Chapter 5

Selection Procedures for Selecting Largest Mean Residual Life Distribution

5.1 Introduction

In this chapter, we develop procedures for selecting a life distribution having largest mean residual life of a given age \( t \) from the \( k \) \( (k \geq 2) \) available distributions, using two-stage selection procedure. In reliability and engineering an experimenter may face the problem of selecting a firm from several competitors, manufacturing components under specific brands with almost identical guarantee life say \( t \), having largest mean residual life beyond a specified age \( t \). The performance of the proposed procedures, when the underlying life distributions of components of different firms belong to (i) the same family
and (ii) the different families, have been checked through simulation study.

Suppose there are \( k \) \((k \geq 2)\) competing firms, each supplying the components under specific brands meant for the same purpose, such that \( F_i \) be the cumulative distribution function (c.d.f.) of the life lengths of components of the \( i \)-th firm, \( i = 1, \ldots, k \). Let the variance \( \sigma_i^2 \) associated with c.d.f. \( F_i \) be finite and \( F_i(0) = 0, \ i = 1, \ldots, k \). The mean residual life of age \( t \) corresponding to c.d.f. \( F_i \) is defined as

\[
\mu_i(t) = E_{F_i}(X - t | X > t) = \int_{F_i(t)}^{\infty} F_i(u) du, \tag{5.1.1}
\]

where \( F_i = 1 - F_i \) and \( E_{F_i} \) is the expectation with respect to the distribution having c.d.f. \( F_i \). We say that mean residual life of given age \( t \) of components of \( i \)-th firm is larger than the mean residual life of age \( t \) of components of firm \( j \) if \( \mu_i(t) \geq \mu_j(t) \).

On the basis of random samples \((i.e., \text{life lengths of components of } k \text{ firms}) \ X_{ij}; \ i = 1, \ldots, k; \ j = 1, \ldots, n, \) we propose selection procedures to meet the goals: (a) to select the firm (population) associated with the largest value of \( \mu(t) \) for a given value of \( t \) and (b) to select a random size subset ‘\( S \)’ of the \( k \)
firms which contains the firm with largest value of mean residual life for a given value of $t$. A motivation for this problem is that suppose there are $k$ competing firms with almost same guarantee life say $t$ of their components and the experimenter is interested to select either: (a) the firm with largest mean residual life of age $t$ from these competing firms or (b) a random size subset of the $k$ firms which includes the firm with largest mean residual life of age $t$.

In the past not much of the work has been done on such problems since the exact distribution of the estimators, used to propose the said selection procedures in the non-parametric set-up, is not known. However, their asymptotic distribution is normal with variance depending upon the underlying unknown distribution. By using asymptotic normality of such estimators, a two-stage procedure, proposed by Dudewicz (1971) and Dudewicz and Dalal (1975) for the normal probability models with different variances, is a way to circumvent such situation. Park and Saxena (1987) defined NBU-ness of a life distribution and proposed a two-stage selection procedure to select the least new better than used (NBU) life distribution.

The proposed selection procedure, based on the estimator $R_{in}(t)$ of $\mu_i(t)$, $i=1,\ldots,k$ for the goal (a) is given in section 5.2,
its modification to consider the goal (b) is given in section 5.3 and modification for the goals (a) and (b) with right censored data is given in section 5.4. A statistical simulation carried out to check the performance of the proposed selection procedures in section 5.5.

5.2 Proposed Selection Procedure

Our goal (a) is to select the firm (population) corresponding to the largest mean residual life $\mu_{[k]}(t)$ of age $t$, where $\mu_{[1]}(t) \leq \ldots \leq \mu_{[k]}(t)$ denote the ordered mean residual lives of age $t$ ($t$ being a pre-assigned or pre-specified positive number).

The selection of any population associated with $\mu_{[k]}(t)$ is called the correct selection (CS). Let $Q_t = \{p(t) = (p_1(t), \ldots, p_k(t))\}$ denote the parameter space, which is partitioned into a preference-zone $Q_t(\delta^*)$ and indifference-zone $Q_t - Q_t(\delta^*)$, where $Q_t(\delta^*)$ is defined as $Q_t(\delta^*) = \{p(t): \mu_{[k]}(t) - \mu_{[k-1]}(t) \geq \delta^*\}$.

For pre-assigned values of constants $\delta'(\delta^* > 0)$ and $P'(k^{-1} < P' < 1)$, a selection procedure (say) $R$ is required to satisfy the probability requirement (called $P'$-condition)
\[ P[CS|R] \geq P^* \quad \text{for all } \mu(t) \in \Omega_i(\delta^*). \quad (5.2.1) \]

The selection procedure is based on the estimators \( R_{in}(t) \) of \( \mu_i(t), \ i = 1, \ldots, k \), where the estimator \( R_{in}(t) \) for a given value of \( t \), using random sample \( X_{i1}, \ldots, X_{in} \) of life lengths of \( n \) components from \( i \)-th firm, can be obtained from (5.1.1) by replacing \( F_i \) by the corresponding empirical c.d.f. \( F_{in}, \ i = 1, \ldots, k \) as follows:

\[
R_{in}(t) = \frac{\int_0^t F_{in}(u) du}{F_{in}(t)} - tF_{in}(t) + \int_t^\infty u dF_{in}(u) = \frac{\int_0^\infty u dF_{in}(u)}{F_{in}(t)} - t, \quad (5.2.2)
\]

where \( F_{in}(x) = \frac{1}{n} \sum_{j=1}^n I(X_{ij} \leq x) \),

\[
\int_0^\infty u dF_{in}(u) = \frac{1}{n} \sum_{j=1}^n X_{ij} I(X_{ij} > t),
\]

and \( I(A) \) is the indicator function of event \( A \). The asymptotic distribution of \( R_{in}(t) \), given by Li (1997), is stated in the following lemma.
Lemma 5.2.1: Assume that variance $\sigma_i^2$ associated with c.d.f. $F_i$ be finite and $F_i(0) = 0$. Then, for fixed $t$

$$n^{1/2}\{R_{in}(t) - \mu_i(t)\} \to N(0, \sigma_i^2(F_i)),$$  

$i = 1, \ldots, k$  \hspace{1cm} (5.2.3)

where

$$\sigma_i^2(F_i) = \frac{1}{F_i(t)} \left\{ \int_0^\infty x^2 dF_i(x) - \left[ \int_0^\infty x dF_i(x) \right]^2 \right\}$$

$$+ \frac{F_i(t)}{F_i(t)} \left\{ \left[ \int_0^\infty F_i(x) dx \right]^2 - 2 \int_0^\infty F_i(x) x^2 dF_i(x) \int_0^\infty F_i(x) dx \right\}.$$  \hspace{1cm} (5.2.4)

In other words, for sufficiently large $n$, the distribution of $R_{in}(t)$ can be approximated by a normal distribution with mean $\mu_i(t)$ and variance $\sigma_i^2(F_i)/n$, $i = 1, \ldots, k$. Consequently, the problem of selecting the population associated with the largest mean residual life of age $t$, if the sample sizes from each of the $k$ populations are large, can again be looked at as the problem of selecting a normal distribution with largest mean. However, the variances $\sigma_i^2(F_i)/n$ of the limiting normal distributions of $R_{in}(t)$, $i = 1, \ldots, k$, being dependent on unknown $F_i$, are unknown and unequal for the underlying $k$ populations. In such a situation there does not exist a single-stage procedure, which can satisfy the $P'$-condition (5.2.1). A solution to the problem is provided by a two-stage procedure as explained below.
Chapter 5

The consistent estimator $\hat{\sigma}_i^2(F_{in})$ of $\sigma_i^2(F_i)$ [see Li (1997)], based on a random sample of size $n$ from the distribution $F_i$, is obtained by replacing $F_i$ by the corresponding empirical c.d.f. $F_{in}$, $i=1,...,k$ in the expression of $\sigma_i^2(F_i)$ given in (5.2.4).

We now propose the following two-stage selection procedure $R$, assuming large sample sizes. Let $h=h(k,P^*),>0$ be the unique solution of the equation

$$\int_{-\infty}^{\infty} \Phi^{-1}(z+h)d\Phi(z) = P^*, \quad (5.2.5)$$

where $\Phi(.)$ is the standard normal c.d.f..

Procedure $R$: Take an initial sample $X_{i1},...,X_{in_0}$ of size $n_0(\geq 2)$ from the population $\pi_i$ and for a given $t$ compute $R_{in_0}(t)$ and $\hat{\sigma}_i^2(F_{in_0})$, $i=1,...,k$.

Define

$$n_i = \max\left\{n_0+1, \left\lfloor \frac{1}{c_i} \right\rfloor \right\}, \quad (5.2.6)$$

where $[x]$ denotes the smallest integer $\geq x$ and

$$\frac{1}{c_i} = \hat{\sigma}_i^2(F_{in_0}) \left( \frac{h^2}{\hat{\delta}^2} \right). \quad (5.2.7)$$

The second-stage sample $n_i$ from $\pi_i$ is determined as follows.
Chapter 5

Compute \( R_{n_i}(t) \), based on \( n_i \) additional sample observations taken from \( \pi_i \), and define

\[
S_i(t) = a_i R_{n_0}(t) + (1-a_i) R_{n_i}(t),
\]

(5.2.9)

where \( 0 < a_i \leq 1 \) are determined to satisfy

\[
\sigma^2_i(F_{n_0}) \left[ \frac{a_i^2}{n_0} + \frac{(1-a_i)^2}{n_i} \right] = \left( \frac{\delta_i^*}{h} \right)^2.
\]

(5.2.10)

The left-hand side of (5.2.10) is a strongly consistent estimator of the asymptotic variance of \( S_i(t) \). The existence of such \( a_i \)'s follows from Lemma 5.2.2, given below. Finally we select the population which yields \( S_{[k]}(t) \) for a given \( t \), where \( S_{[1]}(t) \leq \ldots \leq S_{[k]}(t) \) denote the ordered value of \( S_1(t), \ldots, S_k(t) \), computed from (5.2.9) for \( i = 1, \ldots, k \). It is assumed that the additional samples are independent of the initial samples. The equation (5.2.10) means that all the \( S_i(t) \)'s have the same asymptotic variance.
Remark 5.2.1: If \( n_i = 0 \), i.e., no second-stage sample is taken from the \( i \)-th population, we define \( a_i = 1 \).

Lemma 5.2.2: There exists

\[
1 \quad \text{if } \left[ \frac{1}{c_i} \right] \leq n_0
\]

\[
\frac{n_0}{(n_0 + 1)^{-1}} \left[ 1 + \left( \frac{c_i + c_i n_0 - 1}{n_0} \right)^{1/2} \right] \quad \text{if } n_0 < \left[ \frac{1}{c_i} \right] \leq n_0 + 1
\]

\[
n_0 + \left[ n_0 \left( \frac{1}{c_i} \right) - n_0 \right] \left( c_i \left[ \frac{1}{c_i} \right] - 1 \right)^{1/2}
\]

\[
\frac{1}{c_i}
\]

\[
\quad \text{if } \left[ \frac{1}{c_i} \right] > n_0 + 1,
\]

so as to satisfy (5.2.10).

Proof: From (5.2.10) and (5.2.7), we have

\[
\frac{a_i^2}{n_0} + \frac{(1 - a_i)^2}{n_i} = c_i,
\]

that is

\[
n_i a_i^2 + (1 + a_i^2 - 2a_i) n_0 - c_i n_0 n_i = 0 \quad (5.2.11)
\]

which is a quadratic equation in \( a_i \).

From (5.2.8), we have

\[
n_i = 0 \quad \text{if } \left[ \frac{1}{c_i} \right] \leq n_0.
\]

Therefore, by Remark 5.2.1,
a_i = 1 \quad \text{if} \quad \left[ \frac{1}{c_i} \right] \leq n_0. \quad (5.2.12)

Again from (5.2.8), we have

\[ n_i = n_i - n_0 \quad \text{if} \quad \left[ \frac{1}{c_i} \right] > n_0 \]

and from (5.2.6), we have

\[ n_i = n_0 + 1 \quad \text{if} \quad \left[ \frac{1}{c_i} \right] \leq n_0 + 1. \]

Hence,

\[ n_i = n_i - n_0 = n_0 + 1 - n_0 = 1 \quad \text{if} \quad n_0 < \left[ \frac{1}{c_i} \right] \leq n_0 + 1. \]

Therefore from quadratic equation (5.2.11), we have

\[ a_i = \frac{n_0}{n_0 + 1} \left[ 1 + \left( \frac{c_i + c_i n_0 - 1}{n_0} \right)^{1/2} \right] \quad \text{if} \quad n_0 < \left[ \frac{1}{c_i} \right] \leq n_0 + 1. \]

(5.2.13)

Again from (5.2.6), we have

\[ n_i = \left[ \frac{1}{c_i} \right] \quad \text{if} \quad \left[ \frac{1}{c_i} \right] > n_0 + 1, \]

thus (5.2.8) gives

\[ n_i = \left[ \frac{1}{c_i} \right] - n_0. \]

Therefore, by taking \( n_i = \left[ \frac{1}{c_i} \right] - n_0 \) in (5.2.11), we get
Chapter 5

\[ a_i = \frac{n_0 + \left[n_0 \left(1 - \frac{1}{c_i} \right) - n_0 \left(\frac{1}{c_i} - 1\right)\right]^{1/2}}{\frac{1}{c_i}} \]

if \[ \frac{1}{c_i} > n_0 + 1. \]

(5.2.14)

So, from (5.2.12), (5.2.13) and (5.2.14) we can write

\[
\begin{cases}
1 & \text{if } \frac{1}{c_i} \leq n_0 \\
\frac{n_0}{(n_0 + 1)^{-1}} \left[1 + \left(\frac{c_i + c_i n_0 - 1}{n_0}\right)^{1/2}\right] & \text{if } n_0 < \frac{1}{c_i} \leq n_0 + 1 \\
n_0 + \left[n_0 \left(1 - \frac{1}{c_i} \right) - n_0 \left(\frac{1}{c_i} - 1\right)\right]^{1/2} & \text{if } \frac{1}{c_i} > n_0 + 1.
\end{cases}
\]

(5.2.15)

Hence the proof.

**Theorem 5.2.1:** For any \( P^* \in (k^{-1}, 1) \), there exists \( n_0 \) large enough such that for a given \( t \), \( \inf_{\Omega(n^*)} P[CS|R] \approx P^* \).

**Proof:** Let \( S_{(i)}(t) \) denote the statistic associated with the population having parameter \( \mu_{(i)}(t) \), \( i = 1, \ldots, k \). Then for a given \( t \), we have
\[ P[C_i \mid R] = P[S_{(i)}(t) \leq S_{(k)}(t), \quad i = 1, \ldots, k - 1] \]
\[ = P[S_{(i)}(t) - \mu_{(i)}(t) \leq S_{(k)}(t) - \mu_{(k)}(t) + \mu_{(k)}(t) - \mu_{(i)}(t), \quad i = 1, \ldots, k - 1] \]
\[ = P[(S_{(i)}(t) - \mu_{(i)}(t))h/\delta^* \leq (S_{(k)}(t) - \mu_{(k)}(t))h/\delta^* \quad + (\mu_{(k)}(t) - \mu_{(i)}(t))h/\delta^*, \quad i = 1, \ldots, k - 1] \]
\[ = P[Y_{(i)} \leq Y_{(k)} + (\mu_{(k)}(t) - \mu_{(i)}(t))h/\delta^*, \quad i = 1, \ldots, k - 1] \]
\[ = \prod_{i=1}^{k-1} G\left( y + (\mu_{(k)}(t) - \mu_{(i)}(t))h/\delta^* \right) G(y), \quad (5.2.16) \]

where, for a fixed value of \( t \), \( Y_{(i)} = (S_{(i)}(t) - \mu_{(i)}(t))h/\delta^* \), \( i = 1, \ldots, k \) are independent and identically distributed (i.i.d.) random variables with some unknown distribution \( G(.) \). Below we have shown that for large \( n_0 \), \( G(y) \) can be approximated by a standard normal cumulative distribution \( \Phi(y) \). Without loss of generality we can ignore the ordering with respect to (w.r.t.) \( i \).

Thus for any \( i \) and fixed \( t \), we can write
\[ G(y) = P[Y_i \leq y] \]
\[ = P \left[ \frac{S_i(t) - \mu_i(t)}{\sigma_i^2(F_{in_0}) \left( a_i^2 n_0^{-1} + (1-a_i)^2 n_i^{-1} \right)} \leq y \right] \]
\[ = E_{\sigma_i^2(F_{in_0})} P \left[ \frac{S_i(t) - \mu_i(t)}{\sigma_i^2(F_{in_0}) \left( a_i^2 n_0^{-1} + (1-a_i)^2 n_i^{-1} \right)} \leq y \right] \]
\[ = 126 \]
Chapter 5

\[ E_{\sigma_i(F_{i0})} \left[ \frac{S_i(t) - \mu_i(t)}{\sigma_i^2(F_{i0})(a_{i}^2n_0^{-1} + (1-a_i)^2n_i^{-1})^{1/2}} \right] \leq \Phi \left[ \frac{\hat{\sigma}_i^2(F_{i0})^{1/2}}{\sigma_i^2(F_{i0})^{1/2}} \right] \]

(by using similar arguments given in Dudewicz and Dalal (1975))

\[ \approx E_{\sigma_i(F_{i0})} \left[ \Phi \left( \frac{\hat{\sigma}_i^2(F_{i0})^{1/2}}{\sigma_i^2(F_{i0})^{1/2}} \right) \right] \]  
\[ (5.2.17) \]
\[ \approx \Phi(y). \]  
\[ (5.2.18) \]

Approximations in (5.2.17) and (5.2.18) follow, respectively, from Lemma 5.2.1 and consistency of $\hat{\sigma}_i^2(F_{i0})$. Now using (5.2.18) in (5.2.16), we get

\[ P[CS|R] \approx \prod_{i=1}^{k-1} \Phi(z + (\mu_{[i]}(t) - \mu_{[i]}(t))h/\delta^*)d\Phi(z) \]  
\[ (5.2.19) \]
\[ \geq \int_{-\infty}^{\infty} \Phi(z+h)d\Phi(z) \]  
\[ (5.2.20) \]
\[ = P^*. \]  
\[ (5.2.21) \]

The inequality in (5.2.20) is true since the right hand side of (5.2.19) is minimized when $\mu_{[i]}(t) = ... = \mu_{[k-i]}(t) = \mu_{[k]}(t) - \delta^*$, and (5.2.21) follows from (5.2.5).

Remark 5.2.2: The procedure R can also be modified to select the population corresponding to least mean residual life of age t. In this case the preference-zone for fixed $\delta^*$, will be defined.
as \( \{ \mu(t): \mu_{[2]}(t) - \mu_{[1]}(t) \geq \delta^* \} \), the population corresponding to \( S_{[1]}(t) \) is selected and the probability of correct selection will be

\[
P[S_{(i)}(t) \geq S_{(i)}(t), \ i = 2, \ldots, k].
\]

\section*{5.3 Probability of Correct Selection and Expected Subset Size}

Procedure R can be modified to meet the goal (b), i.e., to select a random size subset \( S \) of \( k \) populations (firm) such that the selected subset has probability at least \( P'(k^{-1} < P' < 1) \) of containing the population associated with \( \mu_{[k]}(t) \). This approach of selection due to Gupta (1956,65) is termed as the subset selection approach. The modified rule is

\[ R_i: \text{Include } i\text{-th population (firm) in } S \text{ iff } S_i(t) \geq S_{[k]}(t) - \delta^*, \]

where the selection constant \( \delta^* \) and other details regarding the modified procedure \( R_i \), for selecting a subset of \( k \) populations, one may refer to Dudewicz and Dalal (1975).

\textbf{Theorem 5.3.1:} For any \( P'(k^{-1} < P' < 1) \), there exists \( n_0 \) large enough such that for a given \( t \),

\[
\inf_{\Omega} P[S \mid R] \approx P',
\]

where \( \Omega = \{ \mu(t): \mu(t) = (\mu_1(t), \ldots, \mu_k(t)) \} \).
Chapter 5

**Proof:** Let $S_{(i)}(t)$ denote the statistic associated with the population having parameter $\mu_{[i]}(t)$, $i = 1, ..., k$. Then for a given $t$, we have

$$P[CS|R_1] = P[S_{(k)}(t) \geq S_{[k]}(t) - \delta^*]$$

$$= P[S_{(k)}(t) \geq S_{(i)}(t) - \delta^*, \ i = 1, ..., k - 1]$$

$$= P[S_{(i)}(t) \leq S_{(k)}(t) + \delta^*, \ i = 1, ..., k - 1]$$

$$= P[S_{(i)}(t) - \mu_{[i]}(t) \leq (S_{(k)}(t) - \mu_{[k]}(t))$$

$$+ (\mu_{[k]}(t) - \mu_{[i]}(t)) + \delta^*, \ i = 1, ..., k - 1]$$

$$= P\left[(S_{(i)}(t) - n)N(t)h/\delta^* \leq (S_{(k)}(t) - n)N(t)h/\delta^*$$

$$+ (\mu_{[k]}(t) - \mu_{[i]}(t))h/\delta^* + h, \ i = 1, ..., k - 1\right]$$

$$\approx \int \prod_{-\infty}^{\infty} \Phi(z + (\mu_{[k]}(t) - \mu_{[i]}(t))h/\delta^* + h) d\Phi(z) \quad (5.3.1)$$

$$\geq \int \Phi^{-1}(z + h) d\Phi(z) \quad (5.3.2)$$

$$= P' \quad (5.3.4)$$

The approximate equality in (5.3.1) follows from (5.2.19) and independence of $S_i(t)$'s. The inequality in (5.3.2) is true since the right hand side of (5.3.1) is minimized when $\mu_{[i]}(t) = ... = \mu_{[k]}(t)$, and (5.3.4) follows from (5.2.5).
**Remark 5.3.1:** The procedure $R_1$ can also be modified to select a subset of $k$ populations containing the population associated with $\mu_{[i]}(t)$. The modified rule is

$$R_2: \text{Include i-th population (firm) in } S \text{ iff } S_j(t) \leq S_{[i]}(t) + \delta^*$$

and the corresponding probability of correct selection will be $P[S_{[i]}(t) \leq S_{[i]}(t) + \delta^*]$.

### 5.4 Modification of Procedures $R$ & $R_1$ with Right Censored Data

Suppose the population $\pi_i$ has the c.d.f. $F_i$, $i=1,...,k$. The life lengths of $k$ firms $X_{ij}^0$, $i=1,...,k; \ j=1,...,n$ from populations $\pi_1,...,\pi_k$ are subjected to right censorship by i.i.d. random variables $Z_{ij}$, $i=1,...,k; \ j=1,...,n$ with c.d.f. $G_i$, $i=1,...,k$. Assume that the random variables $X_{ij}^0$ and $Z_i$ are mutually independent. The observable random variables are $X_{ij} = \min(X_{ij}^0, Z_{ij})$, $i=1,...,k; \ j=1,...,n$ along with the indicator function.
Chapter 5

\[ \delta_{ij} = \begin{cases} 
1 & \text{if } X_{ij} = X_{ij}^0 \\
0 & \text{if } X_{ij} = Z_{ij}.
\end{cases} \]

The indicator variable thus states whether or not \( X_{ij} \) is censored.

Let \( X_{i(1)} \leq \ldots \leq X_{i(n)} \) be the order statistics corresponding to \( X_{i1}, \ldots, X_{in} \) and let \( \delta_{i(j)} \) be the value of \( \delta \) associated with \( X_{i(j)} \).

Define

\[ \tau_{X_i} = \max \{ X_{ij} \mid \delta_{ij} = 1, \ 1 \leq j \leq n \} \quad i = 1, \ldots, k. \]

Then the commonly used estimator of \( \bar{F}_i \) is the product limit estimator (PLE) \([Kaplan and Meier (1958)]\) which, when there are no tied observations, is

\[ \frac{\bar{F}_{in}}{F_{in}} = \begin{cases} 
\prod_{j=1}^{n-j \mid X_{i(1)} \leq X_{i(j)}} \left( \frac{n-j}{n-j+1} \right)^{\delta_{i(j)}} & \text{if } t \leq X_{i(n)} \\
0 & \text{if } t > X_{i(n)}.
\end{cases} \]

Thus, an estimator of \( \mu_i(t) \), the mean residual life of age \( t \), is

(5.4.1)

Thus, an estimator of \( \hat{R}_{in}(t) \), the mean residual life of age \( t \), is

\[ \hat{R}_{in}(t) = \frac{1}{\bar{F}_{in}(t)} \int_{\bar{F}_{in}(u)du} \]
for $0 \leq t < \tau_{\eta^*}, \ i = 1, \ldots, k$.

Since it is almost impossible to estimate $F_i$ if the support of $G_i$ is smaller than that of $F_i$, we assume that

$$\tau_{F_i} = \text{Sup}\{x \mid x \text{ is in the support of } F_i\}$$

$$\leq \text{Sup}\{x \mid x \text{ is in the support of } G_i\}, \ i = 1, \ldots, k.$$  

[For detail one may refer to Li (1997)].

Below we state a theorem due to Li (1997).

**Theorem 5.4.1:** For fixed $t$, the distribution of

$$n^{-1/2} \left[ \dot{R}_{in}(t) - \mu_i(t) \right]$$

is normal with mean zero and variance

$$\sigma^2(F_i, G_i, U_i) = \frac{1}{F(t)^2} \left\{ \int_U \left[ F(x) \right] dx + \int_{U(t)} \int U(x) \dot{F}(v) dv \right\}.$$
Here \( U_i(t) = \int_0^t \frac{dF_i(x)}{F_i(x)H_i(x)} \) for \( t < \tau_i \), \( i = 1, \ldots, k \),

where \( \overline{H}_i(t) = 1 - H_i(t) = \overline{F}_i(t)\overline{G}_i(t) \), \( i = 1, \ldots, k \).

The consistent estimator \( \hat{\sigma}^2(F_{in}, G_{in}, U_{in}) \) is obtained by replacing \( F_i \) by its product limit estimator \( \hat{F}_{in} \) and \( U_i \) by \( \hat{U}_{in} \) where,

\[
\hat{U}_{in} = n \sum_{j=1}^{\infty} \frac{\delta_{i(j)}}{(n-j)(n-j+1)}, \quad i = 1, \ldots, k.
\]

Having obtained the asymptotic distribution of \( \hat{R}_{in}(t) \), procedures \( R \) and \( R_1 \) can easily be modified by replacing \( R_{in}(t), R_{in_0}(t), \sigma^2(F_{in}) \) and \( \hat{\sigma}^2(F_{in_0}) \) by \( \hat{R}_{in}(t), \hat{R}_{in_0}(t), \sigma^2(F_{in_0}, G_{in_0}, U_{in_0}) \) and \( \hat{\sigma}^2(F_{in_0}, G_{in_0}, U_{in_0}) \), respectively.

### 5.5 A Simulation Study

In this section we present the result of a simulation study, carried to see the performance of procedures \( R \) and \( R_1 \) when the underlying distributions belong to (i) the same family and (ii) the different families. Tables 5.5.1 and 5.5.2 list the distributions considered for simulation study. The entries given in brackets along with the values of \( \theta_i \) represent the corresponding mean residual life \( (\mu_i(t)) \) of age \( t \) (in Table 5.5.2
order of $\theta_i(\mu_i(t))$ is same as that of the distributions mentioned there).

### TABLE 5.5.1

<table>
<thead>
<tr>
<th>Family</th>
<th>Exponential: $F(x) = e^{-x}$</th>
<th>Weibull: $F(x) = e^{-x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k \rightarrow$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$t \rightarrow$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\theta_i(\mu_i(t)) \rightarrow$</td>
<td>0.8(0.8), 1.0(1.0), 2.4(2.4)</td>
<td>0.4(7.67, 0.67(2.07), 0.92(1.16)</td>
</tr>
<tr>
<td>Initial sample size ($n_0$) $\rightarrow$</td>
<td>15,20,25,30,35,40,45,50</td>
<td>15,20,25,30,35,40,45,50</td>
</tr>
</tbody>
</table>

### TABLE 5.5.2

<table>
<thead>
<tr>
<th>Families $\rightarrow$</th>
<th>$F(x) = e^{-x}$</th>
<th>$F(x) = e^{-x}$</th>
<th>$F(x) = 1 - (1 - e^{-x})^{k+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k \rightarrow$</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$t \rightarrow$</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\theta_i(\mu_i(t)) \rightarrow$</td>
<td>2.4(2.4), 0.9(1.19), 1.0(1.1127)</td>
<td>1.2(1.2), 0.57(2.89), 2.0(1.2272)</td>
<td></td>
</tr>
<tr>
<td>Initial sample size ($n_0$) $\rightarrow$</td>
<td>15,20,25,30,35,40,45,50</td>
<td>15,20,25,30,35,40,45,50</td>
<td></td>
</tr>
</tbody>
</table>

In the simulation procedure the value of $\delta^*$ is chosen to be near the true difference between the two largest $\mu(t)$ values. IMSL subroutines are used to generate random samples from different distributions. In each case, for parameters ($\theta_i$) of $k$ distributions we first generate a random sample of size $n_0$ and then compute $R_{in_0}(t)$ and $\delta^2_i(F_{in_0})$ given by (5.2.2) and (5.2.4), respectively. Then by using (5.2.8) we determine the additional sample size $n_i$ and generate a fresh random sample of $n_i$ observations from distribution $F_i$. Then $R_{in_i}(t)$ is computed.
Using Bechhofer (1954) table we use $h=1.6524$ for $k=3$ and $P^*=0.80$. $S(t)$ is obtained from (5.2.9). For a particular set of parameters considered in Tables 5.5.1 and 5.5.2 and initial sample size $n_0$, we repeated each simulation procedure 10,000 times by generating fresh random samples. The values of relative frequencies of correct selection i.e., relative frequency of selecting the population associated with largest mean residual life of age $t=1$, estimated expected additional sample size ($\hat{E}(N)$) and estimated expected subset size ($\hat{E}(S)$) are computed from 10,000 repetitions. $\hat{E}(N)$ and $\hat{E}(S)$ are respectively the averages of additional sample sizes and subset sizes among 10,000 repetitions. The numerical results obtained from the simulation study are given below in the Tables 5.5.3 and 5.5.4.
TABLE 5.5.3
Estimates of expected additional sample size (\(\hat{E}(N)\)), the probabilities of correct selection (\(\hat{P}_L\) for indifference zone) and \(\hat{P}_S\) (for subset selection)) and estimated expected subset size (\(\hat{E}(S)\)) with \(P^* = 0.80\), \(k = 3\), \(h = 1.6524\), \(t = 1\).

<table>
<thead>
<tr>
<th>Families (\rightarrow)</th>
<th>(n_0)</th>
<th>(\theta^*_1)</th>
<th>(\hat{P}_L)</th>
<th>(\hat{P}_S)</th>
<th>(\hat{E}(S))</th>
<th>(P_L)</th>
<th>(P_S)</th>
<th>(\hat{E}(S))</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>(\mu(t)\rightarrow)</td>
<td>2.4, 1.0, 0.80</td>
<td>0.9052</td>
<td>0.9600</td>
<td>1.5067</td>
<td>0.9052</td>
<td>0.9586</td>
<td>2.2751</td>
</tr>
<tr>
<td></td>
<td>(\hat{E}(N))</td>
<td>13.5082, 1.3535, 1.0328</td>
<td>0.9052</td>
<td>0.9600</td>
<td>1.5067</td>
<td>0.9052</td>
<td>0.9586</td>
<td>2.2751</td>
</tr>
<tr>
<td>20</td>
<td>(\hat{E}(N))</td>
<td>9.3096, 0.4982, 0.5468</td>
<td>0.9187</td>
<td>0.9642</td>
<td>1.6781</td>
<td>0.9187</td>
<td>0.9782</td>
<td>2.2449</td>
</tr>
<tr>
<td>25</td>
<td>(\hat{E}(N))</td>
<td>6.1978, 0.1980, 0.2558</td>
<td>0.9421</td>
<td>0.9695</td>
<td>1.8151</td>
<td>0.9421</td>
<td>0.9833</td>
<td>2.2430</td>
</tr>
<tr>
<td>30</td>
<td>(\hat{E}(N))</td>
<td>3.4308, 0.0716, 0.0872</td>
<td>0.9623</td>
<td>0.9788</td>
<td>1.8769</td>
<td>0.9623</td>
<td>0.9910</td>
<td>2.2622</td>
</tr>
<tr>
<td>35</td>
<td>(\hat{E}(N))</td>
<td>1.9633, 0.0263, 0.0394</td>
<td>0.9746</td>
<td>0.9852</td>
<td>1.8935</td>
<td>0.9746</td>
<td>0.9919</td>
<td>2.2700</td>
</tr>
<tr>
<td>40</td>
<td>(\hat{E}(N))</td>
<td>0.9973, 0.0049, 0.116</td>
<td>0.9880</td>
<td>0.9921</td>
<td>1.881</td>
<td>0.9880</td>
<td>0.9934</td>
<td>2.2679</td>
</tr>
<tr>
<td>45</td>
<td>(\hat{E}(N))</td>
<td>0.4193, 0.0022, 0.0039</td>
<td>0.9918</td>
<td>0.9958</td>
<td>1.8787</td>
<td>0.9918</td>
<td>0.9946</td>
<td>2.2677</td>
</tr>
<tr>
<td>50</td>
<td>(\hat{E}(N))</td>
<td>0.1809, 0.0037, 0.0026</td>
<td>0.9956</td>
<td>0.9971</td>
<td>1.8768</td>
<td>0.9956</td>
<td>0.9945</td>
<td>2.2470</td>
</tr>
</tbody>
</table>
### TABLE 5.4.4
Estimates of expected additional sample size ($\hat{E}(N')$), the probabilities of correct selection ($\hat{P}_L$ (for indifference zone) and $\hat{P}_S$ (for subset selection)) and estimated expected subset size ($\hat{E}(S)$) with $P^* = 0.80$, $k = 3$, $h = 1.6524$, $t = 1$.

<table>
<thead>
<tr>
<th>Families $\rightarrow$</th>
<th>Exponential, Weibull, Lehmann, $\delta^* = 1.2$</th>
<th>Weibull, Lehmann, Exponential, $\delta^* = 1.6628$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0$</td>
<td>$\hat{E}(N')$</td>
<td>$\hat{E}(N')$, $\hat{E}(S)$</td>
</tr>
<tr>
<td>$\mu_{(1)} \rightarrow$</td>
<td>2.40, 0.90, 1.00</td>
<td>0.57, 2.00, 1.20</td>
</tr>
<tr>
<td>15</td>
<td>$\hat{P}_L$</td>
<td>0.8791</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9327</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>1.2998</td>
</tr>
<tr>
<td>20</td>
<td>$\hat{E}(N')$</td>
<td>18.1205, 3.8796, 0.2801</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.8931</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9447</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>1.4419</td>
</tr>
<tr>
<td>25</td>
<td>$\hat{E}(N')$</td>
<td>13.5240, 2.4654, 0.0488</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9037</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9504</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>1.6478</td>
</tr>
<tr>
<td>30</td>
<td>$\hat{E}(N')$</td>
<td>10.1422, 1.3950, 0.0068</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9195</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9570</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>1.8831</td>
</tr>
<tr>
<td>35</td>
<td>$\hat{E}(N')$</td>
<td>6.8135, 0.6553, 0.0004</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9381</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9685</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>2.1123</td>
</tr>
<tr>
<td>40</td>
<td>$\hat{E}(N')$</td>
<td>4.3224, 0.3557, 0.0002</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9576</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9777</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>2.3234</td>
</tr>
<tr>
<td>45</td>
<td>$\hat{E}(N')$</td>
<td>2.6784, 0.1859, 0.0000</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9706</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9852</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>2.4647</td>
</tr>
<tr>
<td>50</td>
<td>$\hat{E}(N')$</td>
<td>1.3758, 0.0787, 0.0000</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_L$</td>
<td>0.9775</td>
</tr>
<tr>
<td></td>
<td>$\hat{P}_S$</td>
<td>0.9883</td>
</tr>
<tr>
<td></td>
<td>$\hat{E}(S)$</td>
<td>2.5660</td>
</tr>
</tbody>
</table>
Chapter 5

From Tables 5.5.3 and 5.5.4 we see that when all the distributions are from the same family the values of $\hat{E}(N)$, $\hat{P}_L$, $\hat{P}_S$ and $\hat{E}(S)$ start attaining stability when $n_0 \geq 30$ (see Table 5.5.3). But when the different distributions are considered the said stability is attained for $n_0 \geq 35$ (see Table 5.5.4). The results are not surprising that initial sample sizes are large but are expected so since in simulation we have considered tight parametric configurations and moreover approximately 50% sample observations were seen to be less than $t$ in each simulation implying that only remaining 50% observations greater than $t$ contributed in estimating $R_{in_0}(t)$ and $R_{in_i}(t)$, $i = 1, \ldots, k$. 