NONPARAMETRIC ESTIMATION OF LIFE DISTRIBUTIONS

by

MING-SHU HSU

B.S., Soochow University, 1976

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1980

Approved by:

R Nassar

Major Professor

SPEC COLL LD 2668 .R4 1980 H78 c.2

TABLE OF CONTENTS

I. INTRODUCTION

In the usual formulation of statistical decision theory the probability distribution of the observations is assumed to be a member of some specified class of distribution functions. Under those assumptions, many estimators, such as estimators of mean, median or variance, were derived and discussed. However, in many instances, we do not know if the assumed conditions are appropriate. To overcome this difficulty, other methods that are nonparametric in nature must be considered.

Recently, a great deal of research has been undertaken in nonparametric methods of estimation of life distributions from which the probability of failure at any given time can be estimated.

Many authors, for example, Kaplan and Meier (10), Ferguson (6,7), Hollander and Korwar (9,12), Susarla and Van Ryzin (21,22,23), Barlow and Scheuer (2), Breslow and Crowley (3), Shaked, et al. (18), Ferguson and Phadia (8), consider nonparametric estimation of the life distribution function for many kinds of data. The development of nonparametric analysis in the area of reliability can be found in Shimi and Tsokos (19).

In Section 2 the development of nonparametric estimation of life distribution functions is discussed. Definitions and results of the Dirichlet process and of a process that is neutral to the right, which are very useful with respect to some nonparametric decision theoretic problems, are also presented.

Section 3 includes a summary of several types of data in life testing and of 7 methods of nonparametric estimation of life distribution functions.

In Section 4 we discuss each method given in Section 3 according to its usefulness, comprehension, and accuracy.

The final section provides some proposals for future work.

II. REVIEW

In terms of the nonparametric estimation of life distribution, the classical approach is to use the sample distribution function with non-accelerated type II censored data (see Fig. 1). Recently, nonparametric estimation methods have been developed for use with other types of data.

In 1958, Kaplan and Meier (10) developed several nonparametric estimators for incomplete observations. Among those estimators, the most commonly used one is the product limit (PL) estimator. Breslow and Crowley (3) in 1974 derived properties of the PL estimator.

Another approach to the development of nonparametric estimation of life distribution requires the use of accelerated data. This method was first introduced by Barlow and Schener (2) in 1971 by assuming stochastic ordering. Steck, et al. (20) in 1974 used the functional relationship method. Shaked et al. (18), in 1979, pointed out that both papers suffer from the disadvantage that at least a small sample of nonaccelerated observations are needed and suggested a further improvement by the use of only accelerated data.

In 1973 Ferguson (6) suggested the Bayesian approach in solving nonparametric decision problems. He introduced a class of random probabilities called Dirichlet processes (5). According to Ferguson, the Dirichlet process has the following two desirable properties as a prior distribution for nonparametric problems:

(I) It has a large or nonparametric class of probabilities as its support in the topology of weak convergence.

(II) Posterior distribution given a sample of observations from the Dirichlet process is manageable analytically, and is also a Dirichlet process.

Using the concept of the Dirichlet process, Ferguson (6), Korwar and Hollander (9,12), Susarla and Van Ryzin (21,22,23) developed several useful nonparametric estimators of distribution functions and investigated their properties.

Since the Dirichlet process is, with probability one, a discrete probability, a more general process called the process neutral to the right has been presented by Doksum (5) in 1974. In 1979, Ferguson and Phadia (8) applied the process neutral to the right as a prior to estimate the survival function.

The following is a description of some basic definitions and results of the Dirichlet process and the process neutral to the right. [See (1,5, 6,7,11) for more comprehensive coverage.]

The Dirichlet Process

Definition 2.1. [Ferguson (6,9)] Let Z_1, Z_2, ..., Z_k be independent random variables with Z_j having a gamma distribution with shape parameter $\alpha_j \geq 0$ and scale parameter 1 for all j. Let $\alpha_j > 0$ for some j. The Dirichlet distribution with parameter $(\alpha_1,\ldots,\alpha_k)$ denoted by $D(\alpha_1,\ldots,\alpha_k)$ is defined as the distribution of $(Y_1,\ldots,Y_k)$, where $Y_j = Z_j/\Sigma_{i=1}^{k} Z_i$, j=1,2,...,k.

This distribution is always singular with respect to Lebesque measure in k-dimensional space since $Y_1 + \ldots + Y_k = 1$. Besides, if any $\alpha_j = 0$, the corresponding Y_j is degenerate at zero. However, if $\alpha_i > 0$ for all i=1,2,...,k, the (k-1) dimensional distribution of $(Y_1,\ldots,Y_{k-1})$ is absolutely continuous with density

$$f(Y_1,\ldots,Y_{k-1}|\alpha_1,\ldots,\alpha_k) \tag{2.1}$$

$$= \frac{(\alpha_1 + \ldots + \alpha_k)}{(\alpha_1) \;\ldots\; (\alpha_k)} \left(\prod_{i=1}^{k-1} Y_i^{\alpha_i - 1}\right) \left(1 - \prod_{i=1}^{k-1} Y_i\right)^{\alpha_k - 1} I_S\,(Y_1,\ldots,Y_{k-1}),$$

where S is the simplex

$$S = \{(Y_1,\ldots,Y_{k-1}):Y_i \geq 0, \sum_{i=1}^{k-1} Y_i \geq 1\}$$

For k=2, (2.1) becomes the Beta distribution, $Be(\alpha_1,\alpha_2)$.

For ease of exposition, we restrict attention, unless otherwise specified, to prior distributions on the space of all probability measures on (R,B) where R is the real line and B is the σ-algebra of Boreal subsets of R.

Definition 2.2 [Ferguson (7)] Let $\alpha(\cdot)$ be a finite non-null measure (nonnegative and finitely additive set function) on (R,B), and let $P(\cdot)$ be a stochastic process indexed by elements of B. Then P is a Dirichlet process on (R,B) with parameter α write $P\varepsilon D(\alpha)$ if for every finite measurable partition $\{B_1,\ldots,B_m\}$ of R (i.e., the B_i are measurable, disjoint, and $(\cup_1^m B_i = \mathrm{R})$, the random vector $(P(B_1),\ldots,(B_m))$ has a Dirichlet distribution with parameter $(\alpha(B_1),\ldots,\alpha(B_m))$.

In particular, for every $B\varepsilon\mathrm{B}$, $P(B)\varepsilon Be(\alpha(B),\alpha(\mathrm{R})-\alpha(B))$ and therefore $E[P(B)] = \alpha(B)/\alpha(\mathrm{R})$.

Definition 2.3 *[Ferguson (6)] Let P be a random probability measure on* (R,B). We say that $X_1,\ldots,X_n$ is a sample of size n from P if for any m=1,2... and measurable sets $A_1,\ldots,A_m$, $C_1,\ldots,C_n$

$$P\ \{X_1\varepsilon C_1,\ldots,X_n\varepsilon C_n|P(A_1),\ldots,P(A_m),P(C_1),\ldots,P(C_n)\} \qquad (2.2)$$

$$= \Pi_{j=1}^n\ P(C_j) \text{ a.s.}$$

Ferguson (6,7) and Korwar et al. (9,12) derived some useful theorems governing the properties of the Dirichlet process. We shall list those theorems without proof.

<u>Theorem 2.4</u> [Ferguson (6)] If $F\varepsilon D(\alpha)$ and if $X_1,\ldots,X_n$ is a sample from F, then the posterior distribution of F given $X_1,\ldots,X_n$ is $D(\alpha+\Sigma_1^n\delta_{X_i})$, where $\delta_x(A)=1$ if $X\varepsilon A$, and is 0 otherwise.

<u>Theorem 2.5</u> [Korwar and Hollander (9,12)] Let P be a Dirichlet process on (R,B) with parameter α and let $X_1,\ldots,X_m$ be a sample of size m from P. Then

$$P\{X_1 \leq x_1,\ldots,X_m \leq x_m\} \tag{2.3}$$

$$= \{\alpha(A_{x_{(1)}})\ldots(\alpha(A_{x_{(m)}})+m-1)\}/\{\alpha(\mathrm{R})\ldots(\alpha(\mathrm{R})+m-1)\},$$

where $X_{(1)}\ \ldots\ X_{(m)}$ is the ordered values among $X_1,\ldots,X_m$, and $A_x=(-\infty,x]$.

<u>Theorem 2.6</u> [Ferguson (6,7)] If $P\varepsilon D(\alpha)$, then P is discrete with probability one.

<u>Process neutral to the right.</u>

We present one of the definitions of neutral to the right which is rather easy to comprehend. [For more details, see (5,7,8).]

<u>Definition 2.7</u>. A process F(t) is said to be a random distribution function (i.e., (a) F(t) is nondecreasing a.s., (b) F(t) is right-continuous a.s., (c) $\lim_{t\to-\infty}F(t)=0$ a.s. and (d) $\lim_{t\to\infty}F(t)=1$ a.s.) neutral to the right if *it can be written in the form* $F(t)=1-e^{-Y_t}$ *where* Y_t is a process with independent increments such that (a) Y_t is nondecreasing a.s., (b) Y_t is right continuous a.s., (c) $\lim_{t\to-\infty}Y_t=0$ a.s., and (d) $\lim_{t\to\infty}Y_t=\infty$ a.s.

A process such as Y_t, described in Definition 2.7, has at most countably many fixed points of discontinuity $t_1,t_2,\ldots$. Let $S_1,S_2,\ldots$ be the random heights of the jump in Y_t at $t_1,t_2,\ldots$ respectively. Then $S_1,S_2,\ldots$ are independent nonnegative (possibly infinite-valued) random variables with corresponding densities $f_{t_1},f_{t_2},\ldots$ Let Z_t denote the same random variable as Y_t but with the jumps removed. Then $Z_t=Y_t-\Sigma_j S_j I_{(t_j,\infty)}(t)$ and Z_t is a nondecreasing process with independent increments and Z_t has no fixed points of discontinuity, and therefore has an infinitely divisible distribution with Levy formula for the log of the moment generating function.

$$\text{Log } E[e^{-\theta Z_t}] = -\theta b(t)+\int_0^\infty (e^{-\theta z}-1)\, dN_t(z) \tag{2.4}$$

where b is a nondecreasing continuous function with $b(t)\to 0$ as $t\to-\infty$, and where N_t is a continuous Levy measure; that is,

(i) for every Borel set $B\epsilon\mathcal{B}$, $N_t(B)$ is nondecreasing and continuous.

(ii) for every real t, $N_t(\cdot)$ is a measure on the Borel subsets of $(0,\infty)$

(iii) $\int_0^\infty z(1+z)^{-1}\, dN_t(z)\to 0$ as $t\to-\infty$

From the above definition, we can see that the process neutral to the right is specified by the four quantities $\{t_1,t_2,\ldots\}$, $\{f_{t_1},f_{t_2},\ldots\}$, b, and N_t.

The main results of Doksum (5) for the process neutral to the right are presented in the following theorems.

<u>Theorem 2.8</u> If F is a random distribution function which is neutral to the right then the posterior distribution of F given $X_1,\ldots,X_n$ is neutral to the right. Ferguson and Phadia (7,8) gave an alternative description of Doksum's result in terms of the distribution of the process Y_t for the sample size n=1. The general case of arbitrary sample size would follow by repeated application.

<u>Theorem 2.9</u> Let F be a random distribution function neutral to the right, $F(t)=1-e^{-Y_t}$, and let X be a sample of size one from F. Then the posterior distribution of Y_t given X=x is best treated in two cases.

Case 1. If x is one of the prior fixed points of discontinuity, say $x=t_k$, then the posterior density of the jump in Y_t at x given X=x may be found by multiplying the prior density of the jump by $(1-e^{-s})$ and renormalizing.

Thus,

$$dH_x(s) = (1-e^{-s})dG_x(s)/\int_0^\infty(1-e^{-s})dG_x(s) \tag{2.5}$$

Case 2. If x is not one of the prior points of discontinuity, then the posterior distribution of an increment in Y_t to the left of x may be found by multiplying the prior density of the increment by e^{-y} and renormalizing; that is:

$$dH(s) = e^{-s}dG(s)/\int_0^\infty e^{-s}dG(s) \tag{2.6}$$

Where G is the prior distribution and H is the posterior distribution given X=x.

In the process neutral to the right, there are two cases, one is homogeneous, the other is nonhomogeneous. The definition of the neutral to the right homogeneous process is as follows.

Definition 2.10 (8) A random distribution function F neutral to the right is said to be homogeneous if the independent process Y_t=-log(1-F(t)) has Levy function independent of t; that is, if the MGF has the form

$$M_t(\theta)=e^{\nu(t)\int_0^\infty(e^{-\theta z}-1)\ dN(z)} \tag{2.7}$$

where $\nu(t)$ is continuous nondecreasing, $\lim_{t\to-\infty}\nu(t)=0$, $\lim_{t\to+\infty}\nu(t)=+\infty$ and where N is any measure on $(0,\infty)$ such that $\int_0^\infty z(1+z)^{-1}dN(z)<\infty$.

The following theorem describes the relationship between the Dirichlet process and the process neutral to the right.

Theorem 2.11 (5,7) If $F\epsilon D(\alpha)$, then F is a nonhomogeneous process neutral to the right, and if α is continuous, then Y_t=-log(1-F(t)) has no fixed points of discontinuity.

This implies that if $X\epsilon Be(\alpha,\beta)$ then Y=-log(1-X) is infinitely divisible. The density of Y is:

$$f_y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{-\beta y}(1-e^{-y})^{\alpha-1} I_{(0,\infty)}(y) \tag{2.8}$$

and the moment generating function of Y is:

$$M_y(u) = E[e^{uy}] = \frac{\Gamma(\alpha+\beta)\Gamma(\beta-u)}{\Gamma(\beta)\Gamma(\alpha+\beta-u)} \quad \text{for } u < \beta \tag{2.9}$$

III. ESTIMATION MODELS

3.1 Type of data. Life testing has the following common sampling forms. (See Figure 1 for classification.) (I) Accelerated sample: Samples of certain devices are subject to conditions of greater stress than that encountered under normal operation, and from the results for those high-stress environments (may or may not include normal stress), an estimate of performance of the device under normal operation is obtained. This sampling method is used when lifetime tends to be long and the time consumed in testing a sample of a certain device may be excessive. (II) Nonaccelerated sample: Samples are tested under conditions of normal operation only.

The above sampling schemes are distinguished by the following types of data.

(1) Type I censored data: A test is conducted on n items, as as each failure occurs, the time is recorded. $X_{(1)},X_{(2)},\ldots,X_{(r)}$ are the observed ordered failure times of the r items, $r \leq n$. The test terminates at a preassigned time.

(2) Type II censored data: A test is conducted on n items and as each failure occurs, the time is recorded. $X_{(1)},X_{(2)},\ldots,X_{(r)}$ are the observed ordered lifetimes of the r items, $r \leq n$. The test terminates when a preassigned number of failures, r, has occurred.

(3) Mixed censored data: A test is conducted on n items and as each failure occurs, the time is recorded. $X_{(1)},X_{(2)},\ldots,X_{(r)}$ are observed lifetimes of the r items, $r \leq n$. The test terminates when a preassigned number of failures, r, has occurred or a preassigned time has been reached, which ever comes first.

In either type of data, we have two methods of sampling.

(i) With replacement: Items that fail are immediately replaced by new items having the same expected life distribution.

(ii) Without replacement: Items that fail are not replaced.

Moreover, in each operating method of Type I censored data there are three types of observations.

(i) Real observation: $X_i = x_i$

(ii) Right censored data: $X_i > x_i$ (exclusive censoring) or

$X_i \geq x_i$ (inclusive censoring)

This is usually encountered when one preassigns a different time (t_i) for each different sample, X_i.

(iii) Left censored data: $X_i < x_i$ (exclusive censoring) or

$X_i \leq x_i$ (inclusive censoring)

3.2 Estimation. In this section we shall review some useful approaches in nonparametric estimation of life distribution developed in the last decade or so. (See Table 1 for classification.)

(1) Ferguson's method (6). Suppose a random sample $X_1,\ldots,X_n$ is taken from a distribution F that is a random sample function of a Dirichlet process P with parameter $\alpha(\cdot)$. Take the loss function to be $L(F,\hat{F}) = \int_R (F(t)-\hat{F}(t))^2 dW(t)$ where W is a given finite measure on $(\mathbb{R},\mathcal{B})$ (a weight function) and $\hat{F}$ is an estimator of F. Then,

$$\hat{F}_n(t|X_1,\ldots,X_n) = P_n F_0(t) + (1-P_n) F_n(t|X_1,\ldots,X_n) \tag{3.1}$$

where

$$P_n = \alpha(\mathbb{R})/(\alpha(\mathbb{R})+n)$$

$$F_0(t) = \alpha((-\infty,t))/\alpha(\mathbb{R})$$

and $F_n(t|x_1,\ldots,x_n) = 1/n \sum_{i=1}^{n} \delta_{x_i}((-\infty,t])$ is the empirical distribution function of the sample.

(2) Hollander and Korwar's method (9,12). Let $(F_i, \underline{X}_i)$ i=1,2,...,n be a sequence of pairs of independent random elements. The F's are random probability measures which have a common prior distribution given by a Dirichlet process on (R,B). Assume $\alpha(\mathrm{R})$ is known. Given $F_i=F'$ (say), $\underline{X}_i=(X_{i1},\ldots, X_{im_i})$ is a random sample of size m_i from F'. (In Korwar and Hollander's paper (12) they assume $\underline{X}_i$ has equal sample size.) Under the same loss function as Ferguson's method the proposed sequence of estimator is, for i=1, 2,...,n

$$H_i(t) = P_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \hat{F}_j(t)/(n-1)+(1-P_i)\hat{F}_i(t) \tag{3.2}$$

where

$$P_i = \alpha(\mathrm{R})/(\alpha(\mathrm{R})+m_i) \tag{3.3}$$

and $\hat{F}_i$ is the empirical distribution function of $\underline{X}_i$, i=1,...,n.

Hollander and Korwar illustrated the use of the estimators defined by (3.2) by applying their methodology to the data from Proschan (17). The data consist of intervals between successive failures of the air conditioning systems of three jet airplanes. (9, p. 98)

From the data, n=3, m_1=30, m_2=27, and m_3=24. They considered the case where $\alpha(\mathrm{R})$ is specified to be 7. Then from (3.3) they obtained P_1 = 7/(7+30)=.19, P_2 = 7/(7+27)=.21, P_3 = 7/(7+24)=.23; so that,

$$H_1(t) = .19(\hat{F}_2(t)+\hat{F}_3(t))/2+.81(\hat{F}_1(t)),$$

$$H_2(t) = .21(\hat{F}_1(t)+\hat{F}_3(t))/2+.79(\hat{F}_2(t)),$$

$$H_3(t) + .23(\hat{F}_1(t)+\hat{F}_2(t))/2+.77(\hat{F}_3(t)).$$

(3) Kaplan and Meier's PL method (3,10). Let $T_1,\ldots,T_N$ be a random sample of values of the random variable T (called the lifetime), and $L_1,\ldots,L_N$ be a sample of the random variable L (called limits of observation) where T and L are assumed independent. We observe $t_i=\min(T_i,L_i)$ $i=1,2,\ldots,N$. For each item it is known whether one has

$$T_i \le L_i \quad t_i = T_i \text{ (a death)}$$

or

$$L_i < T_i \quad t_i = L_i \text{ (a loss)}$$

Let N be the total sample size. If one lists and labels the N observed lifetimes (whether to death or loss) in order of increasing magnitude $0 \le t_1' \le t_2' \le \ldots \le t_N'$, then the estimator of survival function is

$$P(t) = \prod_r [(N-r)/(N-r+1)] \tag{3.4}$$

where r assumes those values for which $t_r' \le t$, and t_r' measures the time to death.

As an example, consider the observed data:

Deaths at 0.8, 3.1, 5.4, 9.2 months

Losses at 1.0, 2.7, 7.0, 12.1 months

Here N=8 and the construction of the function $\hat{P}(t)$ proceeds as follows

$$\hat{P}(t) = \begin{cases} 1 & 0\le t<0.8 \\ (8-1)/(8-1+1)=7/8 & 0.8\le t<3.1 \\ (7/8)\times(8-4)/(8-4+1)=7/10 & 3.1\le t<5.4 \\ (7/8)\times(4/5)\times(3/4)=21/40 & 5.4\le 5<9.2 \\ (21/40)\times(1/2)=21/80 & 9.2\le t<12.1 \end{cases}$$

(4) Susarla and Van Ryzin's method (21). Let $X_1,\ldots,X_n$ be the true survival times of n individuals which are censored on the right by n follow-up times, $Y_1,\ldots,Y_n$. It is assumed that the X_i are independent identically distribution function $F(u)$, where F is distributed as a Dirichlet process on $R^+=(0,\infty)$, and that the parameter $\alpha(\cdot)$ is known. The observable data are:

$$Z_i = \min\{X_i, Y_i\}$$

$$\delta_i = \begin{cases} 1 \text{ if } X_i \leq Y_i \\ 0 \text{ if } X_i < Y_i \end{cases} \quad i=1,\ldots,n$$

Assume that $Y_1,\ldots,Y_n$ are mutually independent random variables which are also independent of $X_1,\ldots,X_n$ where Y_i is distributed as H_i, $H_i(u)=P_r(Y_i \leq u)$, $i=1,\ldots,n$. Note that if $\delta_i=1$, the Z_i in the pair (Z_i,δ_i) which is observed is a true lifetime; and if $\delta_i=0$, then Z_i is an exclusive right censored data. Let $Z_1,\ldots,Z_k$ be the real observations and $Z_{k+1},\ldots,Z_n$ be the exclusive right censored observations. Also, let $Z_{(k+1)},\ldots,Z_{(m)}$ denote the distinct observations among the exclusive right censored observations $Z_{k+1},\ldots,Z_n$. Let λ_j denote the number of exclusive right censored observations that are equal to $Z_{(j)}$, for $j=k+1,\ldots,m$, and let $N(u)$ and $N^+(u)$ denote the number of observations greater than or equal to u and the number greater than u, respectively. Then the nonparametric estimator $\hat{S}(u)$ of survival function $S(u)$ under the squared errors loss

$$L(\hat{S},S) = \int_0^\infty (\hat{S}(u)-S(u))^2 \ dw(u)$$

with w being a weight function, is

$$\hat{S}(u) = \frac{\alpha(u,\infty)+N^+(u)}{\alpha(R^+)+n} \prod_{j=k+1}^{\ell} \left\{\frac{\alpha[Z_{(j)},\infty)+N(Z_{(j)})}{\alpha[Z_{(j)},\infty)+N(Z_{(j)})-\lambda_j}\right\} \tag{3.5}$$

in the interval $Z_{(\ell)} \leq u < Z_{(\ell+1)}$, for $\ell=k,\ldots,m$ with $Z_{(k)}=0$, and $Z_{(m+1)}=\infty$.

The authors used the same data given in Kaplan and Meier (10) (and listed under method 3) to obtain the estimate of survival function. Let α be given by $\alpha(u,\infty)=\beta e^{-\theta u}$, and $\theta=.12$, $\beta=4$, 8, and 16. In their notations, $\delta_i=1$ for $i=1,\ldots,4$ and $\delta_i=0$ for $i=5,\ldots,8$ with $Z_1=0.8$, $Z_2=3.1$, $Z_3=5.4$, $Z_4=9.2$, $Z_5=1.0$, $Z_6=2.7$, $Z_7=7.0$, and $Z_8=12.1$. Also that, $Z_{(i)}=Z_i$ for $i=5,\ldots,8$, $m=8$, and $\lambda_j=1$ for $j=5,\ldots,8$. $\alpha(R^+)=\beta$ and $k=4$ then,

$$\hat{S}(u) = \frac{\beta e^{-\theta u}+N^+(u)}{\beta+8} \prod_{5}^{\ell} \left\{\frac{\alpha[Z_{(j)},\infty)+N(Z_{(j)})}{\alpha[A_{(j)},\infty)+N(Z_{(j)})-1}\right\}$$

where $\alpha(Z_{(5)},\infty)=\beta e^{-\theta}, \alpha(Z_{(6)},\infty)=\beta e^{-2.7\theta}, \alpha(Z_{(7)},\infty)=\beta e^{-7.0\theta}$

$\alpha(Z_{(8)},\infty)=\beta e^{-12.1\theta}$ and $N(Z_{(5)})=7$, $N(Z_{(6)})=6$, $N(Z_{(7)})=3$,

$N(Z_{(8)})=1$

and

U in	$N^+(u)$	ℓ
(0,.8)	8	4
(.8, 1.0)	7	4
(1.0, 2.7)	6	5
(2.7, 3.1)	5	6
(3.1, 5.4)	4	6
(5.4, 7.0)	3	6
(7.0, 9.2)	2	7
(9.2, 12.1)	1	7
(12.1, ∞)	0	8

(5) <u>Susarla and Van Ryzin's method</u> (22,23). Let (F_n, X_n, Y_n) be a sequence of independent stochastic processes where for each n, $1-F_n$ is a random distribution function on $R=(-\infty,\infty)$ and distributed according to the Dirichlet process with common parameter α with $\alpha(R)$ known, $X_n \sim$ right sided distribution function (i.e., $F_n(t)=P(X_n>t|T_n)$ and finally, Y_n is a random variable independent of (F_n, X_n) and distributed according to the right sided

distribution H. (Y may be defective, in which case H=0.) We observe only $\delta_i=[X_i\le Y_i]$ and $Z_i=\min\{X_i,Y_i\}$ for $i=1,\ldots,n$. Susarla and Van Ryzin gave the estimator of survival function in two cases:

(i) If $\{H_n\}$ is known then $\hat{S}_i$ is defined by

$$(\alpha(R)+1)\quad \hat{S}_i(u,(\delta_i,Z_i)) \tag{3.6}$$

$$= \begin{cases} 1+\min\{\hat{\alpha}_i(u),\alpha(R)\} & \text{if } u<Z_i \\ \min\{\hat{\alpha}_i(u),\alpha(R)\} & \text{if } \delta_i=1 \text{ and } u\ge Z_i \\ (1+\min\{\hat{\alpha}_i(Z_i),\alpha(R)\})\ \min\{\frac{\hat{\alpha}_i(u)}{\hat{\alpha}_i(Z_i)},\ 1\} & \text{if } \delta_i=0 \text{ and } u\ge Z_i \end{cases}$$

where

$$\hat{\alpha}_i(x) = \alpha(R)\Sigma_{\substack{j=1\\ j\ne i}}^{n}\ (H_j(x))^{-1}(Z_j>x)/(n-1) \tag{3.7}$$

An example involving survival times of melanoma patients was given and expression (3.6) was applied to obtain the survival curve estimator. The authors listed the survival times (in weeks) of 81 participants from a melanoma study conducted by the Central Oncology Group with headquarters office at the University of Wisconsin-Madison. They assumed $H_n(u)$ was known as $e^{-\beta u}$; $u>0$, $\alpha(u)=ce^{-\theta u}$ for $u>0$, $\theta>0$ and $c>0$; and used $\hat{\beta}=(1-\bar{\delta})/\bar{Z}$ as an estimator of β, where $\bar{Z}=(n-1)^{-1}\Sigma_{\substack{j=1\\ j\ne i}}^{n}Z_j$ and $\bar{\delta}=(n-1)^{-1}\Sigma_{\substack{j=1\\ j\ne i}}^{n}\delta_j$. In this example, $\hat{\beta}$ was shown to be a consistent estimator of β.

From this data, $\delta_i=0$, $Z_i=16$, $\alpha(R)=c$, $n=81$, $1-\bar{\delta}=1-46/80=.425$, $\bar{Z}=7055/80=88.1875$ and $\hat{\beta}=.00482$. Applying (3.6), they obtained:

$$\begin{aligned} \hat{S}_i(u)&=1 && \text{if } u<16 \\ &= \min\left(\frac{\exp(.00482(u-16))\hat{G}_i(u)}{78/80},\ 1\right) && \text{if } u\ge 16 \end{aligned} \tag{3.8}$$

where:

$$(n-1)\cdot\hat{G}_i(\cdot)=\Sigma_{\substack{j=1\\ j\neq i}}^{n}\ [Z_j>\cdot].$$

(ii) When $\{H_n\}$ is not known. Assume that $-\alpha'(x)/\alpha(x)\leq r(x)$ is a known function where $\alpha'=d\alpha/dx$ and that K is a known real value bounded function on $\mathbb{R}$ vanishing off $(0,u_1)$, $u_1<\infty$, such that $\int u^j K(u)\, du=0$ for $j=1,\ldots,\ell-1$ with ℓ a fixed positive integer and $\int K(u)\, du=1$ and ξ_n is a function of n with $0<\xi_n<1$. Then the estimator $\hat{\hat{S}}_i$ is defined by (3.6) with $\hat{\alpha}_i$ replaced by $\hat{\hat{\alpha}}_i$ where:

$$\hat{\hat{\alpha}}_i(t) = \exp(-\int_0^t \hat{\phi}_i(x)dx) \tag{3.9}$$

with

$$\hat{\phi}_i(x) = \max\{\min\{\frac{\Sigma_{\substack{j=1\\ j\neq i}}^{n}[\delta_j=1]\ K((Z_j-x)/\xi_n)}{\xi_n\ \Sigma_{\substack{j=1\\ j\neq i}}^{n}\ [Z_j>x]},\ r(x)\},\ 0\} \tag{3.10}$$

(6) Ferguson and Phadia's method (8). Let $F=1-e^{-Y_t}$ be a random distribution function neutral to the right, and let $X_1,\ldots,X_n$ be a sample of size n from F. Assume that the observational data has three forms, m_1 real observations $X_1=x_1,\ldots,X_{m_1}=x_{m_1}$, m_2 exclusive censorings $X_{m_1+1}>X_{m_1+1}\ldots,$ $X_{m_1+m_2}>X_{m_1+m_2}$, and m_3 inclusive censorings $X_{m_1+m_2+1}\geq x_{m_1+m_2+1},\ldots,X_{m_1+m_3}\geq x_{m_1+m_2+m_3}$ where $m_1+m_2+m_3=n$. Let $u_1,\ldots,u_k$ be the distinct values among $x_1,\ldots,x_n$, ordered so that $u_1<\ldots<u_k$. Let $\delta_1,\ldots,\delta_k$ denote the number of real observations at $u_1,\ldots,u_k$ respectively, let $\lambda_1,\ldots,\lambda_k$ denote the number of exclusive censorings at $u_1,\ldots,u_k$ respectively, and let $\mu_1,\ldots,\mu_k$ denote the number of inclusive censorings at $u_1,\ldots,u_k$ respectively so that

$\Sigma_1^k \delta_i = m_1$, $\Sigma_1^k \lambda_i = m_2$, and $\Sigma_1^k \mu_i = m_3$. Let $h_j = \Sigma_{i=j+1}^k (\delta_i + \lambda_i + \mu_i)$ denote the number of the x_i greater than u_j, and $j(t)$ denote the number of u_i less than or equal to t.

Since the form of the process neutral to the right is too general, they derived the estimator of survival function under three types of the process neutral to the right. Each one is rather general and easy to evaluate.

(i) The gamma process. Assume that the independent increments of the process Y_t has gamma distribution with shape parameter $\nu(t)$ and scale parameter τ independent of t, and that $\nu(t)$ is continuous. Then,

$$\hat{S}(t) = \left(\frac{h_{j(t)}+\tau}{h_{j(t)}+\tau+1}\right)^{\nu(t)} \tag{3.11}$$

$$\Pi_{i=1}^{j(t)} \left[\left(\frac{(h_{i-1}+\tau)(h_i+\tau+1)}{(h_{i-1}+\tau+1)(h_i+\tau)} \right)^{\nu(u_i)} \frac{\zeta_G(h_i+\lambda_i+\tau+1, \delta_i)}{\zeta_G(h_i+\lambda_i+\lambda, \delta_i)} \right]$$

where

$$\zeta_G(\alpha, \beta) = \Sigma_{i=0}^{\beta-1} \binom{\beta-1}{i} (-1)^i \log\left(\frac{\alpha+i+1}{\alpha+1}\right) \tag{3.12}$$

If our prior guess at the shape of S(t) is given by $S_0(t)$, then for fixed τ, $\nu(t)$ is

$$\nu(t) = \log(s_0(t))/\log(\tau/(\tau+1)) \tag{3.13}$$

(ii) Simple homogeneous process. Let Y_t be a homogeneous process with MGF of the form:

$$M_t(\theta) = E[e^{-\theta Y_t}] = e^{\nu(t)\int_0^\infty (e^{-\theta z}-1)e^{-\tau z}(1-e^{-z})^{-1}\, dz} \tag{3.14}$$

where ν is continuous, nondecreasing and $\tau > 0$ is a parameter, then the estimator of survival function is:

$$\hat{S}(t)=e^{-\nu(t)/(h_{j(t)}+\tau)} \tag{3.15}$$

$$\Pi_{i=1}^{j(t)}[e^{\nu(u_i)(h_{i-1}-h_i)/((h_{i-1}+\tau)(h_i+\tau))}\ (\frac{h_i+\lambda_i+\tau}{h_i+\lambda_i+\delta_i+\tau})^{(h_i+\lambda_i+\tau)}]$$

If we fix the prior guess at S to be S_0 so that $E(S(t))=S_0(t)$ then we may express (3.15) in an alternate form:

$$\hat{S}(t)=S_0(t)^{\tau/(h_{j(t)}+\tau)} \tag{3.16}$$

$$\Pi_{i=1}^{j(t)}[S_0(u_i)^{-\tau(h_{i-1}-h_i)/((h_{i-1}+\tau)(h_i+\tau))}(\frac{h_i+\lambda_i+\tau}{h_i+\lambda_i+\delta_i+\tau})^{(h_i+\lambda_i+\tau)}]$$

(iii) Dirichlet process. The Dirichlet process, $D(\alpha)$, can be defined as the random distribution function neutral to the right for which the MGF of $Y_t=-\log(1-F(t))$ is:

$$M_t(\theta)=E[e^{-Y_t\theta}]$$

$$=\frac{\Gamma(\alpha(R))}{\Gamma(\alpha(t))\Gamma(\alpha(R)-\alpha(t))} \text{ X } \int_0^\infty e^{-(\alpha(R)-\alpha(t)+\alpha)y}(1-e^{-y})^{\alpha(t)-1}\ dy$$

$$=\frac{\Gamma(\alpha(R))\Gamma(\alpha(R)-\alpha(t)+\theta)}{\Gamma(\alpha(R)-\alpha(t))\Gamma(\alpha(R)+\theta)}$$

$$=e^{\int_0^\infty(e^{-\theta z}-1)\ d\ N_t(z)}$$

where the Levy measure is expressed as:

$$d\ N_t(z)=\frac{e^{-\alpha(R)z}(e^{\alpha(t)z}-1)}{z(1-e^{-z})}\ dz$$

The survival function can then be estimated as:

$$\hat{S}(t)=\frac{\alpha(R)-\alpha(t)-h_{j(t)}}{\alpha(R)+n}$$

$$\Pi_{j=1}^{j(t)}\frac{(\alpha(R)-\bar{\alpha}(u_i)+h_{i-1})\ (\alpha(R)-\alpha(u_i+h_i+\lambda_i)}{(\alpha(R)-\alpha(u_i)+h_i)\ (\alpha(R)-\bar{\alpha}(u_i)+h_i+\lambda_i+\delta_i)}$$

where $\alpha^{-}(u) = \lim_{s \to u} \alpha(s)$.

The authors presented the application of the results by reworking the example of Kaplan and Meier (10). Their data are the same as was illustrated in section 3.2 for Kaplan and Meier's PL method. They take the prior guess at S to be:

$$S_0(t) = e^{-0.1t} \text{ for } t > 0 \tag{3.19}$$

and chose the intensity parameter τ in formulae (3.11) and (3.16) to be 1. From Kaplan and Meier's data, it is seen that u_1=0.8, u_2=1.0, u_3=2.7, u_4=3.1, u_5=5.4, u_6=7.0, u_7=9.2 and u_8=12.1. Furthermore, $\delta_1=\delta_4=\delta_5=\delta_7=1$, $\lambda_2=\lambda_3=\lambda_6=\lambda_8=1$, and the rest of the δ_i's and λ_i's and all of the μ_i's are equal to zero so that $h_i=8-i$ for $i=0,1,\ldots,8$.

(a) The gamma process. Substituting (3.19) and τ=1 into (3.13), one obtains $\nu(t)$=0.1443t. From (3.11), the estimate of survival function is:

$$\hat{S}_G(t) = \left(\frac{9-j(t)}{10-j(t)}\right)^{0.1443t}$$

$$\Pi_{j=1}^{j(t)}\left[\left(\frac{(10-i)^2}{(11-i)(9-i)}\right)^{0.1443u_i}\left[\frac{\ln(\frac{11-i}{10-i})}{\ln(\frac{10-i}{9-i})}\right]^{\delta_i}\right]$$

where j(t) is the number of observations less than or equal to t.

(b) The simple homogeneous process. From (3.16) with $S_0=e^{-0.1t}$ (t>0) and τ=1, one obtains the estimate of survival function as:

$$\hat{S}_H(t) = e^{-0.1t/(9-j(t))}$$

$$\Pi_{j=1}^{j(t)}\left[e^{0.1u_i/((10-i)(9-i))}\left(\frac{10-i-\delta_i}{10-i}\right)\right]$$

(c) The Dirichlet process. From (3.18) with $\alpha(t,\infty)=e^{-0.1t}$, the estimate of survival function is:

$$\hat{S}_D(t)=\frac{e^{-0.1t}+8-j(t)}{9}\,\pi_{i=1}^{j(t)}\frac{e^{-0.1u_{i}}+8-i+\lambda_i}{e^{-0.1u_i}+8-i}$$

(7) <u>Shaked, Zimmer and Ball's method</u> (2,18). Let B be a set in a finite dimensional Euclidean space such that every $V\varepsilon B$ corresponds to one and only one stress level under which an item can operate. Let $V_0 \varepsilon B$ be the normal stress and let $V_1,\ldots,V_k$ be accelerated (greater) stresses under which k life tests are being performed. Assume that k and $V_1,\ldots,V_k$ are determined before the life test begins and remain constant throughout. Thus, without loss of generality, assume that $\beta = \{V_0,V_1,\ldots,V_k\}$. (See Mann (13) for the selection of the accelerated stress levels.)

Suppose that a known function m exists such that for every $V_i \varepsilon B$ and $V_j \varepsilon B$,

$$F_{V_j}(t) = F_{V_i}(m(\alpha,V_j,C_i,t)),t\geq 0 \tag{3.20}$$

where F_V denotes the distribution function of the lifetime of a device subject to stress V and α is an unknown parameter. (α may be a vector.) The set of all possible α's will be denoted by A and the function m will be called a time transformation. Moreover, assume that m of (3.19) is of the form:

$$m(\alpha,V_j,V_i,t)=\frac{g(\alpha,V_j)}{g(\alpha,V_i)}\,t \tag{3.21}$$

$g(\alpha,V)$ 0, $V\varepsilon B,\alpha\varepsilon A$.

The set of data that is obtained from accelerated life tests is the set of observations $T_{i\ell}$, $\ell=1,\ldots,n_i$, $i=1,\ldots,k$ where $T_{i\ell}$ is the time of failure of the ℓth item in the sample of size n_i that is run under stress level V_i, i=1,...,k. In addition, assume that the sample size n_i are fixed in advance. If nonaccelerated data are also available then the procedure which follows can still be used by augmenting 0 to the range of the indices i and j. However, for the application of the procedure no nonaccelerated data are needed. Denote the scale factor between F_{V_j} and F_{V_i} by:

$$\theta_{ij} = g(\alpha,V_j)/g(\alpha,V_i); \quad i\neq j, \ \alpha\varepsilon A; \ V_i,V_j\varepsilon B \tag{3.22}$$

The first step in the procedure for estimating $F_{V_0}(t)$ is to estimate θ_{ij} for given i and j ($i\neq j$). Let $\bar{T}_i=n^{-1}\Sigma_{\ell=1}^{n} T_{i\ell}$, i=1,...,k, then an estimator of θ_{ij} is:

$$\hat{\theta}_{ij} = \bar{T}_i/\bar{T}_j \tag{3.23}$$

Next, for every i, j($i\neq j$) estimate α_{ij} from the equation:

$$\hat{\theta}_{ij} = g(\hat{\alpha}_{ij},V_j)/g(\hat{\alpha}_{ij},V_i) \tag{3.24}$$

then estimate α as a weighted average:

$$\hat{\alpha} = \sum_{i\neq j} W_{ij}\hat{\alpha}_{ij} \tag{3.25}$$

where W_{ij}'s are determined by:

$$W_{ij} \propto \frac{(A(\hat{\alpha}_{ij},V_i,V_j))^2}{Var(\theta_{ij})} \tag{3.26}$$

where:

$$A(\alpha,V_i,V_j) = \frac{(\partial/\partial\alpha)g(\alpha,V_j)g(\alpha,V_i)-(\partial/\partial\alpha)g(\alpha,V_i)g(\alpha,V_j)}{(g(\alpha,V_i))^2}$$

$$\text{and} \ \sum_{i\neq j} W_{ij}=1$$

If $\hat{\theta}_{ij}$ is the estimator of Sen (20), then using the expression for the asymptotic variance of Sen's estimate (p.536) one can approximate $\text{var}(\hat{\theta}_{ij})$ and substitute it in (3.26). Similar remarks hold for other estimators of θ_{ij}.

Define the rescaled values:

$$\tilde{T}_i = \frac{g(\hat{\alpha},V_i)}{g(\hat{\alpha},V_0)} \; T_{i\ell} \quad \ell=1,\ldots,n_i;\ i=1,\ldots,k \tag{3.27}$$

Then:

$$\hat{F}_{V_0}(t) = (\text{Number of } \tilde{T}_i \text{ less than } t)/N \tag{3.28}$$

where $N = \Sigma_{i=1}^{k} n_i$.

If the time transformation of (3.20) is of the form:

$$m(\alpha,V_j,V_i,t)=t^{g(\alpha,V_j)/g(\alpha,V_i)} \tag{3.29}$$

and if θ_{ij} is again defined as in (3.22), one can use the same method as before, but estimate θ_{ij} by $\bar{S}_i/\bar{S}_j$ where $\bar{S}_i$ is the mean of $\log T_{i1},\ldots,$ $\log T_{in_1}$ and $\bar{S}_j$ is similarly defined.

The authors illustrated their method by a numerical example, using the real data reported by Nelson (15,16). The data consisted of times to breakdown of an insulating fluid subjected to seven constant elevated test voltages: 26 kv, 28 kv,..., 28 kv. The normal voltage is V_0 = 20 kv. Assume the model g of (3.21) is:

$$g(\alpha,V) = V^{\alpha},\alpha>0 \tag{3.30}$$

Then from (3.24) and (3.23)

$$\hat{\alpha}_{ij} = \frac{\log \hat{\theta}_{ij}}{\log(V_j/V_i)} = \frac{\log(\bar{T}_i/\bar{T}_j)}{\log(V_j/V_i)} \tag{3.31}$$

To obtain W_{ij} they used (3.26) and from a result of Cramer (4, p.366) they obtained:

$$W_{ij} \propto (\log(V_j/V_i))^2 \tag{3.32}$$

thus from (3.31), (3.32) and (3.25), it follows that:

$$\hat{\alpha} = \frac{\Sigma_{i=1}^{k} \Sigma_{j=i+1}^{k} (\log(V_j/V_i)(\log(\bar{T}_i/\bar{T}_j))}{\Sigma_{i=1}^{k} \Sigma_{j=i+1}^{k} (\log(V_j/V_i))^2} \tag{3.33}$$

Also, from (3.33) they obtained the value of $\hat{\alpha}$ to be 17.9286. The rescaled variables are from (3.27).

$$\tilde{T}_{i\ell} = (V_i/V_0)^{\hat{\alpha}}\, T_{i\ell} = (V_i/20)^{17.9286}\, T_{i\ell},\quad \ell=1,\ldots,n;\; i=1,\ldots,k. \tag{3.34}$$

and the empirical distribution based on $T_{i\ell}$'s is the estimate $\hat{F}_{v_0}(t)$.

IV. DISCUSSION

The feature of the nonparametric estimation of life distribution is to use a weak set of assumptions, as compared to the more restrictive parametric models, to get the estimate of the distribution. Once we have the estimate of the distribution, we can predict the probability of failure at any given time. Besides, nonparametric estimation techniques have the advantage of being relatively insensitive to outliers in the data.

In this section we consider the properties of the estimators we described in Section 3.

(1) <u>Ferguson's method</u> (6). This estimator is a weighted average of our prior guess of F and of the sample distribution function, with respective weights P_n and $(1-P_n)$. Ferguson gave a reasonable interpretation to $\alpha(R)$ as the prior sample size.

If $\alpha(R)$ is large compared to n, little weight is given to the observations; if $\alpha(R)$ is small compared to n, little weight is given to the prior guess of F. As $\alpha(R)$ approaches zero, the estimator converges to the sample distribution function which is a ML estimator.

In theory the concept of the Dirichlet process is not easy to understand. However, in application, the estimator of the distribution function is quite reasonable, and useful. The Bayes risks $R_n(\alpha)$ of Ferguson's estimator (12) is:

$$R_n(\alpha) \overset{\text{def}}{=} R(\hat{F}_n,\alpha) = E_{\underline{X}}[\int\{E_{F(t)|\underline{X}}(F(t)-\hat{F}_n(t))^2\}dW(t)] \tag{4.1}$$

$$= [\alpha(R)/\{(\alpha(R)+1)(\alpha(R)+n)\}]\int F_0(t)(1-F_0(t))\, dW(t)$$

Moreover, Ferguson's estimator has a very nice property (6), that is, no matter what the true distribution is, Ferguson's estimator converges to it uniformly almost surely. This follows from the Glivenko-Centelli theorem and the observation that $P_n \to 0$ as $n \to \infty$.

(2) Hollander and Korwar's method (9,12). This method is a modification of Ferguson's method. It is more useful since we can estimate all n distribution functions simultaneously regardless of whether sizes are equal or not. Besides, Hollander and Korwar's estimator requires less prior information about $\alpha(\cdot)$. Only $\alpha(R)$ needs to be specified. After the sample size is fixed, $\alpha(R)$ is not hard to determine.

Similar to Ferguson's estimator, Hollander and Korwar's estimator of F_i, $i=1,\ldots,n$ is a weighted average of the sample distribution function of $\underline{X}_i$ and of the past samples $\underline{X}_1,\ldots,\underline{X}_{i-1},\underline{X}_{i+1},\ldots,\underline{X}_n$. It is also very easy to apply.

Hollander and Korwar (9,12) showed that even though one needs to specify $\alpha(R)$, the procedure is asymptotically as good as though α were known exactly, and that the difference between the risk of Ferguson's estimator $\hat{F}_{m_i}$ (when based on $\underline{X}_i$) and the overall expected loss using H_i converges to zero as $n \to \infty$. Also, Hollander and Korwar gave a necessary and sufficient condition that their estimator does better than the sample distribution function $\hat{F}_1$ (based on $\underline{X}_i$). This condition is expressed as:

$$\frac{1}{m_i} > \frac{\alpha(R) \sum\limits_{\substack{t=1 \\ t \neq i}}^{n} m_t^{-1} + n - 1}{(n-1)^2\{\alpha(R)+m_i\}} \tag{4.2}$$

Furthermore, they showed that if:

$$(n-1)\cdot\min(m_1,\ldots,m_n) > \max(m_1,\ldots,m_n) \tag{4.3}$$

or

$$(2n-3)\min(\alpha(R),m_1,\ldots,m_n) > \max(\alpha(R),m_1,\ldots,m_n), \tag{4.4}$$

then the Hollander and Korwar's estimator is better than the sample distribution function.

(3) Kaplan and Meier's PL method (3,10). Kaplan and Meier's paper (10) presented not only the PL estimator but also the RS estimator and the actual estimator. We shall only describe the PL estimator since the RS estimator does not utilize all the information from the sample and the PL estimator is a limiting case of the actual estimator. In their papers, they describe the observations as either time to death or time to loss. Here we interpret them as type I censored data and assign to each sample a time scheme L_i (referred to as the limits of observations in Section 3) if death occurs after L_i, otherwise T_i (time of death) is assigned to t_i.

The PL estimate is very easy to calculate, is consistent and of negligible bias (3). The asymptotic expression for its variance is:

$$V|\hat{P}(t)| = P^2(t)\sum_r[(N-r)(N-r+1)]^{-1} \tag{4.5}$$

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

The disadvantage of the PL estimate is that if the greatest observed lifetime corresponds to a loss (t*), then for t>t*, $\hat{P}(t)$ is undefined though bounded by 0 and $\hat{P}(t^*)$. It may be time-consuming and expensive or impossible to overcome this disadvantage.

When no loss occurs at ages less than t, the PL estimate of P(t) reduces, in all cases, to the usual binomial estimate, namely, the observed proportion of survivors.

(4) Susarla and Van Ryzin's method (21). This method gives the nonparametric solution to the estimation of the life distribution function under the squared error loss using the notion of the Dirichlet process prior. The resulting

estimator shown in (21) reduces to the PL estimator in the case where $\alpha(R^+)\to 0$. $\alpha(\cdot)$ is a parameter of the Dirichlet process prior.

This estimator is a function of the sufficient statistics. Unlike the PL estimator, one can obtain an estimate, $\hat{S}(u)$, for any value of u.

(5) Susarla and Van Ryzin's method (22,23). This estimator is useful if individuals respond differently to the same treatment, but on the average have the same survival distribution. In that case, one can estimate the distribution function for each individual survival distribution.

Susarla and Van Ryzin showed that their estimator has: 1) mean-square consistency, 2) almost sure consistency, and 3) asymptotic normality assuming that the observations are i.i.d with right cdf F_0 and that the censoring random variables are i.i.d with a continuous distribution function.

The properties of this estimator is lacking for small samples. When the distribution H_n is unknown, the authors suggest the use of two functions, K(u) and ξ_n. There are certain conditions that K(u) and ξ_n must satisfy, which limit their applicability. When H_n is known, it was found that the probability of survival until 180 weeks ($\hat{S}_i(180)=.198$) was larger than the probability of survival until 160 weeks ($S_i(160)=.18$). Susarla and Van Ryzin attributed this undesirable feature of the estimator to the use of $(H_j(x))^{-1}$ in equation (3.7). Hence it would be desirable in the future to investigate alternative estimators to $\hat{\alpha}_i(X)$.

(6) Ferguson and Phadia's method (8). Ferguson and Phadia's estimators, using the process neutral to the right as prior, are very general and useful.

The intensity parameter τ measures, in some sense, the prior "strength of belief" in the process neutral to the right. In the simple homogeneous process and in the Dirichlet process, the estimators converge to the sample distribution function as τ tends to zero.

(7) Shaked, Zimmer and Ball's method (2,18). Shaked, Zimmer and Ball's method is used under the condition that the lifetime is long, and the time spent on testing a sample of devices is excessive. Accelerated data for this method is obtained from observations collected on the devices under consideration when subject to conditions of greater stress than that under normal operating condition. From the results for these high-stress environments, an estimate of performance of the device under normal operating conditions is obtained. This method has the advantage in that lifetime observations under normal operating condition (which are difficult to obtain) are not required.

The disadvantage of this method is that it is difficult, in the procedure of estimation, to obtain the asymptotic variance of $\hat{\theta}_{ij}$. The authors used simulation to compare their estimator with the power rule method (14, p. 425) under the assumption of exponential lifetime. It turned out that, in some instances, their estimator was asymptotically equivalent to the maximum likelihood method for estimating the nonaccelerated mean lifetime.

V. SUGGESTIONS FOR FUTURE STUDIES

Comparisons between the parametric and the nonparametric methods.

It was suggested in Section 4 that the nonparametric estimation of the distribution function is robust. This point, however, needs further investigation especially for small samples. In accelerated sampling, Shaked et al. (18), using simulation, compared their method with the power rule method [Mann et al. (14)] under the assumption of exponential lifetimes, which revealed that the Shaked et al. (18) method was asymptotically equivalent to the maximum likelihood method for estimating the nonaccelerated mean lifetime. However, when Shaked et al. (18) compared their method with Nelson's parametric method (16) by using the real data reported by Nelson (15,16), they found that there was a significant difference between the nonparametric estimates and the parametric estimates. They did not, however, explain which method was better or under what conditions a parametric or nonparametric method should be used. Also, there has been no comparisons between the parametric and nonparametric methods or among the different nonparametric methods for nonaccelerated sampling. Such comparisons would be desirable, especially for small samples. It would also be desirable to compare nonparametric and parametric methods for different sample sizes in order to determine how the size of the sample might affect the choice of the method.

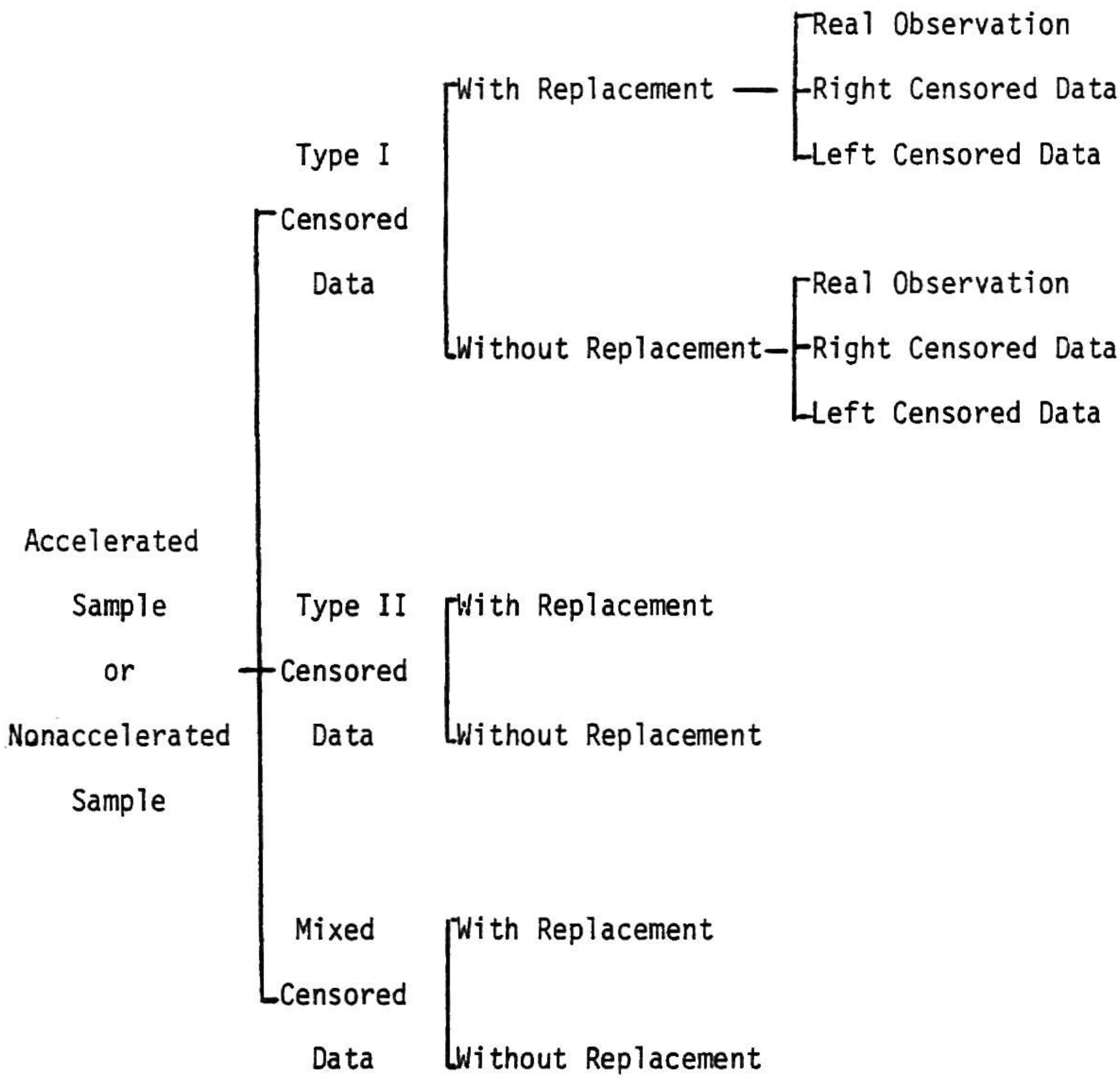

Figure 1. Different forms of sampling that may arise in life testing.

Table 1. Different nonparametric estimation methods and data classification for which each method is suitable.

Purpose	Methods	Conditions	Ref.
Estimate the distribution	Ferguson's method	Nonaccelerated Type II without replacement censored data with r=n	6
Simultaneous estimation of n distribution	Hollander and Korwar's method	Nonaccelerated Type II without replacement censored data. Samples are taken from n distribution functions, and sample sizes may or may not be equal	9 12
	Kaplan and Meier's PL method	Nonaccelerated Type I without replacement censored data	10 3
Estimation of the survival function	Susarla and Van Ryzin's method	Nonaccelerated, without replacement with Type I censored data	21
	Ferguson and Phadia's method	Nonaccelerated Type I censored data without replacement	8
Estimation of the nth distribution function	Susarla and Van Ryzin's method	Nonaccelerated, without replacement and with mixed censored data	22 23
Estimation of the distribution function under operating conditions	Shaked, Zimmer and Ball's method	Accelerated Type II censored data without replacement. The lifetime of the device is very long.	2 18

REFERENCES

1. C. E. Antoniak, "Mixture of Dirichlet process with Application to Bayesian nonparametric problems," Ann. Statist., Vol. 2, pp. 1152-1174, 1974.

2. R. E. Barlow and E. M. Scheuer, "Estimation from accelerated life tests," Technometrics, Vol. 13, pp. 145-159, 1971.

3. Breslow and J. Crowley, "A large sample study of the life table and product limit estimates under random censorship," Ann. Statist., Vol. 2, pp. 437-453, 1974.

4. H. Cramer, Mathematical Methods of Statistics, N.J.: Princeton University Press, 1946.

5. K. Doksum, "Tailfree and neutral random probabilities and their posterior distributions," Ann Prob., Vol. 2, pp. 183-201, 1974.

6. T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," Ann. Statist., Vol. 1, pp. 209-230, 1973.

7. T. S. Ferguson, "Prior distributions on spaces of probability measures," Ann. Statist., Vol. 2, pp. 615-629, 1974.

8. T. S. Ferguson and E. G. Phadia, "Bayesian nonparametric estimation based on censored data," Ann. Statist., Vol. 7, pp. 163-186, 1979.

9. M. Hollander and R. M. Korwar, "Nonparametric estimation of distribution functions," The Theory and Applications of Reliability, Vol. 1, pp. 85-107, 1977.

10. E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," J. Amer. Statist. Assoc., Vol. 53, pp. 457-481, 1958.

11. R. M. Korwar and M. Hollander, "Contributions to the theory of Dirichlet process," Ann. Prob., Vol. 1, pp. 705-711, 1973.

12. R. M. Korwar and M. Hollander, "Emprical Bayes estimation of a distribution function," Ann. Statist., Vol. 4, pp. 581-588, 1976.

13. N. R. Mann, "Design of over-stress life-test experiments when failure times have the two-parameter Weibull distribution," Technometrics, Vol. 14, pp. 437-451, 1972.

14. N. R. Mann, R. E. Schafer and N. D. Singpurwalla. Methods for Statistical Analysis of Reliability and Life Data, New York: Wiley, 1974.

15. W. Nelson, "Graphical analysis of accelerated life test data with the inverse power law model," IEEE Trans. Rel., Vol. R-21, pp. 2-11, Correction p. 195, 1972.

16. W. Nelson, "Analysis of accelerated life test data -- least squares methods for the inverse power law model," IEEE Trans. Rel., Vol. R-24, pp. 102-107, 1975.

17. F. Proschan, "Theoretical explanation of observed decreasing failure rate," Technometrics, Vol. 5, pp. 375-383, 1963.

18. M. Shaked, W. J. Zimmer and C. A. Ball, "A nonparametric approach to accelerated life testing," J. Amer. Statist. Assoc., Vol. 74, pp. 694-699, 1979.

19. I. Shimi and C. P. Tsokos, "The Bayesian and nonparametric approach to reliability: A survey of recent work," The Theory and Application of Reliability, Vol. 1, pp. 5-47, 1977.

20. G. P. Steck, W. J. Zimmer and R. E. Williams, "Estimation of parameters in accelerated models," Proceedings, 1974 Annual Reliability and Maintainability Symposium, IEEE, Piscatany, N.J., pp. 428-431, 1974.

21. V. Susarla and J. Van Ryzin, "Nonparametric Bayesian estimation of survival curves from incomplete observations," J. Amer. Statist. Assoc., Vol. 71, pp. 897-902, 1976.

22. V. Susarla and J. Van Ryzin, "Empirical Bayes estimation of a distribution (survival) function from right censored observations," Ann. Statist., Vol. 6, pp. 740-754, 1978.

23. V. Susarla and J. Van Ryzin, "Large sample theory for a Bayesian non-parametric survival curve estimator based on censored samples," Ann. Statist., Vol. 6, pp. 755-768, 1978.

ACKNOWLEDGMENTS

The author wishes to extend her appreciation to her major professor, Dr. R. Nassar, for his patience and valuable guidance in the preparation of this work. Appreciation is also extended to Dr. Rubison and Dr. Dayton for their critical reading of this report.

Thanks are also due to Dr. Way Kuo for helpful suggestions during the course of this review.

The author also wishes to express her sincere gratitude to her fiance and to her parents for their steady encouragement.

NONPARAMETRIC ESTIMATION OF LIFE DISTRIBUTIONS

by

MING-SHU HSU

B.S., Soochow University, 1976

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1980

ABSTRACT

The aim of this report is to survey recent work in nonparametric estimation of life distribution. My presentation consists of seven methods of nonparametric estimation of life distribution functions and a summary of several types of data in life testing. The definitions and results of the Dirichlet process and of a process that is neutral to the right, which are very useful with respect to some nonparametric decision theoretic problems, are also presented. Some proposals for future work are outlined.