NONPARAMETRIC ESTIMATION FROM INCOMPLETE OBSERVATIONS*

E. L. KAPLAN
University of California Radiation Laboratory
AND
PAUL MEIER
University of Chicago

In lifetesting, medical follow-up, and other fields the observation of the time of occurrence of the event of interest (called a *death*) may be prevented for some of the items of the sample by the previous occurrence of some other event (called a *loss*). Losses may be either accidental or controlled, the latter resulting from a decision to terminate certain observations. In either case it is usually assumed in this paper that the lifetime (age at death) is independent of the potential loss time; in practice this assumption deserves careful scrutiny. Despite the resulting incompleteness of the data, it is desired to estimate the proportion $P(t)$ of items in the population whose lifetimes would exceed t (in the absence of such losses), without making any assumption about the form of the function $P(t)$. The observation for each item of a suitable initial event, marking the beginning of its lifetime, is presupposed.

For random samples of size N the product-limit (PL) estimate can be defined as follows: List and label the N observed lifetimes (whether to death or loss) in order of increasing magnitude, so that one has $0 \leq t_1' \leq t_2' \leq \cdots \leq t_N'$. Then $\hat{P}(t) = \prod_r [(N-r)/(N-r+1)]$, where r assumes those values for which $t_r' \leq t$ and for which t_r' measures the time to death. This estimate is the distribution, unrestricted as to form, which maximizes the likelihood of the observations.

Other estimates that are discussed are the actuarial estimates (which are also products, but with the number of factors usually reduced by grouping); and reduced-sample (RS) estimates, which require that losses not be accidental, so that the limits of observation (potential loss times) are known even for those items whose deaths are observed. When no losses occur at ages less than t, the estimate of $P(t)$ in all cases reduces to the usual binomial estimate, namely, the observed proportion of survivors.

CONTENTS

1. Introduction 458
   1.1 Formulation 458
   1.2 Nonparametric estimation 459
   1.3 Examples of the RS and PL estimates 459
   1.4 Notation 461
2. The Product-Limit Estimate 462
   2.1 Definition and calculation 462
   2.2 Mean and variance of $\hat{P}(t)$ 465
   2.3 Mean lifetime 467
3. The Reduced-Sample Estimate vs. the PL 469
   3.1 Alternatives to the PL estimate 469
   3.2 Dependence of deaths and losses 470

* Prepared while the authors were at Bell Telephone Laboratories and Johns Hopkins University respectively. The work was aided by a grant from the Office of Naval Research.

4. Actuarial Estimates 471
4.1 Estimates using only n, δ, λ 471
4.2 Estimates using average ages 473
5. Maximum Likelihood Derivation of the PL 475
6. Means and Variances 476
6.1 The PL estimate $\hat{P}(t)$ 476
6.2 Covariances and mean lifetimes 478
7. Consistency; Testing with Replacement 479

1. INTRODUCTION

1.1 *Formulation.* In many estimation problems it is inconvenient or impossible to make complete measurements on all members of a random sample. For example, in medical follow-up studies to determine the distribution of survival times after an operation, contact with some individuals will be lost before their death, and others will die from causes it is desired to exclude from consideration. Similarly, observation of the life of a vacuum tube may be ended by breakage of the tube, or a need to use the test facilities for other purposes. In both examples, incomplete observations may also result from a need to get out a report within a reasonable time.

The type of estimate studied here can be briefly indicated as follows. When a random sample of N values, $T_1, T_2, \cdots, T_N$ of a random variable is given, the sample distribution function $\hat{F}(t)$ is naturally defined as that which assigns a probability of $1/N$ to each of the given values, so that $\hat{F}(t)$ equals $1/N$ times the number of sample values less than the argument t. Besides describing the sample, this $\hat{F}(t)$ is also a nonparametric estimate of the population distribution, in the sense indicated in 1.2 below. When the observations are incomplete, the corresponding estimate is still a step-function with discontinuities at the ages of observed deaths, but it can no longer be obtained as a mere description of the sample.

The samples considered in this paper are incomplete in the sense that one has given, not a random sample $T_1, \cdots, T_N$ of values of the random variable T itself (called the lifetime), but the *observed lifetimes*

$$t_i = \min (T_i, L_i), \qquad i = 1, 2, \cdots, N. \tag{1a}$$

Here the L_i, called *limits of observation*, are constants or values of other random variables, which are assumed to be independent of the T_i unless otherwise stated (in Sections 3.2 and 7). For each item it is known whether one has

$$T_i \leqq L_i, \qquad t_i = T_i \text{ (a } death\text{)} \tag{1b}$$

or

$$L_i < T_i, \qquad t_i = L_i \text{ (a } loss\text{)}.$$

Ordinarily the T_i and L_i are so defined as to be necessarily nonnegative.

The items in the sample are thus divided into two mutually exclusive classes, namely deaths and losses. A loss by definition always precludes the desired knowledge of T_i. On the other hand, a death does not always preclude the knowledge of the corresponding L_i, in case the limits of observation are nonrandom and foreseeable. Such knowledge of the L_i may have value; for example,

it makes available the reduced-sample estimate (Section 3), if one chooses to use it.

The type of sample described is a generalization of the censored sample defined by Hald [17], and a specialization of the situation considered by Harris, Meier, and Tukey [18].

The term death has been adopted as being at least metaphorically appropriate in many applications, but it can represent any event susceptible of random sampling. In particular, the roles of death and (random) loss may be interchangeable. By redefining the classification of events into deaths and losses, it may be possible to approach the same data from various points of view and thus to estimate the survivorship functions $P(t)$ that would be appropriate to various categories of events in the absence of the others. This is familiar enough; see for example [6], [13], [16].

1.2 *Nonparametric estimation.* Most general methods of estimation, such as maximum likelihood or minimum chi-square, may be interpreted as procedures for selecting from an admissible class of distributions one which, in a specified sense, best fits the observations. To estimate a characteristic (or parameter) of the true distribution one uses the value that the characteristic has for this best fitting distribution function. It seems reasonable to call an estimation procedure *nonparametric* when the class of admissible distributions from which the best-fitting one is to be chosen is the class of all distributions. (Wolfowitz [28] has used the term similarly in connection with the likelihood ratio in hypothesis testing). With a complete sample, it is easy to see that the sample distribution referred to in 1.1 is the nonparametric estimate on the maximum likelihood criterion. The same result is true of the product-limit estimate for incomplete samples, as will be demonstrated in Section 5.

The most frequently used methods of parametric estimation for distributions of lifetimes are perhaps the fitting of a normal distribution to the observations or their logarithms by calculating the mean and variance, and fitting an exponential distribution $e^{-t/\mu}dt/\mu$ by estimating the mean life μ alone. Such assumptions about the form of the distribution are naturally advantageous insofar as they are correct; the estimates are simple and relatively efficient, and a complete distribution is obtained even though the observations may be restricted in range. However, nonparametric estimates have the important functions of suggesting or confirming such assumptions, and of supplying the estimate itself in case suitable parametric assumptions are not known. An important property of these nonparametric estimates is that if the age scale is transformed from t to $t^*=f(t)$, where f is a strictly increasing function, then the corresponding estimated distribution functions are simply related by $\hat{F}^*(f(t)) = \hat{F}(t)$.

1.3 *Examples of the RS and PL estimates.* We will consider the following situation. A random sample of 100 items is put on test at the beginning of 1955; during the year 70 items die and 30 survive. At the end of the year, a larger sample is available and 1000 additional items are put on test. During 1956, 15 items from the first sample and 750 from the second die, leaving 15 and 250 survivors respectively. As of the end of 1956, it is desired to estimate the proportion $P(2)$ of items in the population surviving for two years or more.

The survival probabilities are supposed to depend on the age (the duration of the test) rather than on the calendar year, and hence the data are arranged as in Table 460.

TABLE 460

Samples	I	II
Initial numbers	100	1000
Deaths in first year of age	70	750
One-year survivors	30	250
Deaths in second year of age	15	
Two-year survivors	15	

This particular example is such that it is easy to form an estimate $P^*(2) = 15/100 = 0.15$ from the first sample alone. This is called the *reduced-sample* (RS) estimate because it ignores the 1000 items tested only during 1956. It is a legitimate estimate only when the reduced sample is itself a random sample; this will be the case only when (as assumed here) the observation limits (two years for the first sample, and one year for the second) are known for all items, deaths as well as losses. In the absence of this information, one would have no basis for discriminating among the 835 deaths observed before the age of two years. One cannot simply ignore the 250 losses at age one year; since only 15 items have survived for two years, $P(2)$ would then be estimated as $15/850 = .018$, an absurd result. The point is discussed further in 3.1 below.

We now inquire whether the second sample, under test for only one year, can throw any light on the estimate of $P(2)$. Clearly it will be necessary to assume that both samples have been drawn from the same population, an assumption that the RS estimate $P^*(2)$ avoided. At any rate, the estimates of $P(1)$ from the two samples, namely 0.30 and 0.25, are not sufficiently different to contradict the assumption. By combining the two samples, the estimate

$$\hat{P}(1) = P^*(1) = (30 + 250)/(100 + 1000) = .255$$

is obtained for $P(1)$. (In this case the RS has the same value as the other estimate to be discussed, the product-limit or PL.) This result exhausts the usefulness of the second sample for the present purposes; how does it help to estimate $P(2)$?

The answer is that there are advantages to using the first (the smaller) sample for estimating $P(2)/P(1)$, the conditional probability of survival for two years given survival for one year, rather than $P(2)$ itself. This estimate is

$$\hat{P}(2)/\hat{P}(1) = 15/30 = 0.50, \text{ whence}$$

$$\hat{P}(2) = 0.255 \times 0.50 = 0.127,$$

a very simple example of the *product-limit* (PL) estimate. The outstanding advantage of this strategy is that it works just as well if we are not privileged to know that the 750 deaths in the second sample had observation limits of one year, because these items are irrelevant to the estimation of $P(2)/P(1)$ in any

case. Other considerations for deciding between the two estimates will be set forth in Section 3.

The discussion of the PL estimate will be continued shortly, in Sections 2 and 3. Section 3 is equally concerned with the RS estimate, while Section 4 is devoted to the actuarial estimates. The remaining three sections consist principally of mathematical derivations. Though much older, the actuarial estimates are essentially approximations to the PL; they are products also, but typically they aim to reduce the number of factors by grouping. (Grouping may or may not be possible for the PL itself.) The distinguishing designation product-*limit* was adopted because this estimate is a limiting case of the actuarial estimates. It was proposed as early as 1912 by Böhmer [6] (referred to by Seal [26]), but seems to have been lost sight of by later writers and not further investigated.

1.4 *Notation.* The survival function

$$P(t) = \Pr(T > t), \tag{1c}$$

giving the population probability of surviving beyond t, will be used in place of the distribution function $F(t) = 1 - P(t)$ because of its convenience where the product-limit estimate and its actuarial approximations are concerned. In addition the following functions are defined:

$\hat{P}(t)$ = product-limit (PL) estimate of $P(t)$.
$P^*(t)$ = reduced-sample (RS) estimate of $P(t)$.
$n(t)$ = the number of items observed and surviving at age t, when deaths (but not losses) at t itself are subtracted off.
$N(t)$ = the expectation of $n(t)$, for fixed observation limits.
$N^0(t)$ = the number of items having observation limits L such that $L \geq t$. In practice this function is not necessarily known.

For the first reading of the paper it may be desirable to suppose that the death of one item and the loss of the same or any other item never occur at the same age, and never coincide with an age t at which any of the above functions are to be evaluated. This condition can always be met by fudging the ages a little when necessary. On the other hand, a regular user of the techniques will probably come to regard overt fudging as naive; he will prefer to formalize his notation and record-keeping by adopting the conventions already insinuated into the definitions of death and loss, $P(t)$, and $n(t)$ above. These conventions may be paraphrased by saying that deaths recorded as of an age t are treated as if they occurred slightly before t, and losses recorded as of an age t are treated as occurring slightly after t. In this way the fudging is kept conceptual, systematic, and automatic.

The convention that deaths precede losses in case of ambiguity is based on the following sequence of operations, which is clearly more efficient than the reverse sequence: Examine a group of items of age t_0, observe the number δ of deaths since the last examination, and then remove (or lose contact with) a number λ of the survivors. It may then be convenient simply to record t_0 as the age of death or loss of the $\delta+\lambda$ items, especially if t_0 is always an integral multiple of a fundamental time interval; in fact, however, the deaths will have preceded the losses.

The chief exception to the immediate applicability of this convention occurs when the losses are random but cannot affect items that have already died. Then the possible sequences of occurrence of deaths and losses between examination times (assumed to be close together) are approximately equally likely, and a reasonable compromise is to assume that half the losses in the interval precede and half follow the deaths, as in (4b) below. If the loss of an item is compatible with the possibility of its having died (unknown to the experimenter) between the time when it was last examined and the time of loss, an item lost in this way is effectively lost just after it was last examined, and the convention is entirely appropriate. The disappearance of individuals subject to medical follow-up is a case in point.

The remainder of the conventions concern the treatment of discontinuities in the functions listed above. Should the value assigned to $n(t)$, say, for an argument t for which it is discontinuous be the right-hand limit $n(t+0) = \lim n(t+h)$ as $h \to 0$ with $h > 0$, the left-hand limit $n(t-0)$ (the same but with $h < 0$), or something else? The superior expressiveness of the notation adopted is illustrated by the relations

$$\begin{aligned} n(t-0) - n(t) &= \text{the number of deaths at } t, \\ n(t) - n(t+0) &= \text{the number of losses at } t, \end{aligned} \tag{1d}$$

or equivalently by the formula (3a): $P^*(t) = n(t)/N^0(t)$, which otherwise would not be valid at discontinuities. Analogous conventions are adopted for all the above-mentioned functions of t, so that $P(t)$ and $\hat{P}(t)$ are right-continuous, $N^0(t)$ is left-continuous, and $n(t)$ and $P^*(t)$ are neither. Other advantages of the convention for $P(t)$ and its estimates are the following: (a) According to the sequence of operations assumed above, one may record deaths as of age t although they actually occurred slightly earlier. (b) It makes $P(0) = 1$ if and only if no item dies at birth (age zero). This is convenient and natural.

One other possibility may be mentioned briefly. The assumption that $P(\infty) = 0$, so that the lifetimes are finite with probability one, is necessary only for parts of Section 7 and for the calculation of a finite mean lifetime. However, in practice there is no apparent need to contradict the assumption either. If half of a sample dies in one day and the other half is still alive after 1000 days, one should still report $\hat{P}(1000)$ (not $\hat{P}(\infty)$) $= 0.50$, since the argument 1000 is not an arbitrary large number, but the actual duration of the test.

2. THE PRODUCT-LIMIT ESTIMATE

2.1 *Definition and calculation.* Both the PL and the actuarial estimates of Section 4 are based on the following general procedure:

(a) The age scale is divided into suitably chosen intervals, $(0, u_1)$, (u_1, u_2), $\cdots$, as described below. (In the example of 1.3, there were only two such intervals, namely (0, 1) and (1, 2).)

(b) For each interval (u_{j-1}, u_j), one estimates $p_j = P_j/P_{j-1}$, the proportion of items alive just after u_{j-1} that survive beyond u_j.

(c) If t is a division point (it may be introduced specially if necessary), the proportion $P(t)$ in the population surviving beyond t is estimated by the product of the estimated p_j for all intervals prior to t.

If step (a) is left relatively arbitrary and approximations or parametric assumptions are accepted in step (b), one arrives at the actuarial estimates. The PL is obtained by selecting the intervals in (a) so that the estimation in (b) is a simple binomial, without any recourse to assumptions of functional form. The condition for this is that within each interval, deaths and losses be segregated in a known fashion. As a beginning, it may be assumed that no interval contains both deaths and losses. Then if the number under observation just after u_{j-1} is denoted by n_j, and δ_j deaths are observed in the interval (u_{j-1}, u_j), the estimate is clearly

$$\hat{p}_j = (n_j - \delta_j)/n_j = n_j'/n_j, \tag{2a}$$

where n_j' is the number under observation just after the δ_j deaths. However, if the interval contains only losses (but at least one item survives throughout the interval), the estimate is $\hat{p}_j = 1$.

In the product of conditional probabilities formed in step (c), unit factors may as well be suppressed; and we need not be concerned with the manner in which the losses are distributed among the intervals, so long as n_j and n_j' are correctly evaluated in (2a), and no losses occur at ages intermediate to the δ_j deaths. The situation is illustrated by the following scheme:

TABLE 463

No. of items	N		n_1		n_1'		n_2		n_2'	$\cdots$
No. of deaths or losses		λ_0		δ_1		λ_1		δ_2		λ_2
Division points	$u_0 = 0$					u_1				u_2

Here N is the initial number of items, and the braces join the numbers whose ratios are the conditional probabilities (2a). The numbers in the second line are the differences of those in the first, the λ's counting losses and the δ's deaths; some of these could be zero. The division points u_j are placed in the third line to show that the δ_j deaths occur between u_{j-1} and u_j, while u_j is located anywhere among the λ_j losses. The relation $n_j' \leq n(u_j) \leq n_{j+1}$ holds.

The PL estimate is now given by

$$\hat{P}(t) = \prod_{j=1}^{k} (n_j'/n_j), \quad \text{with} \quad u_k = t,\ n_j' = n_j - \delta_j. \tag{2b}$$

If the greatest observed lifetime t^* corresponds to a loss, (2b) should not be used with $t > t^*$; in this case $\hat{P}(t)$ can be regarded as lying between 0 and $\hat{P}(t^*)$, but is not more closely defined.

If it is desired to permit the entrance of items into the sample after the commencement of their lifetimes, this can be done by treating such entrances as "losses" that are counted negatively in λ_j. The same items can of course disappear again at a later age and so yield ordinary losses as well. It is assumed that nothing is known of the existence of any such item that dies before it becomes available for observation; that is, the observation is censored on the right but truncated on the left, in the terminology of Hald [17].

The form (2b) was selected for the PL estimate to give the minimum number of elementary factors and the maximum grouping of the observations. Nevertheless, the number of deaths δ_j in an interval can easily be as small as unity. The resulting estimate $\hat{p}_j$, though of limited value by itself, is none the less acceptable as a component of $\hat{P}(t)$. In fact, one is at liberty to take the intervals as short and as numerous as one pleases, and to regard each death as occupying an interval by itself. To specify the resulting expression, one relabels the N ages t_i of death or loss in order of increasing magnitude, and denotes them by $t_1' \leq t_2' \leq \cdots \leq t_N'$. Then

$$\hat{P}(t) = \prod_r [(N - r)/(N - r+1)], \tag{2c}$$

where r runs through those positive integers for which $t_r' \leq t$ and t_r' is the age of death (not loss). The cancellation of like integers in numerator and denominator where they occur reduces (2c) to (2b). If there are no losses, everything cancels except the first denominator N and the last numerator $n(t)$, say, and the PL reduces to the usual binomial estimate $n(t)/N$. (2c) shows that $\hat{P}(t)$ is a step-function which changes its value only at the observed ages of death, where it is discontinuous.

In analyzing data on lifetimes by the multiplication of conditional probabilities, one of the following three procedures will usually suffice:

(1) If the number of deaths is relatively small, these deaths may be arranged in order of age without grouping, and the numbers of losses in the intervening age intervals counted. The PL estimate is calculated by (2c).

(2) If (1) is too time-consuming but the number of distinct ages of loss is relatively small, these ages may be arranged in order, additional division points inserted as desired, and the numbers of deaths in the resulting age intervals counted. If some of these intervals are shorter than necessary and are found to contain no deaths, they can be combined with adjacent intervals. The PL estimate is calculated by (2b).

(3) If neither (1) nor (2) is compact enough, then division points are chosen without close consideration of the sample, deaths and losses are counted in each interval, and an actuarial approximation to the PL, such as (4b), is used.

As a miniature example of case (1), suppose that out of a sample of 8 items the following are observed:

Deaths at 0.8, 3.1, 5.4, 9.2 months.
Losses at 1.0, 2.7, 7.0, 12.1 months.

The construction of the function $\hat{P}(t)$ then proceeds as follows:

TABLE 464

u_j	n_j	n_j'	λ_j	$\widehat{P}(u_j)$
0.8	8	7	2	7/8
3.1	5	4	0	7/10
5.4	4	3	1	21/40
9.2	2	1	0	21/80
(12.1)	1	1	1	21/80

Each value of $\hat{P}(u_j)$ is obtained by multiplying n_j'/n_j by the preceding value $\hat{P}(u_{j-1})$. The age 12.1 is recorded in the last line to show the point at which $\hat{P}(t)$ becomes undefined; since it is a loss time, the 12.1 is enclosed in parentheses. It is to be inferred from the table that $\hat{P}(5.3) = 7/10$, for example. The third and fourth columns could be omitted since $n_j' = n_j - 1$ (except in the last line, which corresponds to a loss) and $\lambda_j = n_j' - n_{j+1}$.

A rudimentary illustration of case (2) has already been given in 1.3. A little more elaborate example of a similar sort with $N = 100$ is given in Table 465. Here 1.7, 3.6, and 5.0 are assumed to be the only ages at which losses occur; they are prescribed as division points. The other division points (1, 2, 3, 4) are selected at pleasure, with the object of interpolating additional points on the curve of $\hat{P}(t)$ vs. t. $\hat{N}_{Ej}$ is the effective sample size defined in (2j) below. In practice four columns headed u_j, n_j, n_j', $\hat{P}_j$ will suffice. From the table one infers that $.74 < \hat{P}(2.5) < .87$, for example.

TABLE 465

Interval	Factor						
u_{j-1}, u_j	j	n_j	δ_j	λ_j	$\hat{p}_j$	$\hat{P}(u_j)$	$\hat{N}_{Ej}$
0 –1	1	100	3	0	97/100	.97	100
1 –1.7	2	97	5	20	92/97	.92	100
1.7–2	3	72	4	0	68/72	.87	88
2 –3	4	68	10	0	58/68	.74	83
3 –3.6	5	58	9	12	49/58	.63	80
3.6–4	6	37	6	0	31/37	.52	73
4 –5.0	7	31	15	16	16/31	.27	51

2.2 *Mean and variance of* $\hat{P}(t)$. The important facts here, derived in Section 6.1, are that $\hat{P}(t)$ is consistent and of negligible bias (unless excessive averaging is done; see Section 3.1), and that an asymptotic expression for its variance can be obtained. Like the estimate itself, the sample approximation to its variance proves to be independent of the limits of observation of items not actually lost. However, the variance derived from population values does depend on all the limits of observation, which are assumed to be fixed during the sampling.

It has been noted that if the greatest observed lifetime t^* corresponds to a loss, then for $t > t^*$, $\hat{P}(t)$ is undefined though bounded by 0 and $\hat{P}(t^*)$. Unless the probability of this ambiguous situation is quite small, however, a nonparametric estimate of $P(t)$ will not be very informative in any case. The ambiguity cannot occur unless the $N^0(t)$ items observable to t all die at ages less than t. The probability of this event is

$$[1 - P(t)]^{N^0(t)} \leq e^{-N^0(t)P(t)} = e^{-N(t)} \tag{2d}$$

This is already less than 0.01 when $N(t)$ is only five.

It is shown in Section 6.1 that if one can supplement the ambiguous case by ascertaining the age of death of the item lost at t^*, or of one or more other randomly selected items alive at t^*, and defines $\hat{P}(t)$ for $t > t^*$ as $\hat{P}(t^*)$ times the survival function for the supplementary sample, then the expected value of

$\hat{P}(t)$ is precisely the population value $P(t)$. In practice this supplementation would often be neither feasible nor worthwhile, but with otherwise adequate data the resulting bias in one or a few samples will be too small to have any practical importance.

It will be shown in Section 6.1 that the variance of $\hat{P}(t)$ is given approximately by

$$V[\hat{P}(t)] \doteq P^2(t) \sum_1^k (q_j/N_j p_j), \tag{2e}$$

where the distinct limits of observation L_j' are now used as the division points; L'_{k-1} is the greatest preceding t; and $p_j = 1 - q_j = P(L_j')/P(L'_{j-1})$, with $L_0' = 0$, $L_k' = t$. After dividing (2e) by $P^2(t)$, one sees that the square of the coefficient of variation (CV) of $\hat{P}(t)$ is set equal to the sum of the squares of the CV's of the estimates of the p_j, the usual approximation for the variance of a product.

If the sample estimates are inserted in (2e) one obtains

$$\hat{V}[\hat{P}(t)] \doteq \hat{P}^2(t) \sum_1^k [\delta_j/n_j(n_j - \delta_j)] = \hat{P}^2(t) \sum_1^k \left(\frac{1}{n_{j'}} - \frac{1}{n_j}\right). \tag{2f}$$

A very similar formula was derived by Greenwood [15] and later by Irwin [19] in connection with actuarial estimates. It is easily verified that (2f) remains valid when the number of intervals is reduced to those used in (2b), or expanded to one interval for each death as in (2c). In the latter case it may be written

$$\hat{V}[\hat{P}(t)] \doteq \hat{P}^2(t) \sum_r [(N - r)(N - r + 1)]^{-1}, \tag{2g}$$

where r runs through the positive integers for which $t_r' \leq t$ and t_r' corresponds to a death.

In terms of integrals (2e) can be written

$$\begin{aligned} V[\hat{P}(t)] &\doteq P^2(t) \int_0^t \frac{|dP(u)|}{N^0(u)P^2(u)} \\ &\doteq P^2(t) \int_0^t \frac{|dP(u)|}{N(u)P(u)} \end{aligned} \tag{2h}$$

In case of ambiguity the first form should be referred to and interpreted in accordance with 1.4. If $P(u)$ alone is discontinuous within the range of integration, it should be regarded as a (continuous) independent variable, so that $\int P(u)^{-2} |dP(u)| = 1/P(u)$.

Since $N^0(t) \leq N^0(u) \leq N$, it is clear that (2h) is greater than the complete sample variance $P(t)[1-P(t)]/N$ but less than the reduced-sample variance $P(t)[1-P(t)]/N^0(t)$ of (3b). If losses are random and the instantaneous rates of death and loss among survivors are in the ratio 1 to ρ at all ages, then one has $E[N^0(u)^{-1}] \doteq [NP^\rho(u)]^{-1}$ for large samples, and (2h) reduces to

$$V[\hat{P}(t)] \doteq [P^{1-\rho}(t) - P^2(t)]/(1 + \rho)N \tag{2i}$$

It is often instructive to estimate an effective sample size $\hat{N}_E(t)$, which in the absence of losses would give the same variance $\hat{V}[\hat{P}(t)]$. Evidently

$$\hat{N}_E(t) = \hat{P}(t)[1 - \hat{P}(t)]/\hat{V}[\hat{P}(t)]. \tag{2j}$$

It can be shown that $\hat{N}_E(t)$ is nonincreasing as t increases, and that

$$n(t)/\hat{P}(t) \leq \hat{N}_E(t) \leq D(t)/[1 - \hat{P}(t)], \tag{2k}$$

where $D(t)$ is the number of deaths observed at ages not exceeding t. The upper bound was proposed as an approximation to $\hat{N}_E(t)$ by Cornfield [8]. The lower bound corresponds to the reduced-sample variance.

In Table 464 of 2.1 we had $\hat{P}(6) = 21/40 = 0.525$. By (2f) the variance of $P(6)$ is estimated as

$$\hat{V}[\hat{P}(6)] = (0.525)^2\left(\frac{1}{8 \times 7} + \frac{2}{5 \times 3}\right) = 0.042.$$

By (2g) the same result appears in the form

$$(0.525)^2\left(\frac{1}{7 \times 8} + \frac{1}{4 \times 5} + \frac{1}{3 \times 4}\right).$$

The effective sample size is estimated as

$$\hat{N}_E(6) = (0.525)(0.475)/0.042 = 6.0.$$

The bounds for $\hat{N}_E(6)$ in (2k) are 5.7 and 6.3. Values of $\hat{N}_E$ are also indicated in Table 465.

2.3 *Mean lifetime.* The PL estimate $\hat{\mu}$ of the mean lifetime μ is defined as the mean of the PL estimate of the distribution. It is well-known (and easily proved by integrating by parts) that the mean of a nonnegative random variable is equal to the area under the corresponding survivorship function. Hence

$$\hat{\mu} = \int_0^\infty \hat{P}(t)dt.$$

Of course, if $\hat{P}(t)$ is not everywhere determined, $\hat{\mu}$ is undefined. In cases where the probability of an indeterminate result is small, $\hat{P}(t)$ is practically unbiased and the same is true of $\hat{\mu}$.

If we "complete" $\hat{P}(t)$ in the example of Table 464 by following the longest observed individual (with $t_r' = 12.1$) to death at $t = 14.3$ we have

$$\begin{aligned}\hat{\mu} &= (1.000)(0.8) + (0.875)(3.1 - 0.8) + (0.700)(5.4 - 3.1) \\ &\quad + (0.525)(9.2 - 5.4) + (0.2625)(14.3 - 9.2) \\ &= 0.800 + 2.012 + 1.610 + 1.995 + 1.339 = 7.76.\end{aligned}$$

If $\hat{P}(t)$ were "completed" by setting $\hat{P}(t) = 0$ for the indeterminate range, the last term in the above sum would be replaced by $(0.2625)(12.1 - 9.2) = 0.761$, and $\hat{\mu}$ would be estimated as 7.18.

Of course, if the probability of an indeterminate result is high, there is no satisfactory way to estimate μ. In such cases Irwin [19] has suggested that in

place of estimating the mean itself, one should estimate the "mean life limited to a time L," say $\mu_{[L]}$. This is the mean of min (T_i, L), with L chosen at the investigator's convenience. Naturally, one would choose L to make the probability of an indeterminate result quite small. If one chooses to use this procedure he should give an estimate of $P(L)$ along with $\hat{\mu}_{[L]}$. If we take $L=10$ in our example,

$$\hat{\mu}_{[10]} = 0.800 + 2.012 + 1.610 + 1.995 + (0.2625)(10 - 9.2) = 6.63,$$

and $\hat{P}(10)=0.2625$.

In Section 6.2 an approximate formula is given for the variance of $\hat{\mu}$:

$$V(\hat{\mu}) \doteq \int_0^\infty \frac{A^2(t) \mid dP(t) \mid}{N(t)P(t)} = \int_0^\infty \frac{A^2(t) \mid dP(t) \mid}{N^\circ(t)P^2(t)}, \tag{2l}$$

where $A(t)=\int_t^\infty P(u)du$. Upon making the obvious substitutions we find, after some reduction, the following estimate of $V(\hat{\mu})$:

$$\hat{V}(\hat{\mu}) = \sum_r \frac{A_r^2}{(N - r)(N - r + 1)} \tag{2m}$$

where r runs over those integers for which t_r corresponds to a death, and $A_r=\int_{t_r}^\infty \hat{P}(u)du$. If there are no losses,

$$A_r = \sum_{i=r+1}^{N} (t_i - t_r)/N,$$

and it can be shown that $\hat{V}(\hat{\mu})$ reduces to $\sum (t_i-\bar{t})^2/N^2$. This fact, plus the impossibility of estimating the variance on the basis of only one observed death, suggests that $(2m)$ might be improved by multiplication by $D/(D-1)$, where D is the number of deaths observed.

For our first estimate of $\hat{\mu}$ above we have

$$\begin{aligned} A_1 &= 2.012 + 1.610 + 1.995 + 1.339 = 6.956, \\ A_4 &= \quad 1.610 + 1.995 + 1.339 = 4.944, \\ A_5 &= \quad\quad 1.995 + 1.339 = 3.334, \\ A_7 &= \quad\quad\quad 1.339 = 1.339. \end{aligned}$$

(Obviously the A_r and $\hat{\mu}$ are best calculated in reverse order.) We then have for the estimated variance,

$$\hat{V}(\hat{\mu}) = \frac{(6.956)^2}{7 \times 8} + \frac{(4.944)^2}{4 \times 5} + \frac{(3.334)^2}{3 \times 4} + \frac{(1.339)^2}{1 \times 2} = 3.91$$

The estimated standard deviation of $\hat{\mu}$ is $\sqrt{3.91}=1.98$. If the factor $D/(D-1)$ $=4/3$ is included, these results become 5.21 and 2.28 respectively.

If one is limited to grouped data as in Table 465 it is necessary to use actuarial-type assumptions (e.g., the trapezoidal rule) to estimate $\hat{\mu}$ and its variance. Thus, we may estimate the mean life limited to 5 time units as follows:

$$\begin{aligned} \hat{\mu}_{[5]} = (1/2)[&(1.00 + 0.97)(1.0 - 0.0) + (0.97 + 0.92)(1.7 - 1.0) \\ &+ \cdots + (0.52 + 0.27)(5.0 - 4.0)] = 3.76. \end{aligned}$$

We observe in conjunction with this estimate that the estimated proportion surviving the limit is $\hat{P}(5) = 0.27$.

3. THE REDUCED-SAMPLE ESTIMATE VS. THE PL

3.1 *Alternatives to the PL estimate.* The acceptable nonparametric estimation procedures known to the writers in the situation considered are variants of the PL and of the reduced-sample (RS) estimate defined by

$$P^*(t) = n(t)/N^\circ(t) = n(t)/[n(t) + D^\circ(t)]. \tag{3a}$$

Here $N^\circ(t) = n(t) + D^\circ(t)$ is the number of items with observation limits $\geq t$; of these, $D^\circ(t)$ die at ages $\leq t$ and $n(t)$ survive beyond t. A simple example of the RS estimate has been given in 1.3. Berkson and Gage [3] have called it the "ad-hoc" method.

If the lifetimes are independent of the observation limits, the reduced sample of $N^\circ(t)$ items will be a random sample and $P^*(t)$ a simple binomial estimate, unbiased, and with the variance

$$V[P^*(t)] = P(t)[1 - P(t)]/N^\circ(t). \tag{3b}$$

It has already been noted that this is equal to or larger than the approximation (2h) to the PL variance. As with the PL, the estimate

$$\mu^* = \int_0^\infty P^*(t)dt \tag{3c}$$

of the mean life may have to be truncated at the greatest of the observation limits. In Section 6.2 it is shown that the variance of μ^* is given exactly by

$$V(\mu^*) = 2\int_0^\infty \frac{1 - P(t)}{N^\circ(t)} \int_t^\infty P(u)dudt.$$

In practice one substitutes $P^*(t)$ for $P(t)$ in this formula and perhaps approximates the integrals.

Whereas the PL estimate has discontinuities only at observed ages of death, it is apparent from (3a) that the RS generally has discontinuities at losses and observation limits also. Furthermore, suppose that one of the observation limits L is such that all (or a sufficient number of) the corresponding items are observed to die prior to L, while other items survive and are observed beyond L. Then the decrease in $N^\circ(t)$ (without an equivalent decrease in $n(t)$) will cause $P^*(t)$ to *increase* as t increases through L. Thus $P^*(t)$, unlike $\hat{P}(t)$ and the true $P(t)$, is not necessarily monotonic decreasing. However, this appears to be a disadvantage only in a psychological sense. It does not seem advisable to avoid it, as one could, by a further reduction of the sample, basing the estimated $P(t)$ for all $t \leq$ a suitable L on the fixed sample of items having observation limits $\geq L$.

The expression $D^\circ(t) = N^\circ(t) - n(t)$ in (3a) is not the total number $D(t)$ of deaths observed prior to or at age t, but the (usually smaller) number of such deaths having observation limits $\geq t$. Since $P^*(t)$ is unbiased, the estimate obobtained by replacing $D^\circ(t)$ by $D(t)$ is too small on the average (except in

special cases where $D(t) = D^0(t)$ with probability one), as has been pointed out many times (e.g., [3] and [23]). Examples such as that of 1.3 show that the resulting bias can be very great, being in fact limited only by the necessity that the expectation be positive. Since $D^0(t)$ is often unknown while $D(t)$ is known, an investigator who is unacquainted with the PL or one of its approximations is all too likely to fall into the trap.

Another insidious but instructive characteristic of the illegitimate estimate based on $D(t)$ is the fact that it is simply the average of the PL estimates $\hat{P}(t)$ (those that are defined at t) obtained by regarding each item as a sample in its own right. For a sample of one item with observed lifetime t_1, $\hat{P}(t)$ is unity for $t < t_1$, and for $t \geq t_1$ is zero or undefined according as the observation ends with death or loss. This is also the RS estimate unless one insists that $P^*(t)$ is undefined for $t > L_1$ (the observation limit of the item), even though it may have been estimated as zero for smaller values of t.

The moral of this discussion is that arithmetic averages of independent PL estimates from very small samples will tend to be biased. On the other hand, the averaging of independent RS estimates is quite satisfactory, especially if the values of $N^0(t)$ are used as weights. To avoid bias one must at least assign zero weight when $N^0(t) = 0$, thus supplying each estimate with a limit of observation (the maximum of those attached to its items) beyond which it is regarded as undefined, even though a value of zero may be indicated at some earlier age.

To summarize, the advantages of the PL over the RS are the following:

(a) The observation limits need not be known for items observed to die.
(b) The sampling variance of the PL is usually a little smaller.
(c) The PL has fewer discontinuities, and its monotonicity may be comforting.

On the other hand, the RS has the following advantages over the PL:

(a) It is perfectly unbiased, and hence estimates from many small samples can be combined by (weighted) averaging.
(b) It facilitates the estimation of an isolated value of $P(t)$.
(c) It may be preferred in some cases of dependence between lifetimes and observation limits (see 3.2 below).

3.2 *Dependence of deaths and losses.* The assumption that the full lifetimes T_i are independent of the observation limits L_i is sometimes violated, as a result of a change in the population sampled, or the conditions leading to the event called death, or the method of sampling the population. For example, in a study of survival after an operation, a change in surgical technique five years before the data are analyzed will affect the survival times only of those with observation limit less than five years. When loss is due to unforeseen circumstances, such as patients moving out of the state, the possibilities for dependence are obvious. Merrell [23] and Sartwell [24] have emphasized this point.

The estimate of $P(t)$ from the sample as a whole, whether PL or RS, involves arbitrary assumptions whose danger is peculiar to the dependent case and hence easily overlooked. The PL estimate in effect assumes that for items having an

observation limit L less than t, the conditional survival probability $P(t)/P(L)$ is the same as that for items whose observation limits exceed t, while the RS estimate makes the same assumption concerning the absolute survival probability $P(t)$ itself. In many applications, the PL assumption may seem as plausible as any; on the other hand, the fact that the results are expressed in terms of absolute probabilities may lead one to prefer the RS assumption (when there is no feeling to the contrary), if only because its operation is more easily visualized. For example, the RS estimate $P^*(t)$ (apart from sampling fluctuations) can never fall outside the range of the true values for the populations sampled, because it is a weighted average, whereas in unfavorable cases the PL estimate $\hat{P}(t)$ may fall outside this range.

If prior observation limits are given for all the items and the sample is large enough, dependence may be inquired into by grouping the items according to these limits, making separate estimates of $P(t)$ for each group, and comparing the results. If some unanticipated losses occur in advance of the a priori limits, the PL method can be used within each group. The group estimates can then be averaged with the initial numbers of items as weights to give something similar to the RS estimate, if that is desired.

On the other hand, no dependence can be demonstrated if one's information is limited to the values of the $t_i = \min(L_i, T_i)$ and their classification as deaths or losses; in this case the observed rates of death and loss can always be represented by a model in which deaths and losses occur independently, as well as by many models in which this is not so. However, one usually has or can obtain other information that is more or less useful.

What one would like is to obtain for each item the value of an auxiliary variable V (which may be either quantitative or qualitative), such that the T_i and L_i are more nearly independent within subsamples defined in terms of V, than they are in the sample or population as a whole. This will generally be true if V is strongly related to the cause of loss, and independent of the cause of death, or vice versa. In this case the subsamples are formed so as to reduce the variability of the L_i (or the T_i) within them. Even if the auxiliary variable does not have the desired properties, it may be worthwhile to get an estimate of $P(t)$ by a different route for the sake of comparison. Also, the auxiliary variable may be of interest for its own sake, for example, classification of the items by starting date would be aimed at detecting a temporal change in the population. As pointed out in 3.1, the procedure in the general case is not satisfactory if the estimates depend on very few items.

The innocuous form of dependence that results from testing with replacement is considered briefly in Section 7.

4. ACTUARIAL ESTIMATES

4.1 *Estimates using only n, δ, λ.* It has already been indicated (in 2.1) that actuarial estimates, like the PL, are formed by multiplying together a sequence of estimates of conditional probabilities of survival through intervals $(0, u_1)$, (u_1, u_2), $\cdots$. Unlike the PL, the actuarial estimates will generally be somewhat dependent on the selection of these intervals. It remains to consider how a typical factor p of this sort can be estimated when the number of items n at the

beginning of the interval is known to be depleted by δ deaths and λ losses within the interval, but the order in which these occur is not known or used. The PL estimate is denoted by $\hat{p}$.

As in 2.1, the estimates

$$\bar{p} = \frac{n - \delta}{n}, \qquad \underline{p} = \frac{n - \lambda - \delta}{n - \lambda} \tag{4a}$$

would be used if all the deaths were known to precede all the losses, and if all losses preceded all deaths, respectively. Evidently $\underline{p} \leq \hat{p} < \bar{p}$. If $\hat{p}$ is not known, another intermediate value is clearly supplied by the well-known "adjusted-observed" estimate (cf. [3], [12], [18], [22])

$$p^{(1)} = \frac{n - \lambda/2 - \delta}{n - \lambda/2} . \tag{4b}$$

This can be recommended for its simplicity even though

$$\frac{n - \lambda/2 - \delta}{n - \lambda/2} \cdot \frac{n - \delta/2 - \lambda}{n - \delta/2} \neq \frac{n - \delta - \lambda}{n} \tag{4c}$$

Here the second fraction is the corresponding estimate of the probability of not being lost (if losses are also random events), and the third fraction is the observed probability of escaping both death and loss. The only other estimate in this section that fails to give equality here is $p^{(4)}$ below.

The "joint risk" estimate (cf. [6] and [12]) is

$$p^{(2)} = \left[\frac{n - \delta - \lambda}{n}\right]^{\delta/(\delta+\lambda)} \tag{4d}$$

It is the maximum likelihood estimate when the losses are random, the instantaneous event-rate for losses is a constant times that for deaths, and the only data given are the values of n, δ, and λ. The above estimates satisfy the relation $\underline{p} \leq p^{(2)} \leq p^{(1)} \leq \bar{p}$.

One can obtain the PL estimate $\hat{p}$ from any of the foregoing by dividing the given interval into smaller intervals no one of which contains both deaths and losses. Since there are reasons (full use of information, absence of arbitrary assumptions, essential uniqueness of the estimate) for preferring small intervals to large, it seems reasonable to regard $\hat{p}$ as the standard to which the large-interval actuarial methods are approximations.

The error incurred in an actuarial estimate of p can be attributed to two sources, namely, the sampling error of $\hat{p}$, reflecting the fact that δ is a random variable, and the discrepancy between $p^{(1)}$ and $\hat{p}$, reflecting the unknown arrangement of deaths and losses within the interval. These sources of error will be compared by calculating the variance of $\log \hat{p}$ in each case, on the assumption that n is large compared with δ and λ, and that factors of the form $1+0(n^{-2})$ can be neglected. On this basis the effect of variation in δ is indicated sufficiently by the square of the coefficient of variation of $p^{(1)}$, namely

$$\delta/(n - \lambda/2)(n - \delta - \lambda/2) \doteq 4\delta(2n - \delta - \lambda)^{-2}. \tag{4e}$$

To evaluate the second source of error we note first that the total range of possible variation in $\log \hat{p}$ (given δ and λ) is by (4a)

$$\log \bar{p} - \log \underline{p} \doteq 4\delta\lambda(2n - \delta - \lambda)^{-2}. \tag{4f}$$

However, if the arrangement is random, the extreme values may be unlikely. Therefore it will now be assumed that all permutations of the δ deaths and λ losses are equally likely (which happens to be a consequence of the proportional event-rates leading to (4d)). In this case $\log \hat{p}$ is minus the sum of a random sample of δ of the $\delta+\lambda$ numbers

$$\log\left(1 + \frac{1}{n-1}\right), \log\left(1 + \frac{1}{n-2}\right), \cdots, \log\left(1 + \frac{1}{n-\delta-\lambda}\right).$$

Sampling theory and the Euler-Maclaurin theorem lead to the approximate variance

$$4\delta\lambda(\delta + \lambda + 1)/3(2n - \delta - \lambda)^4. \tag{4g}$$

Dividing the sampling variance (4e) into the square of half the range (4f) of the grouping error gives

$$\delta\lambda^2(2n - \delta - \lambda)^{-2} \tag{4h}$$

which may be either >1 or <1. Dividing (4e) into the grouping variance (4g) gives

$$\lambda(\delta + \lambda + 1)/3(2n - \delta - \lambda)^2, \tag{4i}$$

which is always <1 (when $n>0$).

The above results suggest that rather large intervals could be used if enough was known about the mechanisms of death and loss, and that the principal source of error is the probable failure of the various permutations of deaths and losses within the interval to be equally likely, or to fit any other scheme that may be assumed. Evidence on this point could be obtained by examining the permutations that actually occur. For example, this is what is done in using (2a). Otherwise one should probably require at the least that (4h) be suitably small. This still assumes that grouping errors, like the sampling errors, are independent from interval to interval. If not, one is thrown back on the ultimate in conservatism, which is to assert only that $\prod_j \underline{p}_j \leq \hat{P} \leq \prod_j \bar{p}_j$.

4.2 *Estimates using average ages.* Several estimates will now be considered that make use of the individual ages of death and loss, although arbitrary age intervals continue to be used. The individual ages enter via the average age Δ of the δ deaths, and the average age Λ of the λ losses, both measured from the beginning of the age interval under consideration, which is of length h.

The first estimate assumes nothing about the losses, but assumes that the instantaneous death-rate is constant throughout the interval. The total exposure to the risk of death in the interval is then $(n-\delta-\lambda)h+\delta\Delta+\lambda\Lambda$, the maximum-likelihood estimate of the death-rate is $\delta/[(n-\delta-\lambda)h+\delta\Delta+\lambda\Lambda]$, and the corresponding estimate of the survival probability p is

$$p^{(3)} = \exp\left(-\delta/[n - \delta - \lambda + (\delta\Delta + \lambda\Lambda)/h]\right). \tag{4j}$$

By using $p^{(3)}$, Harris, Meier, and Tukey [18] have been able to give an estimation procedure for the more general and more difficult case in which the age of death is never known exactly, but only known to fall in some interval, semi-infinite or finite, bounded by one or two observation points which vary from item to item.

The PL estimate $\hat{p}$ can be derived from $p^{(3)}$ as well as from $p^{(1)}$ or $p^{(2)}$. Although the limit of the product of the estimates $p^{(3)}$ for the subintervals is not quite unique, the average value of $\log p^{(3)}$ corresponding to δ deaths occurring at the same age is

$$-\delta \int_0^1 \frac{du}{n - \delta(1 - u)} = \log \frac{n - \delta}{n}, \tag{4k}$$

in agreement with $\hat{p} = (n - \delta)/n$. Here $\Delta (= uh)$ has been regarded as uniformly distributed between 0 and h. However, in other respects the behavior of $p^{(3)}$ does not parallel that of $\hat{p}$. If the position of one or more of the deaths is shifted toward the end of the interval, $p^{(3)}$ is increased, while $\hat{p}$ is decreased (or else not changed). Again, $p^{(3)}$ does not necessarily agree with $\hat{p}$ in the simple case of no losses. Finally, the extreme values of $p^{(3)}$, obtained by putting $\Delta = \Lambda = 0$ or h, fall outside the interval $\underline{p}$, $\bar{p}$ within which $\hat{p}$ is confined; in fact one can have $p^{(3)} < \underline{p}$ while $\hat{p} = \bar{p}$, or $p^{(3)} > \bar{p}$ while $\hat{p} = \underline{p}$.

The following estimates have been constructed with the object of using the mean ages Δ and Λ to simulate the behavior of $\hat{p}$:

$$p^{(4)} = (n - c\lambda - \delta)/(n - c\lambda), \tag{4l}$$

$$p^{(5)} = \left[\frac{n - \lambda - \delta}{n - \lambda}\right]^c \left[\frac{n - \delta}{n}\right]^{1-c} \tag{4m}$$

where

$$\begin{aligned} c &= 1/2 + (\Delta - \Lambda)/h \quad \text{for} \quad |\Delta - \Lambda| \leq h/2, \\ &= 0 \qquad\qquad\quad \text{for} \quad \Delta - \Lambda \leq -h/2, \\ &= 1 \qquad\qquad\quad \text{for} \quad \Delta - \Lambda \geq h/2. \end{aligned} \tag{4n}$$

The value of c is immaterial when $\delta = 0$ or $\lambda = 0$.

These estimates are two different averages of the bounding estimates $\underline{p}$ and $\bar{p}$, taken with weights c and $1 - c$ respectively. The quantity c is so defined that the appropriate one of the bounding estimates is assigned as soon as $|\Delta - \Lambda| \geq h/2$. This is intuitively reasonable, and is justified analytically by the fact that $\log p^{(4)}$ and $\log p^{(5)}$ have essentially the variance (4g) of $\log \hat{p}$, if each death and each loss is uniformly and independently distributed in the interval. To estimate this variance, one can multiply

$$\text{var } c \doteq \text{var } [(\Delta - \Lambda)/h] = (\delta + \lambda)/12\delta\lambda$$

by the square of the range (4f), since (4f) gives the change in $\log p^{(4)}$ and $\log p^{(5)}$ corresponding to a unit change in c.

Experience may indicate which estimate is preferable in a given type of situation. If relevant experience is lacking, it would seem advisable to rely

primarily on adequately small intervals, and to minimize the computation time by using simple estimates. This suggests that $p^{(1)}$ be used when the deaths and losses are thought to be arranged at random within the interval, $\bar{p}$ when the interval is chosen so that most of the deaths precede most of the losses, and $\underline{p}$ when losses precede deaths.

This completes the material that seems to be most important for applications. The remaining three sections are devoted primarily to mathematical derivations.

5. MAXIMUM LIKELIHOOD DERIVATION OF THE PL

In accordance with 1.2 the method of maximum likelihood is viewed as a means for selecting the best fitting distribution from a class of admissible distributions. It will now be shown that if all possible distributions are admitted, the product-limit is the one for which the sample likelihood is a maximum.

Let $T_1<T_2<\cdots, <T_k$ denote the distinct ages of death, with δ_j deaths observed at T_j. Let λ_j denote the number of losses in the interval $[T_j, T_{j+1})$, including any losses at T_j but none at T_{j+1}. (Here $j=0, 1, \cdots, k$; $T_0=0$; $T_{k+1}=\infty$.) The ages of the λ_j losses are denoted by $L_i^{(j)}$, $i=1, 2, \cdots, \lambda_j$. Only in this section, $P(t)$ will represent not the "true" survivorship but rather an arbitrary one that is to be determined so as to maximize the likelihood. The likelihood may now be written

$$\begin{aligned} &\left[\prod_{i=0}^{\lambda_0} P(L_i^{(0)})\right] \\ &\cdot [P(T_1-0)-P(T_1)]^{\delta_1}\left[\prod_{i=1}^{\lambda_1} P(L_i^{(1)})\right] \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdot [P(T_k-0)-P(T_k)]^{\delta_k}\left[\prod_{i=1}^{\lambda_k} P(L_i^{(k)})\right]. \end{aligned} \tag{5a}$$

Here $P(T_j-0)$ includes, whereas $P(T_j)$ excludes, the probability of death at age T_j exactly.

Clearly (5a) will be maximized by making the $P(L_i^{(j)})$ and the $P(T_j-0)$ as large, and the $P(T_j)$ as small, as is consistent with the monotonicity of the function and the assumption $T_j\leq L_i^{(j)}<T_{j+1}$. This means that $P(L_i^{(0)})=1$ and

$$P(T_j) = P(L_i^{(j)}) = P(T_{j+1}-0) \tag{5b}$$

for all i and j. If the common value is denoted by P_j, (5a) reduces to

$$\prod_{j=1}^{k}(P_{j-1}-P_j)^{\delta_j}P_j^{\lambda_j} \quad \text{with} \quad P_0=1. \tag{5c}$$

Now writing $p_j=1-q_j=P_j/P_{j-1}$ gives $P_j=p_1p_2\cdots p_j$ and $P_{j-1}-P_j=p_1\cdots p_{j-1}q_j$. With these substitutions (5c) becomes

$$\prod_{j=1}^{k} p_j^{n_j-\delta_j}q_j^{\delta_j}, \tag{5d}$$

where

$$n_j = N - \sum_{i=0}^{i=j-1} \lambda_i - \sum_{i=1}^{i=j-1} \delta_i = \sum_{i=j}^{i=k} (\lambda_i + \delta_i), \quad \text{or} \quad n(T_j - 0)$$

in the notation of 1.4. In this form each factor with a fixed index j is maximized individually by the binomial estimate $p_j = \hat{p}_j = (n_j - \delta_j)/n_j$, in agreement with the PL in (2a). If $\lambda_k > 0$, so that the greatest observed lifetime t^* corresponds to a loss, then $\hat{P}(t^*) > 0$ and the likelihood of the sample is independent of the values of $P(t)$ for $t > t^*$. Thus $\hat{P}(t)$ and the maximum likelihood estimate are indeterminate under the same circumstances, and the identity between them is complete.

6. MEANS AND VARIANCES

In this section the mean of $\hat{P}(t)$ and the variances of $\hat{P}(t)$ and of the mean-life estimates $\hat{\mu}$ and μ^* are derived. The observation limits are regarded as fixed from sample to sample. In principle, unconditional results for random loss times could be obtained by integration.

6.1 *The PL estimate* $\hat{P}(t)$. Let $L_1 < L_2 < \cdots < L_{k-1}$ be the distinct observation limits that are less than the age $t = L_k$ at which $P(t)$ is being estimated. Let $n_{j+1} = n(L_j + 0)$ be the number of survivors[1] observed beyond L_j; $N_{j+1} = En_{j+1} = N^0(L_j + 0)P(L_j)$; $p_j = 1 - q_j = P(L_j)/P(L_{j-1})$; and δ_j the number of deaths observed in the interval $(L_{j-1}, L_j]$, excluding nonzero values of the left endpoint, with $L_0 = 0$. Then

$$\hat{P}(t) = \prod_{j=1}^{k} \hat{p}_j \quad \text{with} \quad \hat{p}_j = (n_j - \delta_j)/n_j. \tag{6a}$$

Let E_j denote a conditional expectation for $n_1, \cdots, n_j$ (and hence also $\delta_1, \cdots, \delta_{j-1}$) fixed. Then $E_j\hat{p}_j = p_j$, provided that $n_j > 0$. If the condition $n_j > 0$ could be ignored, one could write

$$\begin{aligned} E\hat{P}(t) &= E[\hat{p}_1 \cdots \hat{p}_{k-1}E_k\hat{p}_k] = E[\hat{p}_1 \cdots \hat{p}_{k-1}p_k] \\ &= p_k E[\hat{p}_1 \cdots \hat{p}_{k-2}E_{k-1}\hat{p}_{k-1}] \\ &= \cdots = p_k p_{k-1} \cdots p_1 = P(t), \end{aligned} \tag{6b}$$

so that $\hat{P}(t)$ would be an unbiased estimate. The flaw in this demonstration is the indeterminacy resulting when (for some j) $n_{j+1} = 0$ but $\hat{P}(L_j) > 0$, an event whose probability is bounded by (2d), and which could be obviated in principle by obtaining one or more supplementary observations. In the derivations of approximate formulas that follow, any bias that $\hat{P}(t)$ may have is neglected.

Both Greenwood [15] and Irwin [19] when giving formulas for the variance of an actuarial estimate of $P(t)$ actually treat a special case in which the actuarial estimator and $\hat{P}(t)$ coincide. Translated into our notation their common argument is the following.

Suppose we consider conditional variances for n_j held fixed at the values N_j.

[1] The expression $L_j + 0$ means that losses at L_j itself have been subtracted off.

Then the $\hat{p}_j$ may be treated as independent quantities and

$$E[\hat{P}^2(t)] = \prod_{j=1}^{k}\left(p_j^2 + \frac{p_j q_j}{N_j}\right) = p_1^2 \cdots p_k^2 \prod_{j=1}^{k}\left(1 + \frac{q_j}{N_j p_j}\right)$$
$$= P^2(t) \prod_{j=1}^{k}\left(1 + \frac{q_j}{N_j p_j}\right). \tag{6c}$$

If we ignore terms of order N_j^{-2} we have as an approximation for the variance

$$V[\hat{P}(t)] \doteq P^2(t) \sum_{j=1}^{k} \frac{q_j}{N_j p_j} \cdot \tag{6d}$$

This last expression we will refer to as "Greenwood's formula." The calculation of the variance with $n_1, \cdots, n_k$ fixed can apparently be justified only on the ground that it doesn't matter a great deal; the fixed n_j are unrealistic, and inconsistent with the previous assumption of a sample of fixed size and fixed limits of observation.

The authors have therefore rederived (6d) by means of successive conditional expectations. The procedure is to verify by induction the approximate relation

$$E_{j+1}[\hat{P}^2(t)] \doteq \hat{p}_1^2 \cdots \hat{p}_j^2\left(1 + \sum_{i=j+1}^{i=k} \frac{q_i}{p_i E_{j+1} n_i}\right) p_{j+1}^2 \cdots p_k^2 \tag{6e}$$

which reduces to (6d) when $j=0$. Applying E_j to (6e) has the effect of replacing j by $j-1$ provided that one sets

$$E_j \frac{\hat{p}_j^2}{E_{j+1} n_i} \doteq E_j\left[\hat{p}_j \cdot \frac{E_j \hat{p}_j}{E_j E_{j+1} n_i}\right]$$
$$\doteq p_j^2 / E_j n_i. \tag{6f}$$

wherein $\hat{p}_j / E_{j+1} n_i$ has been replaced by the ratio of the expectations of numerator and denominator. The fact that $\hat{p}_j$ and $E_{j+1} n_i$ are positively correlated improves the approximation.

In the special case in which $k=2$ (all observation limits less than t are equal to L_1) and any indeterminacy is resolved by one supplementary observation, the exact variance of $\hat{P}(t)$ can be calculated to be

$$V[\hat{P}(t)] = P^2(t)\left(\frac{q_1}{N p_1} + \frac{q_2}{N' p_1 p_2}\left[1 + \left(1 - \frac{N'}{N}\right)^2 \frac{N' p_1 - [N', p_1]}{[N', p_1]}\right.\right.$$
$$\left.\left. + \frac{N'}{N}\left(1 - \frac{N'}{N}\right)\left(\frac{q_1}{[N', p_1]} - 2q_1^{N'}\right)\right]\right), \tag{6g}$$

where N is the total sample size; $N' = N^0(t)$, the number of items observable to t or beyond; and

$$[N', p_1] = 1/E(X^{-1}) \tag{6h}$$

with $X = \max(Y, 1)$ and Y distributed binomially (N', p_1). The methods used by Stephan [27] show that $[N', p_1] = N' p_1 + 0(1)$ for N' large. Thus the term

in square brackets is of order $1+0(1/N')$, and (6g) approaches Greenwood's formula (6d) for large N'.

6.2 *Covariances and mean lifetimes.* The method of Irwin [19] can be used to derive the variance of $\hat{\mu}=\int_0^\infty \hat{P}(t)dt$ by writing

$$\begin{aligned} E(\hat{\mu}^2) &= E\int_0^\infty\int_0^\infty \hat{P}(u)\hat{P}(v)dvdu \\ &= 2\int_0^\infty\int_u^\infty E[\hat{P}(u)\hat{P}(v)]dvdu. \end{aligned} \tag{6i}$$

If $u<v$ and L_h and L_k denote u and v respectively, then apart from bias due to indeterminate cases one has

$$\begin{aligned} E[\hat{P}(u)\hat{P}(v)] &= E[\hat{P}^2(u)E_{h+1}(\hat{p}_{h+1}\hat{p}_{h+2}\cdots\hat{p}_k)] \\ &\doteq p_{h+1}p_{h+2}\cdots p_k E[\hat{P}^2(u)] \\ &\doteq [P(v)/P(u)]E[\hat{P}^2(u)] \doteq P(u)P(v)[1+U(u)], \end{aligned} \tag{6j}$$

where by (2h) and (2g)

$$U(u)=\int_0^u \frac{|\,dP(t)\,|}{N(t)P(t)} \doteq \sum [(N-r)(N-r+1)]^{-1};$$

the summation is over deaths at ages not exceeding u, and gives the sample estimate. Thus one has approximately

$$\begin{aligned} V(\hat{\mu}) &\doteq 2\int_0^\infty\int_u^\infty P(u)P(v)U(u)dvdu \\ &= 2\int_0^\infty A(u)P(u)U(u)du = \int_0^\infty A^2(u)dU(u), \end{aligned} \tag{6k}$$

where $A(u)=\int_u^\infty P(v)dv$. The result is discussed in 2.3.

To obtain analogous (but in this case exact) results for the RS estimates $P^*(t)$ and μ^*, let $u<v$ as before, and split off two independent subsamples consisting of the $N'=N^0(v)$ items observable to v or beyond, and the $N''=N^0(u)-N^0(v)$ additional items observable to u or beyond, but not to v. Let the first subsample yield δ' deaths in $(0, u)$ and ϵ' in (u, v), while the second yields δ'' deaths in $(0, u)$. Also let $P_1=P(u)$ and $P_2=P(v)$. Then

$$\begin{aligned} & E[P^*(u)P^*(v)] \qquad (\text{with } u \le v) \\ &= E[(N'+N''-\delta'-\delta'')(N'-\delta'-\epsilon')]/N'(N'+N'') \\ &= E[(N'+N''-\delta'-\delta'')(N'-\delta')]P_2/P_1N'(N'+N'') \\ &= [E(N'-\delta')^2+E(N''-\delta'')\cdot E(N'-\delta')]P_2/P_1N'(N'+N'') \\ &= [N'^2P_1^2+N'P_1(1-P_1)+N'N''P_1^2]P_2/P_1N'(N'+N'') \\ &= P(u)P(v)+P(v)[1-P(u)]/N^0(u). \end{aligned} \tag{6l}$$

Substituting this result in the analogue of (6i) and subtracting μ^2 gives $V(\mu^*)$ as in (3d).

7. CONSISTENCY; TESTING WITH REPLACEMENT

The RS estimate $P^*(t)$, being binomial, is of course consistent. In view of the approximate nature of the formula for the variance of $\hat{P}(t)$, however, a proof of its consistency is in order. In fact, by giving closer attention to the errors of the approximations in the variance derivation, it can be shown that $\hat{P}(t)$ does have the limit $P(t)$ in probability provided only that $N^0(t) \to \infty$. However, the proof is too lengthy to be given here. Of course, the consistency is obvious enough in the special case in which the number k of conditional probabilities to be estimated remains bounded as $N^0(t) \to \infty$.

Thus far the independence of deaths and losses has been assumed. The question of consistency may also be raised in the differing context of "testing with replacement." (Cf. [7], [10], [11], [14].) This is the common life-testing situation in which a fixed number of items, say ν of them, are always under test; when one dies it is replaced by another. The only preassigned constants are the duration of the test, S, and the number, ν, of items on test at all times. The number of losses is then fixed at ν, one for each item on test at the conclusion of the experiment, while the number of deaths observed, and hence the total number of items, is a random variable. The first ν items on test have S as their limit of observation; if one of these dies at age t, with $t<S$, its successor has $S-t$ as its limit of observation, and so on. Consistency will mean an approach to $P(t)$ as $\nu S \to \infty$, which implies that either $\nu \to \infty$ or $S \to \infty$ (or both).

The RS estimate will be considered first. Let $T_1, T_2, \cdots$, be an infinite sequence of independent lifetimes with survivorship $P(t)$ and let $P_k(t)$ be the proportion of $T_1, \cdots, T_k$ that exceed t. By the strong law of large numbers, $\lim P_k(t) = P(t)$ with probability one as $k \to \infty$, for any specified t. Now the RS estimate $P^*(t)$ (which exists only for values of t in the interval 0, S) can be represented as a value of $P_k(t)$, with k a random variable. This can be done by assigning the lifetimes $T_1, \cdots, T_\nu$ to the ν items initially tested. If some of these items have replacements whose limits of observation $S-T_i$ $(i=1, \cdots, \nu)$ are not less than t, then succeeding lifetimes $T_{\nu+1}, T_{\nu+2}, \cdots$, are assigned to these replacements. The process is repeated until no more observation intervals of at least the duration t are left. To prove $P^*(t)$ consistent, it only remains to prove that for any M, $\Pr(k>M) \to 1$ as $\nu S \to \infty$. If $\nu \to \infty$ the result is evident. If $S \to \infty$, then even for $\nu = 1$,

$$\begin{aligned} \Pr(k > M) &= \Pr(T_1 + \cdots + T_M \leq S - t) \\ &\geq \Pr[T_i \leq (S-t)/M \text{ for } i = 1, \cdots, M] \qquad (7a) \\ &= \{1 - P[(S-t)/M]\}^M. \end{aligned}$$

As $S \to \infty$ the last expression has the limit unity and consistency is proved, provided that $P(\infty) = 0$ (eventual death is certain).

When $S \to \infty$, ν is fixed, and $P(\infty) = 0$, the consistency of the PL estimate $\hat{P}(t)$ follows from the relation (cf. (4a))

$$\frac{n(t)}{N^0(t) + D'} \leq \hat{P}(t) \leq \frac{n(t) + \nu}{N^0(t) + D' + \nu} = \frac{n(t) + \nu}{N} \qquad (7b)$$

since the number of losses is ν, $P^*(t) = n(t)/N^0(t)$ is consistent, $N^0(t) \to \infty$, and the number of deaths D' having observation limits less than t is a random variable that does not depend significantly on S. $\hat{P}(t)$ is undoubtedly consistent also when S is fixed and $\nu \to \infty$, but a rigorous proof of this has not been constructed.

ACKNOWLEDGMENT

The viewpoint adopted in this paper owes much to discussions with John W. Tukey. It also incorporates a number of suggestions kindly forwarded by readers of a preliminary draft.

REFERENCES

[1] Bailey, W. G., and Haycocks, H. W., "A synthesis of methods of deriving measures of decrement from observed data," *Journal of the Institute of Actuaries*, 73 (1947), 179–212.

[2] Bartlett, M. S., "On the statistical estimation of mean lifetime," *Philosophical Magazine* (Series 7), 44 (1953), 249–62.

[3] Berkson, J., and Gage, R. P., "Calculation of survival rates for cancer," *Proceedings of the Staff Meetings of the Mayo Clinic*, 25 (1950), 270–86.

[4] Berkson, J., and Gage, R. P., "Survival curve for cancer patients following treatment," *Journal of the American Statistical Association*, 47 (1952), 501–15.

[5] Berkson, J., "Estimation of the interval rate in actuarial calculations: a critique of the person-years concept," (Summary) *Journal of the American Statistical Association* 49 (1954), 363.

[6] Böhmer, P. E., "Theorie der unabhängigen Wahrscheinlichkeiten," Rapports, Mémoires et Procès-verbaux de Septième Congrès International d'Actuaires, Amsterdam, 2 (1912), 327–43.

[7] Brown, G. W. and Flood, M. M., "Tumbler mortality," *Journal of the American Statistical Association*, 42 (1947), 562–74.

[8] Cornfield, J., "Cancer illness among residents in Atlanta, Georgia," Public Health Service Publication No. 13, Cancer Morbidity Series No. 1, 1950.

[9] Davis, D. J., "An analysis of some failure data," *Journal of the American Statistical Association* 47 (1952), 113–150.

[10] Epstein, Benjamin, and Sobel, Milton, "Lifetesting," *Journal of the American Statistical Association*, 48 (1953), 486–502.

[11] Epstein, Benjamin, and Sobel, Milton, "Some theorems relevant to lifetesting from an exponential distribution," *Annals of Mathematical Statistics*, 25 (1954), 373–81.

[12] Fix, Evelyn, "Practical implications of certain stochastic models on different methods of follow-up studies," paper presented at 1951 annual meeting of the Western Branch, American Public Health Association, November 1, 1951.

[13] Fix, Evelyn, and Neyman, J., "A simple stochastic model of recovery, relapse, death and loss of patients," *Human Biology*, 23 (1951), 205–41.

[14] Goodman, L. A., "Methods of measuring useful life of equipment under operational conditions," *Journal of the American Statistical Association*, 48 (1953) 503–30.

[15] Greenwood, Major, "The natural duration of cancer," Reports on Public Health and Medical Subjects, No. 33 (1926), His Majesty's Stationery Office.

[16] Greville, T. N. E., "Mortality tables analyzed by cause of death," *The Record*, American Institute of Actuaries, 37 (1948), 283–94.

[17] Hald, A., "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skandinavisk Aktuarietidskrift*, 32 (1949) 119–34.

[18] Harris, T. E., Meier, P., and Tukey, J. W., "Timing of the distribution of events between observations," *Human Biology*, 22 (1950), 249–70.

[19] Irwin, J. O., "The standard error of an estimate of expectational life," *Journal of Hygiene*, 47 (1949), 188–9.

[20] Jablon, Seymour, "Testing of survival rates as computed from life tables," unpublished memorandum, June 14, 1951.
[21] Kahn, H. A., and Mantel, Nathan, "Variance of estimated proportions withdrawing in single decrement follow-up studies when cases are lost from observation," unpublished manuscript.
[22] Littell, A. S., "Estimation of the T-year survival rate from follow-up studies over a limited period of time," *Human Biology*, 24 (1952), 87–116.
[23] Merrell, Margaret, "Time-specific life tables contrasted with observed survivorship," *Biometrics*, 3 (1947), 129–36.
[24] Sartwell, P. E., and Merrell, M., "Influence of the dynamic character of chronic disease on the interpretation of morbidity rates," *American Journal of Public Health*, 42 (1952), 579–84.
[25] Savage, I. R., "Bibliography of nonparametric statistics and related topics," *Journal of the American Statistical Association*, 48 (1953), 844–906.
[26] Seal, H. L., "The estimation of mortality and other decremental probabilities," *Skandinavisk Aktuarietidskrift*, 37 (1954), 137–62.
[27] Stephan, F. F., "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *Annals of Mathematical Statistics*, 16 (1945), 50–61.
[28] Wolfowitz, J., "Additive partition functions and a class of statistical hypotheses," *Annals of Mathematical Statistics*, 13 (1942), 247–79.