[|\hat{q}-q| \leq \epsilon]$. Figures 10, 11, and 12 display the results. Notice that figure 12 does not contain values for *PL*, since $\alpha = \beta = 0$ implies $a > 0$, $b < 1$, and, consequently, *PL* does not exist.

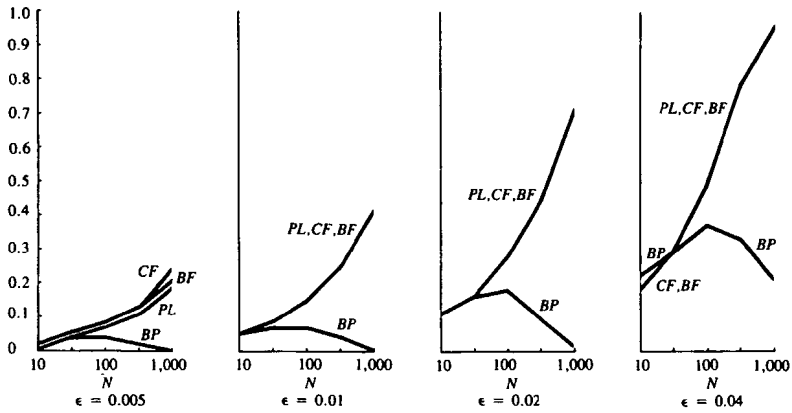


Fig. 10—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.3$, $r = 0.3$, $\alpha = \beta = 0.5$, Distribution A.

APPENDIX A
SUMMARY OF MAXIMUM LIKELIHOOD ESTIMATORS

MODEL/DATA	DISTRIBUTION	MAXIMUM LIKELIHOOD ESTIMATORS
RANDOM/ PARTIAL	A	$\hat{q}$ implicitly defined If $a_i \equiv 0$, $b_i \equiv 1$, $\hat{q} = \dfrac{(2N+D-W) - \sqrt{(2N+D-W)^2-8ND}}{2N}$
	C	$\hat{q}$ implicitly defined If $a_i \equiv 0$, $b \equiv 1$, $\hat{q} = 1 - \left(\dfrac{S}{N}\right)^{\frac{D}{D+W}}$ If $a_i \equiv 0$, $b_i \equiv 1$ for N_1 lives, $b_i \equiv \dfrac{1}{2}$ for N_2 lives, where $\hat{q} = 1 - v^{\frac{D}{D+W}}$ $v = \left[\dfrac{-(D_{\cdot 2}+W_{\cdot 2}) + \sqrt{(D_{\cdot 2}+W_{\cdot 2})^2+4(2N_1+N_2)(2S_{\cdot 1}+S_{\cdot 2})}}{2(2N_1+N_2)}\right]^2$
RANDOM OR FIXED/ FULL	A*	$\hat{q}$ implicitly defined If $a_i \equiv 0$, $b_i \equiv 1$, $T_i = \dfrac{1}{2}$ for W/D's, $\hat{q} = \dfrac{(2N+D-W)-\sqrt{(2N+D-W)^2-8ND}}{2N}$ If $a_i \equiv 0$, $b_i \equiv 1$, $T_i = 0$ for half the W/D's and $T_i = 1$ for half the W/D's, $\hat{q} = \dfrac{D}{N - \frac{1}{2}W}$
	C*	$\hat{q} = 1 - e^{-D/\sum_1^N (T_i - a_i)}$
FIXED/ PARTIAL	A	$\hat{q}$ implicitly defined If $a_i \equiv 0$, $b_i \equiv 1$, $c_i = \dfrac{1}{2}$ for W/D's, $\hat{q} = \dfrac{(2N+D-W)-\sqrt{(2N+D-W)^2-8ND}}{2N}$
	C	$\hat{q}$ implicitly defined If $a_i \equiv 0$, $b_i \equiv 1$, $\hat{q} = \dfrac{D}{N-W+\sum_{\mathcal{W}} c_i}$ (this is the $a \equiv 0$, $b \equiv 1$ actuarial estimator) If $a_i \equiv 0$, $b_i \equiv 1$, $\sum_{\mathcal{W}} c_i = \dfrac{1}{2}W$, $\hat{q} = \dfrac{D}{N-\frac{1}{2}W}$.

*No distribution assumption is needed on time to withdrawal.

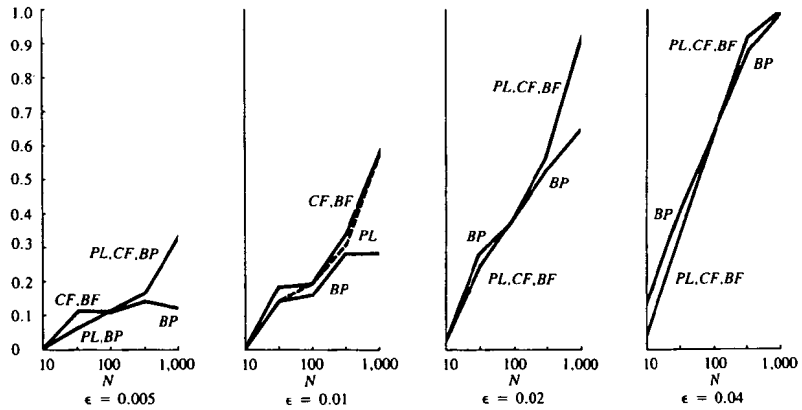


Fig. 11—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.1$, $r = 0.3$, $\alpha = \beta = 0.5$, Distribution A.

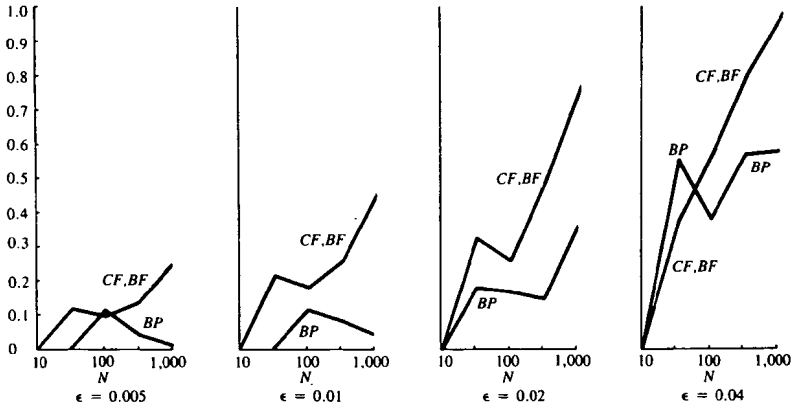


Fig. 12—$P[|\hat{q}-q| \leq \epsilon]$ versus N, $q = 0.1$, $r = 0.3$, $\alpha = \beta = 0$, Distribution A.

We now notice a substantial difference in the estimators for $N \leq 1{,}000$, especially in figures 10 and 12. *BF*, *CF*, and *PL* provide similar results and are superior to *BP*. The poor performance of *BP* is to be expected, since it is not designed for the $a \neq 0$, $b \neq 1$ case.

APPENDIX B

This appendix contains developments of the asymptotic distributions of *BP*, *BF*, *CF*, and *PL*.

For *PL* the asymptotic distribution is a special case of theorem 5 in Breslow and Crowley [4], and is given by

$$\sqrt{N}\,(PL - m_{PL}) \xrightarrow{D} N(0, \tau_{PL}^2) \tag{B.1}$$

where

$$\begin{aligned} m_{PL} &= q \\ \tau_{PL}^2 &= (1-q)^2 \int_0^1 \frac{f(t)}{[1-F(t)]^2[1-H(t)]}\,dt. \end{aligned} \tag{B.2}$$

The notation in (B.1) indicates that the distribution function of $\sqrt{N}\,(PL - m_{PL})$ converges to the distribution function of a normal random variable with mean 0 and variance τ_{PL}^2, as $N \to \infty$.

Notice that the other three estimators may be written in terms of averages. That is, for $a \equiv 0$, $b \equiv 1$,

$$BP = \frac{\overline{D}}{1 - \frac{1}{2}\overline{W}}, \tag{B.3}$$

$$BF = \frac{\overline{D}}{1 - \overline{V}}, \tag{B.4}$$

and

$$CF = 1 - e^{-\overline{D}/\overline{T}}, \tag{B.5}$$

where

$$\overline{D} = \sum_1^N D_i/N,\ \overline{W} = \sum_1^N W_i/N,\ \overline{T} = \sum_1^N T_i/N,$$

$$V_i = (1 - Z_i)W_i, \text{ and } \overline{V} = \sum_1^N V_i/N.$$

We may then take advantage of the following theorem to determine their asymptotic distributions.

THEOREM: *Let* $\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}, \cdots, \begin{bmatrix} X_N \\ Y_N \end{bmatrix}$ *be independent and identically distributed bivariate random variables with mean* $E\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix}$ *and dispersion matrix* $D\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$. *Then*

$$\text{(i) } \sqrt{N}\left(\begin{bmatrix} \overline{X} \\ \overline{Y} \end{bmatrix} - \begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix}\right) \xrightarrow{D} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}\right),$$

$$\text{(ii) } \sqrt{N}\,[g(\overline{X},\overline{Y}) - g(\theta_x,\theta_y)] \xrightarrow{D} N(0,\tau^2),$$

where

$$\tau^2 = g_1^2\sigma_x^2 + g_2^2\sigma_y^2 + 2g_1g_2\sigma_{xy},$$

$$g_1 = \frac{\partial}{\partial\theta_x}g(\theta_x,\theta_y)\ ,\ g_2 = \frac{\partial}{\partial\theta_y}g(\theta_x,\theta_y).$$

Result (i) is simply the bivariate form of the usual central limit theorem. Result (ii) tells us how to get the asymptotic distribution of a function of two sample means.

Applying result (ii) we obtain the asymptotic distributions of (B.3), (B.4) and (B.5). Let $\theta_D = E(D_i)$, $\theta_W = E(W_i)$, $\theta_V = E(V_i)$, $\theta_T = E(T_i)$, $\sigma_V^2 = \text{Var}(V_i)$, $\sigma_T^2 = \text{Var}\ (T_i)$, and $\sigma_{TD} = \text{Cov}(T_i,D_i)$. Then

$$\sqrt{N}(BP - m_{BP}) \xrightarrow{D} N(0,\ \tau_{BP}^2),$$

$$\sqrt{N}(BF - m_{BF}) \xrightarrow{D} N(0,\ \tau_{BF}^2),$$

and

$$\sqrt{N}(CF - m_{CF}) \xrightarrow{D} N(0,\ \tau_{CF}^2),$$

where

$$m_{BP} = \theta_D/(1-\frac{1}{2}\theta_W),$$

$$m_{BF} = \theta_D/(1-\theta_V),$$

$$m_{CF} = 1-e^{-\theta_D/\theta_T},$$

$$\tau^2_{BP} = \frac{\theta_D(1-\theta_D)}{(1-\frac{1}{2}\theta_W)^2} + \frac{\theta_D^2\theta_W(1-\theta_W)}{4(1-\frac{1}{2}\theta_W)^4} - \frac{\theta_D^2\theta_W}{(1-\frac{1}{2}\theta_W)^3},$$

$$\tau^2_{BF} = \frac{\theta_D(1-\theta_D)}{(1-\theta_V)^2} + \frac{\sigma_V^2\theta_D^2}{(1-\theta_V)^4} - \frac{2\theta_D^2\theta_V}{(1-\theta_V)^3},$$

and

$$\tau^2_{CF} = e^{-2\theta_D/\theta_T}\left[\frac{\theta_D(1-\theta_D)}{\theta_T^2} + \frac{\theta_D^2\sigma_T^2}{\theta_T^4} - \frac{2\sigma_{DT}\theta_D}{\theta_T^3}\right].$$

For extreme values of q, the distributions of these estimators will be highly skewed unless N is very large. Hence, the normal approximation should be used with caution.

The following formulas provide the means for computing m and τ^2 for *BP*, *BF*, and *CF*. For *PL*, $m_{PL} = q$ and τ^2_{PL} is given in (B.2).

$$E(D_i) = \int_0^1 [1 - H(t)]f(t)dt \tag{B.7}$$

$$E(W_i) = \int_0^1 [1 - F(t)]h(t)dt \tag{B.8}$$

$$E(V_i^k) = \int_0^1 (1-t)^k[1 - F(t)]h(t)dt \tag{B.9}$$

$$E(T_i^k) = k\int_0^1 t^{k-1}[1 - F(t)][1 - H(t)]dt \tag{B.10}$$

$$E(T_iD_i) = \int_0^1 t[1 - H(t)]f(t)dt. \tag{B.11}$$

Under A and C the values of (B.7)-(B.11) and (B.2) are given below.

Distribution Assumption A:

$$E(D_i) = q - \frac{1}{2}rq$$

$$E(W_i) = r - \frac{1}{2}rq$$

$$E(V_i) = \frac{1}{2}r - \frac{1}{6}rq$$

$$E(V_i^2) = \frac{1}{3}r - \frac{1}{12}rq$$

$$E(T_i) = 1 - \frac{1}{2}r - \frac{1}{2}q + \frac{1}{3}rq$$

$$E(T_i^2) = 1 - \frac{2}{3}r - \frac{2}{3}q + \frac{1}{2}rq$$

$$E(T_iD_i) = \frac{1}{2}q - \frac{1}{3}rq$$

$$\tau_{PL}^2 = q(1-\frac{1}{2}q) \qquad \text{, if } q = r$$

$$= \frac{(1-q)^2\, qr}{(q-r)^2} \ln\left(\frac{1-q}{1-r}\right) + \frac{q^2\,(1-q)}{q-r} \qquad \text{, if } q \neq r.$$

Distribution Assumption C:

$$E(D_i) = \frac{\mu}{\mu + \upsilon}\left[1 - e^{-(\mu+\upsilon)}\right]$$

$$E(W_i) = \frac{\upsilon}{\mu} E(D_i)$$

$$E(V_i) = \frac{\upsilon}{(\mu+\upsilon)^2}\left[\mu+\upsilon-1+e^{-(\mu+\upsilon)}\right]$$

$$E(V_i^2) = \frac{\upsilon - 2E(V_i)}{\mu+\upsilon}$$

$$E(T_i) = \frac{1}{\mu} E(D_i)$$

$$E(T_i^2) = \frac{2}{(\mu+\upsilon)}\left[1 - e^{-(\mu+\upsilon)} - (\mu+\upsilon)e^{-(\mu+\upsilon)}\right]$$

$$E(T_iD_i) = \frac{\mu}{2} E(T_i^2)$$

$$\tau_{PL}^2 = \frac{\mu}{\mu+\upsilon}(e^{\mu+\upsilon}-1)e^{-2\mu}.$$

REFERENCES

1. BATTEN, ROBERT W. *Mortality Table Construction*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1978.
2. BÖHMER, P.E. "Theorie der Unabhängigen Wahrscheinlichkeiten," *Rapports, Mémories et Procès-verbaux de Septième Congrès International d'Actuaries*, Amsterdam, II, 327-43, 1912.
3. BROFFITT, J.D., AND KLUGMAN, S.K. "Method of Moments Derivation of the Balducci Exposure Formula," *ARCH*, Society of Actuaries, 8-9, 1983.1
4. BRESLOW, N., AND CROWLEY, J. "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," *The Annals of Statistics*, II, 437-53, 1974.
5. CHIANG, CHIN LONG. "A Stochastic Study of the Life Table and Its Applications: III. The Follow-up Study with the Consideration of Competing Risks," *Biometrics*, XVII, 57-78, 1961.
6. CHIANG, CHIN LONG. *Introduction to Stochastic Processes in Biostatistics*, New York, NY: John Wiley & Sons, Inc., 1968.
7. ELANDT-JOHNSON, R. C., AND JOHNSON, N.L. *Survival Models and Data Analysis*, New York, NY: John Wiley & Sons, Inc., 1980.
8. ELVEBACK, LILA. "Estimation of Survivorship in Chronic Disease: The 'Actuarial' Method," *Journal of the American Statistical Association*, LIII, 420-40, 1958.
9. JORDAN, CHESTER W. JR. *Life Contingencies*, Chicago, IL: Society of Actuaries, 1975.
10. KAPLAN, E.L., AND MEIER, PAUL. "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, LIII, 457-81, 1958.
11. LINDLEY, D.V. "Analysis of Life Tables with Grouping and Withdrawals," *Biometrics*, XXXV, 605-612, 1979.
12. MILLER, R.B., AND HICKMAN, J.C. "A Comparison of Several Life Table Estimation Methods," Technical Report No. 627, Department of Statistics, University of Wisconsin, Madison, WI, 1980.
13. STEELMAN, JAMES M. "Statistical Approaches to Mortality Estimation," Master of Science Thesis, Department of Statistics, University of Iowa, Iowa City, IA, 1968.

DISCUSSION OF PRECEDING PAPER

RALPH E. EDWARDS:

In this paper, we have an exposition of the theory of mortality measurement, presented with reasonable brevity and expounded so that those with outdated mathematics can (with the use of some imagination as to the meaning of unfamiliar terminology) review the modern approach. Many practicing actuaries would benefit from studying this paper to understand the path by which future students are likely to be educated.

I mentioned unfamiliar terminology. Professor Broffitt specifies "mild" and "asymptotically efficient" conditions, but "in most instances" (even if wrong) may be adequate for the nonstudent reader. I have more trouble, however, with maximum likelihood estimators (MLE). In elementary statistics, the variance is smallest when measured from the average or mean. There is no reason for the mean not to be derived from the variance by minimization, but if MLE really is the mean, perhaps it should be so stated.

My out-of-date mathematics encountered particular difficulty with the development from equation (2.1) to the equation following (2.3). The latter derives from differentiation of (2.3) with respect to q, which provides $d(1-q) - (N-d)q$ as the quantity to be set equal to zero. This result was interesting because the Balducci Hypothesis was not used in that derivation. The MLE equaling D/N and not involving q seems to mean that q is a theoretical concept associated with the deaths provided that $1-q$ be associated with the survivors. It would seem that Balducci ought to be applicable, since it assumes that $p = 1-q$.

This notion about equation (2.3) led to the observation that N is supposed to be large, so that $L(q)$ has a very small numeric value. This seems anomalous where the associated name implies likelihood. To understand this, I considered the expansion of $k(p+q)^m$ with k large. This is a representation of k samples of m items each. We are concerned with just one of the k samples, and this is represented by one of the terms of the expansion which contains $p^v \cdot q^{m-v}$ as a factor. The multiplier of this factor (in the expansion) is of no significance except as it relates this sample to the other $k-1$ samples. By stopping at this point, we have an explanation of the small numeric value, since we are accustomed to the restriction that $p + q = 1$.

Tracing back from (2.3) to (2.2) led to the paper's preceding material. An arithmetic approach suggested letting $q = u = .01$ and $dt = 1/10{,}000$. The contribution to the likelihood function by a life that survives is .99, but

.00000099 for a life that dies. In examining the survivors, or rather, one of them, let the survivors of l_x be measured over successive small intervals of time, these being Δt. The successive values are

$$\frac{l_{x+\Delta t}}{l_x}, \frac{l_{x+2\Delta t} - l_{x+\Delta t}}{l_x}, \frac{l_{x+3\Delta t} - l_{x+2\Delta t}}{l_x},$$

and so on. If we sum and substitute integration, the result is $\int_0^1 (1-q)\,dt$. I did not work out a similar progression for a single death and suspect it can only be obtained for a number of them, but analogy suggests that the result in the paper requires an integration from zero to one. It also seems necessary that the suggested removal of differential elements not occur. I am not sure what this does for the left side of (2.2), but the right side becomes

$$\left[\int_0^1 (1 - q_x)\,dt\right]^{N-d} \prod_{\mathfrak{D}} \int_0^1 (1 - {}_tq_x)\,u_{x+t}\,dt\,.$$

In this result, the first integral has the value of $1-q_x$. The foregoing is purely conjecture, but this version of (2.2) yields (2.3) and also does so for the Balducci Hypothesis.

One of the fundamental relationships relating to this paper is the equation $D = N{\cdot}q - W{\cdot}k$. D is the number of observed deaths, N the number of lives at the beginning of the observation year, W the withdrawals during the observation year, and k an unknown such that $W{\cdot}k$ is the number of deaths in the observation year among those who first withdrew. (More precisely the number desired in $W{\cdot}k$ is based on the assumption that the incidence of mortality is not altered by withdrawal.)

The unknowns in the equation are $W{\cdot}k$ (and not just k) and the value sought is q. Mathematically, a single equation with two unknowns cannot be solved. Additional information about the D and W components enables the Product Limit Estimator method to be used. A solution which begs the question requires the investigator to obtain $W{\cdot}k$. One I do not recall seeing in print could exist if the death and withdrawal rates were related, as, for example, two withdrawals for each concurrent death.

Investigators have assumed that this is as far as theory can go. They turned to see what happens if the withdrawals are concentrated at a particular time during the observation year. Even then, as Gershenson[1] says "it is evident that some assumption needs to be introduced if we are to determine q_x." But if an assumption needs to be introduced when withdrawals are concentrated, it is obvious that a solution requires assumptions to be made when withdrawals occur at a few or many different times. The fact that a contin-

[1] H. Gershenson. *Measurement of Mortality*. (Chicago: Society of Actuaries, 1961).

uous function might apply to withdrawals is, then, obviously of no help except where the incidence of death and withdrawal violates the assumption of independence.

The reason for bringing the foregoing out in great detail is that the paper attributes the unidentifiability of q to a 1979 publication. Admittedly the unidentifiability of q relates to two equations with three unknowns, but this can hardly be significant when they are derived from one equation with two unknowns. The unidentifiability of q, except that it was called indeterminance, was known and discussed when I was a student several decades ago.

In view of Professor Broffitt's giving credit to Lindley, it seems appropriate to note Professor Broffitt's use of A and C to identify two of his required assumptions, taking his source as Bratten's text. In *ARCH* 1978.2, which dates after Bratten, Dr. Thomas N. E. Greville attributes C to Mr. Frank Weck, a current Fellow of our Society. Bratten suggests that A is an offshoot of de Moivre's hypothesis that d_x is constant for all values of x. However, Dr. Greville (*TASA* XLIV, 33–35) more recently attributes the origin of this not to de Moivre, but to a particular characteristic differentiating it from Balducci.

It is traditional that the denominator of the observed mortality rate (where D is the numerator) be a function called the exposure. An obvious modification of Professor Broffitt's formula (2.9) seems to provide N as the exposure. However, the formula immediately preceding (2.9) may be solved for D/q, giving an exposure of $N - \frac{1}{2}W(1/(1 - \frac{1}{2}q))$. Professor Broffitt states that formula (2.8), from which (2.9) is derived, can be solved by iteration. A few years ago, I wrote a piece dealing with the solution of (2.8) printed in *ARCH* 1983.1. In that article I suggest an approximation such that the need for iteration can be avoided.

Even if there is merit to some of my conjectures, this paper is a valuable contribution to our literature. To assemble and prepare so much material must have been a long and arduous task. We are greatly in Professor Broffitt's debt.

JAMES M. ROBINSON:

ABSTRACT

My compliments go to Mr. Broffitt for his timely introduction of maximum likelihood estimators of mortality rates to the actuarial community. Other professional disciplines, such as biostatistics, have been concerned with the development of statistical inference techniques aimed at extracting survival probabilities from past data. Mr. Broffitt's paper goes a long way toward establishing constructive communication lines to tap this external wealth of ideas and techniques.

This discussion expands upon the implications of possible correlation between the times until death and lapsation for an insured population. Speculation about the structure of the mortality and lapsation processes and the impact of certain external influences leads to a possible model of the joint process in which these two components are not independent. Problems of model identifiability are reviewed. The mortality estimators developed in the preceding paper under the independence assumption are analyzed assuming the proposed dependent model holds. Finally, pricing decision implications are discussed in light of the previous findings.

ANTISELECTIVE LAPSATION

In the introduction to his paper, Mr. Broffitt warns against careless use of the independence assumption. The usual form of experience data available from insurance company records does not provide a basis for measuring the degree of dependence between the mortality and lapsation processes. Lacking testable data, we are left with intuition and common sense to assess this problem and properly adjust the business decisions required to manage the mortality risks of insurance companies. The last section of this discussion will provide an example of the potential penalties of avoiding this analysis.

It is reasonable to assume that the participation or nonparticipation of a given individual in an insurance contract has little or no effect upon the mortality process. That is, there is no direct connection between the rate at which one physically deteriorates and the status of one's insurance policy. It is much more likely that an insured's physical condition and level of exposure to the risk of death have a significant effect upon the possibility of policy lapsation, however. Healthy individuals should be more vulnerable to the various forces acting to encourage lapsation than would those in poor health. Likewise, insureds exposed to lower levels of risk of accidental death are logically more mobile than those at higher levels. Healthy individuals in nonhazardous occupations should hold a lower perceived value for their insurance coverage relative to an insured with a higher probability of death, assuming both are paying the same premium. Also, the low-risk insured should be able to satisfy underwriting requirements for alternative coverage elsewhere. The high-risk insured is not likely to be able to find replacement coverage at a favorable premium rate. So, the main ingredients for antiselective lapsation have been formulated—a group of insureds paying the same premium rates divided between those who perceive themselves to be low risk and those with a less favorable view of their life expectancy. The attrition in the former subgroup will be greater than that of the latter, and the population of persisting insureds will become more heavily skewed toward the high-risk, uninsurable category.

Most actuaries have considered this problem subjectively and many have adjusted pricing assumptions in some manner to account for this potential adverse mortality if unusually high pressure for lapsation is anticipated. Quantifying this adjustment has been difficult lacking usable past experience from which one might study the degree of antiselection.

This problem most likely will become more prevalent. Recent innovations in policy structure encourage healthy individuals to lapse and seek new coverage at lower rates. In particular, select-and-ultimate premium structures present the actuary with a spiral of difficult pricing assumptions. The graded premium levels enhance the pressure to lapse in favor of low, first-year rates, which in turn increases anticipated mortality, which requires ever steeper premium levels, which further fuel lapsation, and the cycle continues. Only those unable to meet underwriting standards for new coverage will be able to resist the temptation to roll over their coverage. Furthermore, traditional product pricing will have to be adjusted to anticipate higher lapsation as a result of the availability of those low-premium alternatives.

Rate classification considerations cause increases in the possibility of antiselective lapsation. The increasing use of smoking status is a good example. The creation of policies with favorable nonsmoking premium rates should create pressure for nonsmokers to lapse in favor of these tailored rates. Smokers will be satisfied with the status quo, paying rates based on mortality assumptions appropriate only for a mix of smokers and nonsmokers. Again, we have a situation in which the lapsation process is tied directly to perceived mortality.

These situations arise because of a lack of homogeneity in insured groups. In the select-and-ultimate case, this dichotomy develops over time as previously healthy lives migrate into the high-risk category. In the smoker-nonsmoker case, the split arises when a new rate classification parameter is introduced and competitors make lower rates available to the new low-risk subgroup. This partition is the basis for the following model for the joint mortality/lapsation process.

DEPENDENT MODEL

As in Mr. Broffitt's paper, we define Y as the time until death and Z as the time until lapsation of an insured persisting to attained age X (and policy duration N). Since we may not observe the time of death once lapsation has occurred, and since we are only interested in the age interval, $[X,X+1]$, we observe the random variable $T = \min(Y,Z,1)$ and an associated indicator of the cause of failure, death, or lapsation. At this point, most analyses assume that Y and Z are independent, proceed to construct the resulting distribution for T, and estimate the process parameters from the sample

values of T. The following dependent joint distribution for Y and Z is presented as a plausible alternative:

1. Assume that mortality may be well represented by a constant hazard rate (i.e. force of mortality) over any unit age interval.
2. Assume that lapsation arises from exposure to some constant censoring force throughout the year. This is a questionable provision, since most lapsation occurs at premium payment epochs and on policy anniversaries. So, unless premiums are paid frequently, lapse pressure will be uneven over the policy year. However, for exemplary purposes this structure should suffice.
3. Given the force of mortality and lapse intensity, the random variables Y and Z are independent.
4. Within a cohort of insured lives aged X, there is a distribution of forces of mortality and lapse intensities. Furthermore, these parameters are negatively correlated. That is, individuals exposed to high mortality are subject to lower pressures to lapse. Rather than adopting a continuous joint density, however, the cohort is split into two subgroups. We might consider one group smokers and the other nonsmokers or one group select and the other nonselect. Refinements are certainly justified, but for simplicity of calculation and presentation, we limit our partition to two categories.
5. Assume that the overall lapse intensity is a funtion of external factors, including the availability of lower-premium alternative coverages. Variations in these environmental factors are not spread evenly across the cohort of insureds, however. High-risk individuals may even be totally indifferent to such factors if their chances of satisfying competitor underwriting requirements are small. Thus, it may be reasonable to assume that any fluctuation in external lapse forces is concentrated on the low-risk group.

We may formulate these assumptions using the following notation:

u_1 is the force of mortality of the low-risk group.
u_2 is the force of mortality of the high-risk group.
v_1 is the force of lapsation of the low-risk group.
v_2 is the force of lapsation of the high-risk group.
p is the fraction of the cohort in the low-risk group.
$f(y/u)$ is the density of Y, conditional upon the force of mortality, u.
$f(z/v)$ is the density of Z, conditional upon the force of lapsation, v.

Assumption 1 implies $f(y/u) = ue^{-uy}$, $y \geq 0$, and assumption 2 implies $f(z/v) = ve^{-vz}$, $z \geq 0$. So, the unconditional joint density of Y and Z is $f(y,z) =$

pu_1, v_1, $e^{-u_1 y}\, e^{-v_1 z}$ + $(1-p)u_2 v_2 e^{-u_2 y} e^{-v_2 z}$ for $y \geq 0$ and $z \geq 0$. With this density, we may now calculate the probabilities of various observed events.

1. $P\{T = 1\} = P\{y \geq 1, Z \geq 1\} = pe^{-(u_1+v_1)} + (1-p)e^{-(u_2+v_2)}$.

2. $P\{T = t, T = Y\} = P\{Y = t, Z > t\}$

$$= pu_1 e^{-(u_1+v_1)t} + (1-p)u_2 e^{-(u_2+v_2)t},\ 0 \leq t < 1.$$

3. $P\{T = t, T = Z\} = P\{Z = t, Y > t\}$

$$= pv_1 e^{-(u_1+v_1)t} + (1-p)v_2 e^{-(u_2+v_2)t},\ 0 \leq t < 1.$$

Probability 1 corresponds to observing an insured persist throughout the year, probability 2 to observing a death at time t, and probability 3 to observing a lapse at time t. From this, we can construct the likelihood function for a given sample.

$$L(u_1,u_2,v_1,v_2,p) = \{pe^{-(u_1+v_1)} + (1-p)e^{-(u_2+v_2)}\}^{n_s} \times \prod_{i=1}^{n_f} \{pu_1 e^{-(u_1+v_1)y_i} + (1-p)u_2 e^{-(u_2+v_2)y_i}\} \times \prod_{i=1}^{n_c} \{pr_1 e^{-(u_1+v_1)z_i} + (1-p)v_2 e^{-(u_2+v_2)z_i}\}$$

where n_s is the number surviving the year,
n_f is the number dying during the year,
n_c is the number lapsing during the year,
y_i is the time of the ith observed death, and
z_i is the time of the ith observed lapse.

We may now attempt to find maximum likelihood estimates of the unknown parameters by maximizing L over feasible choices of u_1, u_2, v_1, v_2, and p. This appears to be analytically intractible, and some iterative technique is probably the best approach. Once we have these estimates, we can analyze the dependence between Y and Z, say, by calculating the covariance or correlation coefficient for the model.

$\mathrm{Cov}(Y,Z) = p(1-p)(u_2-u_1)(v_2-v_1)/u_1u_2v_1v_2$. If $u_2 > u_1$ and $v_2 < v_1$, then we have negative correlation. The correlation coefficient is obtained by dividing by the square root of the product of the variances of Y and Z, which are given by the following.

$$\mathrm{Var}(Y) = \{(1-p)u_1^2 + pu_2^2 + p(1-p)(u_2-u_1)^2\}/u_1^2u_2^2$$

$$\mathrm{Var}(Z) = \{(1-p)v_1^2 + pv_2^2 + p(1-p)(v_2-v_1)^2\}/v_1^2v_2^2.$$

For example, if we estimate $\hat{p} = .5$, $\hat{u}_1 = .01$, $\hat{u}_2 = .10$, $\hat{v}_1 = .30$, and

$\hat{v}_2 = .05$, then we might estimate $\hat{\text{cov}}(Y,Z) = -375$, $\text{var}(Y) = 7075$, $\hat{\text{var}}(Z) = 275$, and $\hat{\text{corr}}(Y,Z) = -.27$.

The model parameters can be estimated and used to forecast mortality experience for future cohorts of insureds, with the effects of antiselective lapsation properly included if lapse rates are accurately predicted. However, this is acceptable only if we are certain that the form of the model is correct.

IDENTIFIABILITY

If the estimated parameters are only meaningful in the context of the true underlying model, how may we test the validity of the proposed model? Given that we always observe the minimum of the time until death and the time until lapsation, there is no way to judge whether the assumed form of the process better explains the observed than some other model. In fact, it can be shown (appendix A) that any likelihood function obtained from a dependent joint distribution for Y and Z can always be obtained from a different joint distribution in which Y and Z are independent. That is, there is an independent model, which will fit the data as well as the proposed model. For the previous model, the independent counterpart is given by the following.

$$f(y) = u(y) \exp\{-\int_0^y u(t)dt\}, \text{ where}$$

$$u(t) = \frac{pu_1e^{-(u_1+v_1)t} + (1-p)u_2e^{-(u_2+v_2)t}}{pe^{-(u_1+v_1)t} + (1-p)e^{-(u_2+v_2)t}}$$

$$f(z) = v(z) \exp\{-\int_0^z v(t)dt\}, \text{ where}$$

$$v(t) = \frac{pv_1e^{-(u_1+v_1)t} + (1-p)v_2e^{-(u_2+v_2)t}}{pe^{-(u_1+v_1)t} + (1-p)e^{-(u_2+v_2)t}}$$

$$\text{and } f(y,z) = f(y)f(z),\ y \geq 0 \text{ and } z \geq 0.$$

The likelihood function for this model is the same as that for the dependent model. So, if we estimate the parameters of this new model, they will be the same as those of the old model. The interpretation of these parameter estimates and statistics calculated from them, e.g. the "covariance" and "correlation coefficient," are entirely different. We consider the expression $\hat{p}(1-\hat{p})(\hat{u}_2-\hat{u}_1)(\hat{v}_2-\hat{v}_1)/\hat{u}_1\hat{u}_2\hat{v}_1\hat{v}_2$ to be an estimator for the covariance of Y and Z only if the original model is correct. Within the context of the independent model, this odd collection of terms is virtually meaningless. The sample covariance in the latter case will always be zero by design.

All of this indicates that the data will not tell us which model is correct. The data will only give us the best parameterization within a preselected family of distributions for Y and Z. If the family is too broad, this parameterization is not likely to be unique. If we select a very small family of distributions, the fit may be unique, but there will be no empirical support for the selection. In short, if we don't observe the dependent portion of the process, then we can only speculate on its structure. This is exactly what has been done to create the dependent model being used herein. The data will neither support nor dispute the model's accuracy relative to other models with the same likelihood function.

For the remainder of this discussion, we assume that the dependent model is accurate and consider the implications for estimation of mortality rates and their subsequent application in pricing insurance contracts.

PERFORMANCE OF MAXIMUM LIKELIHOOD AND ACTUARIAL ESTIMATORS BASED ON INDEPENDENCE ASSUMPTIONS WHEN DEPENDENT MODEL HOLDS

Several estimators of the mortality rate in the absence of lapsation were discussed by Mr. Broffitt. Two of these will be discussed here.

1. *CF* - The maximum likelihood estimate of q_x given full data, i.e. times of failure, and based upon a model assuming independent exponential distributions for Y and Z, i.e. constant forces of mortality and lapsation.
   Recall that $CF = 1-e^{-\overline{D}/\overline{T}}$, where $\overline{D} = n_f/n$, the fraction of individuals observed to die, and $\overline{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$, the average exposure per insured.
2. *BF* - The traditional actuarial estimator of q_x given full data and a Balducci distribution for deaths within a year.
   Again recall $BF = \frac{\overline{D}}{1-\overline{V}}$, where $\overline{V} = \frac{1}{n}\sum_{i=1}^{n} V_i$ and V_i is $1-Z_i$ if $T_i = Z_i$ and V_i is zero otherwise.

The asymptotic distributions of *CF* and *BF* are calculated using Mr. Broffitt's approach as presented in appendix B of his paper. Recall that:

$$\sqrt{N}\,(CF - m_{CF}) \xrightarrow{D} N(0, \tau^2_{CF}) \text{ as } N \to \infty,$$

and

$$\sqrt{N}\,(BF - m_{BF}) \xrightarrow{D} N(0, \tau_{BF}^2) \text{ as } N \to \infty$$

where

$$m_{CF} = 1 - e^{-\theta_D/\theta_T};$$

$$m_{BF} = \theta_D/(1-\theta_V);$$

$$\tau_{CF}^2 = e^{-2\theta_D/\theta_T}\left\{\frac{\theta_D\,(1-\theta_D)}{\theta_T^2} + \frac{\theta_D^2\,\sigma_T^2}{\theta_T^4} - \frac{2\sigma_{DT}\,\theta_D}{\theta_T^3}\right\};$$

$$\tau_{BF}^2 = \left\{\frac{\theta_D\,(1-\theta_D)}{(1-\theta_V)^2} + \frac{\theta_D^2\,\sigma_V^2}{(1-\theta_V)^4} - \frac{2\theta_D^2\,\theta_V}{(1-\theta_V)^3}\right\};$$

$$\theta_D = E\,(D_i);$$

$$\theta_T = E\,(T_i);$$

$$\theta_V = E\,(V_i);$$

$$\sigma_T^2 = \text{Var}\,(T_i);$$

$$\sigma_{DT} = \text{Cov}\,(D_i, T_i);$$

$$\sigma_V^2 = \text{Var}\,(V_i).$$

These expected values, variances, and covariances can be calculated under the dependent model as follows.

$$E\,(D_i) = \frac{pu_1}{u_1+v_1}\,(1 - e^{-(u_1+v_1)}) + \frac{(1-p)u_2}{u_2+v_2}\,(1-e^{-(u_2+v_2)});$$

$$E\,(T_i) = \frac{p}{u_1+v_1}\,(1 - e^{-(u_1+v_1)}) + \frac{(1-p)}{u_2+v_2}\,(1-e^{-(u_2+v_2)});$$

$$E\,(V_i) = \frac{pv_1}{(u_1+v_1)^2}\,(u_1+v_1-1+e^{-(u_1+v_1)}) + \frac{(1-p)v_2}{(u_2+v_2)^2}\,(u_2+v_2-1+e^{-(u_2+v_2)});$$

$$\text{Var}(T_i) = E(T_i^2) - E(T_i)^2;$$

$$E(T_i^2) = \frac{2p}{(u_1+v_1)^2}\,(1-e^{-(u_1+v_1)} - (u_1+v_1)e^{-(u_1+v_1)}) + \frac{2(1-p)}{(u_2+v_2)^2}\,(1-e^{-(u_2+v_2)} - (u_2+v_2)e^{-(u_2+v_2)});$$

$$\mathrm{Cov}(D_i, T_i) = E(D_iT_i) - E(D_i)E(T_i);$$

$$E(D_iT_i) = \frac{pu_1}{(u_1+v_1)^2}(1-e^{-(u_1+v_1)} - (u_1+v_1)e^{-(u_1+v_1)}) + \frac{(1-p)u_2}{(u_2+v_2)^2}(1-e^{-(u_2+v_2)} - (u_2+v_2)e^{-(u_2+v_2)});$$

$$\mathrm{Var}(V_i) = E(V_i^2) - E(V_i)^2;$$

$$E(V_i^2) = p\left\{\frac{v_1}{u_1+v_1} - \frac{2v_1}{(u_1+v_1)^3}(u_1+v_1 - 1+e^{-(u_1+v_1)})\right\} + (1-p)\left\{\frac{v_2}{u_2+v_2} - \frac{2v_2}{(u_2+v_2)^3}(u_2+v_2 - 1+e^{-(u_2+v_2)})\right\}.$$

The following example is presented to contrast the asymptotic distributions of *CF* and *BF* under Mr. Broffitt's assumption *C*, constant and independent mortality and lapsation, against the dependent model. Consider the following hypothetical situations:

Case 1: Assumption *C* holds and $q=.05$ and $r=.30$, $u=.05129$ and $v=.35667$.

Case 2: The dependent model holds with $p=.5$, $u_1=.02532$, $u_2=.07796$, $v_1=.69135$, and $v_2=.10536$. These values are chosen to make $q=1-pe^{-u_1}-(1-p)e^{-u_2}=.05$ and $r=1-pe^{-v_1}-(1-p)e^{-v_2}=.30$.

In both cases, the probability of death in the absence of lapsation is .05 and the probability of lapsation in the absence of death is .30. We can find the asymptotic means and variances of *CF* and *BF* under case 1 from Mr. Broffitt's table 3. The corresponding values for case 2 are computed manually from the previous formulas. The following table summarizes the results.

Item	Case 1	Case 2
m_{CF}	.0500	.0534
τ^2_{CF}	.0564	.0604
m_{BF}	.0499	.0534
τ^2_{BF}	.0561	.0624

Not surprisingly, both estimators overstate the probability of death in the absence of lapsation, if we assume case 2 holds. The negative correlation between mortality and lapsation creates antiselective increases in mortality rates when $r=.30$, and the estimators assume this higher rate is also applicable when $r=0$.

PRICING IMPLICATIONS

Is this bias a serious problem? It is if the ultimate objective of the study is to predict mortality in the absence of policyholder lapsation. But, in practice, published mortality rates are typically melded with lapse assumptions appropriate for a product being priced. If these lapse rates are comparable to those exhibited in the mortality study, there should be little concern. If, however, the actuary anticipates heavy lapse activity, then the published mortality rates may be dangerously inadequate.

To exemplify this problem, we will initially consider pricing a one-year nonrenewable, continuous-premium, term policy, ignoring expense and profit loadings. If the pricing actuary ascribes to the independence assumption, then the following premium rate calculation might result. The actuary estimates that the average policy will remain in force six months before lapsing. This may seem drastic, but the only people buying nonrenewable coverage are likely those in need of temporary protection for a period of less than a year. If premiums are paid continuously, the individual may lapse at the instant that the protection is not longer needed. It is granted that this is not a common pricing situation, but it provides an extreme example of antiselective lapsation over one year. The actuary assumes a constant force of lapsation of 2.00, which produces an expected time until lapse of .5 years. The actuary now consults an appropriate mortality table and selects q as the mortality rate. Again, the force of mortality is assumed to be constant. The actuary calculates $u = -\ln(1-q)$. Ignoring interest, the formula for the premium rate, GP, is

$$GP \int_0^1 P\{T > t\}dt = \int_0^1 P\{T = t, T = Y\}dt$$

$$GP \int_0^1 e^{-(u+v)t}\,dt = \int_0^1 ue^{-(u+v)t}\,dt \text{ , under the independence model,}$$

$$= u \int_0^1 e^{-(u+v)t}\,dt .$$

So, $GP = u$, regardless of the projected value of v, the force of lapsation.

Let us now speculate that the underlying mortality/lapsation process, which generated the mortality table and which will govern the experience of the policy, is of the dependent form. Assume that the mortality table process arose from the case where $p = .80$, $u_1 = .01$, $u_2 = .10$, $v_1 = .25$, $v_2 = .05$. The asymptotic mean for BF is .02603 in this case. This is the value of q used by our actuary, and the premium will be calculated as $GP = .02637$. Now reconsider the integrals in the pricing formula.

$$\int_0^1 P\{T > t\}dt = E(T_i) \text{ and } \int_0^1 P\{T = t, T = Y\}\,dt = E(D_i).$$

We have previously derived formulas for these quantities in the dependent case. If we assume that only the low risk insureds will lapse at a rate $v_1 = 2.00$ and the high risk lapse at $v_2 = .05$, then $E(T_i) = .53040$ and $E(D_i) = .02202$. The appropriate premium rate is $GP = .02202/.53040 = .04151$. If the dependent model is correct, our actuary will experience a loss of 57 percent of premium revenue over the year.

This example considers only the effect of an incorrect independence assumption over a single year. If we now allow antiselection lapsation to continue from one policy year to another, the effects can be even more pronounced. We will use the same policy, but allow renewal up to the fifth anniversary. Each year, the payment rate increases to the next attained age level. In this setting, we will assume that the population of insureds is composed of select and nonselect lives. Initially, everyone is select and $p = 1.00$. We assume that a select and ultimate mortality table is used in pricing, and this table is based upon a dependent process in which 15 percent of persisting select lives migrate to the nonselect category at the end of each policy year. Assume that the mortality study uses $CF = 1 - e^{-\bar{D}/\bar{T}}$ to estimate the mortality rate for each policy year. The expected tabular mortality rate will be $m_{CF} = 1 - e^{-\theta_D/\theta_T}$. If the actuary again assumes independence, GP will be set to $-\ln(1 - m_{CF}) = \theta_D/\theta_T$.

The following table summarizes the dependent model assumptions underlying the mortality table experience.

Policy Duration	u_1	u_2	v_1	v_2	p	θ_D	θ_T	$GP = \theta_D/\theta_T$
0	.010	.100	.30	.05	1.0000	.00860	.860	.010
1	.011	.110	.20	.05	.8500	.02368	.905	.026
2	.013	.130	.15	.05	.7169	.04228	.921	.046
3	.016	.160	.10	.05	.6123	.06520	.928	.070
4	.020	.200	.10	.05	.5392	.09170	.916	.100
5	.....	.....	.....	.....	.4856	.....	.....	.....

Of particular importance in this example is the evolution of the values of p. These values represent the split between low and high risk groups among the persisting policyholders. For example, $p = .6123$ at the start of the fourth policy year. This is obtained by computing the probability that an insured persists in the select category for three years and dividing by the probability of surviving in either group for three years. The critical concern in this model is the deterioration of the p values if the select group is subject to greater pressure to lapse.

The following table indicates the same data as the previous table, except the select lives are subjected to ever increasing pressure to lapse.

Policy Duration	u_1	u_2	v_1	v_2	p	θ_D	θ_T	θ_D/θ_T
0	.010	.100	.30	.05	1.0000	.00860	.860	.010
1	.011	.110	.70	.05	.8500	.02194	.747	.029
2	.013	.130	1.30	.05	.6508	.04154	.682	.061
3	.016	.160	2.00	.05	.3188	.10050	.752	.134
4	.020	.200	2.50	.05	.0607	.16666	.853	.195
5	.....	.....	.....	.....	.0056	.....	.....	.....

If these lapse assumptions are plausible pricing assumptions, then we see that the initial premium rate is appropriate, but the renewal rates become increasingly inadequate. This is due to the antiselective development of the makeup of the persisting population.

The large differences in appropriate premium levels at latter durations should be discounted to some extent, since few will survive to these durations. If we compute total premium collections and total claim payments over the five year period, we obtain .05468 and .07857, respectively. This produces an overall loss of 44 percent of premium revenue.

It is evident from these examples that significant penalties may arise if the actuary assumes independence too quickly.

SUMMARY

Policy lapsation may be significantly influenced by the perceived risk level of the insured. Evolving policy structure and rate classification schemes have encouraged a growth in antiselective lapsation. However, since we may not directly observe the mortality process following policy termination, we cannot quantify this effect and may only speculate on its significance. Traditional actuarial and maximum likelihood estimators of mortality rates may be safely applied only if anticipated future lapse pressure is similar to that experienced by the cohort of lives upon which the tabulated probabilities of death are based. Substantial losses may be encountered if the actuary blindly applies the independence assumption in pricing a contract. Even a short-term coverage form can be significantly underpriced. Longer-term contracts accumulate the effects of unfavorable lapsation from year to year. Higher external lapse pressure will quickly skew the persisting population toward the high risk policyholder. The challenge to the pricing actuary is to assess this rate of decay and appropriately adjust expected experience factors. The model identifiability problem leaves this problem to the actuary's common sense and sound judgment.

APPENDIX A

It will be shown that if statistical inference is used to obtain the distribution of the minimum of the time until death, random variable Y, and the time until lapsation, random variable Z, and an associated indicator, random variables Δ, which records the cause of termination, then a joint distribution for Y and Z exists which is consistent with the imputed distribution for the minimum and the indicator and for which Y and Z are independent.

Let $f(t,\delta)$ be the known joint density for $T = \min(Y,Z)$ and Δ.

We wish to find $f_Y(y)$ and $f_Z(z)$, densities for Y and Z, such that

$$\text{(i) } P\{T > t, \Delta = 0\} = \int_t^\infty f_Y(s)P\{Z > s\}\, ds \text{ and}$$

$$\text{(ii) } P\{T > t, \Delta = 1\} = \int_t^\infty f_Z(s)P\{Y > s\}\, ds$$

where $\Delta = 0$ implies death and $\Delta = 1$ implies lapsation at time T. Let $F(t,\delta) = P\{T > t, \Delta = \delta\}$, $F_Z(z) = P\{Z > z\}$, and $F_Y(y) = P\{Y > y\}$.

Then (i) plus (ii) yields

$$\begin{aligned} F(t,0) + F(t,1) &= \int_t^\infty [f_Y(s)F_Z(s) + f_Z(s)F_Y(s)]ds \\ &= \int_t^\infty -\frac{d}{ds}[F_Y(s)F_Z(s)]ds \\ &= F_Y(t)F_Z(t) \end{aligned}$$

Differentiating (i) gives us $f(t,0) = f_Y(t)F_Z(t)$.

So, we combine these results to obtain

$$f(t,0) = f_Y(t) = \left[\frac{F(t,0) + F(t,1)}{F_Y(t)}\right].$$

And

$$\frac{f_Y(t)}{F_Y(t)} = \frac{f(t,0)}{F(t,0) + F(t,1)}$$

$$-\frac{d}{dt}\ln F_Y(t) = \frac{f(t,0)}{F(t,0) + F(t,1)}$$

$$F_Y(y) = \exp\left\{-\int_0^y \frac{f(s,0)}{F(s,0) + F(s,1)}\, ds\right\}.$$

Similarly, $F_Z(z) = \exp\left\{ - \int_0^z \frac{f(s,1)}{F(s,0) + F(s,1)}\, ds \right\}$.

We have found marginal distributions for Y and Z which produce any given joint distribution for T and Δ. So, if we hypothesize a dependent joint density for Y and Z and formulate the implied distribution for T and Δ, there exists some distribution for Y and Z, which produces this same distribution for T and Δ but is such that Y and Z are independent.

This development assumes thhe random variables Y, Z, and T have absolutely continuous distribution functions so that the densities exist. For a more extensive discussion of this nonidentifiability issue, see Johnson and Johnson, *Survival Models and Data Analysis*, pages 277–80, John Wiley and Sons, Inc.

H.J. BOOM:

To borrow a phrase from Donald Jones [6], there is no doubt that the Broffitt paper "will be a valuable contribution to the education of future generations of actuaries." With this in mind, I propose to consider the implications that this paper is likely to have on the development of the course of reading for Part 5, rather than enter into a discussion of the contents of the paper itself.

In the literature, numerous papers have appeared in which the traditional methods of constructing mortality tables were severely criticized (see, for instance, Weck [9], Seal [8], Jones [6], Hoem [4] and later [5]). Most of this criticism was directed toward the lack of proper statistical justification for the methods of estimation involved and toward the use of what has become known as Balducci's assumption and is referred to in the current paper as "Assumption B." This assumption, requiring a decreasing force of mortality in each unit age interval even if the force of mortality is increasing from each year of age to the next, is now generally regarded as inferior to alternative assumptions such as those of a constant force of mortality ("Assumption C") or a uniform distribution of deaths ("Assumption A").

In 1946, the Balducci assumption was still regarded as the only one meriting any serious consideration (Marshall [7]); Weck [9] made a strong case for Assumption C in 1946, but both official Society textbooks, Gershenson [2] in 1961 and Batten [1] in 1978, although considering the alternative Assumptions A and C in some detail, still restricted all further development almost exclusively to the use of B.

Broffitt's paper meets the criticism on both counts: the maximum likelihood principle provides a solid basis for the development of the estimators

from the statistical point of view, and, of the three assumptions, only the more realistic ones, A and C, are used in the main body of the paper—B receiving only a passing mention, mostly for the purposes of comparison.

A paper like this, giving a thorough and comprehensive treatment on the subject matter of a major portion of one of the Society's examinations, can hardly fail to affect the prescribed course of reading. Just what form this will take will be one of the responsibilities of the Education and Examination Committee, and it will be quite a while before the corresponding changes will become effective. Meanwhile, it may be fruitful to speculate on some of the implications of Broffitt's paper.

It would seem almost unavoidable that the course of reading will require the MLE principle as the fundamental theoretical basis for estimating mortality measures, and that the Balducci assumption, with its associated "actuarial estimator" (formula 3.7 in the paper), will fade into oblivion. Whether both assumptions A and C or C alone will survive is problematic—A has the advantage that it is the standard basis for approximations in the recommended Part 4 life contingencies text; however, other considerations could make a stronger case for C.

Students should develop a thorough understanding of the random as well as of the fixed withdrawal model; each of these, both in full data and partial data cases, should be capable of deriving the MLE equations. However, in further development, probably less emphasis should be given to those equations whose solution requires an iterative process (even if special cases, with $a \equiv 0$ and $b \equiv 1$, would allow explicit solutions). Thus, only the full data random and fixed withdrawal models under Assumption C and the product limit estimator, i.e. the ones referred to as *CF* and *PL*, would receive major emphasis. This is well deserved, since Broffitt's comparisons show clearly that we may expect these to be consistently at least equal to the others when the "quality" of the estimators is examined.

The product limit estimator, which leads directly to an estimator for the mortality rate q itself, also has the advantage that it does not require any further assumptions about the mortality pattern within the age interval. On the other hand, the full data random and fixed withdrawal models, with Assumption C, lead to estimators for the average force of mortality μ, from which q is then still to be calculated by $q = 1 - \exp(-\mu)$; however, the expression for μ is in the form of a simple fraction of which the numerator is the number of observed deaths. This allows the intuitively attractive interpretation of the denominator being the precise "exposure" to the risk of death (in this case we could speak of exposure to the force of mortality), with each individual's contribution consisting of the exact amount of time that the individual is under actual observation, while alive, and during the year of age being considered. This is intuitively superior to the concept of

exposure familiar to us in the actuarial estimator/Assumption B combination, since the awkward nonsense of having to expose the already dead still further to the "risk" of death is avoided.

Direct application of the resulting formulas requires, of course, that the "full data" will indeed be available. But with the now omnipresent computers and their almost limitless record-keeping capabilities this will, in practice, not prove to be too much of an objection. (In any case, as suggested by Greville [3], page 50, one can always avoid tabulating exact ages at death by making appropriate assumptions as to average age at death). For the *CF* case, this could lead to exposure formulas for μ very much like the familiar Balducci-based formulas for the actuarial estimator $\hat{q}$, distinguished from these only by the absence of the awkward Balducci correction for the deaths.

Formulas of both the *CF* and the *PL* type could be easily adapted to estimating mortality rates from valuation schedules and recorded deaths.

REFERENCES

1. BATTEN, ROBERT W. *Mortality Table Construction*, Englewood Cliffs, N.J.: Prentice Hall, 1978.
2. GERSHENSON, HARRY. *Measurement of Mortality*, Chicago: Society of Actuaries, 1961.
3. GREVILLE, T.N.E. "Estimation of the Rate of Mortality in the Presence of In-and-out Movement," *ARCH* 1978. 2, Chicago: Society of Actuaries, 41–56.
4. HOEM, JAN M. "Exposed-to-risk Considerations Based on the Balducci Assumption and Other Assumptions in the Analysis of Mortality," Working Paper No. 30, Laboratory of Actuarial Mathematics, University of Copenhagen, 1979. (Reprinted in *ARCH*, Chicago: Society of Actuaries 1980.1, 47–51).
5. HOEM, JAN M. "A Flaw in Acuarial Exposed-to-Risk Theory," *Stockholm Research Reports in Demography*, No. 15, University of Stockholm, 1985.
6. JONES, DONALD A. "What Actuaries Need to Know about Constructing Mortality Tables," *The Actuary*, Society of Actuaries, 13.2, 5, 1979.
7. MARSHALL, E.W. "Principles Underlying Exposed to Risk Formulae," *Transactions*, Actuarial Society of America, 46, 10–50, 1945.
8. SEAL, HILARY L. "An Attempt to Convert American Actuaries," *The Actuary*, Society of Actuaries 15.3, 4, 1981.
9. WECK, FRANK A. "The Mortality Rate and its Derivation from Actual Experience," *The Record*, American Institute of Actuaries, 36, 23–54, 1947.

(AUTHOR'S REVIEW OF DISCUSSION)

JAMES D. BROFFITT:

Many of Mr. Edwards's comments refer to the method of maximum likelihood which was popularized by R. A. Fisher approximately sixty years

ago. A simple example may be helpful in clarifying why maximum likelihood estimators are preferred to method of moments estimators.

Suppose $X_1, \ldots, X_N$ is a random sample from a population with *pdf* $f(x;\theta)$ where the value of θ is unknown and is to be estimated. Let $m(\theta) = EX_i$ be the mean of the population. Mr. Edwards suggests that θ could be estimated by minimizing $\Sigma(X_i - m(\theta))^2$. This is equivalent to solving $m(\theta) = \bar{X}$ for θ, which is precisely the method of moments.

As an example, suppose

$$\begin{aligned} f(x;\theta) &= \theta x^{\theta-1}, \quad 0 < x < 1 \\ &= 0 \quad , \text{ otherwise} \end{aligned}$$

where $0<\theta<1$. Then $m(\theta) = \theta/(\theta+1)$ and the method of moments estimator is $\tilde{\theta} = \bar{X}/(1-\bar{X})$. Since the likelihood function is

$$L(\theta) = \theta^N (\pi x_i)^{\theta-1}, \quad 0<\theta<1,$$

the MLE is $\hat{\theta} = N/\Sigma(-\ln X_i)$.

Which estimator is better? Comparisons are often based on asymptotic results since they are usually easier to obtain. In this example, both estimators are asymptotically unbiased so the comparison is in terms of variances. It may be shown that for large N,

$$\frac{\text{Var}(\tilde{\theta})}{\text{Var}(\hat{\theta})} = 1 + \frac{1}{\theta^2+2\theta},$$

so $\text{Var}(\hat{\theta}) < \text{Var}(\tilde{\theta})$. This is true in "most instances," i.e., MLEs are asymptotically more efficient. Even though we are working with asymptotics, often the asymptotic results are accurate for relatively small N.

When our data consist of independent observations—some continuous and some discrete—the likelihood is the product of the *pdf*'s of the continuous observations, times the product of the probability mass functions of the discrete observations. In (2.2) the continuous observations correspond to the observed times of death which have *pdf*

$$f(t) = (1 - {}_tq_x)\mu_{x+t},$$

and thus provide the factor

$$\prod_{\mathcal{D}}[(1 - {}_{t_i}q_x)\mu_{x+t_i}].$$

The discrete observations correspond to those lives for which $T_i = 1$ or equivalently $D_i = 0$. Since there are $N-d$ such observations and $P[D_i = 0] = 1-q_x$, they contribute the factor

$$(1-q_x)^{N-d}$$

My use of the differential element was intended to give a probabilistic interpretation to the development of the likelihood. I thought this was easier since actuaries are quite comfortable with interpreting $(1-{}_tq_x)\mu_{x+t}dt$ as the probability that (x) dies at age $x+t$ (or more accurately between ages $x+t$ and $x+t+dt$). The *pdf* is obtained by forming this probability and then removing dt.

Applying the Balducci assumption to (2.2) produces the "simplification"

$$L = (1-q)^N q^d / \prod_{\mathcal{D}} [1-(1-t_i)q]^2,$$

which does not lead to an explicit form for the MLE.

My congratulations to Mr. Robinson on formulating and analyzing a very interesting model which allows dependence between Y and Z. This clearly displays the consequences of erroneously assuming independence.

The lack of dependence of GP on the force of lapsation may be attributable to other factors besides independence between Y and Z, e.g., continuous premiums or the constant force assumption. Under independence and assumption A

$$GP = \frac{E(D_i)}{E(T_i)} = \frac{q - \frac{1}{2}rq}{1 - \frac{1}{2}r - \frac{1}{2}q + \frac{1}{3}rq},$$

which is an increasing function of r. Under independence and assumption C, and assuming a single premium,

$$GP = E(D_i) = \frac{\mu}{\mu+\nu}[1-e^{-(\mu+\nu)}],$$

which is decreasing in ν.

Professor Boom has expressed, quite skillfully, an opinion shared by many actuaries. The estimation of mortality rates should be placed on a sound statistical foundation which is based on an assumption more reasonable than the Balducci hypothesis.

The comparisons in section 4 did not provide a clear choice between *CF* and *BF*. However, *CF* is a maximum likelihood estimator, is based on assumption C, and is easy to compute. Unless more detailed comparisons prove otherwise, this makes *CF* a clear choice over *BF*. It will indeed be interesting to see how the Part 5 syllabus develops.